Department of Physics and Astronomy

University of Heidelberg

Master thesis

in Physics

submitted by

Benjamin Weinläder

born in Heidelberg

2020

Development of a Monolithic Pixel Sensor with sub-nanosecond Time Resolution in BiCMOS

This Master thesis has been carried out by Benjamin Weinläder

at the

Physikalischen Institut Heidelberg

under the supervision of

Prof. André Schöning

Abstract:

In the field of particle physics, High Voltage Monolithic Active Pixel Sensors (HV-MAPS) are promising candidates to fulfil the high demands on spatial and time resolution of modern detectors. The Mu3e experiment with its development of the MuPix sensor has strongly driven this technology in recent years. As an example of the latest successes, the MuPix8 can be mentioned, which reached a time resolution of $\sigma_t = 6.8\,\mathrm{ns}$, using a 180 nm HV-CMOS technology.
The combination of HV-MAPS with a BiCMOS technology opens up further possibilities, especially for the improvement of time resolution, which was proven in the scope of the TT-PET project at the university of Geneva.
Based on this experience, this thesis aims to improve the time resolution of HV-MAPS to the sub-nanosecond regime, while maintaining a preferable small power consumption. Thereto, the advantages of the 130 nm BiCMOS process SG13S by IHP are utilised and studied.
Within the thesis, the pixel front end electronics, up to the digitisation of a hit signal is designed, with a special focus on the analogue sensor pixel and the charge sensitive amplifier. A characterisation of the pixel layout with Cadence® Virtuoso® resulted in a simulated ToA Jitter of $\sigma_{ToA} = 86.12\,\mathrm{ps}$ for a signal, corresponding to a MIP. Further, a noise of $ENC = 205\,\mathrm{e}^-$ was measured for the $25 \times 25\,\mu\mathrm{m}^2$ large pixel. After the digitisation of the signal a ToA Jitter of $\sigma_{ToA} = 435\,\mathrm{ps}$ is simulated for the whole pixel front end electronics.

Zusammenfassung:

Im Bereich der Teilchen Physik sind die Hochspannungsbetriebenen-Monolithischen Aktiven Pixel Sensoren (HV-MAPS) vielversprechende Kandidaten, die die hohen Anforderungen moderner Detektoren an die räumliche und zeitliche Auflösung erreichen können. In den letzten Jahren wurde diese Technologie vor allem durch das Mu3e Experiment und die zugehörige Entwicklung des MuPix-Sensors vorangetrieben. Nennenswert ist der letzte Erfolg, der MuPix8, mit dem eine Zeitauflösung von $\sigma_t = 6.8\,\mathrm{ns}$ in einem 180 nm HV-CMOS Prozess erreicht wurde.
Eine Kombination von HV-MAPS mit einer BiCMOS Technologie bietet weitere Vorteile, vor allem kann die Zeitauflösung signifikant verbessert werden, wie im Rahmen des TT-PET Projekts an der Universität Genf gezeigt wurde.
Basierend auf dieser Erfahrung ist das Ziel dieser Arbeit, die Zeitauflösung von HV-MAPS in den Sub-Nanosekunden Bereich zu verbessern, während der Stromverbrauch möglichst klein gehalten wird. Es werden dazu die Vorzüge des 130 nm BiCMOS Prozesses SG13S von IHP genutzt und untersucht.
Die Pixel Frontend Elektronik bis zur Digitalisierung des Signals wird designt, dabei wird ein besonderes Augenmerk auf das analoge Sensorpixel und den Ladungsverstärker gelegt. Aus einer Charakterisierung mittels Cadence® Virtuoso® ergibt sich für das analoge Pixel ein simulierter Eingangszeit-Jitter von $\sigma_{ToA} = 86.12\,\mathrm{ps}$ für ein MIP-ähnliches Signal. Des weiteren wurde ein Rauschen von $ENC = 205\,\mathrm{e}^-$ für das $25 \times 25\,\mu\mathrm{m}^2$ große Pixel gemessen. Nach der Digitalisierung des Signals wird ein simulierter ToA Jitter von $\sigma_{ToA} = 435\,\mathrm{ps}$ für die gesamte Pixel Frontend Elektronik gemessen.

Contents

1 Introduction **9**
- 1.1 State of Research . . . 10
  - 1.1.1 The Mu3e Experiment . . . 10
  - 1.1.2 The TT-PET Project . . . 12
  - 1.1.3 Hexagonal Prototype . . . 13

2 Semiconductor Physics **15**
- 2.1 Definition of Semiconductors . . . 15
- 2.2 Transport Properties of Semiconductors . . . 18
- 2.3 Doped Semiconductors . . . 20
- 2.4 pn-Junction . . . 20
  - 2.4.1 pn-Junction in Equilibrium . . . 20
  - 2.4.2 pn-Junction with External Fields . . . 22

3 Transistor **25**
- 3.1 MOSFET . . . 25
- 3.2 Bipolar Junction Transistor . . . 28
  - 3.2.1 Heterojunction Bipolar Transistor . . . 31
- 3.3 Noise Sources in a Transistor . . . 32
  - 3.3.1 Noise of a MOSFET . . . 32
  - 3.3.2 Noise of a BJT . . . 34

4 Electrical Devices and Circuits **35**
- 4.1 Diode as a Sensor . . . 35
- 4.2 Charge Sensitive Amplifier . . . 36
  - 4.2.1 Single Gain Stage . . . 37
  - 4.2.2 Cascode Amplifier . . . 39
- 4.3 Source Follower . . . 40
- 4.4 Current Mirror . . . 41
- 4.5 Comparator . . . 41
- 4.6 Pixel Front End Electronics . . . 43

5 Simulation tools **45**
- 5.1 Cadence . . . 45
  - 5.1.1 Schematic Editor . . . 45
  - 5.1.2 Layout Suite . . . 45
  - 5.1.3 Analog Design Environment . . . 46

5.2 Design Kit . . . 47
5.2.1 IHP SG13S . . . 47

6 Signal Parameters **49**
6.1 Time Walk . . . 49
6.2 Charge Collection Noise . . . 50
6.3 Electronic Noise . . . 51
6.4 Measured Quantities . . . 53

7 Layout **55**
7.1 Layout Requirements . . . 55
7.2 Optimisation Process . . . 56
7.2.1 Workbench . . . 56
7.2.2 Design Guidelines . . . 58
7.2.3 Gain Stage Optimisation . . . 58
7.2.4 Active Feedback Tree . . . 61
7.2.5 Source Follower . . . 62
7.2.6 Comparator . . . 63
7.3 The Analogue Pixel Design . . . 65
7.3.1 Schematic of the Analogue Pixel . . . 65
7.3.2 Layout Architecture . . . 66
7.3.3 Parasitic Components . . . 69
7.4 The Digital Pixel Design . . . 70
7.4.1 Schematic of the Digital Pixel . . . 70
7.4.2 Layout Architecture . . . 71

8 Characterisation **73**
8.1 Characterisation of the Analogue Pixel . . . 73
8.1.1 Transfer Function . . . 73
8.1.2 Transient Response . . . 74
8.1.3 Noise Analysis . . . 76
8.1.4 Monte Carlo Analysis . . . 79
8.1.5 The Detector Capacity . . . 82
8.2 Characterisation of the Pixel Front End . . . 83
8.2.1 Influence of the Connection Line . . . 83
8.2.2 Transient Response of the Pixel Front End . . . 85

9 Discussion and Summary **91**
9.1 Conclusion . . . 93

A MuPix10 Pixel Capacity **I**

B Reference Pixel **II**

C Bibliography **III**

1 Introduction

In the modern physic, new phenomena are searched that can be described by theories well beyond established theories as the Standard Model (SM) of particle physics. Apart from the elaboration of these new theories, an essential part is to verify or falsify them in experiments, thus pushing the known boundaries further. In general, there are two approaches for modern high energy experiments: The first possibility, on the energy frontier, is to search for heavy resonances or new particles. In order to increase the probability to find such events in a moderate time scale, these experiments tend to increase the frequency at which is measured. This reflects in an increase of the luminosity of the associated particle collider. Collisions occur almost simultaneously, resulting in a signal pile-up. Here, a good example are the experiments located at the Large Hadron Collider (LHC) and in particular the planned upgrades. Another approach for high energy experiments are specialised precision experiments, which search for dedicated properties of particles or their rare decay channels. Deviations from the SM expectation can lead to new physics. The planned Mu3e experiment [1] with the search for the muon decay $\mu^+ \rightarrow e^+e^-e^+$ can be mentioned as an example.

Regardless of the nature of the experiment, they set high demands on the spatial and time resolution. Either to achieve the necessary sensitivity or to distinguish single events, despite signal pile-up. Hence, the sensor technology, prevalently implemented as silicon tracking detectors, has to be adapted to the high requirements of new experiments.

A promising candidate is the implementation of High-Voltage Monolithic Active Pixel Sensors (HV-MAPS). Here, the Mu3e experiment with its development of the MuPix sensor has strongly driven this technology in recent years. The combination of HV-MAPS with the Bipolar Complementary Metal Oxide Semiconductor (BiCMOS) technology adds additional advantages. This was proven at the university of Geneva, in the scope of the TT-PET project [2]. Due to the combination of the technologies a significant improvement of the time resolution was achieved.

This thesis studies the improvements, which can be made by using the BiCMOS process SG13S [3] by IHP for the development of HV-MAPS. This process includes a heterojunction bipolar transistor (HBT) with a transient frequency of 240 GHz, besides 130 nm large standard Complementary Metal Oxide Semiconductor (CMOS) devises. The ambition is to design the analogue sensor pixel within the chip matrix and to optimise it for a time resolution in the sub-nanosecond regime. Based on the experience gained with the MuPix a starting point is chosen, e.g. for various circuitries and the power consumption. Therefore, several requirements are based on those of the Mu3e experiment. The performance of the optimised pixel and the

subsequent circuits up to the digitisation of the signal are studied. It is aimed to operate the resulting small pixel layout in a preferably large matrix with a low power consumption. The design process and the following characterisation is done with the software Cadence® Virtuoso®.

A possible fully functional sensor, including a time resolution of a couple tens of picoseconds and a large pixel matrix offers a huge potential for many experiments. Such sensor would approach the time performance of scintillators, while the pixel matrix would also collect spatial information. In that sense, it would meet with the requirements of the phase II of the Mu3e experiment, enabling a precise time measurement on the pixel detector, without the need of additional timing detectors.

1.1 State of Research

A starting point for this study was chosen according to experience from the development of the MuPix sensor. In addition, the work of other research groups was also included, whereby the idea to implement HV-MAPS in a BiCMOS process has already been realised by the TT-PET project. Here, the current state of research will be presented.

1.1.1 The Mu3e Experiment

The Mu3e experiment [1, 4] at the Paul Scherrer Institut (PSI) is a precision experiment, which aims to search for the lepton flavour violating decay $\mu^+ \rightarrow e^+e^-e^+$. In two experimental phases, muons are stopped in a fixed target and their decay products are observed, with the detector presented in figure 1.1. In phase I a sensitivity of the branching ratio of 2×10^{-15} is targeted, which will be increased to 10^{-16} in the second phase. Therefore, in phase II the muon decay rate is increased to approximately 2×10^9. Simultaneously, the detector will be adapted to the high muon intensity.

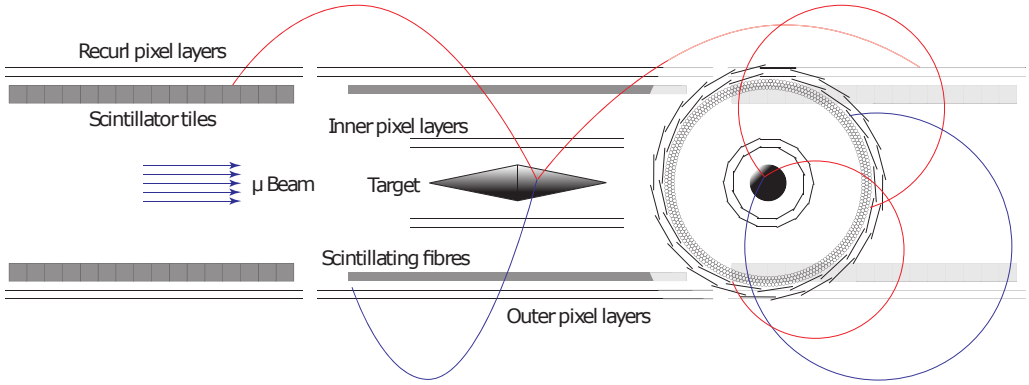


Figure 1.1: Schematic presentation of the Mu3e phase I detector, with an indicated decay $\mu^+ \rightarrow e^+e^-e^+$.

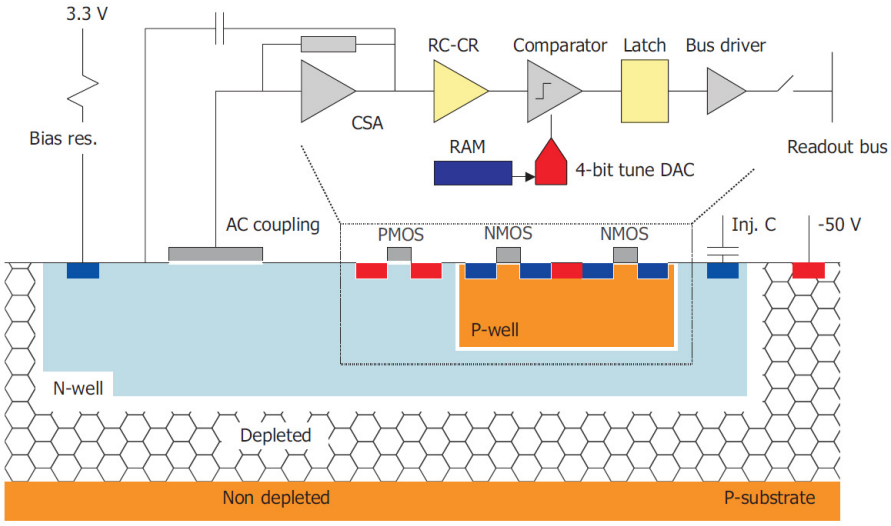


Figure 1.2: Conceptional illustration of an HV-MAPS.

In order to guarantee a successful measurement, several requirements are postulated for the detector system. This includes high constrains for the spatial and time resolution, as it is necessary to distinguish between several possible background signals and the searched decay. Moreover, the spatial resolution is limited by the multiple Coulomb scattering, due to the low energy of the decay particles. For this reason, it is essential to build the detector with a low material budget.

The MuPix sensor is being developed to meet these requirements. As an HV-MAPS it offers various advantages compared to hybrid pixel sensors, which are classically used for silicon pixel tracker. HV-MAPS combines the readout electronics and the sensor diode in a single die, reducing the used material. The electronics is implemented in a state of the art 180 nm high-voltage CMOS process, which is commercially available and thus reduces the cost of the sensor.

The particle detection of the chip is realised as a diode, embedded in the silicon p-typed substrate. As indicated in figure 1.2, the anode of the diode is implemented by a deep n-well, below the readout electronics of the pixel. A more detailed discussion of the detection process is given in section 4.1. As explained there, another feature of the HV-MAPS is the fast charge collection via drift, reinforced by the high voltage. Further, the high voltage increases the detection volume of the sensor diode, as it results in a large depletion zone.

The MuPix sensor, specially developed for the Mu3e experiment, has gone through several iterations. The latest version, the MuPix10 is a fully integrated large prototype with a chip size of $20.66 \times 23.18\,\mathrm{mm}^2$. Thereby, its active matrix contains 256×250 pixels, each of a size of $80 \times 80\,\mu\mathrm{m}^2$. This results in a pixel capacity of approximately 70 fF, calculated from TCAD simulations presented in figure A.1. Based on the geometry of the pixel, the sensor has a spatial resolution of approximately $23\,\mu\mathrm{m}$, given by $1/\sqrt{12}$ of the pixel size. For the time resolution a value

below 20 ns is targeted. In measurements with previous iterations like the MuPix8 this specification could be met, with a measured time resolution of 6.8 ns [5]. Further, due to the limited cooling power of the detector, the chip should not exceed a power consumption of 350 mW/cm^2. Depending on it settings, the MuPix8 sensor typically reaches a power consumption of 250 mW/cm^2. With regard to its geometry this translate in a power consumption per channel of approximately 16 μW/ch [6].

Additional to the MuPix sensor, it is foreseen to have a dedicated time measurement. This is realised by a layer of scintillating fibres [7] and tiles [8], as it can be seen in figure 1.1. These provide a time resolution below 500 ps and 100 ps respectively. If these timing requirements are fulfilled by an updated version of the MuPix, it would be possible to run phase II of the experiment only with pixel sensors.

1.1.2 The TT-PET Project

The aim of the Thin Time-of-flight Positron Emission Tomography (TT-PET) project [2] is to develop a PET scanner with an improved spatial resolution, based on the time-of-flight measurement. According to the basic principle of PET, the ring detector, shown in figure 1.3, measures two gamma rays travelling in opposite directions. The silicon sensor therefore measures the electron shower, emerging from the photons stopped in a lead layer. The additional information from the time difference of both signals, can be used to reconstruct their vertex. Therefore, the project requires a time resolution of 30 ps, which translates into a spatial resolution of approximately 1 mm. Due to a structure of several layers of sensors, the spatial resolution is conserved, even when the origin is not located in the detector centre.

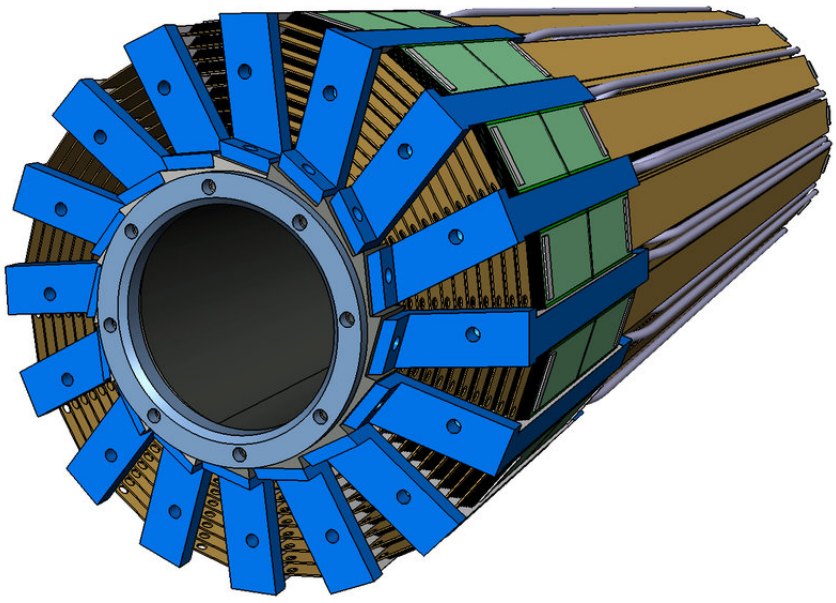


Figure 1.3: CAD drawing of the TT-PET detector [9].

For this project an HV-MAPS is developed, with the idea to implement its front end electronics in a BiCMOS process. Due to the excellent timing and noise performance of the bipolar transistor, it is possible to achieve the required time resolution. Both, timing and noise behaviour of the transistor are explained in more detail later, e.g. in section 3.2.1 and 6.3 respectively. Based on this, a demonstrator chip [10] was produced, using the BiCMOS process SG13S by IHP. It was designed to test the main features of the final TT-PET chip. Thereto, it comprises an active matrix with 10×3 pixels, each of a size of $470 \times 470\,\mu\mathrm{m}^2$ and a spacing of $30\,\mu\mathrm{m}$. This results in a simulated pixel capacity of 750 fF. Signals are read out in a simple architecture, reducing the complexity of the readout logic.

For the demonstrator chip a time resolution of 110 ps was measured in a high-power mode with a power consumption of $375\,\mu\mathrm{W/ch}$. The front end noise of the sensor pixel was estimated to be $350\,\mathrm{e}^-$.

1.1.3 Hexagonal Prototype

Based on this excellent result from the TT-PET demonstrator chip, a prototype chip [11] is produced at the university of Geneva. It is a proof of concept with the intention to achieve a time resolution below 100 ps.

The HV-MAPS is produced in the SG13G2 [12] process by IHP and contains two matrices with hexagonal pixels, either with a side of $130\,\mu\mathrm{m}$ or $65\,\mu\mathrm{m}$. Due to the smaller pixel size and the consequently lower detector capacity of 220 fF or 70 fF, it was possible to reduce the electric noise of the chip. This results in a measured time resolution of 55 ps for the large pixels and 46 ps for the small pixels. The estimated noise charge are $160\,\mathrm{e}^-$ and $90\,\mathrm{e}^-$ respectively.

2 Semiconductor Physics

The main scope of this thesis is to create and discuss circuits on chip-level. Therefore, it is essential to understand the basics of semiconductors physics, starting with the definition of a semiconductor and the behaviour of charge carriers inside the semiconductor.

2.1 Definition of Semiconductors

The physics of condensed matter [13, 14, 15] distinguish between isolators, conductors and semiconductors. Naturally, the simplest definition of these three classes can be done via the conductivity. In this sense, the conductivity of a metal is infinite large, while it goes to zero for insulators. The conductivity of a semi conductor should be located somewhere in between, whereby the borders between the classes blur and are therefore not well defined.

A more elaborate way uses the electronic band structure of solid materials to define semiconductors and further describe their basic behaviour. In order to establish the electronic band model, it is necessary to make some assumptions: the lattice of the solid crystal is expressed by a periodic potential, whereby the positive atomic nuclei are taken into account as a small perturbation. Further, the One-Electron Approximation should be true, meaning there is no interaction between single electrons. It is therefore sufficient to solve the Schrödinger Equation

$$\mathcal{H}\Psi(\boldsymbol{r}) = \Big[-\frac{\hbar^2}{2m}\nabla^2 + V(\boldsymbol{r})\Big]\Psi(\boldsymbol{r}) = E\Psi(\boldsymbol{r}) \tag{2.1}$$

once for a single electron. By summing up all N possible electronic states, considering the Pauli Principle, the solution for the entire solid follows.

Starting with the periodic potential $V(\boldsymbol{r})$: It fulfils the translation symmetry $V(\boldsymbol{r}) = V(\boldsymbol{r}+\boldsymbol{R})$, whereby $\boldsymbol{R}$ is an arbitrary lattice vector between two nuclei. Calculations are done in the reciprocal space, hence the potential is expressed dependent on the reciprocal lattice vector $\boldsymbol{G}$, using the Fourier series.

$$V(\boldsymbol{r}) = \sum_{\boldsymbol{G}} V_{\boldsymbol{G}} \mathrm{e}^{i\boldsymbol{G}\boldsymbol{r}} \tag{2.2}$$

For the electron, the general ansatz of a linear combination of plain waves is used. A more precise description of the electrons was done by Felix Bloch [16]. He found that the plain waves had to be modulated by a lattice periodic function, in order to move in the periodic crystal without scattering.

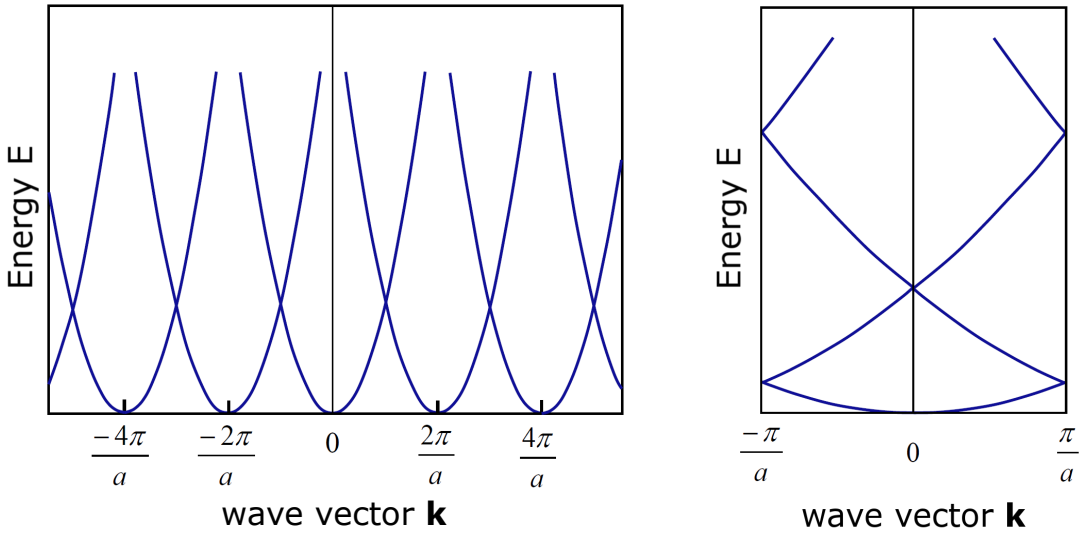
$$\Psi_{\boldsymbol{k}}(\boldsymbol{r}) = u_{\boldsymbol{k}}(\boldsymbol{r})\mathrm{e}^{i\boldsymbol{k}\boldsymbol{r}}, \quad u_{\boldsymbol{k}}(\boldsymbol{r}) = u_{\boldsymbol{k}}(\boldsymbol{r} \pm \boldsymbol{R}) \tag{2.3}$$

Only deviation of the perfect periodicity will lead to scattering.

Further, by explicitly solving the Schrödinger Equation 2.1 it can been shown that the momentum of the electron $\boldsymbol{k}$ is directly connected to $\boldsymbol{k} + \boldsymbol{G}$. Hence, the equation:

$$\Psi_{\boldsymbol{k}+\boldsymbol{G}} = \Psi_{\boldsymbol{k}} \tag{2.4}$$

holds, meaning, the momentum is not unique but can differ by a reciprocal lattice vector $\boldsymbol{G}$. It is therefore sufficient to indicate the solution for the first Brillouin zone to solve the whole system. This method is called reduction to the first Brillouin zone and it also applies to the eigenvalues.



(a) Representation of the energy eigenvalues according to equation 2.5.

(b) The eigenvalues reduced to the first Brillouin zone.

Figure 2.1: Eigenvalues of a one-dimensional system in the empty lattice approximation (adapted from[13]).

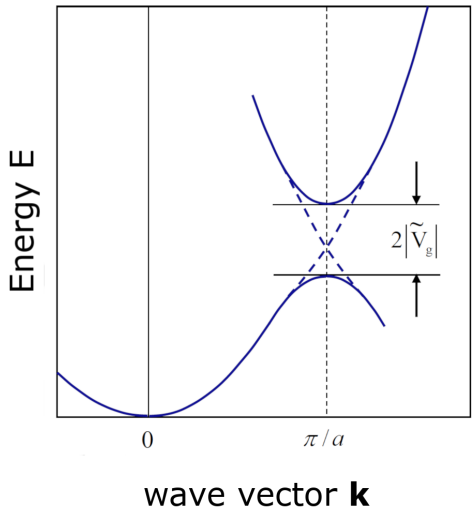
A first simple solution for the eigenvalues can be found using the approach of the empty lattice approximation, whereby the potential $V_{\boldsymbol{G}} \approx 0$ of the lattice is assumed to be small. However, the periodic nature of the lattice will remain. This leads to the following parabolic expression for the energy eigenvalues:

$$E_{\boldsymbol{k}} = \frac{\hbar^2 \boldsymbol{k}^2}{2m} = E_{\boldsymbol{k}+\boldsymbol{G}} = \frac{\hbar^2}{2m} |\boldsymbol{k} + \boldsymbol{G}|^2 \tag{2.5}$$

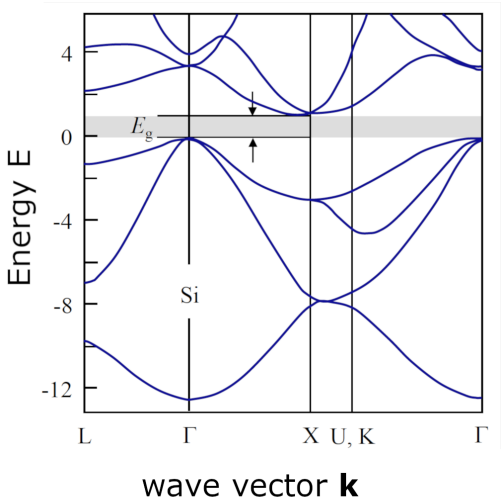
In figure 2.1 these eigenvalues in a one-dimensional system are presented. Furthermore, the reduction to the first Brillouin zone 2.1(b) is shown, which in principle corresponds to folding back the eigenvalue function.

Introducing the potential $V_{\boldsymbol{G}}$ again as a small perturbation leads to the nearly free electron approximation. Due to the influence of the atomic nuclei, a splitting of

the eigenvalue function occurs at crossing points, which is indicated in figure 2.2(a). Thus, individual energy bands are formed, which are separated from each other by gaps E_g. Within these gaps no energy states are allowed, while there is a discreet function describing the bands. It should be noted that the bands are not a consequence of lattice periodicity, but of interactions between atoms.



(a) Influence on the eigenvalues of the potential V_G in the nearly free electron approximation. The dashed line indicates the solution for the emtpty lattice approach.



(b) Band structur of a silicon crystal given in eV. Marked is the energy gap E_G of the indirect semiconductor.

Figure 2.2: Illustration of the band structur in a simplfied theory and for silicon (adapted from [13]).

The energy eigenvalues in the system are occupied until all electrons are distributed, with the electrons satisfying the Fermi-Dirac statistic. Accordingly, a state can only be occupied by two electrons with different spin. The energy of the highest occupied state at $T = 0\,\mathrm{K}$ is given by the so-called Fermi energy:

$$E_F = \frac{\hbar^2}{2m}(3\pi^2 n)^{2/3} \tag{2.6}$$

whereby n is the density of electrons. The highest fully occupied band is called the valence band, the first band not occupied is the conduction band. In the following the upper edge of the valence band is referred to as E_V and the lower edge of the conduction band as E_C.

With this, a semiconductor can be defined as material, where the Fermi energy is located between the valence and the conduction band. In contrast, the Fermi energy of a metal, is located inside the conduction band. Since every state below the Fermi energy is occupied at $T = 0\,\mathrm{K}$ and there is a state for every momentum direction,

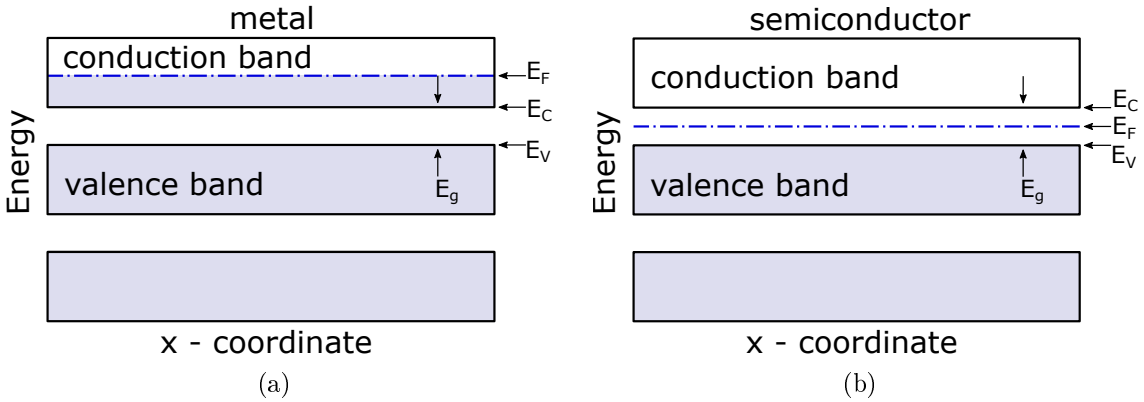


Figure 2.3: Representation of the valence- and conduction band for a metal-like material (a) and a semiconductor (b). Also indicated is the location of the Fermi energy E_F for both cases (adapted from [13]).

the full bands do not contribute to conduction effects. To be more precise, for every electron with a specific momentum and energy there is another electron within the same band, with the contrary momentum and the same energy. Thus, on average, no macroscopic current can flow. In order to obtain conduction, electrons need to be excited into higher states - the conduction band. For a metal-like material, electrons already occupy part of the conduction band. But since the conduction band is not fully filled in the case of metal, a current can flow at $T = 0$. For a semiconductor, an energy equal the energy gap E_g is required to achieve conductivity. The energy gap of insulators is often that large, the required energy is similar to the energy to cause a phase transition of the material.

Semiconductors per se can be further classified into direct and indirect semiconductors, whereby the difference here is given by the minimal band gap $E_G = \min(E_C - E_V)$. Figure 2.2(b), a realistic presentation of the band structure of silicon, is a good example for an indirect semiconductor. As indicated there, the minimal gap is between the valence band at the Γ- point and the X-point of the conduction band, whereby the points refer to different orientations in the reciprocal space. For the indirect transition at E_G, an additional momentum $\boldsymbol{k}$ is necessary. Contrary to this, for direct semiconductors a direct transmission is possible.

2.2 Transport Properties of Semiconductors

In order to understand the mechanism of conduction in a semiconductor, its transport properties should be discussed. Again the band structure plays an important role. As mentioned above, the Fermi energy is located between the valence and the conduction band. Therefore, the valence band is fully occupied at $T = 0\,\mathrm{K}$, while the conduction band is empty. This means the semiconductor is an insulator in first place. However, if energy is added to the system, the Fermi function $f(E, T)$, which

describes the occupation probability of the energy eigenstates, softens at the Fermi edge. In this case electrons are excited into the conduction band, where they can move freely and contribute to a current. The number of excited electrons can be described by the following equation:

$$n = \frac{1}{V} \int_{E_C}^{\infty} D_C(E) f(E,T) dE = 2\Big(\frac{m_e^* k_B T}{2\pi\hbar^2}\Big)^{3/2} \mathrm{e}^{-(E_C - E_F)/k_B T} \tag{2.7}$$

$$D_C = \frac{V}{2\pi^2}\Big(\frac{2m_e^*}{\hbar^2}\Big)^{3/2} \sqrt{E - E_C} \tag{2.8}$$

whereby D_C is the density of states of the electrons in the conduction band, assuming again the nearly free electron approximation with a parabolic eigenfunction. Further, the concept of the effective mass tensor

$$\Big(\frac{1}{m_e^*}\Big)_{ij} = \frac{1}{\hbar^2} \frac{\partial^2 E(\boldsymbol{k})}{\partial k_i \partial k_j} \tag{2.9}$$

is used, which is a mathematical way to introduce the dispersions relation or rather the interaction between the electrons and the atomic nuclei into the equation.

Due to the transition of electrons into the conduction band, empty states are left in the valence band. These empty states are called holes and act similar to the excited electrons, meaning they can move in the $\boldsymbol{k}$- space like a free electron by giving nearby electrons the possibility to change their state. The most important properties are $\boldsymbol{k}_p = -\boldsymbol{k}_n$, $E_p(\boldsymbol{k}) = -E_n(\boldsymbol{k})$ and $m_p^* = -m_n^*$. Externally, the holes appear like positively charged particles, and accordingly they tend to move towards the upper edge of the valence band and also contribute to the charge transportation. In the following holes will be referred to with the index p, while electron depending values are referred to with n.

Hence, the conductivity of the semiconductor

$$\sigma = e(n\mu_n + p\mu_p) \tag{2.10}$$

is given by the density of free electrons n and holes p multiplied by there mobility μ_i with $i = n, p$. For the latter applies $\mu_i = e\tau_i / m_i^*$, which is referable to the classic model of Drude [17]. The mobility of both, electrons and holes is defined by the relaxation time τ_i characterising the friction or in other words the mean scattering time. Typically, electrons scatter at defects or vacancies in the lattice and at phonons. Since holes are free electronic states, so their motion is in fact due to electron motion, their scattering rates are of the same order of magnitude.
In silicon the mobilities at room temperature are $\mu_n(Si) = 1900\,\mathrm{cm}^2/\mathrm{Vs}$ and $\mu_p(Si) = 480\,\mathrm{cm}^2/\mathrm{Vs}$. Other semiconductors, such as germanium, have significantly higher mobilities $\mu_n(Ge) = 3800\,\mathrm{cm}^2/\mathrm{Vs}$ and $\mu_p(Ge) = 1800\,\mathrm{cm}^2/\mathrm{Vs}$.

When the semiconductor is brought into an external electric field, the free electrons drift along the field lines. Thereby, the drift velocity $v_D = \mu\mathcal{E}$ is defined by

the mobility of the semiconductor and the external field $\mathcal{E}$. For very large electric fields the model of Drude no longer holds and the mobility becomes dependent on the electric field. The drift velocity saturates at $v_D = \mathcal{O}(10^5)\,\mathrm{m/s}$ in the case of silicon.

2.3 Doped Semiconductors

In the previous description, only intrinsic semiconductors were considered. Their conductivity is usually too low for useful electrical applications. Thus, in the following a process is described, by which the conductivity of semiconductors can be further modified, the so-called doping. During doping, dopant atoms are introduced into the lattice of the semiconductor, providing additional energy levels between valence and conduction band. These dopant atoms are typically atoms containing one electron more (donors) or one electron less (acceptors) in the outer electron shell. If, for example, a donor is introduced into the semiconductor, it has an additional electron which is not needed for covalent bonding to the neighbouring atoms. In order to release this electron from the donor and lift it as a free electron into the conduction band, a comparably small amount of energy is sufficient. As indicated in figure 2.4(a), the energy level of the donor is located closely to the edge of the conducting band. This type of doping is hereinafter referred to as n-type. Accordingly, in a p-type semiconductor, the acceptor offers a free state that can be occupied very easily by an electron, creating a hole in the valence band. In this sense, doping can be used to modify the number of free electrons or holes in the semiconductor, hence modifying its conductivity. The actual number of charge carriers is highly dependent on the temperature. At $T = 0\,\mathrm{K}$ the doping impurities have no influence, whereas this influence becomes significant when the thermal energy reaches up to E_d. Additionally, the added n- or p-typed charge carriers effect the position of the Fermi energy and shift it in the direction of the donor or respectively acceptor energy.

2.4 pn-Junction

When differently doped semiconductors are combined, a pn-junction [13, 15] is created, which will be discussed in the following. To simplify the system an abrupt transition from p- to n-type semiconductor is assumed. Despite this simplification, the upcoming chapters will show that the description of the junction is helpful to characterize different electrical components.

2.4.1 pn-Junction in Equilibrium

As mentioned above and indicated in figure 2.4, the Fermi Energy is effected by doping. Further, it should be noted that in a single system only a unified Fermi level E_F can exist.

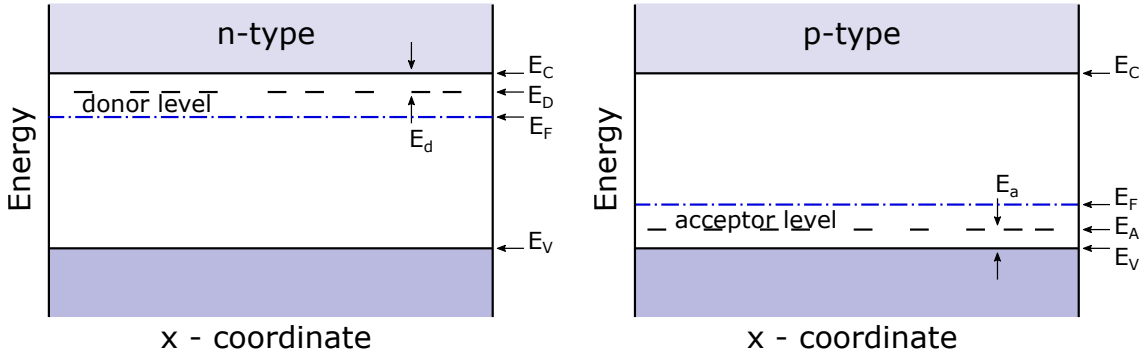


Figure 2.4: Energy levels for an n-type (a) and p-type (b) semiconductor. The Fermi Energy E_F is also represented. Here room temperature was assumed, since E_F varies with additional charge carriers and respectively the temperature (adapted from [13]).

If differently doped semiconductors are brought together, a common Fermi level must be formed. To achieve an equalization of the different Fermi energies, charge carriers diffuse from one side of the semiconductor to the other. Electrons diffuse from the n-type side, while the additional holes migrate from the p-type side. A diffusion voltage V_D is created by the dopant ions left behind, which causes the bands to bend. At the same time, the voltage results in a drift of charge carriers in the opposite direction. Both currents, diffusion- and drift current, cancel each other, forming the new equilibrium state.

The diffusion voltage can be calculated, using the concentration of donor n_D and acceptor n_A atoms:

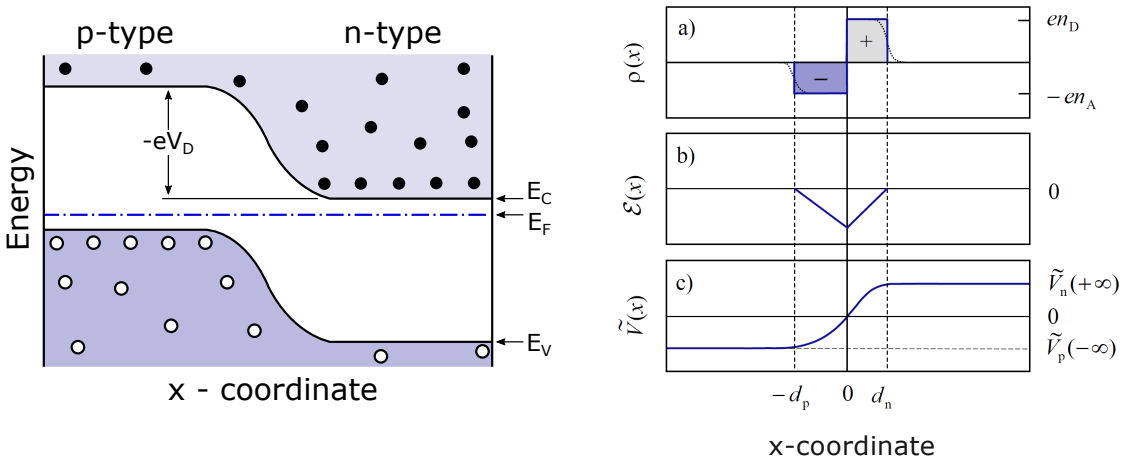
$$eV_D = k_B T \ln\left(\frac{n_D n_A}{n_i^2}\right) \tag{2.11}$$

Here, the law of mass action $n_i^2 = n \cdot p$ applies for the density of charge carriers n and p.

In the junction area, the so-called depletion zone is formed, in which the number of free charge carriers is reduced significantly. The reason is that charges which have diffused into the differently doped side recombine with the dopant atoms. Most of the other charge carriers are expelled by the prevailing electrical fields. The depth of the depletion zone in the n-type side is given by following equation:

$$d_n = \sqrt{\frac{2\epsilon_r \epsilon_0 V_D}{e} \frac{n_A/n_D}{n_A + n_D}} \tag{2.12}$$

The depth on the p-side can be derived by interchanging the concentrations of doping atoms accordingly.



(a) Schematic picture of the band structure of a pn-junction in equilibrium (adapted from [13]).

(b) Representation of the space charge $\rho(x)$ and the resulting electrical field $\mathcal{E}(x)$ and potential $\tilde{V}(x)$. Further, the depth of the depletion zone $d_{n/p}$ is indicated [13].

Figure 2.5: Pictures illustrating the principle of a pn-junction in equilibrium.

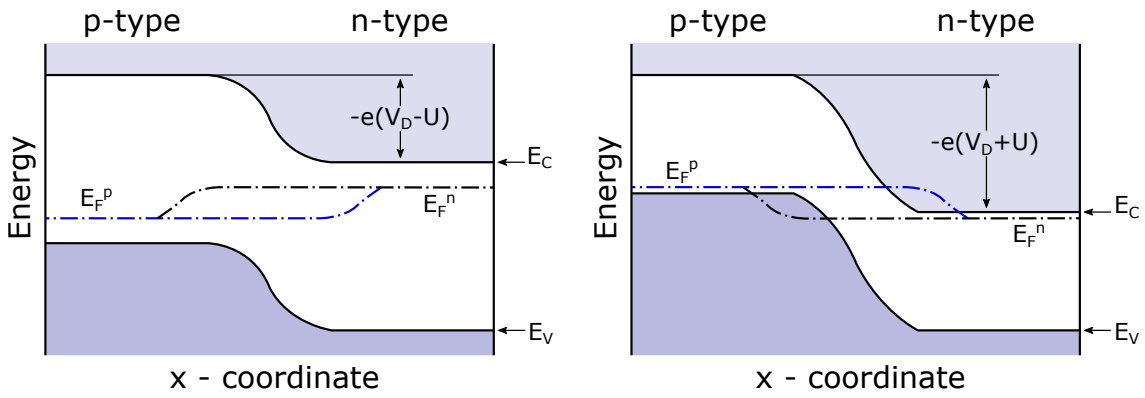
Since in this area no free carriers exist, the whole charge is dominated by the dopant atoms. The depletion zone is charged negatively in the p-doped side and positively in the n-doped side. As indicated in Figure 2.5(b) a box function is assumed. This space charge results in an electrical field $\mathcal{E}(x)$ and again in the bend potential curve $\tilde{V}(x)$, obeying the Poisson equations.

2.4.2 pn-Junction with External Fields

By applying an external voltage to the pn-junction, the state of equilibrium is disturbed, which results in new effects. Thereby, a distinction is made between the forward and the reversed bias mode.

In the forward biasing case a positive potential is connected to the p-type semiconductor. As shown in figure 2.6(a) the potential energy of the electrons on the p-side is reduced by U. Thereby, the potential step is reduced. This disturbs the equilibrium and the two currents that flow in the pn-junction no longer cancel each other out. Simultaneously, it is no longer possible to form a unified Fermi level due to the different charges in the space-charge zones. The same applies in the case of an applied reverse bias voltage (figure 2.6(b)), with the exception of an increase of the potential step.

The two types of currents within the junction are influenced by the change in band structure. As already mentioned, they are out of balance due to the bias voltage. The reason is that the diffusion current is highly dependent on the height of the



(a) Forward biasing (b) Reverse biasing

Figure 2.6: Band structure of a pn-Junction with an external voltage U (adapted from [13]).

barrier, which is reduced by $V_D - U$:

$$j_{diff}(U) \sim \mathrm{e}^{-e(V_D-U)/k_bT} \sim j_{diff}(0)\,\mathrm{e}^{eU/k_BT} \tag{2.13}$$

whereas the drift current is independent:

$$j_{drift}(U) \approx j_{drift}(0) = j_{diff}(0) \tag{2.14}$$

These equations apply likewise to electrons and holes. Hence, the total current is given by the following equation:

$$j(U) = (j^n + j^p)_{diff}(U) - (j^n + j^p)_{drift}(U) = j_0\,(\mathrm{e}^{eU/k_BT} - 1) \tag{2.15}$$

which shows the highly non-linear characteristic of a biased diode.

Furthermore, the influence of the bias voltage on the depletion zone should be discussed, whereby $U < V_D$ applies in the following. In equation 2.12 the potential V_D must be exchanged by $V_D - U$, in order to take the change of the potential step into account. Thus the equation changes as follows:

$$d_n(U) = d_n(0)\sqrt{1 - \frac{U}{V_D}}, \quad d_p(U) = d_p(0)\sqrt{1 - \frac{U}{V_D}} \tag{2.16}$$

Meaning, the depletion zone increases in the reversed biased mode with increasing voltage U.

Since in the pn-junction charge is separated by the depletion zone it can also be seen as a capacitor. Its capacity therefore has to change with the depth of the depletion zone and is dependant on the impurity density n_D.

$$C = \frac{\mathrm{d}Q}{\mathrm{d}U} = en_DA\left(\frac{\mathrm{d}}{\mathrm{d}U}\,d_n(U)\right) = \frac{A}{2}\sqrt{\frac{n_An_D}{n_A+n_D}\,\frac{2e\epsilon_r\epsilon_0}{(V_D-U)}} \tag{2.17}$$

Here A describes the cross section area of the pn-junction.

3 Transistor

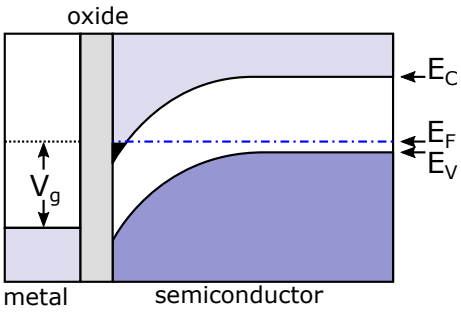
The most important technical application for semiconductors are transistors, which are essential for modern electronics. In principle, transistors can be used to control and amplify currents or voltages. In its simplest application it acts as a switch. Among others it can be distinguished between two main types of transistors, described in the following chapters.

3.1 MOSFET

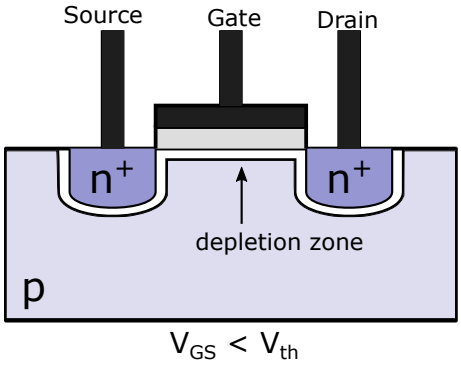
First, the metal-oxide-semiconductor field-effect transistor (MOSFET) [13, 18] will be discussed, of which the structure is slightly more complicated, but the MOSFET is much more common nowadays. The MOSFET has three main terminals and one additional terminal, with the latter one connects the bulk and is often shared between many transistors. The transistor is controlled by the main terminals, which are the gate, source and drain pin. A voltage applied to the gate controls the resistance between drain and source, and thus the current I_{DS} flowing between them. Thereby, a special feature is that the gate connection is insulated by an oxide layer, so in an ideal picture no current can flow into the gate. Meaning, the current is truly controlled by an electric field, which leads to its name.

In order to understand the functionality of the MOSFET, a stack of semiconductor, oxide and metal (see figure 3.1(a)) can be discussed, which represents the behaviour at the gate. By applying a positive potential V_g on the gate, holes from the p-typed semiconductor are repelled and the concentration of free holes is reduced. Similar to the pn-junction, the immobile dopant atoms create a space-charge zone at the boundary surface. Then, as a result, the bands are bend. With the edge of the conduction band approaching the Fermi level, the number of free electrons increase significantly. A conductive area is created, which is isolated from the rest of the semiconductor by the depletion zone. Since electrons are the main type of charge carriers here, this area is called inversion layer. For high voltages V_g the edge of the conduction band is below the Fermi energy, hence the conducting channel has metallic properties.

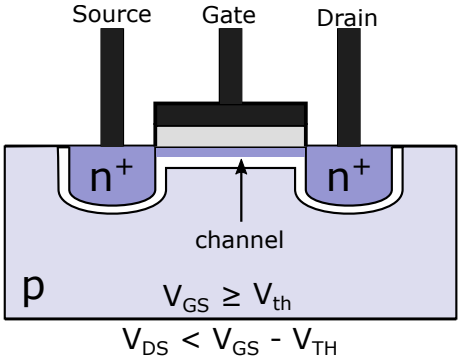
A more realistic layer structure of a MOSFET is presented in figure 3.1(b) to 3.1(d), whereby the voltage V_{GS} and V_{DS} are increased respectively. Furthermore, the pictures represent different modes, starting with the sub-threshold mode in figure 3.1(b). As the name indicates, the gate-source voltage V_{GS} is below the threshold voltage V_{th} in this mode. V_{th} marks the point, where the channel starts to form. Without the inversion layer the conductivity between drain and source pin is low



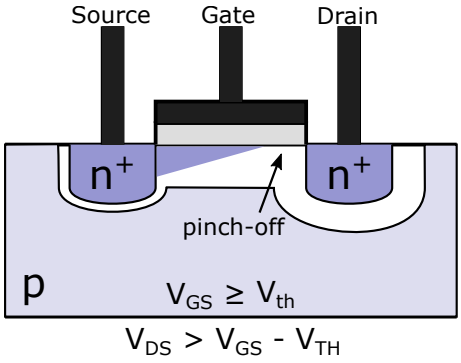
(a) Representation of the influence of a gate voltage on the semiconductor band structure. The boundary layer of metal, oxide and semiconductor is shown, with an external potential V_g applied to the metal side (adapted from [13]).



(b) Schematic layer structure of an n-type MOSFET in sub-threshold mode. For simplification source and bulk share the same potential (adapted from [19]).



(c) MOSFET in linear mode, with the n-channel displayed (adapted from [19]).



(d) MOSFET in the saturation mode after pinch-off (adapted from [19]).

Figure 3.1: Illustrations describing the functionality of the MOSFET and its different operating modes.

and only thermal excitation can cause a small current I_D [20]:

$$I_D = I_0 \mathrm{e}^{\frac{q(V_{GS}-V_{th})}{nk_BT}} \left(1 - \mathrm{e}^{-\frac{qV_{DS}}{k_BT}}\right), \quad I_0 \sim \frac{W}{L} \tag{3.1}$$

here n is the slope factor and is given by $n = 1 + C_{dep}/C_{Ox}$, with the capacity of the depletion zone and the oxide respectively. The parameter W and L describe the width and the length of the transistor gate. Using equation 3.1 another important parameter can be derived:

$$g_m = \frac{\partial I_D}{\partial V_g} \tag{3.2}$$

The transconductance g_m is defined as an inverse resistivity and is often used to describe the current gain of a MOS-transistor.

Increasing the gate-source voltage above the threshold voltage, the MOSFET enters the linear mode, as long $V_{DS} < V_{GS} - V_{th}$ holds. The transistor acts as a voltage controlled resistor in this region, resulting in a linear dependency between I_D and V_{GS}:

$$I_D = \mu_n C_{Ox} \frac{W}{L} \left((V_{GS} - V_{th})V_{DS} - \frac{{V_{DS}}^2}{2}\right) \tag{3.3}$$

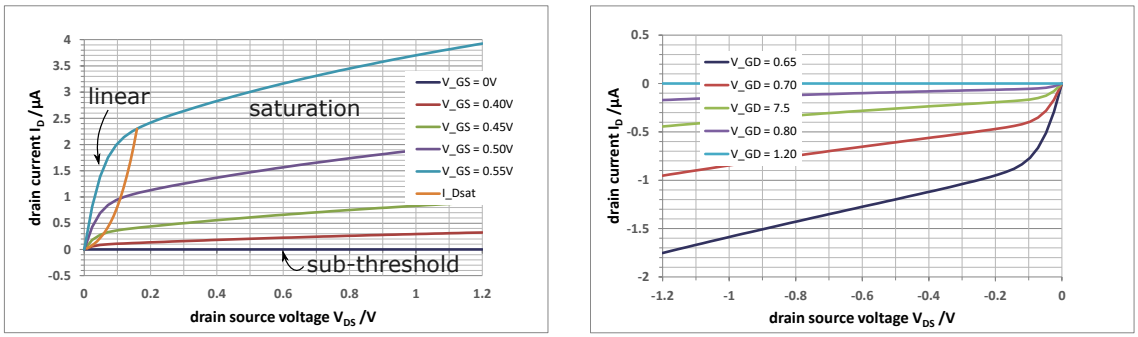
With μ_n being the mobility of electrons.

The last mode, the saturation mode, is reached when $V_{DS} > V_{GS} - V_{th}$ applies, which implies an increase in voltage V_{DS}. Thereby, the voltage between gate and substrate is decreased in the vicinity of the drain connection, to the extent that the threshold voltage is undercut and accordingly the inversion layer is removed. This effect is called pinch-off and is visible in figure 3.1(d). Due to the pinch-off the current cannot rise further. Ideally the current should be independent from V_{DS}, however, with short transistors the channel lengths modulation becomes effective. With increasing V_{DS} the point of the pinch-off is shifted to the source, reducing the effective length of the channel. Thus, the resistance of the channel drops and the current can increase. Thereby, the area between the remaining channel and the drain is depleted. Meaning there are no free charge carriers available , but the large electric fields enable a drift of electrons. The channel length modulation is considered in the equation by λ:

$$I_D = \frac{\mu_n C_{Ox}}{2} \frac{W}{L} (V_{GS} - V_{th})^2 (1 + \lambda V_{DS}) \tag{3.4}$$

The total current characteristic, including all modes, is represented in figure 3.2(a)

Up to this point, a p-typed substrate was assumed, resulting in an enriched electron conducting channel. These type of transistor is called n-type metal-oxide-semiconductor (NMOS). Accordingly there is also a p-type metal-oxide-semiconductor (PMOS), which conducts holes in the channel. In order to obtain a PMOS, the structure from figure 3.1(b) needs to be inverted, meaning the strong doped n^+ regions



(a) (b)

Figure 3.2: Current characteristic of an NMOS (a) and PMOS (b) transistor with indicated operation modes. Both transistors have a size of $L = 130\,\text{nm}$ and $W = 150\,\text{nm}$.

must be p^+ and the p-type substrate must be n-typed. Besides, the functionality remains the same as with NMOS, only the changed sign of the charge carriers must be taken into account. The current characteristic of the PMOS is shown in figure 3.2(b).

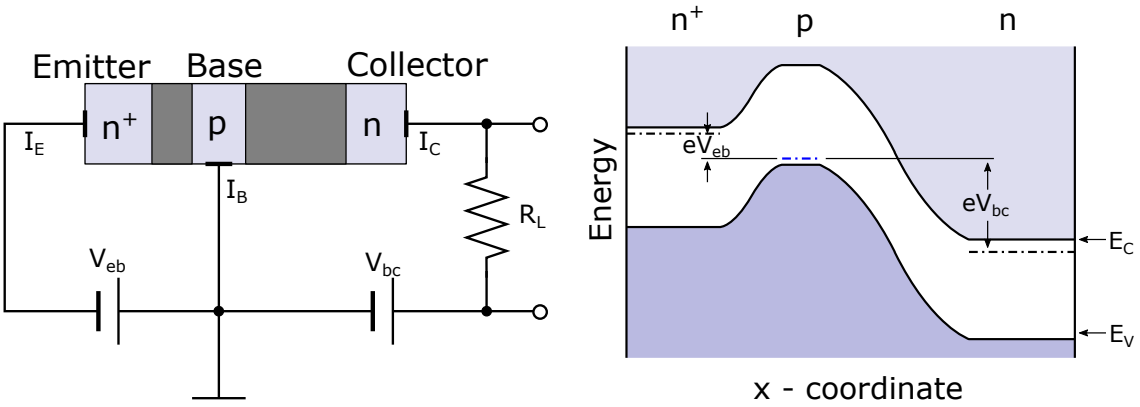
The combination of both types of MOSFET, p- and n-channel device, is called Complementary Metal Oxide Semiconductor (CMOS) topology. In the digital technology, both transistor form an input gate. Depending on a high or low voltage input either one of the transistors is turned of, due to the complementary behaviour of the transistor types. This results in the great advantage of CMOS logic that a current can only flow, while the state is switched. Hence, the power consumption is reduced significantly.

3.2 Bipolar Junction Transistor

A second type of transistor, which will be discussed in the following, is the bipolar junction transistor (BJT) [13, 15, 20]. In principle the bipolar transistor is a combination of two back-to-back pn-junction. Accordingly, there are pnp- and npn-transistors, while in the following the npn-typed bipolar transistor is discussed. Certainly, all results and properties also hold for the pnp-transistor, if the change of main charge carriers is taken into account.

The BJT has three terminals, namely the collector, emitter and base. But in contrast to the MOSFET, the bipolar transistor uses a control current applied to the base to regulate a current between collector and emitter. Another difference to the MOSFET is that here both types of charge carriers contribute to the current transport. Nevertheless, in the npn-BJT electrons are the predominant charge carriers.

The bipolar transistor works as follows: The np-junction from emitter to the base is biased by V_{be} in forward direction, as it is indicated in figure 3.3(a). Thus,



(a) Schematic structure of a BJT with an interconnection plan. Indicated is also the depletion zones between the different semiconductor areas (adapted from [15]).

(b) Band structure of the BJT with connected bias voltage V_{eb} and V_{bc} (adapted from [15]).

Figure 3.3: Illustrations, which describe the function principle of the BJT.

electrons can move from the emitter into the base. The emitter is n^+ doped to further increase the number of free electrons. Electrons, which reached the area of the base can recombine with the large number of holes located in the p-doped domain. To operate the transistor efficiently, it is therefore necessary to keep the width of the base as small as possible, at least in the order of the recombination length. The adjacent pn-junction, base to collector, is reverse biased by the V_{bc} voltage. If electrons that diffuse in the base region randomly enter the extended depletion zone of this junction, they are drawn off towards the collector by the strong electric fields. This behaviour can be traced back easily with the help of Figure 3.3(b). Furthermore, the figure illustrates that the resulting current does not significantly depend on the voltage V_{bc}. In contrast, the voltage V_{eb} has a direct influence on the number of electrons present and thus on the current.

For the BJT again different operation modi can be distinguished: Starting with the cut-off region, which is reached when for the potentials at the terminals $V_b < V_e$ and $V_b < V_c$ holds. In an ideal picture, in this region no current I_C should flow, since both junctions are reverse biased. In contrast, equation 2.15 indicates that a junction with reverse bias still transmits a latching current $I \sim -j_0$. Nevertheless, for most applications the current I_C is negligibly small in this working mode.

In the case of $V_e < V_b < V_c$, the forward active mode is reached. This mode was used to explain the function principle of the BJT above. Here, the collector current can be approximated by the characteristic linear current amplification $I_C = \beta I_B$. Additionally, $I_E \approx I_C$ can be assumed in good approximation. In order to be able to compare the bipolar transistor to the MOSFET, it is helpful to define its transconductance g_m.

$$g_m = \frac{\partial I_C}{\partial V_{be}} \tag{3.5}$$

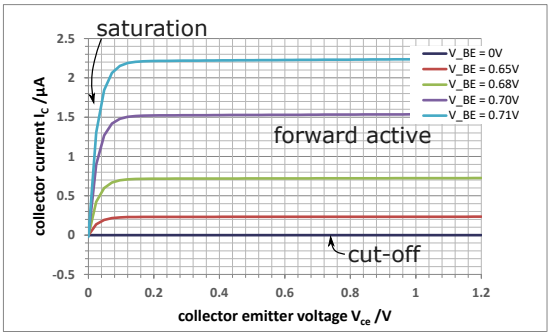


Figure 3.4: Current characteristic of a bipolar transistor with indicated operation modes.

In the saturation mode, with $V_b > V_e$ and $V_b > V_c$, both junctions are active. The region of the base is saturated with more electrons than needed for the current transmission. Thus, the current I_C becomes independent of I_B.

The total current curve, for all operation points (shown in figure 3.4), is described by the Ebers-Moll equation[21]:

$$I_E = I_S \left[\left(\mathrm{e}^{\frac{eV_{be}}{k_B T}} - \mathrm{e}^{\frac{eV_{bc}}{k_B T}} \right) + \frac{1}{\beta} \left(\mathrm{e}^{\frac{eV_{be}}{k_B T}} - 1 \right) \right] \tag{3.6}$$

whereby I_S describes the saturated latch current.

A similar effect as the channel length modulation of the MOSFET also occurs in the BJT. The change in voltage V_{bc} between base and collector, leads to a change in the size of the depletion zone between the two regions. Hence, the width of the base is reduce with increasing V_{bc}. As mentioned above, the width of the base has a significant influence on recombination of electrons and also on the current. This effect is called Early effect [22] and can be described in the forward active region by the following approximation:

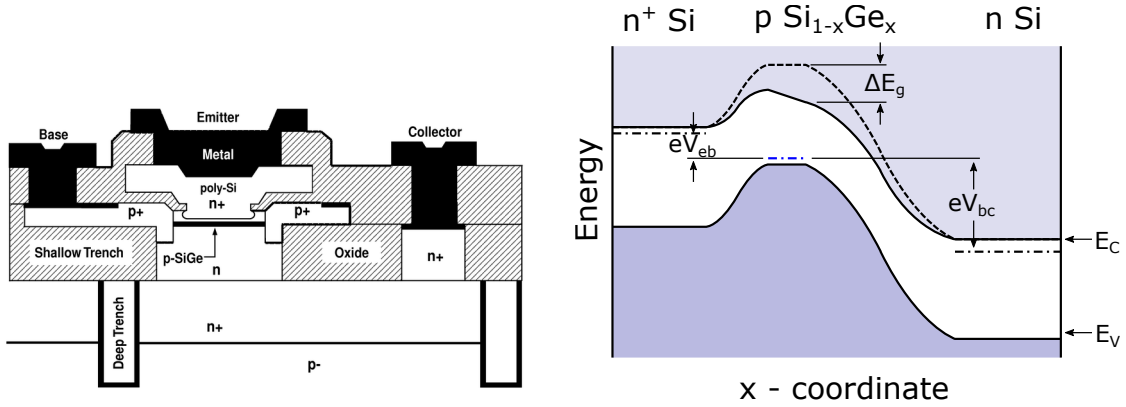
$$I_C = I_S \mathrm{e}^{\frac{eV_{be}}{k_B T}} \left(1 + \frac{V_{ce}}{V_{Early}} \right) \tag{3.7}$$

Here, the Early voltage V_{Early} is a transistor dependent factor, describing the linear dependency on V_{CE}.

The advantages of the BJT are the high current gain β and corresponding transconductance g_m, as well as their typically high cut-off frequency. The cut-off frequency is defined as the frequency, where the gain of the transistor is reduced by 3 dB. Hence, with these attributes a BJT is suitable for high frequency applications. Regardless, the development of the CMOS technology is more advanced, leading to high density circuits with less power consumption. For analogue circuits many applications combine bipolar transistors with the CMOS topology. Such technologies are called Bipolar CMOS (BiCMOS), with the goal to benefit from advantages of both types of transistors [23].

3.2.1 Heterojunction Bipolar Transistor

The heterojunction bipolar transistor (HBT) [24, 25] is a special design of a BJT that is intended to further increase the performance of the transistor. It uses the properties of heterostructures, which are layers of different semiconductors, to improve current amplification and transit time. When different semiconductors are brought together, their band structure is deformed, due to the varying energy gap of the single semiconductors. Thereby, its electrical properties and behaviour are changed.



(a) Schematic structure of a SiGe HBT [26].

(b) Band structure of an HBT. For comparison, the band structure of the BJT is indicated by a dashed line (adapted from [25]).

Figure 3.5: Illustrations describing the HBT and its features.

In the following, the special case of a silicon-germanium HBT will be picked out, although other semiconductor combinations are also possible. A typical structure of a SiGe HBT is shown in figure 3.5(a). As shown there, the base region consists of an alloy of p-doped $Si_{1-x}Ge_x$, while emitter and collector are still pure n^+- and n-typed silicon, respectively. It is particularly important for the base region that the germanium content increases towards the collector. Since the band gap of germanium is smaller than that of silicon, this gradient in concentration causes a reduction of the band gap along the base. This gradient is indicated in figure 3.5(b). It acts like a build in electrical field, changing the transport properties of the electrons in this region. In a normal BJT the electrons diffuse in the base region and enter the depletion zone between base and collector randomly. Due to the curvature of the band structure in HBT, a drift current is introduced into the base, which drives the electrons into the depletion zone. Meaning, the transit time τ_B of the electrons in the base is reduced significantly.

If again the current amplification $\beta = I_C/I_B$ is considered, it can be shown that this is proportional to τ_{Bp}/τ_t. τ_{Bp} defines the part of the base current which results from the recombination of holes in the base. The total transit time of the electrons is described by τ_t which decreases respectively with τ_B. Consequently, higher current

gains can be achieved in the HBT, as well as a higher transit frequency $f_t = 1/\tau_t$. According to its definition the transit frequency indicates the maximum switching speed of the transistor. Besides, it marks the frequency, where the current gain becomes unity.

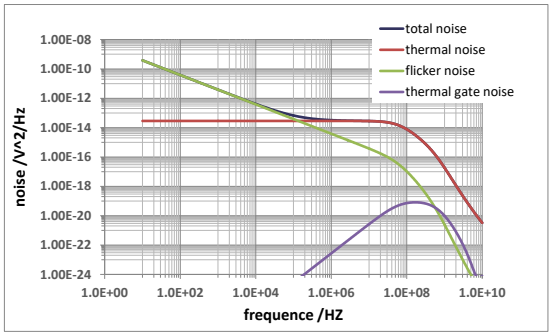
In addition, germanium has a higher mobility, thus reducing the base resistance. This allows for a much higher oscillation frequency f_{max}, because it is bound to the RC-element of base resistance R_B and base-collector capacity C_{BC} as follows [24]:

$$f_{max} \approx \sqrt{\frac{f_t}{8\pi R_B C_{BC}}} \tag{3.8}$$

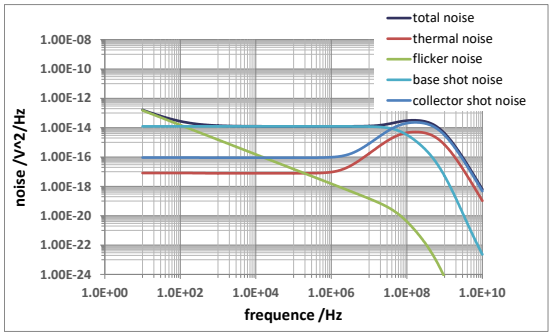
The maximum oscillation frequency is further defined by the power gain, which is unity at the point of f_{max}.

3.3 Noise Sources in a Transistor

To operate a transistor efficiently in a circuit, it is necessary to understand its noise behaviour [23, 18, 27].



(a) NMOS with minimum length $l = 130\,\text{nm}$ and width $w = 150\,\text{nm}$.



(b) Bipolar transistor.

Figure 3.6: Simulated Noise of an NMOS (a) and bipolar transistor (b). Shown are the total noise as well as different components in a range from 10 Hz to 10 GHz.

3.3.1 Noise of a MOSFET

For the MOSFET, there are mainly two types of noise, whereby thermal noise will be discussed first.

The reason for thermal noise are thermal fluctuations in the velocity distribution of the charge carriers. Since, in a simple picture, the channel of a MOS-transistor is an adjustable resistor, thermal voltage noise is a significant factor. A characteristic of thermal noise is its independence of frequency, which is the reason why it is also

called white noise. Its noise frequency spectrum can be expressed by the following equation:

$$\frac{\mathrm{d}\langle v_{therm}^2 \rangle}{\mathrm{d}f} = 4k_B T\,\gamma \frac{1}{g_m} \tag{3.9}$$

It should be noted, this equals the typical thermal noise of a resistor with the resistance γ/g_m. The additional coefficient γ takes into account that the channel thickness decreases towards the drain pin depending on V_{DS}. This indicates a change in the factor depending on the operation mode and thus channel structure. In the saturation mode typically $\gamma = 2/3$ is used.

As shown in the simulation in figure 3.6(a), the characteristic of frequency independence applies over a wide frequency range. Only the natural bandwidth of the MOSFET influences the thermal noise at $f > f_c$, where f_c is the cut-off frequency of the transistor, given by the limited transition time of the charge carriers in the channel.

The thermal fluctuation of the charge carriers in the channel also result in a noise at the gate terminal. Due to the capacitive coupling via the oxide capacitance C_{Ox}, it shows a f^2 dependency and is induced as a gate current. In the following this type of noise will be referred to as gate induced thermal noise [28].

$$\frac{\mathrm{d}\langle i_{gate}^2 \rangle}{\mathrm{d}f} \approx 4k_B T\,\gamma g_m \left(\frac{f}{f_t}\right)^2 \tag{3.10}$$

Here, the transit frequency of the MOSFET can be approximated by $f_t \approx g_m/(2\pi\, C_{GS})$ and furthermore the fraction $(f/f_t)^2$ indicates that this type of noise is only relevant for very high frequencies. The gate induced thermal noise is also presented in figure 3.6(a). However, it only makes a negligible contribution to the overall noise, since a MOS transistor is simulated with minimal size and therefore the capacitive coupling is negligible. Yet the f^2 dependency is clearly visible.

The second main type of noise in a MOSFET is the so called flicker or $1/f$ noise. It is often described by the following model: Moving charge carriers in the semiconductor can be trapped by impurities and defects in the atomic lattice. While bound, the shortly fixed charge influences the potential and thereby the behaviour of the transistor. Typically, according to a characteristic time constant τ, the charge is released again. The release time strongly depends on the energy of the binding state. This noise source is particularly prominent in MOSFETs because the conducting channel is located directly at the boundary between semiconductor and oxide. Accordingly, electrons are trapped in the defects that are often present there, since in general lattices at interfaces are less impeccable. The superposition of a large number of traps with different τ causes the $1/f$ noise, whose spectral noise distribution, described as gate noise voltage, is given by:

$$\frac{\mathrm{d}\langle v_{1/f}^2 \rangle}{\mathrm{d}f} = \frac{K_f}{C_{Ox} W L}\,\frac{1}{f} \tag{3.11}$$

Whereby the parameter K_f is a production process related constant.

3.3.2 Noise of a BJT

The noise sources in a BJT differ from those in a MOSFET. Although both flicker and thermal noise occur in almost every component, they play a minor role in bipolar transistors, as it is indicated in figure 3.6(b). Thermal noise is always associated with a resistivity, which in the case of the BJT would be the resistance of the base. This resistance is typically very low and even lower in HBTs. As described in the previous section, flicker noise is pronounced in the MOSFET, because the current flow is located at the boundary between oxide and semiconductor. This is not the case for the BJT nor the HBT. In both cases the current flows inside the semiconducting substrate, where less traps are located.

A major type of noise sources in the BJT is the so-called shot noise. While equation 3.6 describes an average current through the transistor, the actual process in which electrons overcome the pn-junction occurs randomly. This results in fluctuations in the current flow. Shot noise thus describes the discrete nature of individual charge carriers and always occurs when these have to pass a potential barrier. In the BJT one can distinguish between two types of shot noise. The one, in the case of a npn-structure, which is associated with electrons passing the emitter-base junction generating the collector current [29]:

$$\frac{\mathrm{d}\langle v_{shC}^2 \rangle}{\mathrm{d}f} = 2qI_C \left(\frac{R_B}{\beta} + \frac{k_B T}{qI_C} \right)^2 \tag{3.12}$$

and the one, which originating from holes passing from the base to the emitter above the same junction:

$$\frac{\mathrm{d}\langle v_{shB}^2 \rangle}{\mathrm{d}f} = 2qI_B R_B^2 \tag{3.13}$$

Accordingly, shot noise is a source of white noise, whereby equations 3.12 and 3.13 only hold below the cut-off frequency.

4 Electrical Devices and Circuits

In the following chapter, the individual circuits of the front-end electronics of an HV-MAPS are described in sequence. Starting with the charge collection in the sensor diode and the core part, the charge sensitive amplifier. From there, along the signal line, the source follower and the comparator are described. These circuits together form the pixel front end electronics.

4.1 Diode as a Sensor

One application of the diode can also be found outside of its typical electrical usage. In the scope of particle physics and consequently also the Mu3e experiment, the reversed biased diode is frequently used for particle detection [30]. The operation principle is based on the interaction between charged particles, whereby here only the detection of electrons and positrons is considered.

An electron or positron passing material will deposit part of its energy on the way through, due to the electromagnetic interaction between the particle and the material. The leading contributor to the energy loss is ionization. Thereby, the mean energy loss can be described by the Berger-Seltzer-formula [31]:

$$\left\langle \frac{dE}{dx} \right\rangle = \rho \frac{0.13536}{\beta^2} \frac{Z}{A} \cdot \left(B_0(T) - \log(\frac{I}{m_e c^2}) - \delta \right) \tag{4.1}$$

Here $B_0(T)$ describes the stopping power, dependent on the material and momentum and I the material dependent mean excitation energy. Moreover, there are the material density ρ, the ratio of protons and nucleons Z/A and the density correction δ.

In a semiconductor the deposited energy creates electron-hole-pairs, whereby the average activation energy for a pair is approximately 3.6 eV [32] in silicon. Accordingly, electron hole pairs will be created along the path of the traversing particle. Due to the electric field present in the reversed biased diode, the pair get separated and the single charge carriers drift along the field lines. Hence, holes drift to the p-side and electrons to the n-side of the junction. As the electric field is only present inside the depletion zone, only charge inside this volume is collected. Pairs created in the remaining not depleted substrate are not separated and most probably recombine after a specific decay time.

It should be noted that the charge is mirrored in the electrodes. Therefore, the collection starts as soon as the pairs are separated and continues till all charge carriers reached the electrodes. The mean time of this process can be calculated,

using again the Drude model:

$$v_D = \mu \mathcal{E} \tag{4.2}$$

as discussed in section 2.2. Thus, the collection time of holes is larger, as their mobility μ is smaller.

4.2 Charge Sensitive Amplifier

The most important circuity, in the scope of this thesis is the amplifier, or to be more precise, the charge sensitive amplifier (CSA) [23, 18, 27]. Later this circuit will be optimised in terms of the rise time (defined in section 6.4) of an amplified signal and its noise behaviour. Therefore, it is crucial to understand its operation principal and the expected noise.

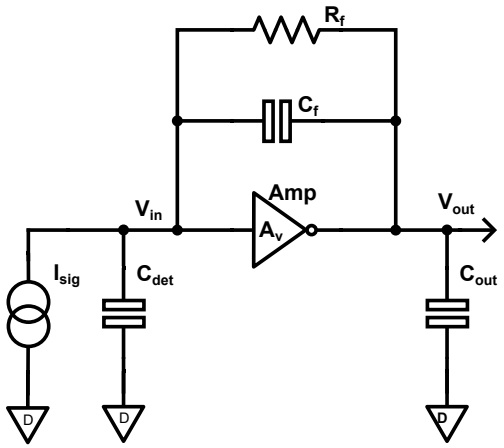


Figure 4.1: Principle of a charge sensitive amplifier.

In general, an amplifier is defined by its open loop gain A_v, which is the pure gain, given by the ratio of the output and the input voltage. In order to influence the voltage gain, typically a feedback circuit is introduced. It connects the output and input node of the amplifier. In this case the gain is referred to as closed loop gain. By using a capacitance as a feedback, the input current is integrated, forming a CSA.

A basic schema of the CSA is shown in figure 4.1. Starting within an ideal picture, the CSA consists of an ideal amplifier with an infinite inverse voltage gain $-A_v$ and an infinite input impedance. In addition, a feedback capacity C_f and a feedback resistance R_f are used. The purpose of R_f is to bring the circuit to its operation point. Hence, the feedback resistance can be assumed to be very high and it does not contribute to the rising edge of a signal. But the falling edge is shaped by the resistance to decay with the time constant $\tau = C_f R_f$. On the input side of the CSA an ideal current source and a capacity C_{det} are connected. These represent a signal

coming from a sensor diode. C_{out} is a load capacity, modelling the circuit connected subsequently.

For an infinite voltage gain A_v the input node of the CSA is a virtual ground, meaning the voltage is fixed by the amplifier and can not change. Charge Q_s, induced by the current source, can therefore only flow to the feedback capacity C_f. There the charge causes a voltage drop and defines the output voltage $V_{out} = -Q_s/C_f$ accordingly.

For a finite voltage gain $A_v = V_{out}/V_{in}$, the assumption of a virtual ground is not accurate. In this case an effective input capacitance $C_{eff} = (1 + A_v)C_f$ can be defined. The virtual increase of the feedback capacity is called the Miller effect[33] and occurs because on the side of the output node, C_f sees a large signal with inverted phase. The signal charge is distributed between the effective capacitance and all other capacitances $C_{in} = C_{det} + C_{amp,in} + C_{para}$ at the input side of the amplifier. These are the capacity of the sensor diode C_{det}, the real input capacity of the amplifier $C_{amp,in}$ and possible parasitic capacities C_{para}. Hence, the signal amplitude at the output is reduced according to [27]:

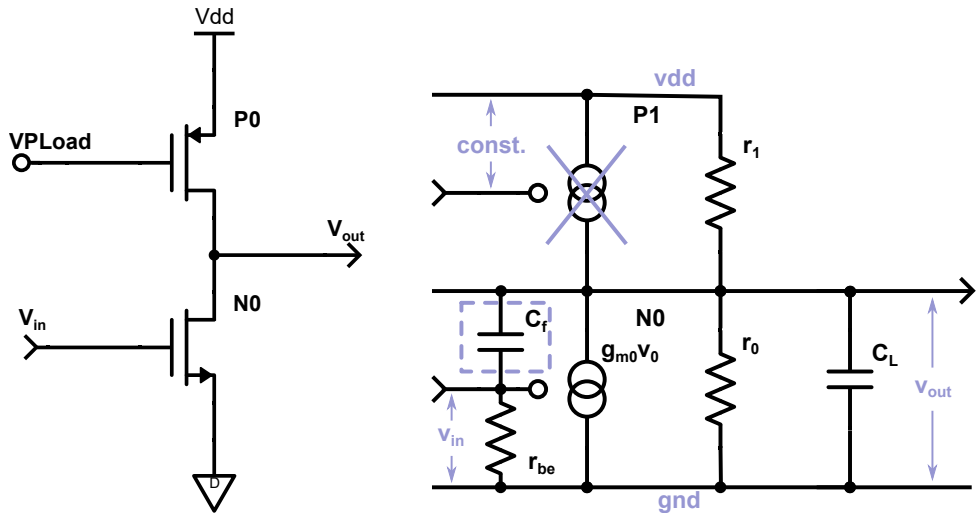
$$V_{out} = -\frac{Q_s}{C_f} \frac{1}{1 + 1/A_v + C_{in}/(A_v C_f)} \tag{4.3}$$

4.2.1 Single Gain Stage

Previously the amplifier was considered a black box with a certain gain. In the following, actual amplifier circuits are discussed, starting with the single gain stage [34]. This type of amplifier is the simplest design, consisting of only two transistors, as shown in figure 4.2(a). In this case the NMOS transistor is the input transistor, connected in the common-source configuration and amplifying the signal accordingly to its transconductance g_m. The second transistor, a PMOS, acts as a current source for the circuit. There are several options to build up a single gain stage: Also possible is to use an npn-BJT as an input transistor or exchange the functionalities of the N- and PMOS in the circuit.

In order to understand the behaviour of the circuit, it is helpful to investigate its small signal model. For this purpose, the operation point of the circuit is fixed and only small variations in voltage v and current i are observed. In this approximation the electrical components of the circuit can be expressed linearly. Hence, a transistor is represented by a voltage controlled ideal current source and the transistor's output resistance. Accordingly, for the MOSFET applies $i_d = g_m v_{gs}$ and the channel resistance is given by r_{ds}. Similar holds for a BJT, with a current $i_c = g_m v_{be}$ and resistance r_{ce}. Additionally, the base current needs to be taken into account in the case of the BJT. Therefore, an additional resistance r_{be} is used.

In the following, all equations, if not marked otherwise, are given independently of the type of transistor used. For this reason, output resistance and current are labelled r_n and $i_n = g_{mn} v_n$ respectively, where n is given by the transistor number in the corresponding circuit.



(a) (b)

Figure 4.2: Represented is the circuit of a single gain stage with an NMOS input (a) and the corresponding small signal model (b). Additionally, a load capacity C_L, a feedback capacity C_f and the base resistance r_{be} are included in (b). Latter only exists in the case of a BJT input.

The small signal model for a single gain stage is represented in figure 4.2(b), including a load capacity C_L. As indicated there, the current source representing $P0$ can be neglected, because its gate-source voltage does not change in the small signal picture.

By using Kirchhoff's first law the circuit, initially without the feedback capacity C_f, in figure 4.2(b) can be solved for its transfer function [34]:

$$H(s) = \frac{V_{out}}{V_{in}} = -\frac{g_{m0} r_0}{1 + s\, r_0 C_L}, \quad \text{for } r_1 >> r_0 \tag{4.4}$$

Whereby $s = \mathrm{i}\omega$ is the complex frequency, with the imaginary unit i and the angular frequency ω. The transfer function models the output V_{out} of the amplifier for all possible input signals V_{in}. Based on the transfer function the dc-gain $H(0) = A_v = -g_{m0} r_0$, the cut-off frequency $\omega_t = 1/(C_L r_0)$ and the unity gain bandwidth $GBW = A_v \omega_t = g_{m0}/C_L$ can be derived. Latter is the product of the gain and the bandwidth, describing the point at which the gain of the amplifier equals one. Meaning it is an indicator for the maximum speed of the amplifier. In this sense, it is comparable to the transit frequency of a single transistor.

Assuming the input transistor is operated in the saturation mode, its transconductance is given by

$$g_{m,NMOS} \approx \sqrt{2 \mu_n C_{ox}\, I_D W/L} \tag{4.5}$$

for the case of an NMOS and

$$g_{m,BJT} = q I_C / k_B T \tag{4.6}$$

for a BJT input. Hence, g_m scales in both cases with the current inside the transistor, while the BJT benefits more from an increasing current. This behaviour also reflects in the unity gain bandwidth. On the other hand the output resistance is in both cases proportional to $r_n \sim 1/I$. Accordingly, the dc-gain does not benefit from a larger current, in the case of a MOSFET it will even decrease.

In order to operate the single gain stage as a CSA, a feedback capacity is necessary. Therefore, the influence of this capacity can be studied, using the small signal model with additional C_f. For the transfer function then applies:

$$H(s) = \frac{sC_f r_0 - g_{m0} r_0}{1 + s\, r_0 (C_L + C_f)}, \quad \text{for } r_1 >> r_0 \tag{4.7}$$

Hence, the feedback capacity reduces the unity gain bandwidth $GBW = g_{m0}/(C_L + C_f)$. Note that C_f is increased by the Miller effect, as discussed above, and is replaced by the effective capacity $C_{eff} = (1 + A_v)C_f$. This results in a significant limitation at high frequencies.

4.2.2 Cascode Amplifier

The cascode [34] is a concept to increase the gain of the amplifier further. Thereto, a second transistor is introduced between the output node and the input transistor, as shown in figure 4.3(a). Again it is possible to use various types of transistors to build up the cascode circuit.

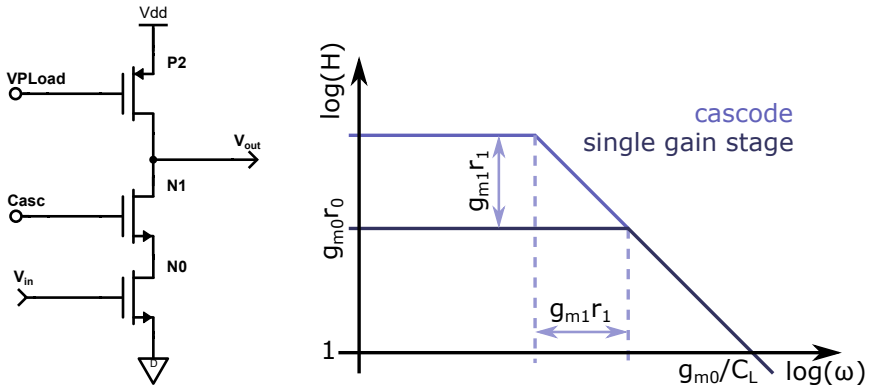


Figure 4.3: Shown is the circuit of a single gain stage with additional cascode transistor (a). On the right side (b), the influence of the cascode on the transfer function is indicated [34].

The cascode adds mainly two advantages to the circuit. First, it acts as a second stage of amplification, which can be proven by investigating the small signal behaviour. The transfer function is given by [34]:

$$H(s) \approx -\frac{g_{m0} r_0 \times g_{m1} r_1}{1 + s\, r_0 g_{m1} r_1 C_L}, \quad \text{for } g_{mn} r_n >> 1 \tag{4.8}$$

Whereby variables marked with the index 1 indicate the cascode transistor.

For the dc-gain, equation 4.8 yields to $A_v = g_{m0} r_{ds0} \times g_{m1} r_{ds1}$, which is a significant increase in comparison to the single gain stage. However, the unity gain bandwidth is not influenced by the additional transistor, which can be seen in figure 4.3(b).

The cascode adds a second advantage, not covered by the small signal model. The cascode transistor fixes the drain voltage of the input transistor to a nearly constant value, making it independent from changes at the output node. Hence, parasitic feedback capacities in a real input transistor are not affected by the Miller effect. This is unfortunately not true for the feedback capacity in the CSA, which is still located between the input and output node.

4.3 Source Follower

The output impedance of the CSA is approximately given by the resistance of the current source, which is typically very high. Therefore, the CSA is suited best to drive high impedance loads, in order to minimize power loss. For the case of a small load impedance, thus an additional circuit is necessary to match the impedance between CSA and its load. For this purpose, the so-called source follower [34, 35] circuit is suitable. Its consist of two transistors, whereby one is operated in the common-drain configuration and the second acts as a current source. The related circuit of the source follower is represented in figure 4.4(a).

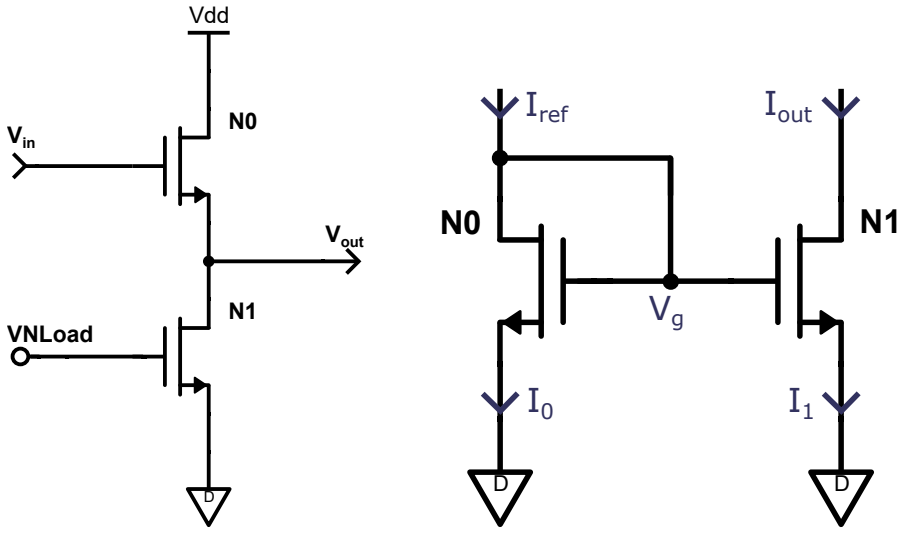


Figure 4.4: Display of a source follower (a) and a current mirror (b) circuit. Both circuits are build up with NMOS transistors.

Essential parameters of the circuit can be calculated within the approximation of the small signal model. The output impedance is given by $r_{out} = r_1/(g_{m0} r_1 + 1) \rightarrow 1/g_{m0}$ for $g_{m0} r_1 >> 1$. Typically, the transistor $N0$ is operated in the saturation

region, leading to the equation $g_m \sim \sqrt{I_D W/L}$. Hence, the output impedance can be adjusted by varying the size of the input transistor and the current flowing inside. Besides, the input is a MOSFET, so in first approximation no input current flows and the input impedance approach infinity. The case of a bipolar transistor as an input will be not discussed.

The dc-gain of the circuit can be again derived from the transfer function:

$$H(s) \approx \frac{g_{m0} r_0}{1 + g_{m0} r_0 + s r_0 C_L}, \quad \text{for } r_1 >> r_0 \tag{4.9}$$

It is $A_v = \frac{g_{m0} r_0}{1+g_{m0} r_0} \to 1$ for $g_{m0} r_0 >> 1$. In summary, the circuit is used to transfer the impedance from a high to a low value without internal loss. Of course this is only true in the first approximation. Effects such as the Body effect [35] cause the gain to fall below unity, leading to losses. The Body effect always occurs when there is a difference between the source and the substrate potential. It results in a shifting of the threshold voltage of the transistor, due to the increase of the depletion zone at the source terminal with increasing source-substrate voltage. In the source follower this effect occurs when a signal is transferred, yield to a non-linearity in the signal response of the circuit.

4.4 Current Mirror

The current mirror is not a circuit for processing a signal but an assistance, used inside other circuits. Nevertheless, it will be explained shortly. Like the name already indicates, the purpose is to mirror a current from one branch into another. In that sense, it is often used to distribute supply currents without power losses.

On the reference side of the circuit a diode connected transistor is used, as it is shown in figure 4.4(b). By connecting the gate and the drain terminal of the transistor, it is fixed in saturation mode, since $V_{DS} > V_{GS} - V_{th}$ is always fulfilled for $V_{DS} = V_{GS}$. Additionally, V_{GS} will automatically adjust such that the reference current I_{ref} can flow through the transistor, due to the diode connection.

The gate potentials of the transistors in the reference and the working branch are shared. Hence both transistors are operated with the same voltage V_{GS} and their current is given by equation 3.4. Thus, the ratio of the currents I_1/I_0 in both branches is determined by the ratios of the lengths L_0/L_1 and the widths W_1/W_0, whereby differences occur only because of the channel length modulation.

4.5 Comparator

An application example for the current mirror within another circuit is the comparator [34]. It is essential circuitry to convert analogue signals into digital signals, by comparing two input signals.

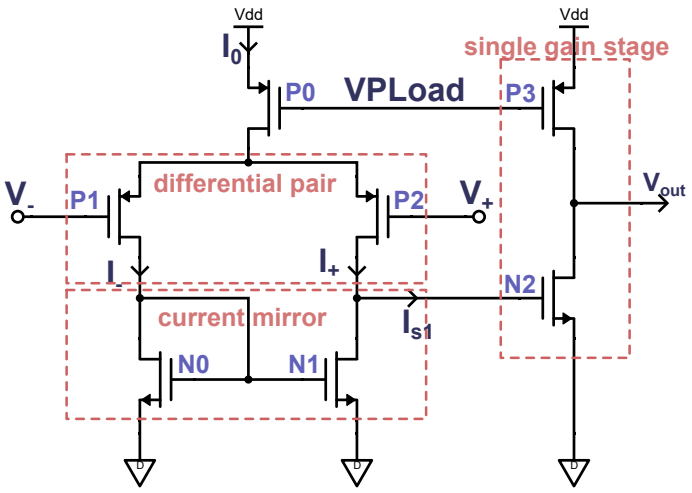


Figure 4.5: Circuit diagram of a comparator, marked are the different parts and bias voltages.

A comparator circuit is shown in figure 4.5, consisting of a differential pair, a current mirror and a single gain stage as output stage. Also possible is a source follower as output stage. The input signals are V_+ and V_-, connected to the differential pair.

To start with, equal input potentials $V_+ = V_-$ are assumed. Both transistors of the differential pair see the same gate-source voltage and therefore conduct the same current. In this case the equation $I_+ = I_- = I_0/2$ holds, whereby I_0 is determined by $P0$, which acts as a current source. The output current of the fist stage, flowing into the gain stage is $I_{s1} = I_+ - I_- = 0$, since the current mirror only allows for a current flow of I_+ in the output branch. The potential at this fist stage output node depends on the actual voltage level of V_+ and V_-. The comparator can only work if the input voltage is sufficiently low, otherwise the differential pair will switch-off. Similar, it hast to be sufficiently high, hence the transistors are not saturated.

A second case applies for $V_+ < V_-$. The current flowing through transistor $P2$ increases with decreasing input voltage V_+, while the current I_- must drop by the same amount. Based on this behaviour the circuit is often also called differential amplifier. In the limiting case of a large difference between the input potentials, the whole current I_0 flows through transistor $P2$. Again the output current of the first stage is given by $I_{s1} = I_+ - I_- > 0$. Therefore, the gate capacity of the gain stage is charged and the voltage at this node rises. Due to the increase of the drain-source voltage of transistor $N1$, its drain current starts to increase. Hence, an equilibrium state is reached after a short time. Simultaneously, the output voltage V_{out} drops close to zero, according to the negative amplification of the single gain stage.

A similar behaviour is observed for $V_+ > V_-$, only with swapped currents, so that the output stage capacity is discharged. Accordingly, the output voltage rises.

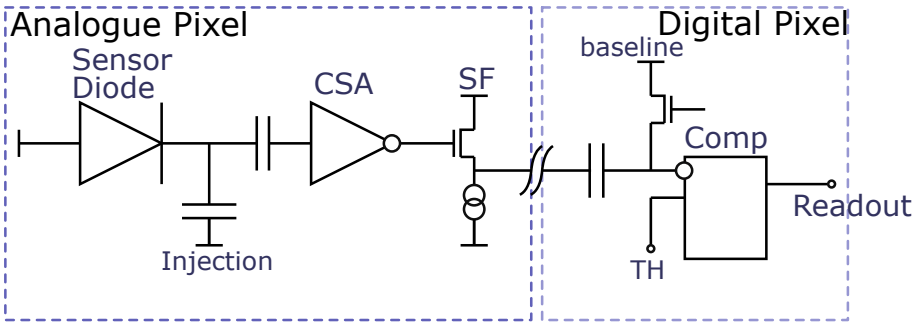


Figure 4.6: Simplified schematic representation of the pixel front end electronics.

4.6 Pixel Front End Electronics

The pixel front end electronics (FE) covers the whole analogue electronics up to the digitisation of the signal. Figure 4.6 shows this part of the readout chain. As indicated there, the schematic is divided in two parts, an analogue and a digital pixel. The analogue pixel consists mainly of the sensor diode, which is implemented as a deep n-well in the p-typed substrate. Accordingly, the diode is located vertically inside the chip. Thereby, the substrate is biased with a negative high voltage, to achieve the charge collection via drift, as described previously.

The remaining electronics of the analogue part are located above the diode, inside the n-well. Thereby, the CSA is the most important device, as it amplifies the signal emerging in the diode and makes it measurable. Furthermore, there is a possibility to inject a test signal into the FE. The output of the analogue pixel is driven by a source follower.

In a full sensor chip many analogue pixels will be combined in a large matrix, giving additional spacial information. The matrix defines the active area of the chip, where traversing particles can be detected. Outside the pixel matrix, the digital pixels are located, with each analogue pixel being assigned to a corresponding digital counterpart. A long line connects each analogue pixel with a digital one with a point-to-point connection, acting as a low pass, due to its parasitic capacity.

For the digital pixel a minimised version is used. Only the comparator is included, which compares the signal to a threshold voltage V_{Th}. Every time the signal exceeds the threshold voltage, a logical one is produced at the output of the comparator. Normally this information will be processed further by the whole readout chain, but this part is not considered in this thesis.

5 Simulation tools

Simulating a circuit is an important tool to understand its behaviour, to optimise it and to determine essential parameters. Therefore, the following chapter is dedicated to introduce the used simulation tools. In the framework of this thesis mainly programmes, which belong to the Cadence® Virtuoso® family were used. Further, it is necessary to connect these tools with a design kit, providing information about the used process.

5.1 Cadence

Cadence Design System Inc. is an American producer for electronic design automatic (EDA) software. Among other products, their focus is a supportive software for the development of chip level analogue and digital circuits, which is called Cadence® Virtuoso®. It is a software package, including several sub-programmes, each optimised for a specific area in the design flow. Additionally, Cadence provides an interface to include libraries from various manufactures of semiconductor elements and integrated circuits (ICs). Thus, appropriate design rules, process dependent parameters and other information can already be taken into account during development. In the following different, sub-programmes and important features are presented.

5.1.1 Schematic Editor

The Virtuoso® Schematic Editor [36] is the core of the circuit design flow, as most of the other tools are centred around it an can be activated from it.

However, basically it is used to define and design circuits. For this purpose, various libraries with different components can be accessed from here. Cadence® also provides a library with ideal standard components, like voltage or current sources, helping to verify a circuit in an ideal environment. This makes it possible to define a workbench with variable input parameters and to use it to examine different components. Here, the hierarchically structured design flow plays an important role, possible due to the integration and storage of all circuits in library form.

5.1.2 Layout Suite

After the schematic is defined, it is necessary to create a design describing how the circuit should be physically implemented on a silicon wafer. Therefore, the Virtuoso® Layout Suite [37] is used. It allows for a multi-layer design, whereby

each layer corresponds to a footprint of a physical layer like a metal layer or a specific type of doping.

Additionally, the Layout Suite offers multiple tools like the Design Rule Check (DRC), which automatically compares a design with a set of rules, given by the manufacturer. This includes for example distances between traces or minimal sizes of shapes, in order to ensure the layout is producible. Another tool is the Layout vs. Schematic (LVS) test, which compares the layout with the circuit defined in the schematic. Thereby, it searches for mismatches in device parameters or their connectivity among themselves.

Translating the ideal schematic into a more real image in the layout leads to further parasitic components. These can be, for example, the line resistance of a metal trace or capacitances between metal planes and the underlying substrate. The parasitic extraction tool Quantus QRC calculates these parasitic components based on the layout and adds them into a schematic. As a result, the parasitics can be taken into account as the circuit is simulated and their influence can be studied. It should be noted that parasitic components between differently doped areas can only be found if a depth profile and the density of doping atoms is known. This is highly dependent on the process and needs to be provided by the manufactures.

5.1.3 Analog Design Environment

In order to verify the performance of the designed circuit, the Virtuoso® Analog Design Environment (ADE) is used. With this tool various simulations can be

Type	Description
dc-analyses	Simulates the operation point of a circuit, depending on different parameters as the bias voltage, temperature and so on.
ac-analysis	Observes the small signal ac-behaviour to find for example the transfer function, gain and phase shift of a circuit.
transient analysis	Gives information about the time depending reaction of a circuit. Optional: inclusion of noise into the simulation.
noise analysis	Simulates the frequency dependent noise behaviour in the small signal model and lists all contributors, divided into different types of noise.
Monte-Carlo simulation	Statistic analysis of the performance of the circuit, depending on variations of parameters due to the manufacturing process. A distinction is made between global process parameters and local device parameters.

Table 5.1: Overview of the possible simulation types accessible from the ADE.

controlled and performed. Thereby, it is possible to simulate a circuit based on its schematic or its layout. In the case of the latter, parasitic components from the parasitic extraction are taken into account. An overview of some of the possible simulation types is listed in table 5.1, together with a short description of them.

Further, the ADE offers a waveform calculator, which can be used to perform measurements and calculations on the simulation data. It provides basic operation and pre-built functions, to find minima and maxima, the crossing point between two waveforms or the variance of a parameter. Additionally, user defined functions can be integrated. Functions used within this thesis are presented in section 6.4.

5.2 Design Kit

In general, a Design Kit provides information about the manufacturing process. It contains basic information, such as the layer stack and a library of all components that can be manufactured in the process. Latter links the component as circuit symbol with a representation in layout form and corresponding models for simulation. Further, design rules are defined.

5.2.1 IHP SG13S

As already introduced in chapter 1 the process SG13S [3, 38] by the institute Innovation for High Performance (IHP) Microelectronics is addressed by this thesis. The SG13S technology offers high performance BiCMOS devices. Meaning the technology allows for both types of transistors, standard MOSFETs , including n- and p-channel devices, as well as SiGe:C based HBTs. The additional use of carbon atoms helps to stabilize the SiGe area of the base for subsequent process steps, so that these can be performed at higher temperatures [39].

Type	Parameter	Value
HBT	DC-Gain β	900
	Transit Frequency f_T	240 GHz
	Oscillation Frequency f_{max}	350 GHz
	Early Voltage V_{Early}	110 V
MOSFET	Supply Voltage	1.2/3.3 V
	Minimum Channel Length	130/330 nm
	Minimum Transistor Width	150/300 nm

Table 5.2: Characteristic parameters of the transistors in the SG13S Technology [38].

Table 5.2 gives an overview of some characteristic parameters of the transistors provided by the technology. As indicated there, two types of MOSFETs with different core supply voltages exist, for low and high voltage application. For simulations the transistors are represented by the VBIC [40] (Rev.1.15) model, in the case of

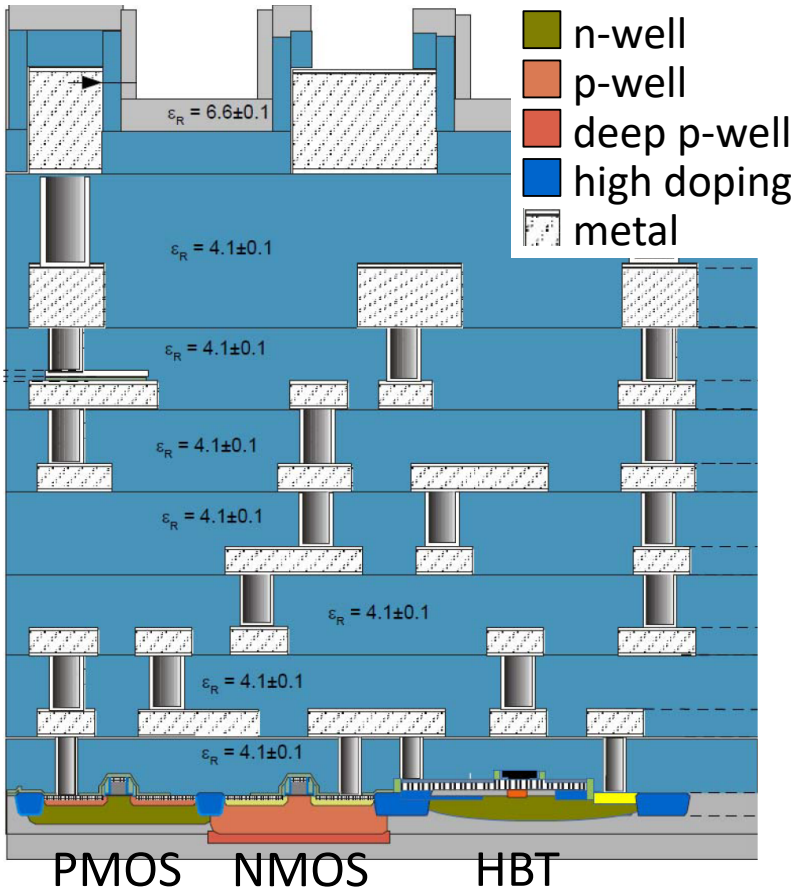


Figure 5.1: Architecture of the SG13S technology. In the silicon substrate a PMOS, an NMOS and an HBT are indicated from left to right(adapted from [43]).

the HBT and the PSP 103.1 [41] model, in the case of a MOSFET. For more information on the used simulation models and the used parameters [38] can be referred. Further, IHP validated the VBIC model for their HBT in [42]. In conclusion, the quality of the model was found to be sufficient up to the transient frequency of the transistor.

A cross section of the layer stack is shown in figure 5.1, displaying the heights of each metal layer. The technology offers 5 thin metal layers and 2 thick top metal layers.Each metal layer (grey, dashed) is made of aluminium and is isolated by an interlayer dielectric of silicon oxide (light blue) and a small nitride layer (light grey). The connection between the layers is realised by vias (grey, gradient). On the top side, the stack is enclosed by a nitride passivation (light grey), only interrupted for pads.

A special feature of this process is that it offers an additional buried n-doped array, located inside the p-doped substrate. This, so-called deep n-well is also indicated as a darker orange area below the NMOS transistor. This allows a complete isolation of the NMOS transistors from the rest of the substrate.

6 Signal Parameters

The signal quality and thus especially the time measurement of it is influenced by various effects. This includes effects as the time walk, as well as fundamental phenomena as the electric noise and the charge collection noise. The following chapter aims to describes such effects and to define parameters, providing information about the signal quality.

6.1 Time Walk

The rising edge of the signal triggers the creation of the digital hit information, the time measurement. This is commonly implemented as a comparator circuit, discriminating the signal to a certain threshold voltage, where the crossing point defines the hit-time.

The signal and the threshold voltage at the comparator input are represented in figure 6.1 and thus the principle of the time walk. As indicated, the signal amplitude is not fixed, but it depends on the deposited charge in the sensor diode. However, the rise time of the signal is independent of the deposited charge, as it only depends on the mobility of the charge carriers. Hence, the signal latency differs depending on the signal amplitude. Meaning for a signal emerging at the same time, the hit time will decrease in time for a larger amplitude. The difference in the hit time is then called time walk.

It should be noted, that the time walk is directly connected to the rise time of the signal, since the latest possible crossing point is given by the rise time. In that sense, the time walk effect is directly reduced by decreasing the rise time. Nevertheless,

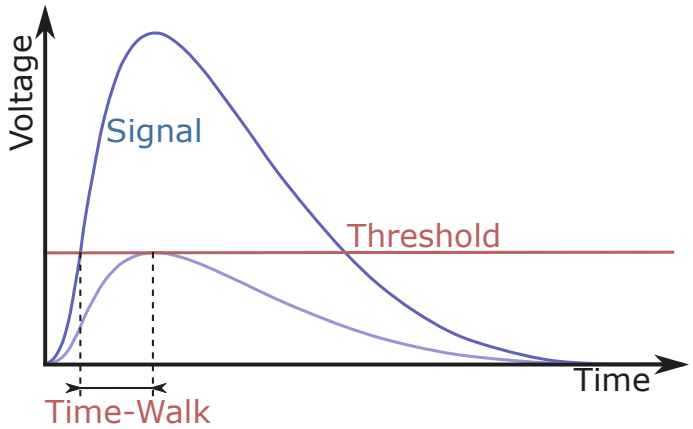


Figure 6.1: Illustration of the time walk effect.

given by the order of magnitude of the rise time, the time walk makes a significant contribution to time resolution, hence it is necessary to correct for it. One example is the time over threshold (ToT) method. It utilizes the correlation between the amplitude and the time needed for the signal to decay again. In order to obtain this additional information only the second crossing point needs to be measured.

6.2 Charge Collection Noise

The charge collecting process was described in section 4.1. There, the Berger-Seltzer-formula was introduced, describing the mean energy loss of a particle passing material. The underlying statistical process can be expressed by the Landau-Vavilov [44] distribution, resulting in large variation in the deposited energy for thin sensors. This variation in energy and thus generated charge inside the sensor results in different signal amplitudes. As mentioned above, passing through the threshold at the comparator yields to time walk.

Additional to this effect, correlated to the energy deposition also charge collection noise [45] occurs. The clusters of charge carriers, emerging along the path of the traversing particle, differ in position, granularity and size. Initially the induced signal current is constant, but as soon as a cluster reaches the diode electrode, their contribution to the signal current vanishes. This results in fluctuations of the signal length. The CSA integrates the signal current and the rise time of the produced output signal depends on the charge collection time. Hence, the charge collection noise results in a variation of the rise time. Depending on the threshold height, this effect translates into a jitter. However, due to smaller distances the charge carriers travel, the absolute value of the charge collection noise decrease for smaller sensors, as it is shown in figure 6.2.

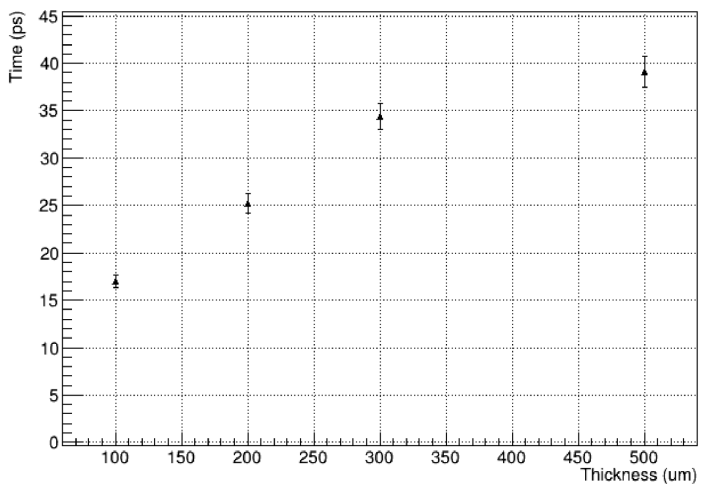


Figure 6.2: Simulated time jitter due to charge collection noise depending on the detector thickness [45].

The effect of charge collection noise can be neglected for most applications, especially in the case of an ideal current readout. For a charge integration measurement, it sets a lower limit for the intrinsic time resolution of the sensor in the order of a couple tens picoseconds.

6.3 Electronic Noise

Every electronic device adds noise to the signal, whereby it can be distinguished between voltage noise, e.g. influencing the signal amplitude and jitter. The latter describes noise in the time domain of the signal, hence a fluctuation of the signal edge. Reasons for jitter are statistical and thermal processes during signal processing in individual devices, resulting in so-called transit time jitter. Further, the voltage noise itself translates directly into jitter, depending on the steepness of the signal edge. The latency, due to the time a signal requires to pass through the circuit is not harmful, as it adds only a constant part to the hit time.

Different contributors for the voltage noise were already discussed in section 3.3, considering single transistors. For the pixel front end electronics, given in section 4.6, the CSA or to be more precise its input transistor is the main contributor to the noise. A typical parameter defining the noise at the output node of an amplifier is the equivalent noise charge (ENC). It is defined as:

$$ENC = \frac{\sigma_{rms}}{A_Q} \tag{6.1}$$

the fraction of the root mean square (rms) output noise σ_{rms} and the charge gain A_Q of the amplifier.

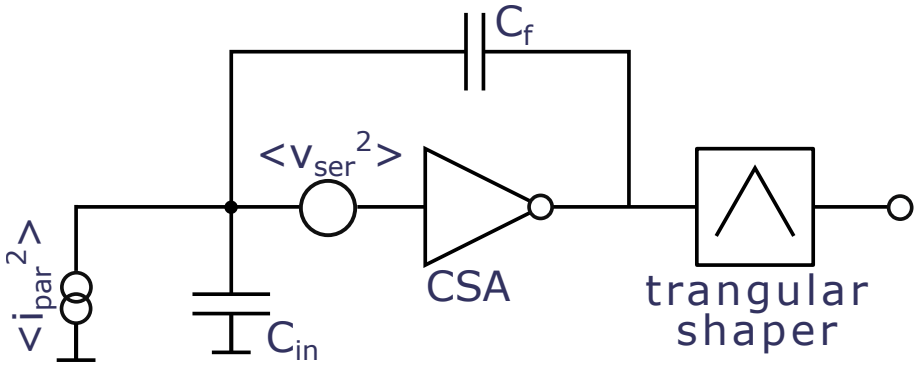


Figure 6.3: Representation of the method to calculate the equivalent noise charge [27].

Figure 6.3 illustrates the typical approach to calculate the ENC. Here, the noise sources are divided into serial and parallel noise. In the following, the shot noise caused by leakage current in the detector diode is additionally considered as parallel

noise. Its spectral characteristic is given by:

$$\frac{\mathrm{d}\langle i_{leak}^2 \rangle}{\mathrm{d}f} = 2qI_{leak} \tag{6.2}$$

with the leakage current I_{leak}. All other noise sources depend on the input transistor of the CSA. In the case of a MOSFET contributors to the serial noise are the thermal channel noise $\langle v_{therm}^2 \rangle$ and the flicker noise $\langle v_{1/f}^2 \rangle$. Further, contributors to the parallel, including the gate induced thermal nose $\langle i_{gate}^2 \rangle$, are not taken into account. In the case of a BJT as input transistor, the serial noise is dominated by the shot noise correlated with the collector current $\langle v_{shC}^2 \rangle$, while the shot noise in the base $\langle i_{shB}^2 \rangle$ contributes to the parallel noise.

Based on this and in the approximation of the virtual ground, the ENC can be calculated [32]:

$$ENC^2 = \frac{\mathrm{e}^2}{4q^2}\left(\frac{1}{2\tau}aC_{in}^2 + \frac{\tau}{2}b + 2cC_{in}^2\right) \tag{6.3}$$

with the euler number e, the input capacity $C_{in} = C_{det} + C_{amp,in} + C_{para}$ and the shaping time τ. Latter is defined by the bandwidth of filters, whereby for equation 6.3 a high and a low pass filter with the same cut-off frequency are assumed. For the FE, presented in figure 4.6 this assumption is not fully satisfied. There, the input capacity marks a high pass and the low pass is given by the parasitic components of the long connection line between analogue and digital pixel. But the cut-off frequencies of both filters does not necessarily match. Further, the parameters a, b and c in equation 6.3 are given by the different types of noise, as given in table 6.1.

A qualitative analysis of equation 6.3 shows that the electronic noise is highly dependent on the given shaping time. For large shaping times the parallel white noise becomes the dominant term, while short shaping times pronounces the serial white noise. A typical MOSFET is dominated by its thermal noise. Hence, in this case larger shaping times help to decrease the electric noise of the circuit. In contrary, the BJTs contribution to serial noise is typically low, so it is preferable to operate it with small shaping times [45, 27].

parameter	noise type	MOSFET	BJT
a	serial white noise	$\langle v_{therm}^2 \rangle$	$\langle v_{shC}^2 \rangle$
b	parallel white noise	$\langle i_{leak}^2 \rangle$	$\langle i_{leak}^2 \rangle + \langle i_{shB}^2 \rangle$
c	serial $1/f$ noise	$\langle v_{1/f}^2 \rangle$	0

Table 6.1: Noise parameters for equation 6.3 depending on the used input transistor of the CSA.

6.4 Measured Quantities

In order to validate the quality of the signal, the time resolution among other parameters needs to be defined and measured in simulations. Figure 6.4 shows a typical waveform of the amplified signal at the output node of the CSA. Indicated there are some basic parameters defining the signal and its quality.

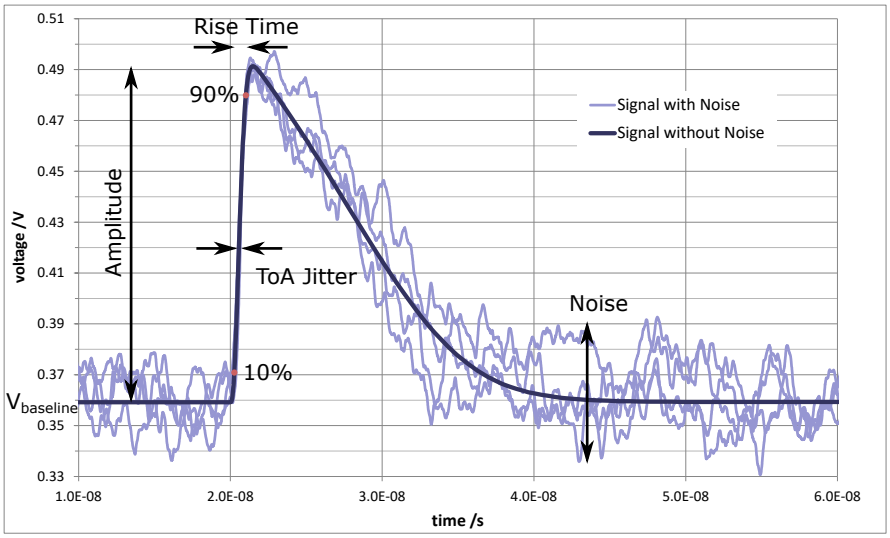


Figure 6.4: Typical waveform of an amplified signal with and without noise. Indicated are basic parameters defining the signal and its quality.

Starting with the amplitude, it can be measured according to:

$$V_{amp} = max(V_{sig}) - avg(V_{baseline}) \tag{6.4}$$

whereby the equation ensures for a uniform definition whether the signal is simulated with or without noise. Here, $V_{baseline}$ describes the typically flat value of the output voltage, while V_{sig} describes the modulation on the baseline voltage, due to a particle hitting the sensor diode. As mentioned, the amplitude does not directly influence the timing behaviour of the signal, but it needs to stay above a certain value, so the pixel front end works properly.

The rise time of the leading edge is defined commonly as

$$t_{rise} = t(0.9V_{amp}) - t(0.1V_{amp}) \tag{6.5}$$

the time it takes for the signal to rise from 10 % to 90 % of the amplitude. These two parameters can be used to approximate the slew rate (SR) of the signal, which reflects the gradient of the inflection point:

$$SR = max\left(\frac{\mathrm{d}V}{\mathrm{d}t}\right) \approx \frac{V_{amp}}{t_{rise}} \tag{6.6}$$

Further, the electrical noise σ_V can be measured in a simulation as the variance of the baseline voltage $V_{baseline}$:

$$\sigma_V^2 = \frac{1}{t_2 - t_1} \int_{t_1}^{t_2} (V_{baseline}(t) - avg(V_{baseline}))^2 \mathrm{d}t \tag{6.7}$$

This can be used to calculate the ENC of the amplifier or the signal to noise ratio (SNR)

$$SNR = \frac{V_{amp}}{\sigma_V} \tag{6.8}$$

With this, the time resolution [45] of the circuit can be defined as

$$\sigma_t = \frac{\sigma_V}{\mathrm{d}V/\mathrm{d}t} \approx \frac{t_{rise}}{SNR} \tag{6.9}$$

whereby the transit time jitter is not taken into account. The equation is a result of a geometrical approach, translating the noise on the base line into a jitter, depending on the steepness of the signal. The transient simulation offers the possibility to directly measure the time resolution of the amplifier as the time of arrival (ToA) jitter. It is illustrated in figure 6.4 and is defined by the variance of the time when the half amplitude is reached.

$$\sigma_{ToA}^2 = \frac{1}{k} \sum_{n=0}^{k} (t(0.5V_{amp,n}) - avg(t(0.5V_{amp})))^2 \tag{6.10}$$

Here, the signal must be simulated k-times to observe differences. While this measurement includes transit time jitter, according to the device model, time walk effects can not be observed, as the input signal is held constant in the simulation.

7 Layout

A main part of this thesis was to design the pixel front end electronics, with the goal to optimise it for its time resolution. Particularly the outstanding performance of IHP's SiGe HBT at high speeds should offer the decisive advantage here. The following chapter describes various design decisions that had to be taken and finally resulted in a finished layout for the analogue and digital pixel.

7.1 Layout Requirements

The previous chapters were constructed to understand the FE and in particular the CSA, with emphasis on the transfer function and noise behaviour. Based on this, several goals can be formulated, which must be fulfilled by the layout. As a signal source a minimum ionizing particle (MIP) is assumed from now on, i.e. the particle deposits the minimal mean energy according to equation 4.1. Therefore, this corresponds to the smallest signal, which needs to be detected by the sensor. In general, the detection threshold for a typical particle sensor is set to 0.4 MIP, in order to increase its detection efficiency.

Equation 6.9 showed, in first approximation, the time resolution of the amplifier depends on the rise time of the signal. Besides, the effect of the time walk is also influenced by the rise time. Hence, it is preferable to achieve a short rise time, whereby the target is to reduce it to the order of 1 ns.

Equation 6.9 indicates further that the SNR needs to be increased in order to improve the time resolution. Accordingly, a large amplitude and low noise is required. The noise can be reduced by a good choice of the shaping time $\tau = R_f C_f$, which can be adjusted in the feedback tree. Therefore, a small feedback resistance should be used, but it must not be too small, otherwise the feedback capacity will be discharged too quickly. Similar holds for the feedback capacity itself. A small C_f reduces τ and simultaneously increases the unity gain bandwidth of the CSA, according to 4.7. Nevertheless, as described in section 4.2.1, the signal charge is distributed between C_f and C_{in}. If C_f is too small, less signal charge is transferred and amplified, even though the feedback capacity is increased by the miller effect. This also results in restrictions for the detector capacity, which should be small, in order to reduce the same effect.

For the amplitude, a value of the order of $\sim 100\,\mathrm{mV}$ is targeted, which ensures a properly working comparator. The gain of the circuit must be adjusted accordingly.

One crucial parameter of the circuit is the current I_0 flowing in the amplifier, as it influences several other parameters. This includes the transconductance g_m and hence the GBW of the input transistor, as well as its transit frequency f_t.

The latter two both indicate the switching speed of the circuit and therefore the rise time. Both benefit from a higher current, same as g_m. The following must be taken into account, a full chip holds a matrix of $m \times n$ analogue pixels, which leads to a current consumption of $I_{analogue} = mnI_0$, dominating the total current. In order to be within the scope of the Mu3e experiment, the specified detector target cooling capacity of $400\,\mathrm{mW/cm^2}$ [4] for phase I should not be exceeded. With the technology node of a supply voltage of 1.2 V, this gives an upper limit for the power consumption. Additionally, larger transistors are needed to operate the circuit with a larger current. This increases parasitic capacities, which may have a negative influence on the signal. Eventually, a current of $I_0 = 20\,\mu\mathrm{A}$ for a single amplifier is targeted, which will be discussed in more detail later.

7.2 Optimisation Process

To get started, the CSA with bipolar input transistor is investigated exclusively in an ideal environment. Thus, all parameters are tuned and the ideal operation point is found.

7.2.1 Workbench

In order to compare different variations of the CSA, a standardised workbench (figure 7.1) is needed, which provides an ideal environment and unified signal. As introduced already, the sensor diode can be expressed by a capacity C_{det} and an ideal current source, providing a signal. The capacity corresponds to the diode capacity, given by equation 2.17, whereby the equation ignores fringe effects and parasitic capacities from electronic devices inside the pixel. Typically, the pixel capacity of an HV-MAPS is in the order of $\sim 100\,\mathrm{fF}$. Since the final size of the analogue pixel and the applied high voltages were not defined yet, $C_{det} = 100\,\mathrm{fF}$ was used to investigate the CSA.

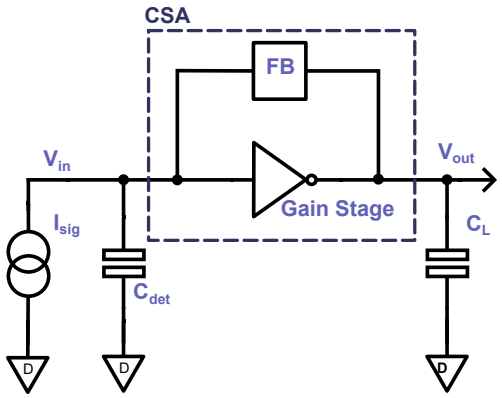
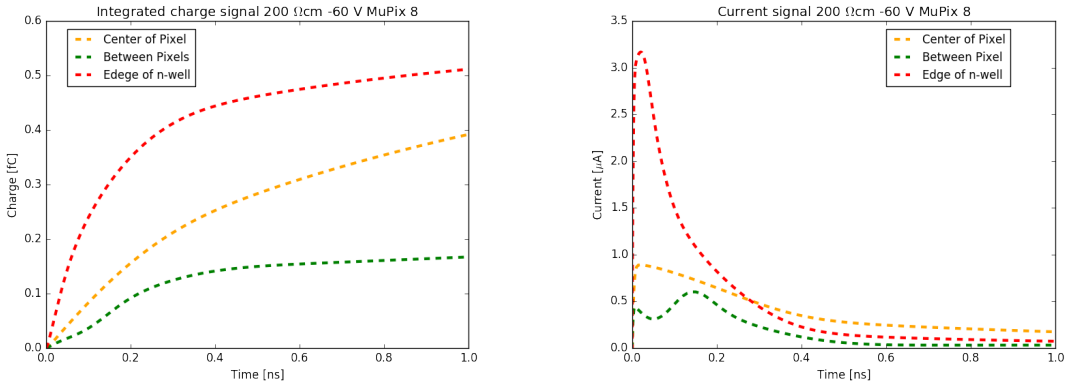


Figure 7.1: Schematic of the workbench, used to simulate and investigate the CSA.



(a) (b)

Figure 7.2: TCAD simulations of the signal charge (a) and the resulting current (b) in a MuPix8-like pixel with an HV of 60 V, a resistivity of 200 Ωcm and a depletion depth of approximately 30 μm created by a passing MIP particle [46].

The charge collection in the sensor diode is described in section 4.1. Here, electrons are the main contributor to the signal charge, as the readout electronics is located in the n-typed side of the diode. The created holes are absorbed by the voltage source, biasing the substrate. Figure 7.2 shows a TCAD simulation of the signal charge and current, corresponding to a traversing MIP particle in a MuPix8-like shaped pixel.

According to this simulation, the test signal is chosen, whereby it is assumed that the particle hits the pixel centre. Further, it is assumed that the signal is discharged fast by the small feedback resistor. Hence, the long tail of the yellow dashed line in figure 7.2(b) is neglected. Eventually a signal current with a triangular shape is used. It rises within 10 ps up to 900 nA, followed by a linear decay within 1 ns. This corresponds to a signal charge of $Q_s = 0.45\,\mathrm{fC} \approx 2800\,\mathrm{e}^-$.

In chapter 4.2.1 it was shown that the load capacitance has a huge influence on the behaviour of the amplifier. Hence, a load capacity needs to be considered for the workbench. Its value should be in the order of the input capacity of the subsequent circuit, which is most like given by the gate capacity of a MOSFET. Therefore, $C_L = 10\,\mathrm{fF}$ was chosen, which represents roughly a MOSFET with a size of $W = 1\,\mu\mathrm{m}$ and $L = 1\,\mu\mathrm{m}$.

Figure 7.1 shows the schematic of the workbench. Not displayed there, is the biasing of the CSA. In order to have a unified environment, all bias voltage are created by ideal voltage sources. Further, in figure 7.1 the CSA is divided into the gain stage and the feedback tree. This allows to optimise the gain stage separately, while a feedback, consisting of ideal components is used. For the ideal feedback, a resistance of $R_f = 400\,\mathrm{k}\Omega$ and a capacity of $C_f = 2.5\,\mathrm{fF}$ are used.

7.2.2 Design Guidelines

There are rough guidelines, which help to size transistors correctly. These can be followed to get a good starting point. Firstly, the transistors are distinguished by their purpose, in transistors acting as a current source and in transistors acting as a switch or gain stage. For the latter, g_m is the decisive parameter, given in equation 4.5 for a MOSFET. In order to increase g_m, these transistor should be as wide and short as possible. However, by increasing the width of the transistor parasitic capacities are added, acting as a feedback. Therefore, the optimum value is not necessarily very large. For transistors serving as a current source, the channel length effect should be minimized. This means long transistors are more suitable.

In addition, bipolar transistors are used. In the IHP technology these have less parameters than a MOSFET. Only the emitter length can be adjusted. It was found that the emitter length has a negative influence on the rise time, probably due to an increase in parasitic capacities and in extant a reduction in the transit frequency. Therefore, the minimum length should be chosen. Based on these rules, simulation can be used to systematically optimize the circuit.

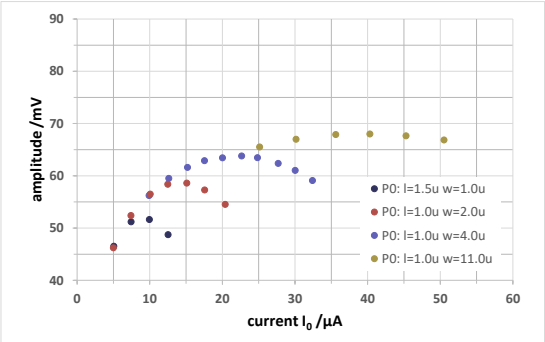
7.2.3 Gain Stage Optimisation

The power consumption of the amplifier is a crucial parameter that must be carefully selected. Other parameters such as the transistor size can be determined on the basis of the current I_0. Further, it must be declared if the gain stage should be operated with a cascode (figure 4.3(a)) or not (figure 4.2(a)). Therefore, a simulation scan was performed, measuring the amplitude and the rise time of the signal depending on the current I_0. This scan was done for both cases, with and without a cascode and is shown in figure 7.3. For the simulation, different sizes of the current source transistor $P0$ were investigated. This is necessary, as the width of the current source must be increased to provide a larger current in the amplifier.

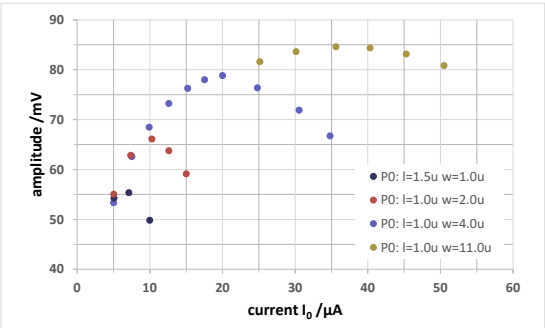
A first observation shows, as expected, the amplitude and the rise time improve with increasing current. Recalling the theoretical description from chapter 4.2.1: Accordingly, the dc-gain should be independent of the current. However, the GBW is increased, helping to increase the amplification of high frequencies and thus the amplitude of the fast signal. As higher frequencies contribute to the signal also the rise time of the signal decreases.

Nevertheless, the simulation shows that the amplitude remains below the required 100 mV, whereby the cascode clearly helps to increase the amplitude of the signal. On the other hand the usage of the cascode, results in a slightly higher rise time. Since this effect is not significant, these results lead to the decision to use the cascoded amplifier.

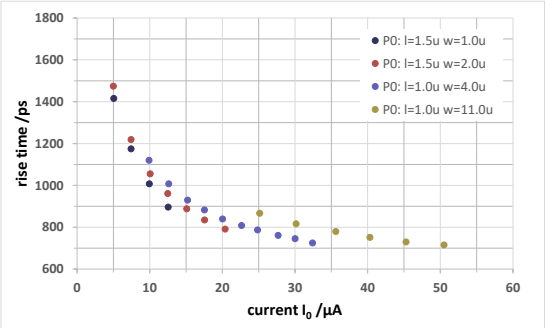
Based on the data, it was decided to use a current of $20\,\mu$A and thus a current source with a size of $L = 1.0\,\mu$m and $W = 4.0\,\mu$m. The amplitude does not increase significantly with a higher current. Same holds for the rise time, which is already well below 1 ns for $I_0 = 20\,\mu$m. To increase the current above $20\,\mu$A a very large current



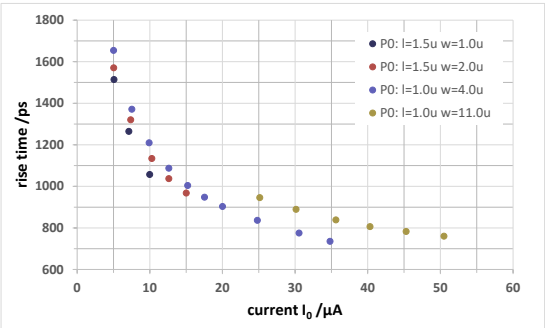
(a) Gain stage without cascode.



(b) Gain stage with cascode.



(c) Gain stage without cascode.



(d) Gain stage with cascode.

Figure 7.3: Simulations of the amplitude (top) and the rise time (bottom) depending on the amplifier current I_0. Investigated are the bipolar gain stage without (left) and with (right) cascode transistor. In order to increase the current the width of the current source $P0$ need to be adjusted. For the simulation the workbench with ideal feedback tree was used.

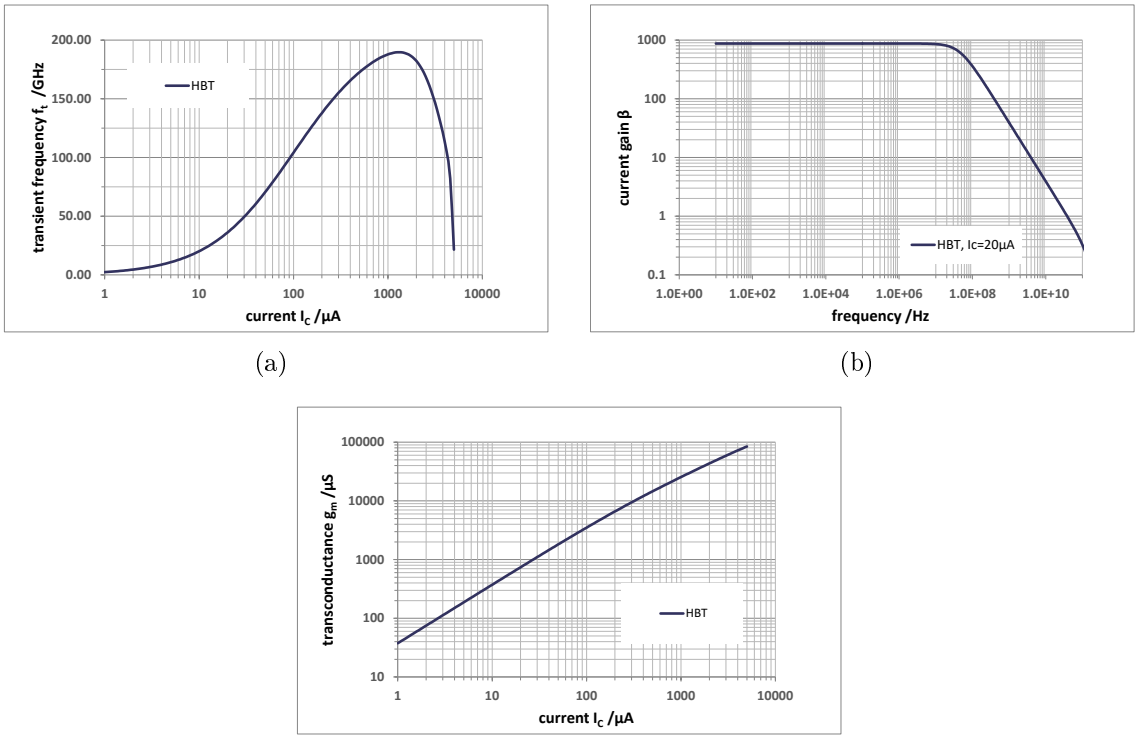


Figure 7.4: Simulated transit frequency (a) and transconductance (c) depending on the collector current I_c of the HBT. Further, the frequency dependency of the current gain β (b) at a current $I_c \approx 20\,\mu$m is presented.

source would be needed, which is not feasible in a small pixel layout. On the other hand, a high power consumption was accepted with this setting. In comparison, for the amplifier in the MuPix8 sensor a current of approximately $5\,\mu$A [47] was provided. But as the simulations show, this current is needed to achieve a good performance.

The choice of the current defines some key parameters, which are simulated in figure 7.4 for the bipolar input transistor. The simulation of g_m in figure 7.4(c) nicely reflects the linear dependency on the current I_c, which was given by equation 4.6. For the chosen operation point a value of $g_m = 740\,\mu$S was found.

For the transit frequency a value of $f_t = 38.9\,$GHz can be extracted from figure 7.4(a), which is way below the specifications of the technology. Thus, the restriction in current and power consumption results in a significant loss of performance. At the same time, this demonstrates the need for a high-end bipolar transistor, else it would not be possible to operate it. The same point can be made by looking at figure 7.4(b). The interesting frequency range to transmit a signal of the length of $\sim 1\,$ns is located at $\sim 1\,$GHz and above. In this area the current gain is in the order of $\mathcal{O}(10)$. Eventually, this is a brief description of the difficulty of designing pixel sensors, where high frequency performance and a low power consumption are both essential. Thus, a trade-off between performance and power consumption is done.

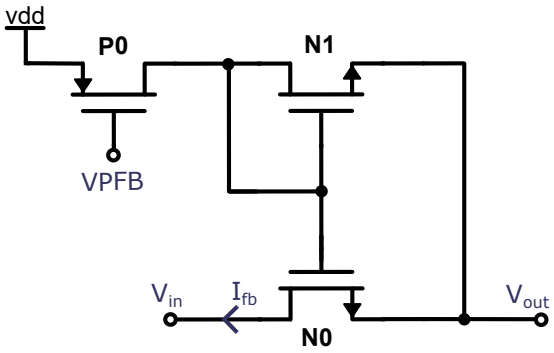


Figure 7.5: Schematic of the active feedback tree.

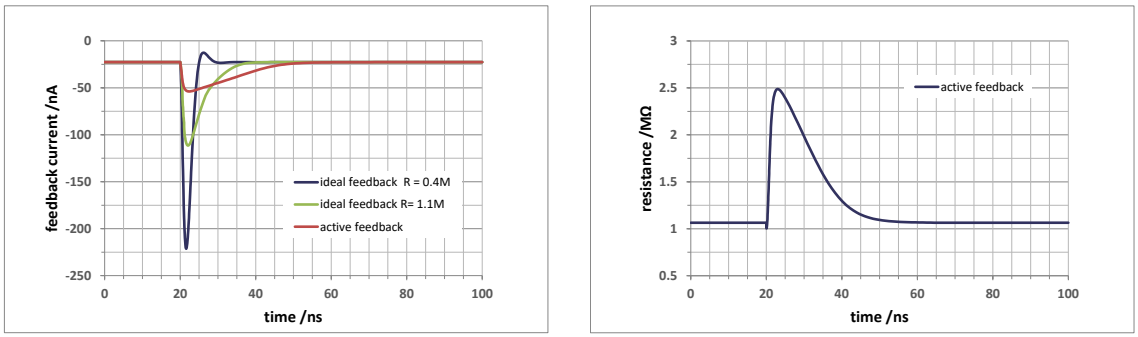
7.2.4 Active Feedback Tree

For the optimisation process, an ideal feedback consisting of a resistance and a capacity was used. On chip level, a resistance is typically implemented by a poly resistance. It utilizes the lower conductivity of polycrystalline silicon, compared to the one of a silicon mono crystal, in order to build up a resistance. However, the size needed to achieve a resistance of the order of $\mathcal{O}(100\,\mathrm{k\Omega})$ is not feasible in a small pixel layout. A commonly used solution is to use the channel resistance of a MOSFET instead.

Implementing an active feedback tree enables additional advantages. By operating the transistor as a current mirror, it results in a constant feedback current I_{fb} rather then a constant resistance. This influences the falling edge of the signal, since the feedback capacity is therefore discharged linearly. In contrast, discharging with a constant resistance results in a e^{-x} behaviour.

The circuit used for the feedback is represented in figure 7.5, it was already used in the Mupix8 in a similar configuration. Here it should be noted that the circuit is not a real current mirror, as both transistors do not share the same gate-source voltage. Besides, the current in the feedback is not really defined by the control branch, but rather by the bipolar input transistor and the equilibrium state, which is reached in the absence of a hit signal. Nevertheless, the control circuit influences the resistance of the feedback transistor. The feedback current characteristic is simulated in figure 7.6(a) and compared to an ideal feedback with a resistance of $0.4\,\mathrm{M\Omega}$ and $1.1\,\mathrm{M\Omega}$. In all cases an ideal feedback capacity $C_f = 2.5\,\mathrm{fF}$ was included. Accordingly, the active feedback reduces the change in the feedback current significantly. In relation to this, a change of the feedback resistance occurs, which is presented in figure 7.6(b).

With the active feedback, also the signal is influenced. Thereto, some signal parameters are represented in table 7.1. As shown there, the higher feedback resistance of the active circuit helps to increase the amplitude of the signal, which exceeds the intended value. This is because the charge was discharged too fast in the case of the



(a) (b)

Figure 7.6: Simulation of the feedback current I_{fb} (a) for an ideal and an active feedback tree, with a hit signal emerging at 20 ns. Further, the resistance (b) of the active feedback is presented. The simulations for the active feedback are done, including a feedback capacity of $C_f = 2.5\,\text{fF}$. It represents parasitic capacities, which occur outside the ideal environment.

	ideal feedback	active feedback
rise time	903 ps	1362 ps
amplitude	78.8 mV	108.2 mV
slew rate	69.81 mV/ns	63.57 mV/ns

Table 7.1: Measurements of the signal with an ideal feedback ($C_f = 2.5\,\text{fF}$, $R_f = 400\,\text{k}\Omega$) and the active feedback. Simulations with the latter, including a feedback capacity of $C_f = 2.5\,\text{fF}$, which represents parasitic capacities.

small feedback resistance and a part of the signal charge was not amplified. Simultaneously, the slope of the rising edge is not influenced significantly by changing the feedback tree. Hence, the rise time of the signal is larger with the active feedback.

Another drawback is that the measure time τ increases with the higher resistance and as a consequence a higher noise is expected. Regardless, due to the size of the poly resistor, it is necessary to use a transistor as feedback.

7.2.5 Source Follower

For the source follower a gain of one is targeted, as well as a large cut-off frequency. The latter is mainly affected by the load capacity, whereby for the optimisation again a load capacity of $C_L = 10\,\text{fF}$ is used. The optimal working point was found at a supply current of $4\,\mu\text{A}$.

In figure 7.7 the gain of the optimised source follower is presented. From this, the dc-gain can be extracted. With a value of $A_v = -1.26\,\text{dB} \approx 0.75$, an amplitude loss of 25 % is expected. Further, the cut-off frequency is at $f_c = 110\,\text{MHz}$, accordingly the loss is even increased for high frequencies. Besides, the rise time will increase if

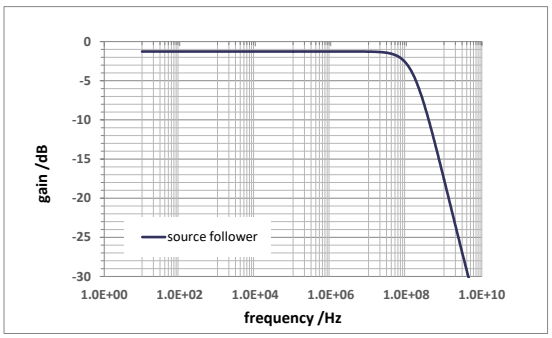


Figure 7.7: Simulated transfer function of the source follower with a load capcity of $C_L = 10\,\mathrm{fF}$.

a signal with a rising edge of 1 ns is transferred. On the other side, later it will be shown that a limited bandwidth helps to decrease the noise in the circuit. Hence, the SF does not necessarily worsen the total performance of the analogue pixel.

7.2.6 Comparator

The comparator is not optimized within this thesis, as it was already used in the MuPix8 in a similar configuration. Nevertheless, its operation point must be found. A cascoded comparator is used, with an additional option to tune the current in one branch of the differential pair. The corresponding circuit is presented in figure 7.8. For simulating the circuit, an input with a variable baseline V_{base} and on top a

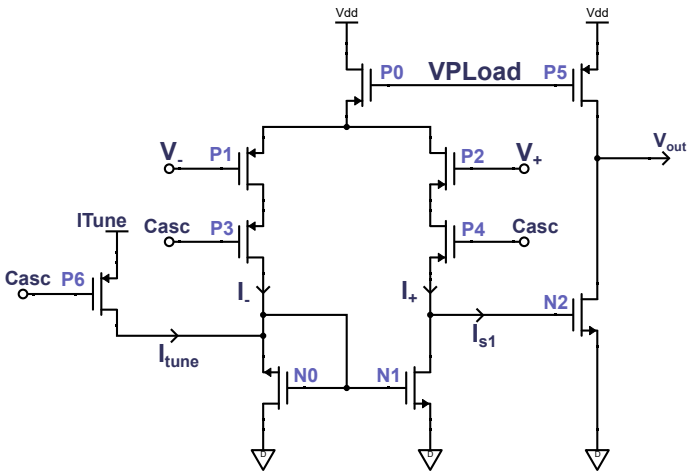
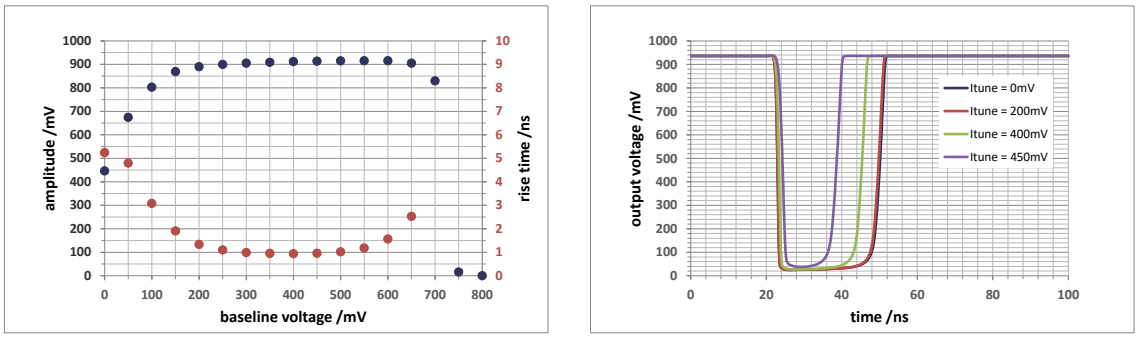


Figure 7.8: Schematic of the cascoded comparator with the option to tune the current in the negative branch of the differential pair.

modulated hit signal with an amplitude of 100 mV is used. Further, the rise time of the signal is set to 2 ns and its decay time to 38 ns. This corresponds to the expected signal shape, coming from the analogue pixel. The threshold voltage for detecting the hit is set to $V_{base} + 50\,\mathrm{mV}$.



(a) (b)

Figure 7.9: Simulation of the baseline dependency (a) of the comparator for $ITune = 0\,\mathrm{V}$ and the influence of the $ITune$ voltage on the signal response of the comparator (b).

With these settings, firstly the dependency on the baseline voltage can be observed, whereby the circuit part to tune the current is turned off, by connecting it to ground. From the theoretical description of the comparator in section 4.5, it is already known that the differential pair does not work over the full range from ground to the supply voltage. This behaviour can be seen in figure 7.9(a). From this simulation a range from 200 mV to 550 mV for the baseline voltage can be extracted. In this range the comparator shows a good signal response with an amplitude of $\sim 900\,\mathrm{mV}$ and a rise time of $\sim 1\,\mathrm{ns}$. Accordingly, nearly the full swing of a digital signal is achieved, which allows for a further procession.

The influence of the circuit, which tunes the current in the minus branch of the differential pair, is shown in figure 7.9(b). The circuit adds a current I_{tune}, which is also mirrored into the second branch. Thus, the current at the first stage output node can be expressed as $I_{s1} = I_+ - (I_- + I_{tune})$, in comparison to section 4.5. The additional current helps to discharge the voltage at this node. Due to the second gain stage and its negative amplification, this behaviour is inverted at the output node of the comparator. As shown in figure 7.9(b), an increase of the current I_{tune} results in an earlier switching back to the default state. Accordingly, with $ITune$ the effective threshold of the comparator can be tuned and thus the length of the digital signal can be adjusted. This is very helpful in order to deal with large signal. In the case of this thesis, considering a FE with a beneficially small measurement time τ, it can be argued that such a feature is not necessary. Nevertheless it could be used to tune the switching point of the comparator and adjust it precisely, in order to deal with large noise levels at the input signal.

7.3 The Analogue Pixel Design

After the optimisation of the individual circuits in an ideal environment is completed, they can be joint in the analogue pixel and a layout can be designed.

7.3.1 Schematic of the Analogue Pixel

First, the whole schematic of the analogue pixel is presented in figure 7.10. The individual parts are indicated by dashed boxes. The bias voltages needed to operate the analogue pixel are listed in table 7.2, with a short description of their purpose and typical value.

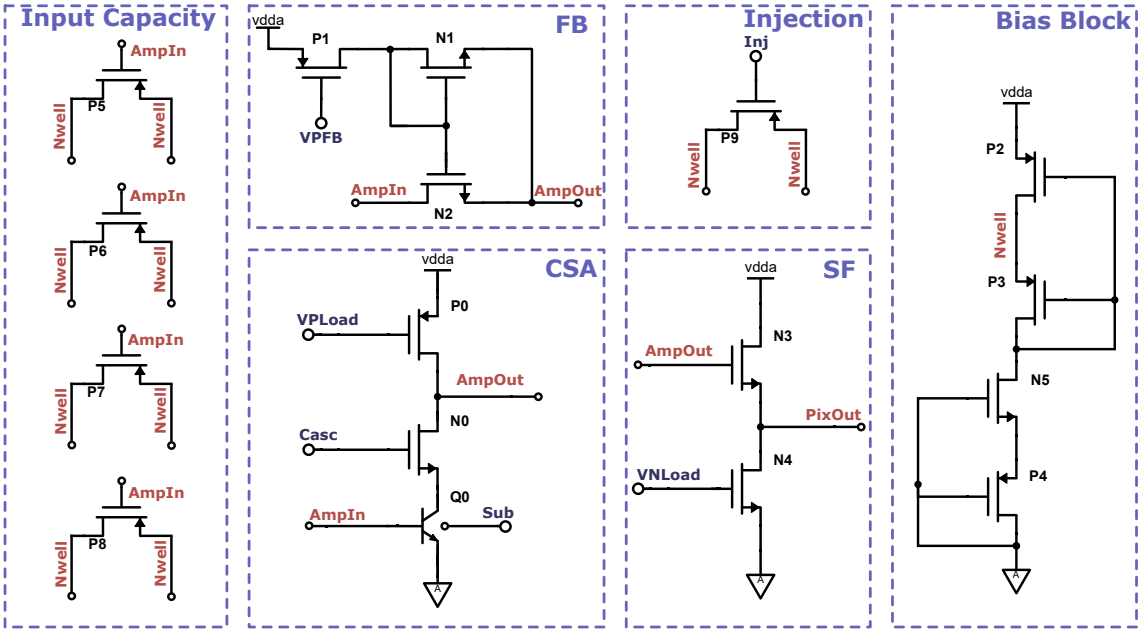


Figure 7.10: Schematic of the analogue pixel, with indicated bias voltages (blue) and net names (red). Different circuit parts are separated and marked by dashed boxes.

The pixel includes some components, which were not discussed yet. These are the input capacity, a small circuit to inject a test signal and a bias block. The former is realised by four large PMOS transistors. Since the signal emerges in the pixel diode and its mirror charge is collected in the n-well substrate, the gate-substrate capacity of these transistors act as the input capacity. Additionally, there is an ohmic connection between the substrate and the drain and source terminal. Hence, both capacities, gate-drain and gate-source, also add to the input capacity. The four transistors forming an input capacity of the order of $C_{in} \approx 1\,\mathrm{pF}$. Here it is important to have a large capacity, since it forms a high pass together with the base-emitter resistance of the bipolar transistor. By increasing the capacity, its cut-off frequency is decreased and therefore the bandwidth of the circuit increased. Besides, a larger

circuit name	bias name	purpose	value
CSA	VPLoad	tunes current in CSA	465 mV
	Casc	controls cascode transistor	1000 mV
	VPFB	tunes feedback resistance	600 mV
	Sub	substrate voltage of the HBT	0 mV
SF	VNSF	tunes current in SF	600 mV
Injection	Pulse	injects test pulse	0 mV

Table 7.2: List of all bias voltages, which are needed to operate the analogue pixel.

input capacity helps to transfer the signal charge from the detector capacity into the amplifier stage.

Simultaneously, these four transistors and their ohmic connections to the substrate are used to bias the n-well. The related bias block was also not introduced yet. It was already used in previous projects of the Mu3e experiment and its purpose is to generate a voltage *NWELL*, as close as possible to the supply voltage of 1.2 V. It is crucial that this potential is separated from *vdda* and *gnda* by large resistance, with the related time constant being large. Hence, there is enough time to process the signal before the pixel is brought back to its operation point by the bias block.

The test signal injection is used to validate and characterise the analogue pixel and the whole readout system, when it is implemented on a chip. Therefore, a voltage signal is induced to the pixel n-well, where it is processes as a real hit signal. It is implemented like the input capacity by a simple PMOS transistor, whereby its gate capacity is used to induce the test pulse. Since this capacity adds to the pixel capacity, it should be small. Here, a transistor size of $W = 1\,\text{nm}$ and $L = 1\,\text{nm}$ is used, resulting in a capacity of $C_{test} \approx 1\,\text{fF}$. If the circuit is not used, its input line should be connected to ground, in order to reduce the noise, which is produced by the transistor.

7.3.2 Layout Architecture

The size of the analogue pixel is given by the size of the n-doped part of the sensor diode. All electronic is located in this n-well or p-wells inside the n-well. Accordingly, the minimum size of the pixel is limited by the electronic it holds. Here the dominant parts are the isolated bipolar transistor in the CSA and the transistors, forming the input capacity. It was found that a pixel size of $25 \times 25\,\mu\text{m}^2$ is sufficient to hold all required circuits, whereby the guard ring, isolating it from the neighbouring pixels, is not considered. It is assumed that for the guard ring an additional area of $7.5\,\mu\text{m}$ on each side of the pixel is needed. Thus, the total size of the pixel is $40 \times 40\,\mu\text{m}^2$, resulting in a spacial resolution of

$$\sigma_x = \frac{40\,\mu\text{m}}{\sqrt{12}} \approx 11.5\,\mu\text{m} \tag{7.1}$$

Further, based on the pixel size, the power consumption of the pixel can be approximated. Due to parasitic components, the current in the CSA is increased to $21.5\,\mu$A. Together with the current inside the source follower, this leads to a total power consumption of

$$P_{analogue} = 1.2\,\mathrm{V}\,\frac{21.5\,\mu\mathrm{A} + 4\,\mu\mathrm{A}}{40\,\mu\mathrm{m} \times 40\,\mu\mathrm{m}} \approx 1910\,\mathrm{mW/cm^2} \tag{7.2}$$

It should be noted that this is the power consumption of the analogue pixel and not a whole chip. The latter is probably smaller, as the digital part consumes less power. Nevertheless, the power restriction of the Mu3e experiment could not be met, due to the trade-off which had to be taken. An adjustment of the power consumption can be made by enlarging the analogue pixel, however the minimum pixel size will be discussed in the following.

The analogue pixel is presented in figure 7.11. As shown there, the n-well holds two p-wells, which are isolated from the substrate by the n-well itself and the underlying deep n-well. One is dedicated to hold the bipolar transistor, while in the other all NMOS devices are placed. Using two separated p-wells allows to adjust the ground and substrate potential of the bipolar transistor and the remaining electronics independently. PMOS transistors must be placed inside the n-well. This structure requires the circuits to be pulled further apart than usual, but introducing more p-wells would unnecessarily increase the pixel capacity.

The HBT is located inside the left p-well, and is surrounded by two guard rings, one for each well. Thus, it is isolated from noise produced by the other devices in the pixel. All four corners of the pixel are populated with the large input transistors, which collect the signal charge and transfer it to the bipolar transistor. In between, surrounding the right p-well, the remaining part of the CSA and the SF are located. Above and below, there is a via stack, which connects the circuit with the supply voltage and ground respectively. The bias block is placed far to the right and the injection is located in the upper right corner.

The wiring of the pixel is done in the metal 1 layer, poly silicon and the substrate exclusively. All shared bias voltages are distributed between the pixels with vertical traces in the metal 2 layer. This allows for transmission lines in the metal layer 3 to 5, connecting the analogue and the digital pixels. The remaining top metal layer 1 and 2 are dedicated to distribute power and ground to all pixels.

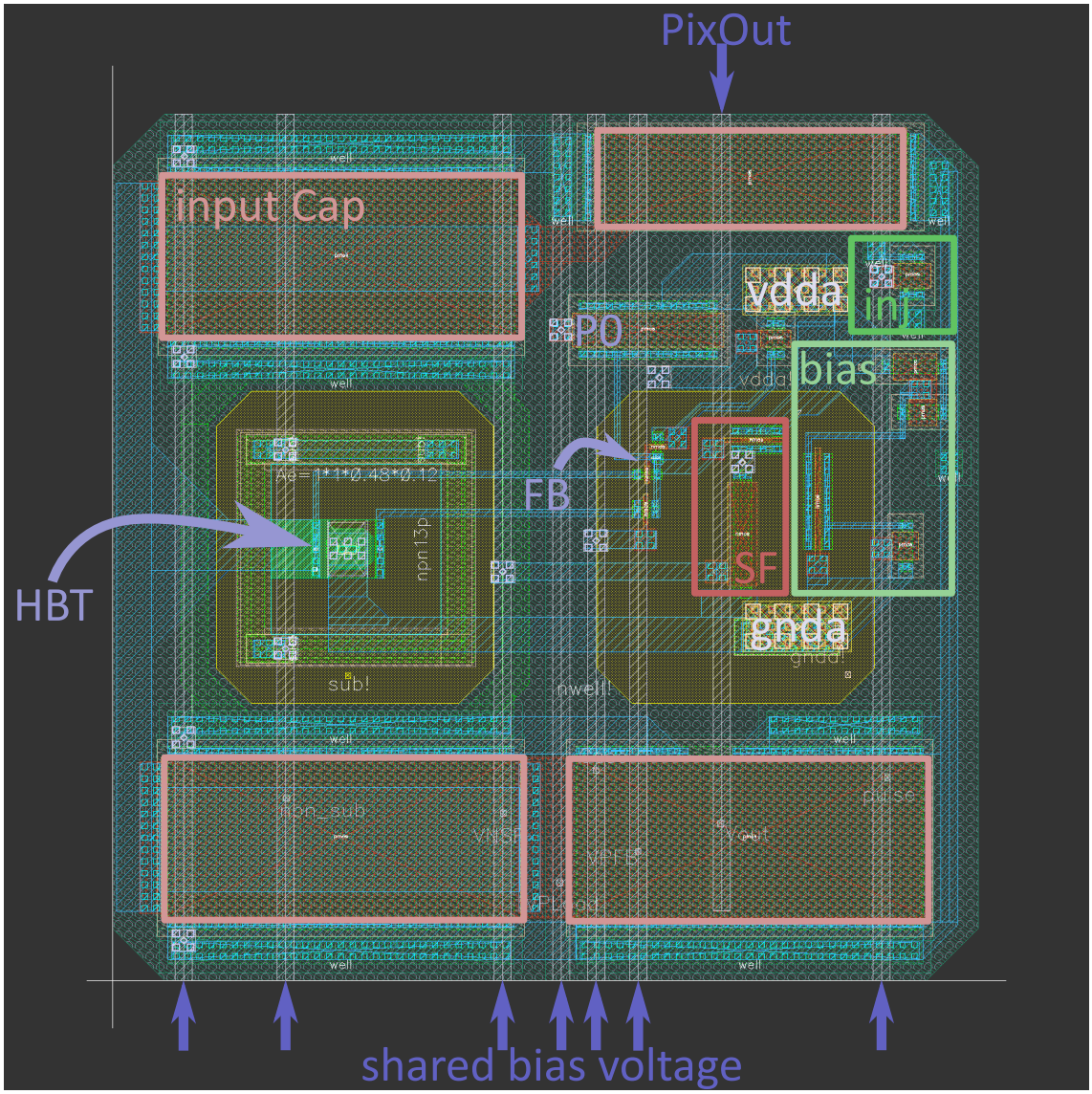


Figure 7.11: Layout of the analogue Pixel.

	n-well		gate poly silicon
	p-well		metal 1
	buried n-well		metal 2
	high doping area		device boundary

Table 7.3: Color scale of the different layers in the layout.

type	from	to	value
C_{det}	*Nwell*	-	10.6 fF
C_{in}	*Nwell*	*AmpIn*	2.8 fF
C_f	*AmpOut*	*Nwell*	0.04 fF
	PixOut	*Nwell*	0.73 fF
C_{Load}	*AmpOut*	-	1.5 fF
	PixOut	-	2.1 fF

Table 7.4: An overview of the parasitic capacities found in the layout of the analogue pixel. Net names marked with - are a combination of various nets.

7.3.3 Parasitic Components

Parasitic components have a significant influence on the behaviour of a circuit. In order to consider them in simulations, they are extracted from the layout. This includes all parasitic components between metal, poly silicon and substrate. Only substrate to substrate parasitics, i.e. those between differently doped structures, are not taken into account, since this option is not given by the design kit. Further, parasitics within electric devices are not extracted, because these are already included in the respective model. Some of the crucial devices are observed explicitly.

Firstly, parasitic capacities are observed, which can be distinguished into four main categories: Capacities, which add to the pixel capacity, to the input capacity, to the feedback capacity and to the load capacity. An overview of the parasitic capacities is given in table 7.4.

Feedback capacities are formed between the amplifier output *AmpOut* or the Source follower output *PixOut* to all nets in front of the amplifier, which are the *AmpIn* and the *Nwell* net. As aforementioned, the feedback capacity has a huge influence on the signal. Therefore, while designing the layout, extra attention was paid to avoid such capacities. For example, overlapping of the above mentioned nets was widely avoided. However, it is not possible to prevent all parasitics. Additionally, the circuit features hidden capacities, which are not revealed by the parasitic extraction. These are capacities within transistors. While the NMOS devices are isolated from the n-well and also *AmpIn* by the p-well, the PMOS transistors are located directly in the n-well. Hence, the drain-bulk capacity $C_{db} \approx 2\,\text{fF}$ of the PMOS current source $P0$ of the CSA forms a direct feedback capacity. This effect could not be observed in the ideal environment, since there its bulk is connected to the supply voltage.

The same transistor also adds to the pixel capacity as well as parasitics between the n-well or its bias net to most of the other nets. However, the largest contribution to the pixel comes from substrate to substrate capacities, or to be more precise from the diode capacity. As mentioned above, these types of parasitics can not be extracted within the used technology. To accommodate the diode capacity anyway, the additional capacity of 100 fF is still used for all simulations.

For the load capacity, a distinction between amplifier load and source follower load can be done. The former is connected to the $AmpOut$ net and is additionally enlarged by the gate capacity $C_{gate} = 0.2\,\text{fF}$ of the input transistor of the source follower. As a result, a total load capacity $C_{Load,Amp} = 1.7\,\text{fF}$ of the amplifier can be found. The load capacity, which needs to be driven by the source follower, is largely dominated by the connection line to the digital pixel.

Similarly, the input capacity is dominated by the gate capacity of the dedicated transistors. As the input capacity should be as large as possible, parasitics do not harm here.

7.4 The Digital Pixel Design

For every analogue pixel of a sensor chip, a digital pixel exists, with the purpose to translate the analogue signal into a digital one. Typically, the digital pixel contains a further processing of the signal and an extraction of different information. For this thesis, a reduced version is used, which only contains the conversion to a digital signal.

7.4.1 Schematic of the Digital Pixel

The schematic of the digital pixel is shown in figure 7.12. Since it is a reduced version, it contains only the comparator circuit as well as an input capacity. The capacity is used to decouple the analogue and the digital part. Similar to the input capacity in the analogue pixel, this one also acts as a high pass and therefore it is beneficial to use a large capacity. It is also realised by the gate capacity of two large transistors with a size of $W = 4.36\,\mu\text{m}$ and $L = 3.5\,\mu\text{m}$. This results in an input capacity of $C_{in} = 64\,\text{fF}$.

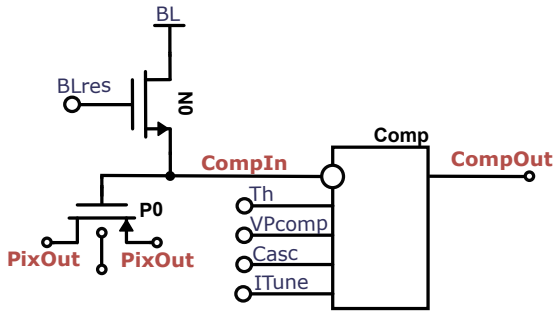


Figure 7.12: Schematic of the digital pixel, with indicated bias voltages (blue) and net names (red).

In order to operate the comparator efficiently, it is necessary to define a baseline of the signal. Due to the decoupling, the dc-component of the signal can be chosen freely, whereby the signal line is connected to the baseline voltage BL via a large

bias name	purpose	value
VPcomp	defines current in comparator	400 mV
Casc	controls the cascode transistor of the comparator	0 mV
ITune	adjusts the effective switching point of the comparator	0 mV
Th	threshold voltage to detect a hit	400 mV
BL	baseline voltage of the signal	350 mV
BLres	tunes the resistance connecting signal line and baseline	400 mV

Table 7.5: List of all bias voltages, which are needed to operate the digital pixel.

resistance. The resistance is implemented as the channel resistance of an NMOS device, which can be tuned by the *BLres* voltage.

An overview of all bias voltages that are needed to operate the pixel is presented in table 7.5. Included are a short description of their purpose and a value, defining the found operation point.

In this operation point, the comparator is operated with a current of 8.6 μA. This results in an approximated power consumption of the digital pixel of $P_{digital} \approx 11\,\mu\text{W}$. Noted that this is the power consumption in the absence of a hit signal. While processing a signal, the second stage amplifier in the comparator is turned on temporarily, doubling the consumed power. The whole channel or pixel front end electronics is operated with a power consumption of

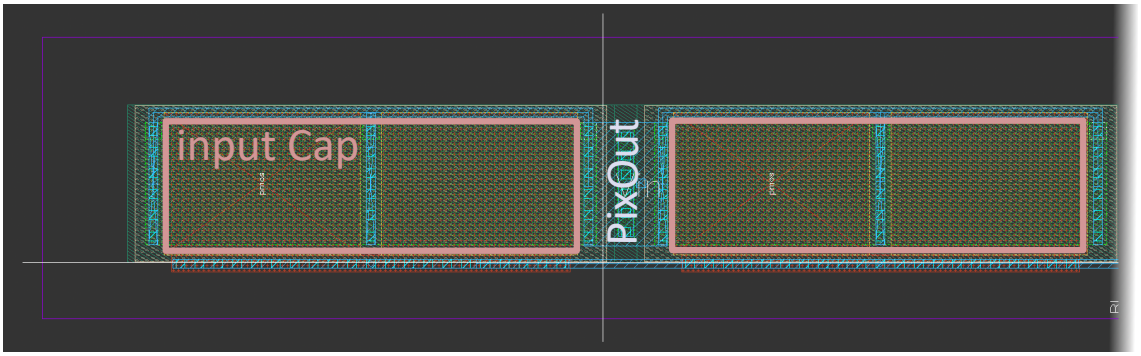
$$P_{ch} = 1.2\,\text{V}(21.5\,\mu\text{A} + 4\,\mu\text{A} + 8.6\,\mu\text{A}) \approx 41\,\mu\text{W} \tag{7.3}$$

7.4.2 Layout Architecture

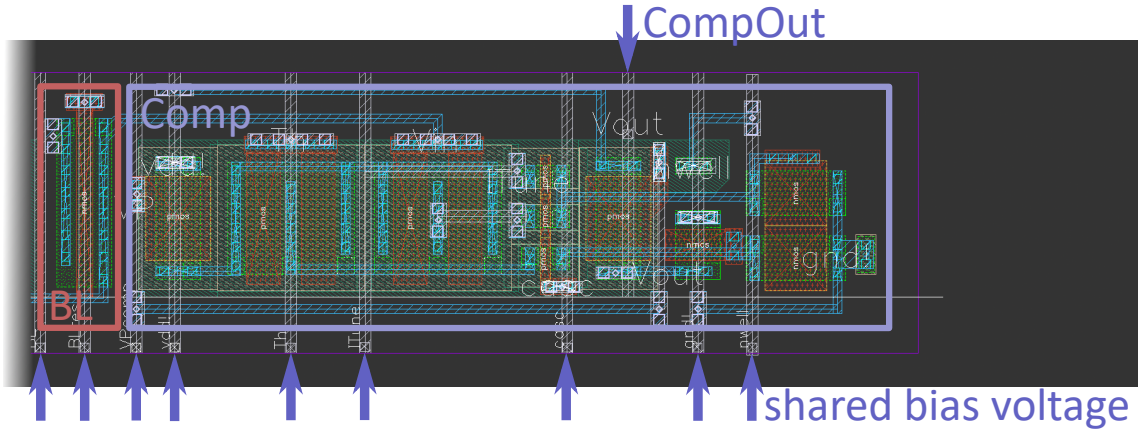
The layout of the digital pixel should match to the analogue one, meaning its size should be an integer fraction in one dimension. Accordingly, it was aimed for a size of $35 \times 5\,\mu\text{m}^2$, which allows 5 digital pixel next to each other for every matrix row. To achieve this, a very dense layout style is necessary. Thus, wide transistors are split into multiple finger devices, which allow for a better matching. Moreover, two metal layers are used for the wiring, to further increase the possible density of the transistors.

The resulting layout is presented in figure 7.13. As shown there, the layout is highly dominated by the input capacities, implemented as two large, two-finger PMOS transistors. They are located in a separated n-well, which is necessary, since the signal is transmitted through it. The comparator is located in a second n-well, decreasing the crosstalk between the digital logic and the input capacities. Further, a separation from the neighbouring pixel is required, increasing the height of the pixel. Its size is therefore indicated by the purple box.

Similar to the analogue pixel, the bias voltages of the digital ones are shared between all pixels. These are distributed by vertical traces in the metal 2 layer.



(a)



(b)

Figure 7.13: Layout of the digital pixel. In Order to display the layout, it is split into two parts, whereby (a) shwos the two large input transistors and (b) the comparator circuit.

8 Characterisation

In the following chapter several simulations are presented, which contribute to a characterisation of the pixel front end electronics and in particular of the analogue pixel. The focus here is on the noise behaviour and the associated time resolution, with regard to all parasitic informations of the layout. Furthermore, the functionality of all circuits of the FE, in a broad operation point, should be ensured and verified.

8.1 Characterisation of the Analogue Pixel

First of all, the analogue pixel is examined once again exclusively, since this allows best to investigate the influence of the bipolar transistor, which is a central part of this work. However, all circuits included in the analogue pixel contribute to the simulations and typically the *PixOut* net will be observed. As mentioned above, thereto an additional detector capacity of $C_{det} = 100\,\mathrm{fF}$ is used, as well as a load capacity of $C_{Load} = 10\,\mathrm{fF}$. For the simulations the signal is produced by an ideal current source, as described in section 7.2.1 and all bias voltages are set by ideal voltage sources.

In order to get a reference value, the pixel is compared with an NMOS amplifier in several measurements. This is the same analogue pixel, where only the bipolar transistor has been replaced by an NMOS transistor. It should be mentioned here that less effort was put into optimizing the reference pixel and thus its results are not necessarily the best possible. The layout of the reference pixel is presented in figure B.1 together with a list of the bias voltages in table B.1.

8.1.1 Transfer Function

The simulation in figure 8.1 shows the ac-response of the analogue pixel, for a test signal induced into the n-well. Here, its transfer function is measured, given by the magnitude of the output voltage divided by the magnitude of the input voltage.

This reflects the expected behaviour of the transfer function, recalling equation 4.7. For very low frequencies the gain is dominated by the constant term $-g_m r_{ce}$, where both parameters are connected to the bipolar transistor.

In the mid-frequency range from 10 kHz to 10 MHz, the transfer function increases by 20 dB per decade. This is the typical slope of a high pass. A comparison with the red curve, the ac-response of the amplifier exclusively, shows a clear difference in this region. As already mentioned, this high pass is caused by the input capacity and the base resistance in the bipolar transistor. Besides, in this region the influence

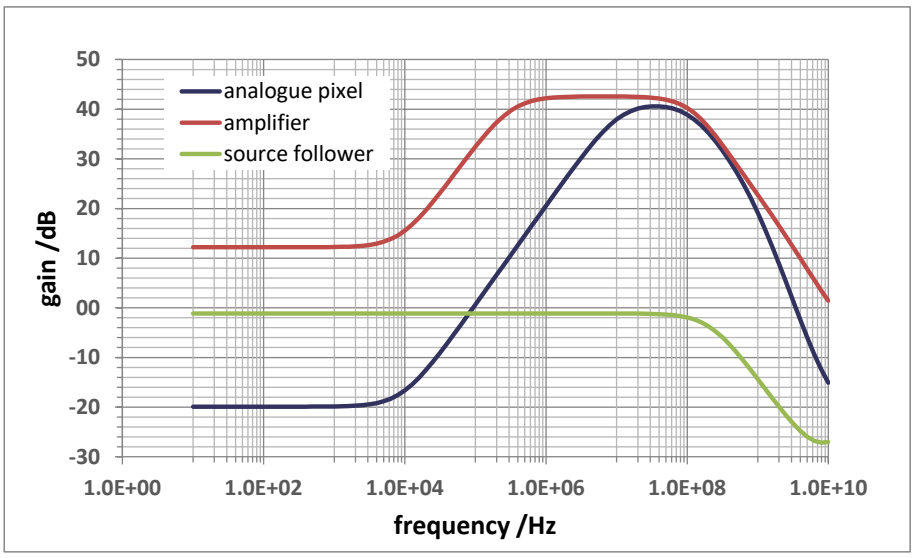


Figure 8.1: Ac-behaviour of the analogue pixel.

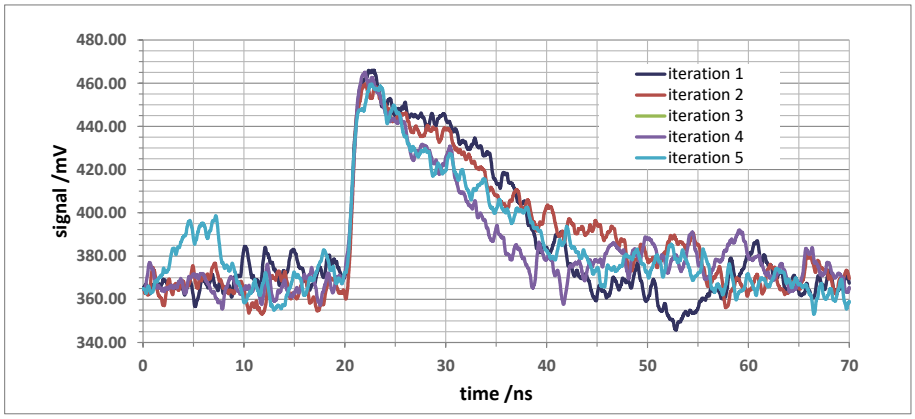
of the feedback capacity, enlarged by the Miller effect, increases, explaining the rise of the red curve.

For high frequency above 100 MHz the gain of the amplifier decreases again with a slope of 20 dB per decade. For the whole pixel, this decrease is enlarged, due to the influence of the source follower, which has a significantly smaller bandwidth than the amplifier. Its ac-response is given by the green curve. Accordingly, the usage of the source follower limits the gain of the circuit at high frequencies. But in the vicinity of 1 GHz it is still in the order of 20 dB, which is sufficient for a good performance. Higher frequencies dominated by noise are cut off, due to the limitation of the bandwidth. The analogue pixel reaches a gain of 0 dB at a frequency of approximately 35 GHz. Its highest gain of 40.6 dB was measured at a frequency of 40 MHz.

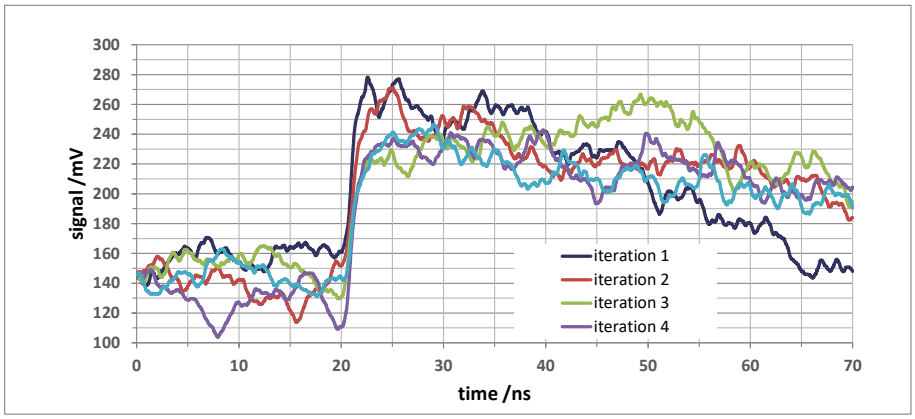
8.1.2 Transient Response

In the following the transient response of the analogue pixel is observed, including the electronic noise of the circuit. A simulation with five iterations is presented in figure 8.2(a). Here, a 70 ns long segment is studied, with a hit signal at $t = 20$ ns, as defined in section 7.2.1. From the length of the segment a minimum noise frequency of 14.3 MHz is given, while the maximum noise frequency is set to 10 GHz. This noise then results in a fluctuation in the signal parameters, which were defined in section 6.4. The measured quantities obtained in $N = 100$ iterations are presented in Table 8.1

It can be seen that the analogue pixel does not fully meet the specifications defined in section 7.1. The requirement of an amplitude of 100 mV and a rise time of 1 ns are not achieved, although they were only just missed. Nevertheless, a very good result



(a)



(b)

Figure 8.2: Transient response at the *PixOut* net of the analogue pixel with bipolar (a) or NMOS (b) input, for a MIP-like signal induced to the n-well. Included are all parasitic components, as well as a detector capacity of $C_{det} = 100\,\text{fF}$ and a load capacity of $C_{load} = 10\,\text{fF}$. Also the electric noise is simulated, in the frequency range from 14.3 MHz to 10 GHz.

		V_{amp} [mV]	t_{rise} [ns]	SR [mV/ns]	σ_V [mV]	SNR	σ_{ToA} [ps]
HBT	mean	95.02	1.413	70.36	-	14.23	-
	sigma	6.90	0.775	18.06	6.95	3.05	86.12
HBT*	mean	115.14	1.240	107.68	-	16.50	-
	sigma	6.51	0.745	22.2	7.23	3.29	61.77
NMOS	mean	114.37	3.215	32.33	-	13.40	-
	sigma	19.49	1.905	16.62	9.39	4.46	301.4

Table 8.1: Signal parameters extracted from the transient simulation, comparing the case of a bipolar and an NMOS input transistor. To calculate the values, $N = 100$ iterations were performed. For the data marked with * the size of the current source in the CSA was changed.

was found for the time of arrival jitter σ_{ToA} with a value of 86.12 ps. It indicates the intrinsic minimum time resolution, which can be achieved by the circuit and is thereby well below the target value.

In section 7.3.3 it was found that there is an additional feedback capacity hidden in the current source $P0$ of the CSA, which is given by its drain-bulk capacity C_{db}. Therefore, a version with a reduced transistor size of $W = 2.0\,\mu$m and $L = 0.5\,\mu$m is also studied. By reducing the width and the length by the factor 2, the operation point should not vary much, while its drain-bulk capacity is reduced to $C_{db} = 0.49$ fF. Due to this adjustment the performance of the pixel is improved again, so that the target amplitude is also reached. However, in this configuration the noise of the circuit also increases. Unfortunately, due to time constraints, this version of the pixel could not be analysed more closely. Nevertheless, it should be kept in mind that this is a possibility to improve gain and rise time further.

In comparison, the signal parameters for the amplifier with NMOS input transistor are also presented in table 8.1. While the reference pixel achieves a higher amplitude, for all other parameters it can not obtain the same results as the HBT. Particularly interesting is the significantly higher value for its ToA jitter, which is roughly 3 times higher compared to the HBT. Thereby, for the calculation of the parameters of the reference pixel several runs had to be excluded. In these cases the pixel lost its operation point due to high noise.

Using equation 6.9 the time resolution of the normal and the reference pixels can be estimated, resulting in a time resolution of $\sigma_{t,HBT} = (99.35 \pm 9.06)$ ps, $\sigma_{t,HBT^*} = (75.12 \pm 30.18)$ ps and $\sigma_{t,NMOS} = (240.1 \pm 62)$ ps. Here it is expected that the calculated value of the time resolution and the simulated ToA jitter roughly match, as both parameters describe the influence of the baseline noise on the rising edge of the signal. The calculation again confirms the improvement by using an HBT instead of an NMOS transistor.

Another parameter indicating the performance of an amplifier is the equivalent noise charge, given in equation 6.1. Therefore, the charge gain A_Q of the circuit needs to be calculated. For the given operation point it can be estimated as the output voltage V_{amp}, which results from amplifying the signal charge. The latter is given by the test signal, with a charge of 2800 e^-. Accordingly, the equivalent noise charge is $ENC_{HBT} = 205\,e^-$, $ENC_{HBT^*} = 176\,e^-$ and $ENC_{NMOS} = 230\,e^-$ for the different pixels.

8.1.3 Noise Analysis

In the next step the noise behaviour of the pixel will be studied in detail. For this purpose the small-signal noise analysis from Cadence is used, where the noise is examined in a frequency range from 10 Hz to 10 GHz. It is possible to observe the noise independent from a signal on which it is normally modulated and to distinguish between different noise sources. A corresponding simulation of the power spectral noise density is shown in figure 8.3. The total noise is represented by the blue curve and is given by the RMS of all noise contributors. Additionally, in fig-

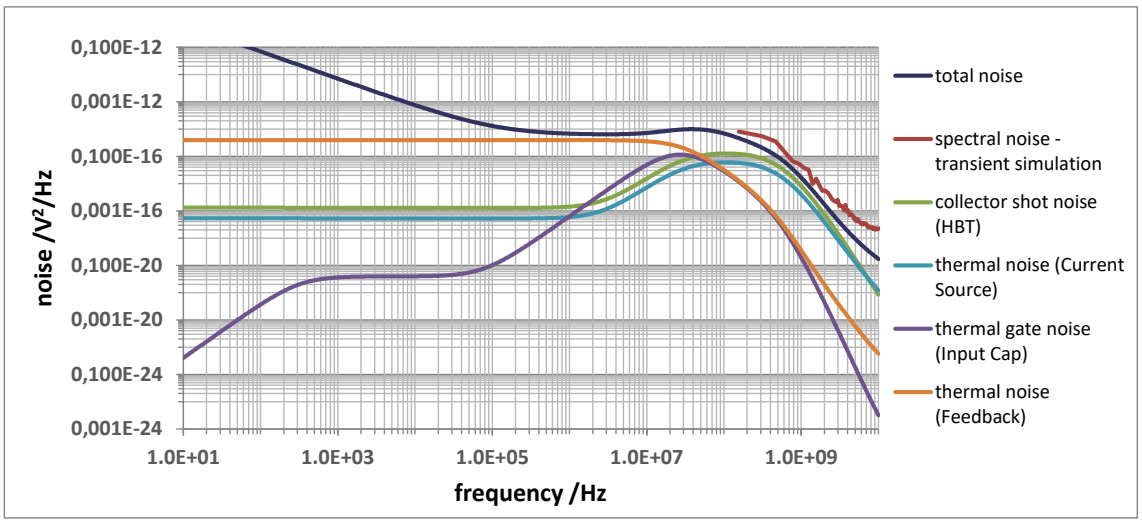


Figure 8.3: Total noise spectrum of the analogue pixel and its largest noise contributors. Included is the noise extracted from the transient simulation.

ure 8.3 the largest contributors to the integrated noise are displayed. Although the flicker noise has a visual effect at small frequencies to the total noise, it does not significantly contribute to the integrated noise. Here, contributors at high frequencies are dominant.

The largest noise contributor is, as expected, the bipolar input transistor and its collector shot noise. This is especially dominant at high frequencies, where the gain of the circuit is at its maximum. In the interesting region from 50 MHz to 10 GHz it causes 30.67 % of the total integrated noise. The theoretical description of this noise source is given in equation 3.12. In the used technology the base resistance $r_B \approx 165\,\Omega$, as well as the current amplification $\beta = 900$ are optimized. Therefore, the noise should be dominated by the second term, which decreases with the current I_C. The current limitation for the pixel, hence results in a larger noise. On the other hand, in the previous section it was proven that a smaller current source is beneficial for most signal parameters. Here only a trade-off between these two phenomena can be made.

The second largest contributor is the thermal noise of the current source in the CSA. According to equation 3.9 this noise source is determined by the channel resistance, here expressed as γ/g_m. As g_m scales with the transistor current and its width (compare equation 4.5), this leads to the same conclusion: The noise would decrease with increasing current and size of the current source. Besides, this also gives an explanation for the higher noise of the analogue pixel with smaller current source, which was found in the previous section.

Due to the large gate capacity of the four transistors, forming the input capacity, their gate induced thermal noise is a dominant factor. However, this cannot be prevented, as the large capacities are necessary for the hit acceptance of the pixel.

type	total noise [mV]	total noise (trans)[mV]	coll. shot noise (HBT) [mV]	therm. noise (CS) [mV]	therm. gate noise (Input Cap.)[mV]
HBT	4.22	5.76	2.34	1.62	4×1.43
NMOS	3.11	7.29	2.05*	1.98	4×0.52

Table 8.2: Total integrated noise of the analogue (HBT) and the reference (NMOS) pixel in the range from 50 MHz to 10 GHz. The * marks the noise of the NMOS input transistor, which is a thermal noise source.

A collection of the largest contributor to the noise is given in table 8.2. Here the integrated noise in a frequency range from 50 MHz to 10 GHz is presented for the analogue pixel, as well as the reference pixel. Firstly, it should be noticed that the total noise of the reference pixel is smaller then the one of the analogue pixel. This does not reflect the behaviour, which was found in the transient simulation. Moreover, the noise simulated here is significantly smaller compared to the calculated noise, given in the previous section.

It leads to the impression, noise is handled differently in this two types of simulations. Therefore, the noise of the transient simulation is again studied in more detail. A transient simulation of the length of 100 ns is performed, without a hit signal. Based on this, the spectral noise can be extracted from the baseline voltage and it is presented in figure 8.3 and table 8.2. In the measured range the transient spectral noise follows the trend of the total noise. Since the noise is calculated as random fluctuations in the transient simulation, it is clear that the spectral noise also shows these fluctuations, enlarged by uncertainties in the calculation of the Fourier transform. Nevertheless, it shows a trend to a higher noise, which is also reflected in the integrated transient noise in table 8.2. However, it is still below the noise measured directly in the transient simulation, via the variance of the baseline.

Eventually, the mismatch between the noise from the dedicated noise simulation and the transient noise could not be fully clarified, whereby the following reasons could be excluded: It was ensured that always the same frequency ranges of the noise were measured. Further, it could be verified that the noise follows in good approximation a Gaussian distribution. Only the reference pixel lost in a few simulations its working point, resulting in a shift of the baseline voltage. This is an indication that the assumption of a small signal noise simulation is not fulfilled. It linearises the circuit around its working point and assumes that the noise is small enough to not alter it. Accordingly, the non-linearity of the amplifier is not taken into account, which may have an impact on its noise behaviour. As stated by Cadence® [48], the direct noise analysis in the time domain, gives excellent accuracy to examine noise response of a non-linear system. Hence, in the following it is assumed that the measurements of the transient simulation are more accurate and therefore used for calculations. However, the small signal noise analysis gives a good inside into the predominant noise sources.

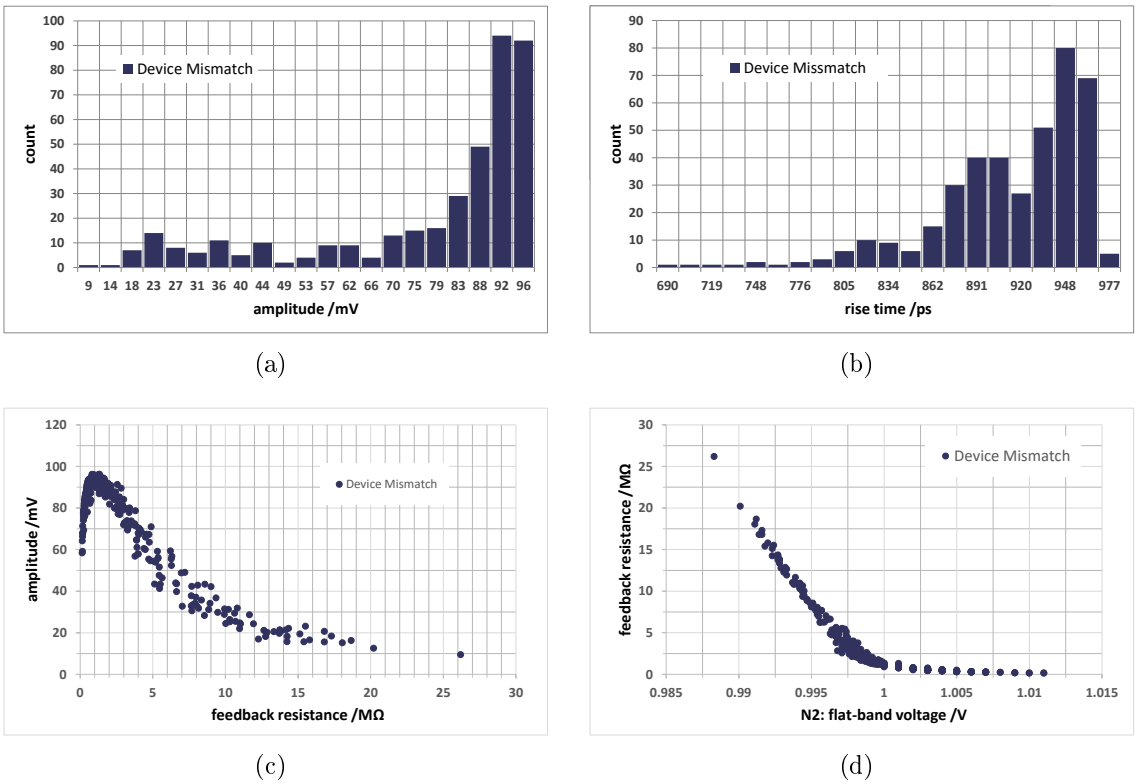


Figure 8.4: Several results of the Monte Carlo simulation with $N = 400$ runs for device mismatch.

8.1.4 Monte Carlo Analysis

The following Monte Carlo analysis is intended less to characterize the pixel than to ensure that the circuit works well over a wide operating range. Thereto, the simulation randomly varies model parameters within a range, characteristic for the used manufacturing process. A distinction is made between device mismatch and variations of process parameters. In general, the device mismatch refers to parameters, which can vary in single devises within the circuit, while the process parameters always effect the whole chip. To perform the Monte Carlo simulation a transient simulation with a length of 70 ns is set up. Again a hit signal, corresponding to a MIP signal is generated at 20 ns. This simulation is repeated $N = 400$ times with varying parameters. The number of runs is kept low, in order to reduce the simulation time. Still the number should be sufficient for a meaningful statistic.

Firstly, device mismatches are observed, with several measurements presented in figure 8.4. Figure 8.4(a) shows that the majority of the runs result in a good signal amplitude in the vicinity of 90 mV. Nevertheless, there is a not negligible number of runs, which result in a small amplitude. So there are parameters, which influence the functionality of the circuit to an extent, where it stops working correctly. Observing the rise time of the signal edge in figure 8.4(b), shows that the circuit is not slowed

down by varying device parameters. But the distribution has a tail in the direction of smaller times, which is probably connected to runs with a small amplitude. At larger rise times there is a sharp cut that is not exceeded.

It was found that those runs, where the circuit stops working correctly, are connected to a variation in the channel resistance of the feedback transistor. The correlation between the resistance and the signal amplitude is shown in figure 8.4(c). From previous studies it is already known that the CSA only works in a rough range from 0.10 MΩ to 2 MΩ for the feedback resistance. For lower resistances, the signal charge is discharged too fast and cannot be amplified by the CSA. Larger resistances result in an increase of the voltage at the output node of the amplifier and thus a limitation of its dynamic range. The reason is, the output baseline voltage is defined by the input voltage via the feedback resistance, whereby the input voltage is fixed by the bipolar transistor.

In fact, a single parameter has caused the change in feedback resistance and hence of the amplitude. The flat-band voltage of the feedback transistor, or to be more precise, its model parameter *vfbo*, describing the geometry independent part of the flat-band voltage. Thereby, the flat-band voltage [49] describes the voltage, which needs to by applied to the transistor gate to have a flat energy band in the semiconductor (compare figure 3.1(a)). The correlation between the model parameter and the feedback resistances is given in figure 8.4(d).

Since the geometry independent part of the flat-band voltage causes the change in the feedback resistance, it would not help to change the size of the feedback transistor.

A similar behaviour can be observed for the Monte Carlo simulation with varying process parameters in figure 8.5. Again there exist runs, where the circuit stops working and its gain drops. In the case of the varying process parameters the number of bad runs is even larger, indicating that several parameters influence the behaviour of the analogue pixel. The rise time is again not influenced, as it stays below 1 ns.

The change of the signal amplitude can be, once more, traced back to a change of the feedback resistance. Figure 8.5(c) shows the correlation of the two variables. In this case the model parameter called *dphibl* and *dphibo*, both connected to NMOS transistors, are accountable for the largest impact. These are the length dependent and the geometry independent part of the *dphib*-parameter, which describes an offset to the bulk potential ϕ_B of the semiconductor. Since it is a process parameter all NMOS devices are affected but the change of the feedback transistor has the most influence on the circuit performance. Figure 8.5(d) shows the correlation between the feedback resistance and the *dphibl*-voltage, which is not as clear as before, because now several parameters are important.

In a short study, with a larger feedback transistor with $L = 300\,\text{nm}$ the *dphibl*-parameter lost its negative influence. However, the geometry independent part still remains and additionally the impact of the parameter *toxo*, connected to the thickness of the NMOS gate-oxide, emerges. Therefore, this seems not to be a good solution in order to increase the broadness of the operation point.

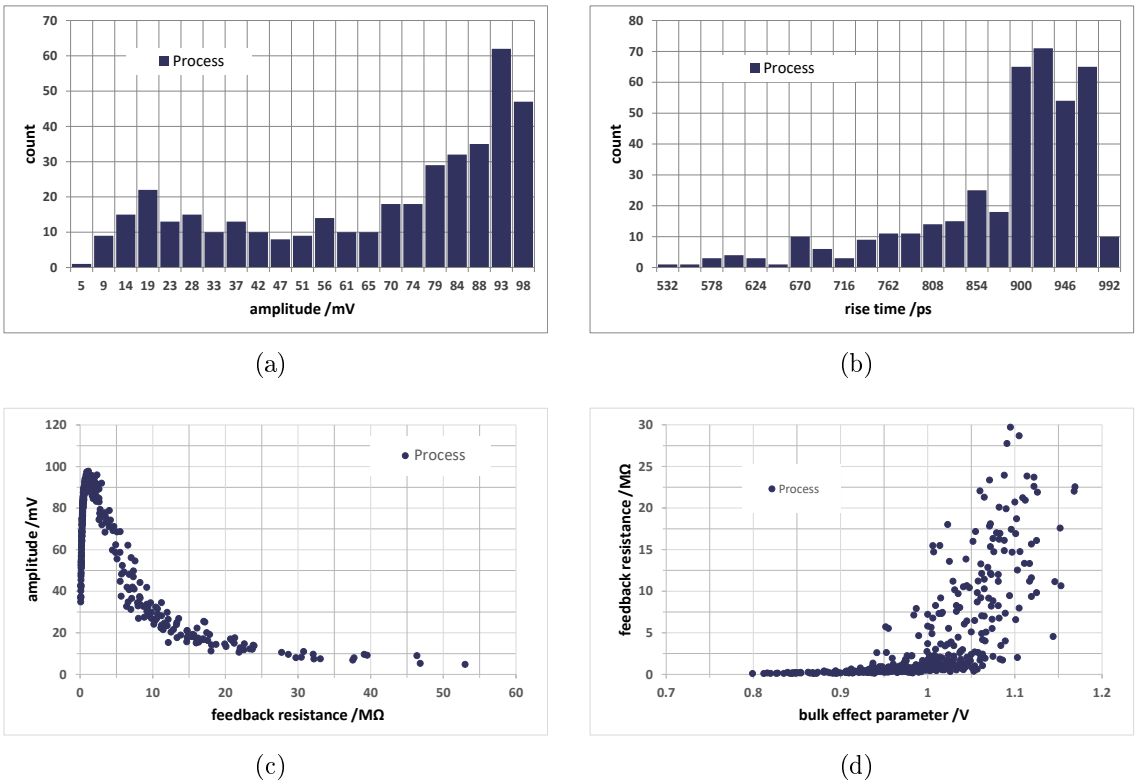


Figure 8.5: Several results of the Monte Carlo simulation for variations in process parameters.

A conclusion is that the feedback transistor is the most sensitive part of the circuit, since it influences the circuit in both simulations significantly. However, to verify that the pixel is still operable, a measure point is studied in more detail. The process parameters are therefore set to $dphibl = 1.1$ and $dphibo = 0.9$, which represents one of the worst points found in the Monte Carlo simulation. A feedback resistance of $R_f \approx 37\,\mathrm{M\Omega}$ was found here. By adjusting the bias voltages, it is possible to operate the analogue pixel in these conditions. The most efficient way to neglect the effect of the offset in the bulk potential, is to shift the ground potential and thereby the potential of the bulk. With a small offset of $gnda = -90\,\mathrm{mV}$ the circuit is approximately at its normal operation point. Another possibility is to change *VPLoad*, to adapt to the change in the output baseline voltage, which is associated with the larger feedback resistance. Thus, the dynamic range and in extent the amplitude increases. But the resistance is not changed significantly, so a large measuring time τ still remains. Finer adjustment of the feedback resistance can be made, using the bias voltage *VPFB*.

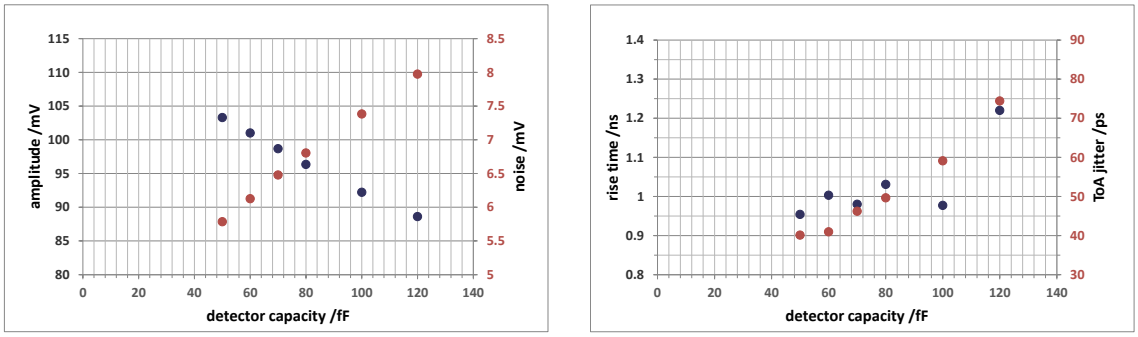
A study for the device mismatch and the parameter *vfbo* could not be made, since it is not possible to change model parameters of a single device in the transient simulation. A similar procedure as for the process parameters should also lead to success, because in both cases the threshold voltage of the feedback transistor

is influenced. It should be considered whether crucial bias voltages as *VPLoad* and *VPFB* should be provided for each pixel individually, in order to better resist negative effects.

8.1.5 The Detector Capacity

The design kit, which was provided by IHP, does not allow to extract substrate to substrate capacities. This lead to the decision to hold the detector capacity at the value of 100 fF, used for the optimisation of the amplifier. In general this is a good approximation for the pixel capacity of an HV-MAPS but then it highly depends on the size of the pixel itself. For the small design of the analogue pixel, the value is most likely too large. In comparison, the pixel of the MuPix10 with a size of $80 \times 80\,\mu\text{m}^2$ has a simulated pixel capacity of approximately 70 fF at a bias voltage of -60 V [46].

As the actual value is unknown, a simulation scan is performed with varying detector capacity C_{det}. Therefore, for each measuring point $N = 10$ transient simulations are evaluated, with the result displayed in figure 8.6.



(a) (b)

Figure 8.6: Simulation of various signal parameters depending on the detector capacity for $N = 10$ runs.

Towards a smaller detector capacity a linear behaviour of the signal amplitude and the baseline noise can be seen, as both parameters improve. This observation is consistent with the simulation from [9], where the ENC depending on the detector capacity is measured. According to the improvement of the electric noise, the ToA jitter also benefits from a smaller detector capacity. Note that this does not mean that a smaller pixel always results in a better time resolution. Effects caused by the charge collection are not taken into account. These could be enhanced because the electric field of a small diode is not necessarily ideally homogeneous.

The rise time of the signal does not show a clear behaviour. This could maybe be traced back to the small number of runs. Additionally, the circuit rise time was optimised, especially for a detector capacity of 100 fF, explaining the dip at this point.

Based on this measurement it can be assumed that the actual performance of the pixel is even better than initially expected. Especially the improvement of the amplitude and the noise behaviour could be very helpful, as will be shown in the following section.

8.2 Characterisation of the Pixel Front End

Since the most important characteristics of the analogue pixel are studied, it can be merged with its digital counterpart. Here, as well, the proper function of the entire signal chain will be guaranteed and verified. Further, the final performance will be studied. For this purpose, the subsequent circuits are gradually brought together with the analogue pixel.

8.2.1 Influence of the Connection Line

It is already known that a long trace, connecting the analogue and digital pixel, will worsen the performance of the amplifier. According to the theoretical description of the CSA in chapter 4.2.1, a load capacity decreases the bandwidth of the amplifier, and thereby the rise time of the leading edge. Furthermore, the line resistance of the trace will decrease the amplitude of the signal. This chapter aims to investigate the maximum line length to maintain the given requirements and hence give an indication for the maximum matrix size, which can be achieved.

Here, a single pixel row is observed: The width of the pixel then defines the space, which can be used to distribute the connection lines. Thereby, the width of 25 μm of the analogue pixel do not include the guard ring between the neighbouring pixels. For this reason in the following calculation, additional 7.5 μm are assumed on each side of the pixel. Minus the size of the via-stack, which connects each pixel to the supply voltage, a width of 38 μm remains, in which the connecting lines can be distributed freely. Additionally, 3 metal layers are available for this purpose.

The capacity and the resistance of the trace are determined by its size and the distance to its neighbouring traces. For a rough estimation of these values the parasitic extraction tool is used. The cross section of the observed layer stack is presented in figure 8.7. It is assumed that metal 2, the layer underneath the connection line, is a solid plain connected to ground. The results for a 1 mm long trace are given in table 8.3, depending on the gap between the traces.

It was found that the line resistance does not effect the signal quality as much as the additional load capacity. The reason is that it is dominated by the output resistance of the source follower, since these two resistances are in series. Therefore, the line width was held at its minimum, resulting in the maximal resistance and likewise the minimal capacity. Table 8.3 shows that the available space is not a huge problem, as the line capacity does not increase significantly if the density of the traces is increased. However, the length of the line needs to be taken into account, whereby capacity and resistance scale linearly with the line length.

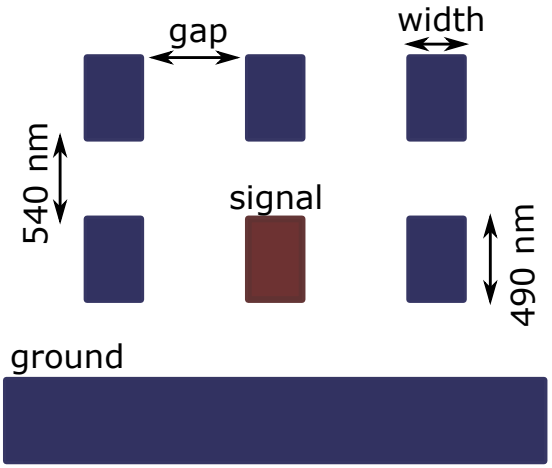


Figure 8.7: Trace stack for the connection line. The resistance and capacity of the red trace is extracted, while all other nets are connected to ground.

width [nm]	gap [nm]	capacity [fF/mm]	resistance [Ω/mm]	number pixels
200	4410	134	440	25
200	1340	142	440	75
200	720	150	440	125
200	420	183	440	180

Table 8.3: Estimated capacity and resistance of a trace line depending on its width and distance to a neighbour line for a length of 1 mm.

In order to study the influence of the connection line on the analogue pixel, it is introduced as ideal resistance R_L and capacity C_L in a low pass configuration at the output of the pixel. The related values are given in table 8.3, whereby the length of the trace has to be considered. The maximal length is approximated by the number of pixels, which can be connected, multiplied by their width. Further, it is considered that after half of the length, half of the pixels are already connected, thus twice the space is available for the remaining lines. Note, the used model gives no prediction regarding cross-talk to neighbouring traces.

The measurement of various crucial signal parameters is presented in table 8.4, whereby for each set $N = 10$ runs were performed. As a reference point, the analogue pixel without the connection line is also given. As the load of the analogue pixel increases, various effects can be observed. According to equation 4.7, the unity gain bandwidth decreases with C_L and thus influences the amplitude and the rise time of the signal. Besides, due to the lower bandwidth the noise level is decreasing. Thereby, the SNR shows clearly, the noise level is reduced faster then the amplitude. The reason is that the noise contains frequencies well above the signal frequency, which form a large amount of the total noise power. While this is in principle a good consequence for the signal quality, the amplitude must not become too small,

length [mm]	R_L [kΩ]	C_L [fF]	V_{amp} [mV]	t_{rise} [ns]	σ_V [mV]	SNR	σ_{ToA} [ps]
-	0	10	92.40	0.997	7.39	12.63	61.11
0.5	0.22	65	88.87	1.283	6.31	14.37	81.33
1	0.44	130	82.92	1.990	5.37	15.80	107.1
2	0.88	260	74.98	2.873	4.18	18.30	175.1
3	1.32	427	66.58	4.226	3.20	21.22	237.0
4	1.76	560	61.10	4.910	2.63	23.70	293.3
5	2.20	719	55.49	5.539	2.22	25.58	379.5

Table 8.4: Characteristic parameters of the analogue pixel for different lengths of the connection line. For each measuring point a set of $N = 10$ runs was performed. For the reference with no trace $N = 100$ runs were used.

in order to ensure a properly working comparator. Furthermore, the simulated time resolution, the ToA jitter, increases significantly. Already at a trace length of 2 mm it has more than doubled.

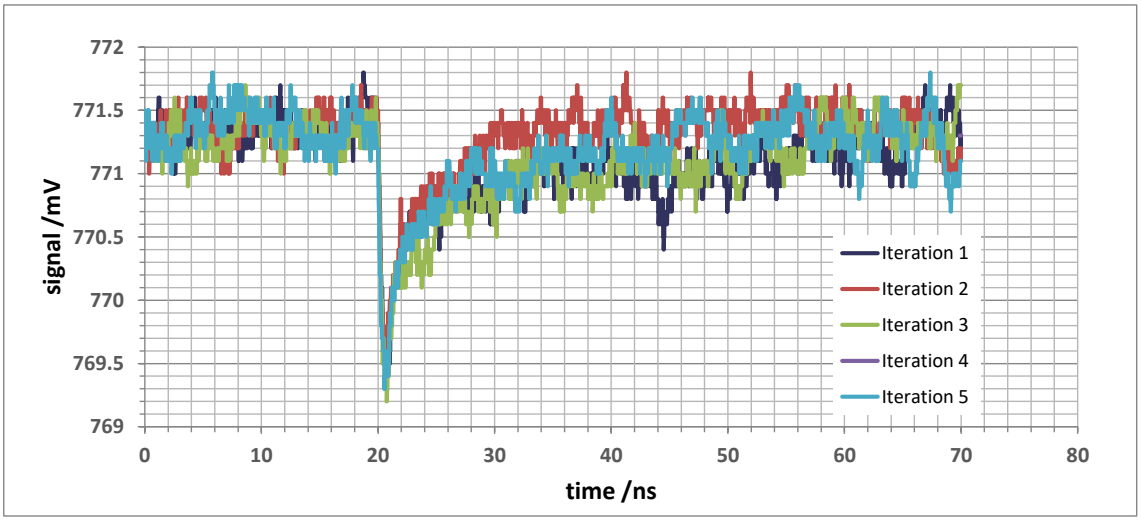
Therefore, to stay within the scope of this thesis to achieve an optimal time resolution, it must be recommended to keep the track length as short as possible. In order to increase the size of the matrix anyway, it is possible to read out the data via two or more edges of the matrix. For all further measurements a matrix length of 0.5 mm is assumed.

8.2.2 Transient Response of the Pixel Front End

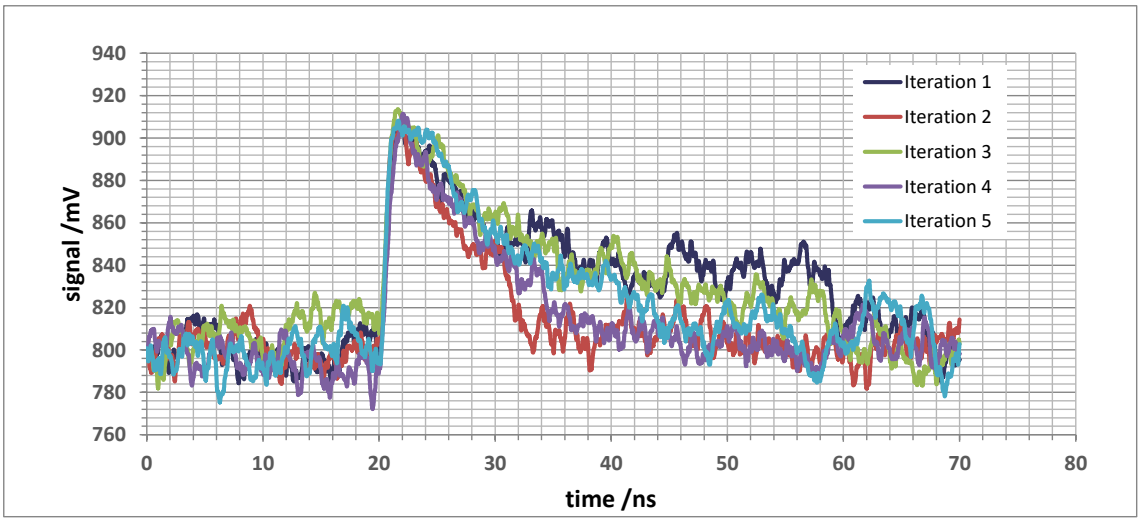
In a last step, the transient response of the whole readout chain (compare figure 4.6) up to the digitisation by the comparator can be studied. Thereto, the analogue pixel is connected via the connection line to the digital pixel. Again the connection line is approximated by an ideal resistor $R_L = 220\,\Omega$ and capacity $C_L = 65\,\text{fF}$, representing a trace length of 0.5 mm. Besides, a workbench is used, which provides both pixels with ideal bias voltages, including the threshold voltage at the comparator. A pixel capacity of $C_{det} = 100\,\text{fF}$ is added and in order to reduce simulation time, only the parasitics of the analogue pixel are considered.

Figure 8.8 shows the signal as it propagates through the different parts of the circuit. The small voltage drop at the amplifier input, due to the induced signal charge, can be seen in figure 8.8(a). In a next step the signal is amplified by the CSA, whereby its output is represented in figure 8.8(b). The amplification results in a signal amplitude of roughly 100 mV and a rise time of 1 ns. Here, it should be noted, this is not the same point discussed as in section 8.1.1, which included the whole analogue pixel.

Proceeding from the amplifier output, the signal is processed by the source follower and driven through the connection line. At the digital pixel it is electrically decoupled from the analogue circuit by the input capacity, implemented as a transistor gate capacity. Figure 8.8(c) shows this point, the comparator input together

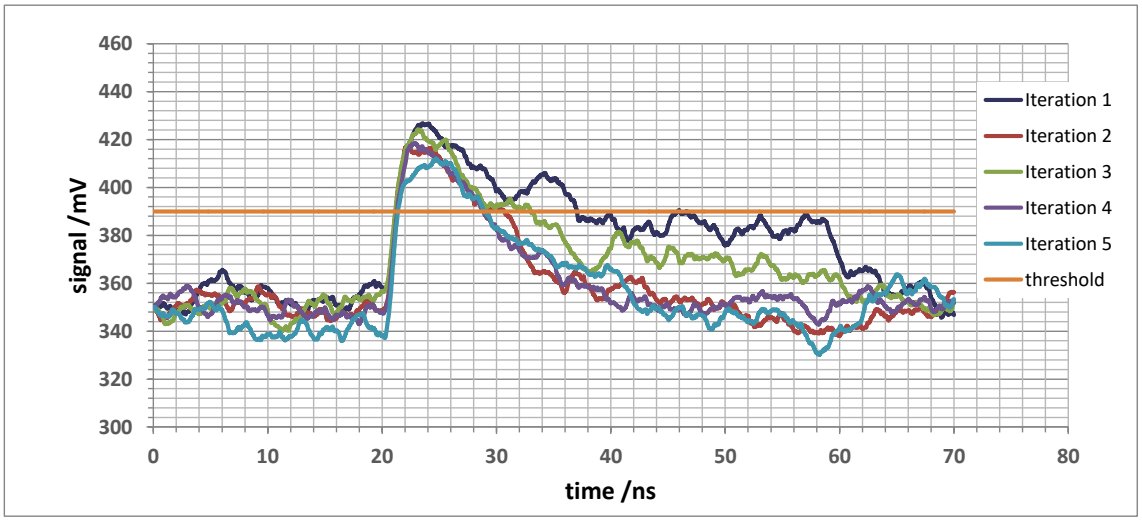


(a)

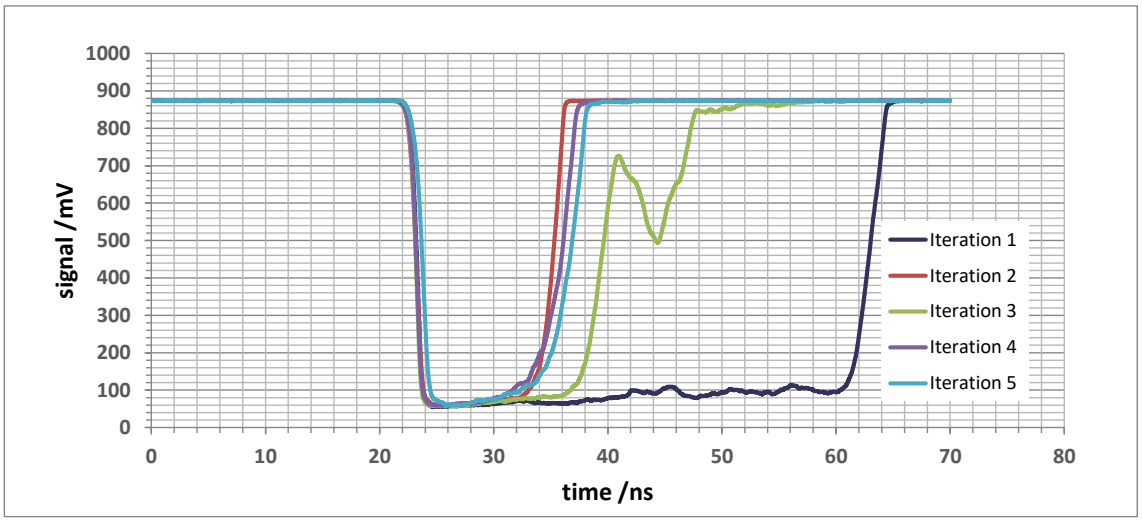


(b)

Figure 8.8: Transient response of the whole pixel front end electronics for a 70 ns long transient simulation, with a test signal induced at 20 ns. The signal is shown as it propagates through the different parts of the circuit, starting with the input at the CSA (a), its output (b), both inputs of the comparator (c) and its output (d). (Continued on the next page)



(c)



(d)

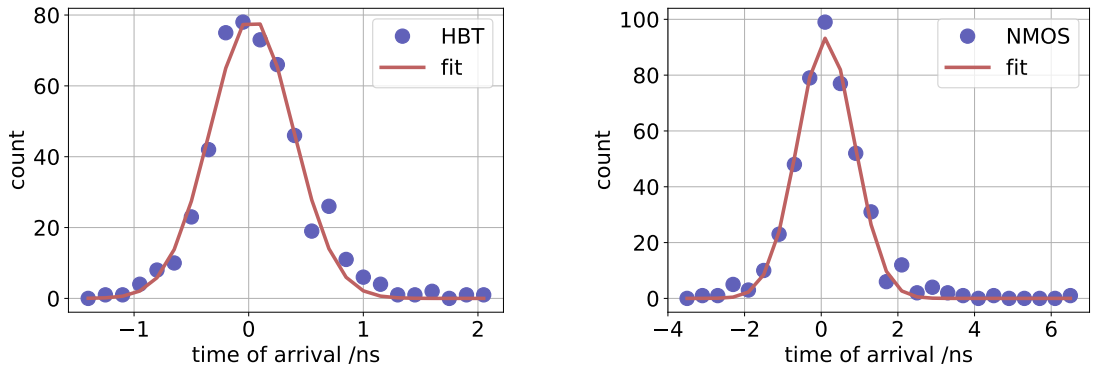
Figure 8.8: Transient response of the whole pixel front end electronics for a 70 ns long transient simulation, with a test signal induced at 20 ns. The signal is shown as it propagates through the different parts of the circuit, starting with the input at the CSA (a), its output (b), both inputs of the comparator (c) and its output (d).

with the ideal threshold voltage. From a brief look it seems like the threshold voltage is set to high, but it must not be forgotten, the switching point of the comparator is not ideal. Hence, it already switches when the signal approximately attains a voltage of 370 mV. An *ITune* bias voltage of 300 mV was used, to adjust the switching point as close as possible to the baseline.

The comparison between figure 8.8(b) and figure 8.8(c) demonstrates that the high-frequency noise component is significantly lower. Besides, the amplitude of the signal decreases. These effects were discussed in the previous section and are the result of the limited bandwidth of the connection line and the source follower. Due to the choice of a short trace, the rise time is not yet influenced. In fact, the measured value will be smaller because the amplitude decreases. In this case the slope remains unchanged.

The response of the comparator to the signal is presented in figure 8.8(d). Here, the analogue signal from the sensor is converted into digital information, whereby possible subsequent circuits can distinguish between the two logic level easily, due to the large amplitude of approximately 900 mV. The slope of the signal depends on the amplitude of the input signal and how far it emerges above the switching point. Hence, the ToA jitter on the leading edge depends not only on the ToA jitter of the input signal, but also on its noise. The same applies for switching back into the normal state, which is highly depend on the noise of the falling edge of the signal.

For a more detailed study of the ToA jitter and thus the time resolution of the whole FE, these transient simulations are repeated $N = 500$ times. The resulting ToA jitter measurement is presented in figure 8.9, whereby the data is corrected for the circuit delay by subtracting the measured mean value. As a reference the measurement is repeated using the NMOS analogue pixel.



(a) (b)

Figure 8.9: Distribution of the corrected hit times from $N = 500$ simulation runs for the bipolar (a) and the NMOS (b) analogue pixel. A Gaussian function is fitted to the data, whereby hits triggered by noise are excluded. A bin size of 0.15 ns is used for (a) and 0.4 ns for (b).

The indicated Gaussian fit curve was not used for further measurements, but to illustrate the distribution of the data. Although it shows a good agreement, the data is not really Gaussian distributed, which manifests itself in a small tail to higher values. As there can be no hit signal before the test signal is induced into the circuit, the distribution is capped at its lower end. In that sense it is more of a Poisson distribution at its transition to a Gaussian distribution.

The variance in the hit time, hence the ToA jitter, is directly measured from the collected data. Thereby, hits which were triggered by noise are not considered and were excluded from the measurement. In the case of the HBT one run out of the 500 had to be excluded. As a result a ToA jitter of $\sigma_{ToA,HBT} = 435.0\,\text{ps}$ was measured. For the reference pixel 42 out of 500 runs had to be excluded, resulting in a ToA jitter of $\sigma_{ToA,NMOS} = 993.1\,\text{ps}$.

This leads to several conclusions: First, the higher number of excluded data points for the reference pixel at the same settings for the threshold voltage indicate a higher noise connected to the NMOS input transistor. This was already verified in section 8.1.1. A higher threshold setting would have also lead to problems, as the noise also effect the amplitude. Thus, several measurements with a high latency would have to be excluded.

Further, the excellent timing behaviour, measured for the analogue pixel could not be reconfirmed for the whole front end electronics. The time resolution is significantly enlarged by the digital pixel, whereby the version with the HBT again shows a clearly better performance. In general, the ToA jitter is for both cases still in the target regime below one nanosecond, but the worsening has to be analysed.

At the comparator input, a signal amplitude in the order of 70 mV is measured. The loss of approximately 25 % of the amplitude, compared to the analogue pixel output, is caused by the connection line and the capacity, decoupling the analogue and digital part. Due to noise, the actual amplitude of the individual runs varies and thus the height of the signal, emerging above the threshold voltage. This results in a large variation in the ToA, as the comparators slew rate highly depends on the height of the signal, especially in the vicinity of the switching point. Here, the characteristics of the comparator as a differential amplifier become apparent. With a small difference between the two input voltages, the currents in the two branches of the differential pair differ only by a small amount. The output node is therefore discharged slower, reducing the switching speed of the comparator.

Accordingly, the ToA jitter at the comparator output is dominated by the variance of the amplitude and not the ToA jitter of the input signal. The excellent timing behaviour of the bipolar transistor is therefore not the decisive factor after the signal passes the comparator. However, the smaller noise related to the HBT clearly helps to improve the time resolution of the whole circuit.

As a consequence, the question must be asked whether a further optimisation of the comparator is possible in order to better utilise the performance of the HBT. An definite answer to this question could not be found within the time frame of this work. But it is possible to give some hints, which could improve the overall time resolution: As discussed, the amplitude of the signal is a crucial factor for the

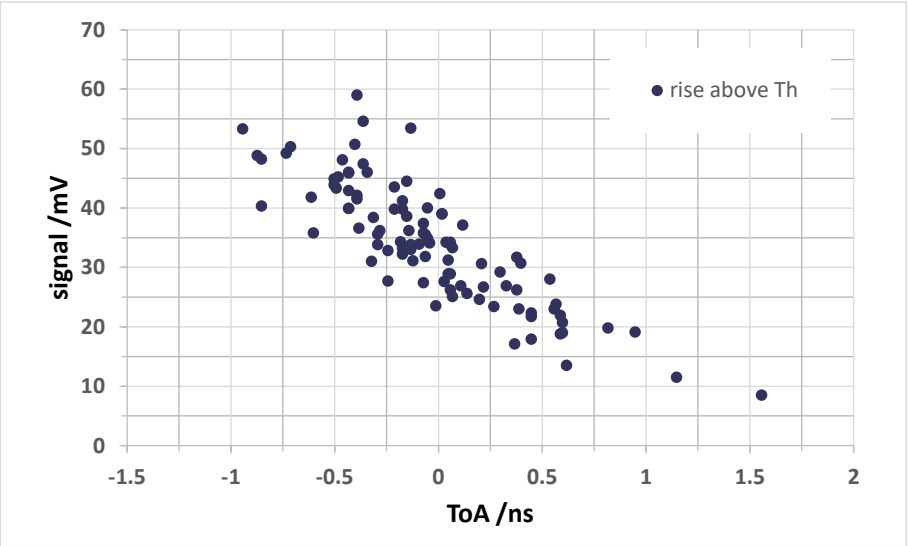


Figure 8.10: Correlation between the height of the signal, emerging above the threshold voltage *Th* and the ToA jitter.

performance of the comparator. The previous studies showed that an increase of the gain of the circuit is possible, e.g. by changing the size of the current source of the CSA. Thereby, it is important not to increase the noise of the signal. Another simple possibility to increase the amplitude at the comparator input is to increase the input capacity of the digital pixel. This makes a significant contribution to the loss of amplitude. Certainly, the size of the corresponding transistors is already dominating the layout of the digital pixel.

A last possibility could be to correct the time resolution offline. The described effect is very similar to the effect of the time walk. Both lead to a deterioration of the time resolution, depending on the signal amplitude. The clear correlation between the simulated ToA jitter and the signal height above the threshold voltage is shown in figure 8.10. A correction, which is anyway necessary for the time walk, would address both effects, as there is no possibility to distinguish them in an experiment.

9 Discussion and Summary

The goal of this thesis was to design and build up the front end electronics for a high-voltage monolithic active pixel sensor. This includes the analogue pixel, with the sensor diode and, in the sense of the monolithic architecture, the amplification circuity. The signal processing is further observed and designed, until the digitisation in the digital pixel. While designing the circuity, the main goal was to achieve a time resolution for the sensor pixel in the sub-nanosecond regime. Thereto, the advantages of a BiCMOS technology, the SG13S process by IHP, are utilised.

From the theoretical description it was known that the time resolution is connected to the rise time of the signal edge, divided by the SNR. Based on this, several requirements for the layout could be formulated: For a signal $Q_s \approx 2800\,\mathrm{e}^-$, corresponding to a hit by a MIP, a rise time of the order of 1 ns and amplitude above 100 mV is targeted. The noise of the CSA was minimised by reducing the measuring time $\tau = R_f C_f$.

With this boundary conditions, the circuits were optimised. For the amplifier, it was found that a single gain stage with cascoded input transistor has the best qualifications. A key factor for the decision to use the cascode was the higher gain provided by it. The feedback is realised as an active feedback, where the feedback resistance is implemented by the transistor channel resistance with a value of $R_f = 1.1\,\mathrm{M\Omega}$. This resistance adjusts automatically when a signal is amplified so that the falling edge of the signal is approximately linear. The feedback capacity is thereby dominated by the drain-substrate capacity $C_f \approx 2\,\mathrm{fF}$ of the current source of the amplifier.

The result was a finalised version of the analogue pixel, presented in figure 7.11, with a size of $25 \times 25\,\mu\mathrm{m}^2$. Besides the amplifier, it holds a source follower, a bias block, an injection circuit and large input capacities. All wiring is done in the metal 1 layer, leaving the metal 2 layer to distribute the shared bias voltages and metal 3 to 5 for the point-to-point connection between analogue and digital pixel. Top metal 1 and 2 were reserved for the supply voltage and ground.

The required low power consumption, given by the Mu3e experiment could not be maintained. It could be shown that the transit frequency of the bipolar transistor is highly dependent on its collector current. With the used value of $21.5\,\mu\mathrm{A}$ it is already operated at its lower limit, way below the peak performance. Besides, the small pixel size even increases the power consumption per area to $P_{analogue} = 1910\,\mathrm{mW/cm^2}$. Regarding the power of one channel, instead of the power per area, it is shown that the designed pixel front end electronics indeed has a low power consumption of $P_{ch} = 41\,\mu\mathrm{W/ch}$. As a comparison the Mupix sensor with typical power consumption of $P_{MuPix} = 16\,\mu\mathrm{W/ch}$ and the TT-PET demonstrator with $P_{TT-PET} = 375\,\mu\mathrm{W/ch}$ in

	σ_{ToA} [ps]	σ_t [ps]	ENC [e$^-$]
analogue pixel	86.12	99.35	205
reference pixel	301.4	240.1	230

Table 9.1: Summary of some crucial values of the analogue and the reference pixel.

the high power mode can be mentioned. Accordingly, the originally set target could not be met, but the power consumption could be held at a minimum. Connected to the power consumption several parameters of the analogue pixel change. In addition to the aforementioned high-frequency performance of the bipolar transistor, its noise contribution also decreases for a higher current. To reach a higher current, the size of the current source has to be increased. But a short study showed that a smaller current source is beneficial for the amplitude and the rise time of the signal, as C_f is reduced. At the same time, however, the noise level rises.

In summary, the operation point with the chosen current and size of the current source, showed an excellent performance, as presented in table 9.1. It could not be verified within this thesis, if it is the optimal design. Therefore, a detailed study concerning the size of the current source has to be made, also allowing for a higher power consumption. It should be noted, thereto an adjustment of the layout is necessary for every test point and in extent the transient noise simulation is time-consuming. For this thesis only a study within an ideal environment was performed, which does not cover all effects.

Table 9.1 summarises the most crucial parameters measured for the analogue pixel and the reference pixel. Accordingly, it could be shown that a time resolution far below one nanosecond can be achieved with BiCMOS technology and the designed layout. In terms of time resolution the performance of the analogue pixel is thereby significant better compared to reference pixel. However, also with the reference pixel a time resolution in the sub-nanosecond regime could be achieved. This performance could be maintained over a broad range. As it was shown, only the feedback resistance is sensitive to changes in process or device parameters. Unfortunately, it is not possible to overcome this sensitivity by changing the size of the feedback transistor, but it is possible to adjust the operation point by changing the bias voltages. In order to achieve the best results, it is advisable to control every pixel separately.

Not measured within this thesis was the true value of the pixel capacity. The value of 100 fF, used for the characterisation, is most probably set too high, as a comparison reveals with the pixel capacity of the MuPix10 with a value of 70 fF. Based on this, it was shown that parameters as amplitude and noise of the analogue pixel scale linearly with the detector capacity. Thus, a small improvement is expected for the small pixel layout, when the true pixel capacity is used. Conversely, it is not advisable to make the pixel significantly larger, because then much more power is needed to achieve the same performance.

Including the connection line to the simulation, resulted in the proposal to keep the matrix size as small as possible. Thus, the capacitive load, which must be driven,

is as small as possible. In order to double the length of the matrix anyway, it is possible to locate digital pixels at two edges of the matrix. Following other projects, like the ATLASPix1 [50], it is also possible to include the comparator into the sensor pixel. Especially the ATLASPix1 showed good results using such a configuration. With the comparator inside the pixel, the load capacity would be minimized and the bandwidth increased. Thus, the excellent signal quality would be maintained until the signal is digitised. Drawbacks would be that the pixel size would have to be increased and the amplifier would see noise from the digital circuit.

In a last step, the whole pixel front end electronics was investigated with a matrix length of 0.5 mm. In a set of 500 simulations, a ToA jitter of $\sigma_{ToA,HBT} = 435.0\,\text{ps}$ was found. In comparison, for the reference pixel a ToA jitter of $\sigma_{ToA,NMOS} = 993.1\,\text{ps}$ was measured. Although this is still a very good result, it shows that the digital pixel cannot maintain the same performance as the analogue one. The question arises, whether the comparator can be further optimised, to make better use of the results of the analogue pixel.

9.1 Conclusion

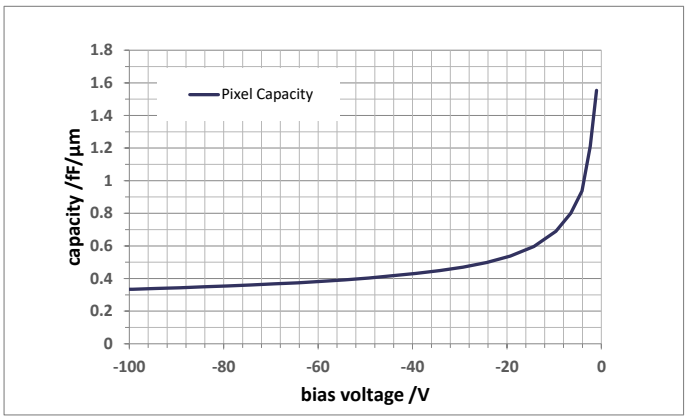
Simulations have successfully shown that a time resolution in the sub-nanosecond range is possible for a HV-MAPS. The use of a BiCMOS technology, such as the SG13S process from IHP, provides clear advantages. Thus, a significant smaller time resolution was measured.

The functionality of the pixel front end electronics with a small pixel layout and a low power consumption has been proven in various simulations. Same holds for the quality of a signal, related to a MIP.

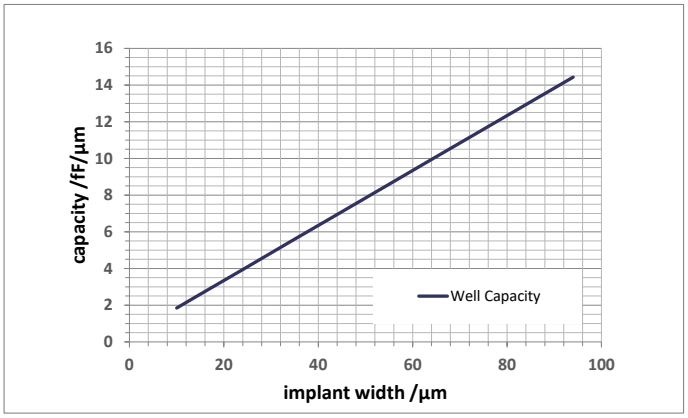
The design has been finalised to the point that it could possibly be implemented into a sensor chip. Yet some questions still remain unanswered: This includes the question of a detailed study of the noise, associated with the size of the current source and the power consumption. Furthermore, it remains unsolved whether the time resolution of the digital pixel can be improved and matched to that of the analogue pixel.

Appendix

A MuPix10 Pixel Capacity



(a)



(b)

Figure A.1: 2D-TCAD simulation of the diode capacity (a) formed by the deep n-well and substrate and the well capacity (b) of the p-well inside the anode. A substrade resistivity of 200 Ωcm was assumed [46].

Based on the TCAD simulations, presented in figure A.1, the capacity of the Mupix10 can be calculated. The simulation was done in two dimensions, thus the result hast to be multiplied by the width of the pixel or p-well. Accordingly, the bare diode capacity is $C_{diode} = 3.8\,\mathrm{fF}/\mu\mathrm{m} \cdot 80\,\mu\mathrm{m} = 30.4\,\mathrm{fF}$ at a bias voltage of $-60\,\mathrm{V}$. For the p-well implant a capacity of $C_{well} = 2.4\,\mathrm{fF}/\mu\mathrm{m} \cdot 15\,\mu\mathrm{m} = 36\,\mathrm{fF}$ is simulated, whereby the size of the implant is roughly $14 \times 15\,\mu\mathrm{m}^2$. This results in a total pixel capacity of $C_{det} = 66.4\,\mathrm{fF}$.

B Reference Pixel

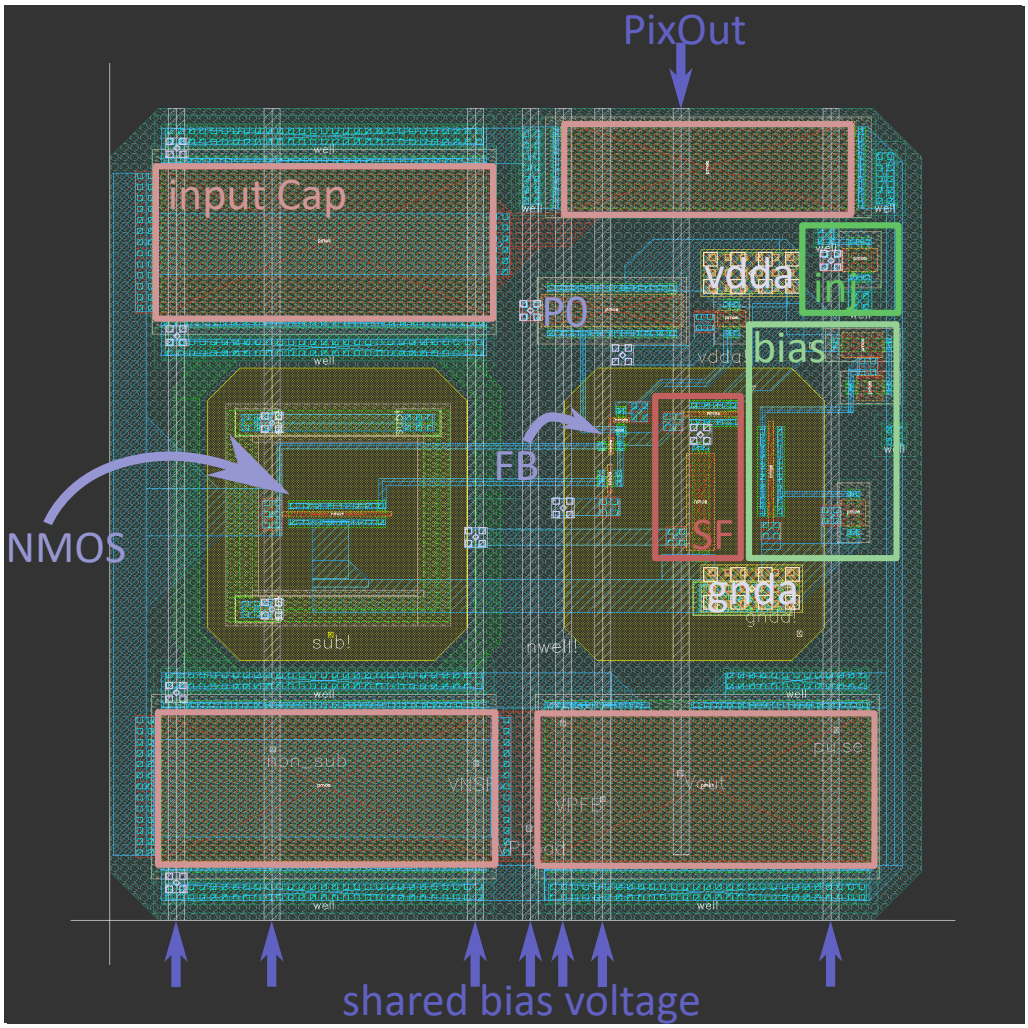


Figure B.1: Layout of the reference Pixel with NMOS input transistor.

circuit name	bias name	purpose	value
CSA	VPLoad	tunes current in CSA	465 mV
	Casc	controls cascode transistor	900 mV
	VPFB	tunes feedback resistance	980 mV
	Sub	substrate voltage of the NMOS	0 mV
SF	VNSF	tunes current in SF	480 mV
Injection	Pulse	injects test pulse	0 mV

Table B.1: List of all bias voltages, which are needed to operate the reference pixel.

C Bibliography

[1] A. Blondel et al. "Research Proposal for an Experiment to Search for the Decay $\mu \to eee$". 1 2013. arXiv:1301.6113.

[2] Y. Bandi et al. "The TT-PET project: a thin TOF-PET scanner based on fast novel silicon pixel detectors". IN: *Journal of Instrumentation*, vol. 13.01 (2018). DOI:10.1088/1748-0221/13/01/c01007.

[3] H. Rücker et al. "A 0.13 μm SiGe BiCMOS Technology Featuring f $_T$/f $_{\max}$ of 240/330 GHz and Gate Delays Below 3 ps". IN: *IEEE Journal of Solid-State Circuits*, vol. 45.09 (2010). DOI:10.1109/JSSC.2010.2051475.

[4] P. Baesso et al. "Technical design of the Phase I Mu3e Experiment". Tech. rep., Mu3e Experiment. In Preparation.

[5] J. Hammerich. "Analog Characterization and Time Resolution of a large scale HV-MAPS Prototype". Master's thesis, Heidelberg University. 2018.

[6] H. Augustin. May 2020. Private Conversation.

[7] S. Bravar et al. "Scintillating fibre detector for the Mu3e experiment". IN: *Journal of Instrumentation*, vol. 12.07 (2017). DOI:10.1088/1748-0221/12/07/c07011.

[8] H. Eckert. "The Mu3e Tile Detector". Ph.D. thesis, Heidelberg University. 2015.

[9] P. Valerio et al. "A monolithic ASIC demonstrator for the Thin Time-of-Flight PET scanner". IN: *Journal of Instrumentation*, vol. 14.07 (2019). DOI:10.1088/1748-0221/14/07/p07013.

[10] L. Paolozzi et al. "Characterization of the demonstrator of the fast silicon monolithic ASIC for the TT-PET project". IN: *Journal of Instrumentation*, vol. 14.02 (2019). DOI:10.1088/1748-0221/14/02/p02009.

[11] G. Iacobucci et al. "A 50 ps resolution monolithic active pixel sensor without internal gain in SiGe BiCMOS technology". IN: *Journal of Instrumentation*, vol. 14.11 (2019). DOI:10.1088/1748-0221/14/11/p11008.

[12] H. Rücker et al. "Half-Terahertz SiGe BiCMOS technology". IN: "IEEE 12th Topical Meeting on Silicon Monolithic Integrated Circuits in RF Systems", 2012 DOI:10.1109/SiRF.2012.6160164.

[13] S. Hunklinger. "Festkörperphysik". De Gruyter, Berlin, 2018, 5 edn. ISBN 978-3-11-056775-5.

[14] R. Gross and A. Marx. "Festkörperphysik". De Gruyter, Berlin, 2018, 3 edn. ISBN 978-3-11-055918-7.

[15] H. Ibach and H. Lüth. "Festkörperphysik". Springer Verlag, Berlin, 2009, 7 edn. ISBN 978-3-540-85794-5.

[16] F. Bloch. "Über die Quantenmechanik der Elektronen in Kristallgittern". IN: *Zeitschrift für Physik*, vol. 52:555–600 (1929). DOI:10.1007/BF01339455.

[17] P. Drude. "Zur Elektronentheorie der Metalle; II. Teil. Galvanomagnetische und thermomagnetische Effecte". IN: *Annalen der Physik*, vol. 308.11 (1900). DOI:10.1002/andp.19003081102.

[18] G. Lutz. "Semiconductor Radiation Detectors". Springer Verlag, Berlin, 2007, 2 edn. ISBN 978-3-540-71678-5.

[19] P. S. Olivier Deleage. "Modes of operation of an N-type MOSFET". 2008. URL: `https://commons.wikimedia.org/wiki/File:MOSFET_functioning.svg` Accessed 23. April 2020.

[20] P. Fischer. "Transistors". IN: "Components, Circuits & Simulations", University Lecture, ZITI Uni Heidelberg, 2018 URL: `https://sus.ziti.uni-heidelberg.de/Lehre/SS20_CCS/CCS_Fischer_08_Transistors.pptx.pdf` Accessed 23. April 2020.

[21] H. K. Gummel and H. C. Poon. "An integral charge control model of bipolar transistors". IN: *The Bell System Technical Journal*, vol. 49.05:827–852 (1970). DOI:10.1002/j.1538-7305.1970.tb01803.x.

[22] J. M. Early. "Effects of Space-Charge Layer Widening in Junction Transistors". IN: *Proceedings of the IRE*, vol. 40.11 (1952). DOI:10.1109/JRPROC.1952.273969.

[23] H. Spieler. "Semiconductor Detector Systems". Clarendon Press, Oxford, 2005, 1 edn. ISBN 978-0-198-52784-8.

[24] H. Rücker and B. Heinemann. "SiGe HBT Technology". IN: "Silicon-Germanium Heterojunction Bipolar Transistors for mm-Wave Systems", (editors) N. Rinaldi and M. Schröter River Publishers, 2018. ISBN 978-8-793-51961-9.

[25] J. C. Bardin. "Silicon-Germanium Heterojunction Bipolar Transistors For Extremely Low-Noise Applications". Ph.D. thesis, California Institute of Technology. 2009.

[26] M. Bellini. "Operation of silicon-germanium heterojunction bipolar transistors on silicon-on-insulator in extreme environments". Ph.D. thesis, Georgia Institute of Technology. 03 2020.

[27] P. Fischer. "Readout Electronics". IN: "Silicon Detectors", University Lecture, ZITI Uni Heidelberg, 2018 URL: https://sus.ziti.uni-heidelberg.de/Lehre/WS1920_Detectors/Part5_Electronics.pptx.pdf Accessed 25. April 2020.

[28] J. D. Schipper. "Noise sources in MOSFET transistors". nikhef Amsterdam. 1999. URL: https://www.nikhef.nl/~jds/vlsi/noise/sansen.pdf Accessed 25. April 2020.

[29] A. Konczakowska and B. M. Wilamowski. "Noise in Semiconductor Devices". IN: "Fundamentals of Industrial Electronics", (editors) B. M. Wilamowski and J. D. Irwin CRC Press, 2017. ISBN 978-1-138-07439-2.

[30] H. Augustin. "Characterization of a novel HV-MAPS Sensor with two Ampli cation Stages and first examination of thinned MuPix Sensors". Master's thesis, University Heidelberg, Physikalischen Institut. 2015.

[31] S. M. Seltzer and M. J. Berger. "Improved procedure for calculating the collision stopping power of elements and compounds for electrons and positrons". IN: *The International Journal of Applied Radiation and Isotopes*, vol. 35.7 (1984). DOI:10.1016/0020-708X(84)90113-3.

[32] P. Fischer. "Signals & Reconstruction". IN: "Silicon Detectors", University Lecture, ZITI Uni Heidelberg, 2018 URL: https://sus.ziti.uni-heidelberg.de/Lehre/WS1920_Detectors/Part3_Signals.pptx.pdf Accessed 29. April 2020.

[33] J. M. Miller. "Dependence of the input impedance of a three-electrode vacuum tube upon the load in the plate circuit". IN: *Scientific Papers of the Bureau of Standards*, vol. 15.367 (1919).

[34] P. Fischer. "Basic Circuits". IN: "Components, Circuits & Simulations", University Lecture, ZITI Uni Heidelberg, 2018 URL: https://sus.ziti.uni-heidelberg.de/Lehre/SS20_CCS/CCS_Fischer_09_Circuits.pptx.pdf Accessed 25. April 2020.

[35] D. K. Shedge et al. "Analysis and Design of CMOS Source Followers and Super Source Follower". IN: *ACEEE International Journal on Control System and Instrumentation*, vol. 4.2 (2013). DOI:01.IJCSI.4.21281.

[36] Cadence Design Systems Inc. "Virtuoso Schematic Editor L and XL". 2014. URL: https://www.cadence.com/content/dam/cadence-www/global/en_US/documents/tools/custom-ic-analog-rf-design/virtuoso-vse-fam-ds.pdf Accessed 02. May 2020.

[37] Cadence Design Systems Inc. "Virtuoso Layout Suite XL". 2016. URL: https://www.cadence.com/content/dam/cadence-www/global/en_US/documents/tools/custom-ic-analog-rf-design/virtuoso-vlsxl-ds.pdf Accessed 02. May 2020.

[38] Innovations for High Performance Microelectronics. "Low-Volume & Multi-Project Service". URL: https://www.ihp-microelectronics.com/en/services/mpw-prototyping/sigec-bicmos-technologies.html Accessed 02. May 2020.

[39] M. Birkholz et al. "Technologiemodule des IHP für die Lebenswissenschaften". Tech. rep., IHP. 2015. URL: http://www.bioprocess.tu-berlin.de/fileadmin/fg187/JLB/Broschüre_Techmod_12.pdf Accessed 04. May 2020.

[40] F.Sischka. "VBIC MODELING HANDBOOK". Tech. rep., Keysight Technologies©. 2017.

[41] X. Li et al. "PSP 103.1". Tech. rep., NXP Semiconductors©. 2009. TECHNICAL NOTE NXP-R-TN-2008/00299.

[42] G. G. Fischer. "VBIC Models for High-Speed and HighVoltage HBTs in IHP's 0.13µm BiCMOS Technology". Tech. rep., IHP. 2010. URL: https://www.iee.et.tu-dresden.de/iee/eb/forsch/AK-Bipo/2010/AKB2010_04_Fischer_VBIC_for_HBTs_of_IHPs_013um_BiCMOS_process.pdf Accessed 08. May 2020.

[43] M. Wietstruck et al. "Development of a Through-Silicon Via (TSV) Process Module for Multi-project Wafer SiGe BiCMOS and Silicon Interposer". IN: *IEEE 68th Electronic Components and Technology Conference (ECTC)*, (2018). DOI:10.1109/ECTC.2018.00341.

[44] M. Tanabashi et al. "Review of Particle Physics: Particle Data Group". IN: *Physical Review D*, vol. 98 (2018). DOI:10.1103/PhysRevD.98.030001.

[45] L. Paolozzi. "Development of particle detectors and related Front End electronics for sub-nanosecond time measurement in high radiation environment." Ph.D. thesis, University of Roma. 12 2014.

[46] A. M. Gonzalez. "Phd Thesis in Preparation". Ph.D. thesis, Heidelberg University.

[47] I. Perić. April 2019. Private Conversation.

[48] Cadence Design Systems Inc. "Application Notes on Direct Time-Domain Noise Analysis using Virtuoso Spectre". 2016. URL: http://www.lumerink.com/courses/ECE614/Handouts/Transient%20Noise%20Simulation.pdf Accessed 15. May 2020.

[49] K. Piskorski and H. Przewlocki. "The methods to determine flat-band voltage VFB in semiconductor of a MOS structure". IN: *MIPRO*, (2010).

[50] M. Kiehn et al. "Performance of the ATLASPix1 pixel sensor prototype in ams aH18 CMOS technology for the ATLAS ITk upgrade". IN: *Journal of Instrumentation*, vol. 14.08 (2019). DOI:10.1088/1748-0221/14/08/c08013.

Danksagung

Ich möchte die letzte Seite dazu nutzen, mich bei allen zu bedanken, die mich bei dieser Arbeit unterstützt haben und deren Tipps und Anregungen meine Arbeit an so mancher Stelle verbessert haben.

Insbesondere bedanke ich mich bei Herrn Prof. André Schöning, der es mir ermöglicht hat, in dem spannenden Feld des Chip Designs zu forschen und damit eine Neugier geweckt hat, die ich hoffentlich in den kommenden Jahren weiter verfolgen kann. Weiter möchte ich meinen Dank aussprechen an Herrn Prof. Norbert Herrmann, der nach meiner Bachelorarbeit nun auch meine Masterarbeit als Zweitkorrektor begleitet.

Besonders zu Beginn der Arbeit wurde ich durch Herrn Prof. Ivan Perić unterstützt, der mir dabei half, den vollen Umfang des Themengebietes zu erfassen und eine sinnvolle Ausrichtung zu finden. Daher auch an ihn meinen herzlichen Dank. Ebenso an Lorenzo Paolozzi, der sich die Zeit nahm, mir viele meiner Fragen zu beantworten.

Danke auch an die Mu3e-Gruppe, in der ich im vergangen Jahr erneut mitarbeiten durfte und dort auf die vielfältige Erfahrung und Unterstützung aller zählen konnte. Hier sei besonders Heiko Augustin genannt, der meine Arbeit betreut hat und mir immer mit Rat und Tat zur Seite stand. Auch sonst soll niemand vergessen sein, da mir die Zusammenarbeit mit allen sehr viel Spaß machte und ich mich bei vielen für Feedback und die Korrektur meiner Arbeit bedanken muss. Insbesondere sollen Jan Hammerich, Sebastian Dittmeier, David Immig, Lars Noehte, Thomas Rudzki und Alena Weber erwähnt werden.

Außerdem möchte ich mich bei Martin Richter und Jan Eberhardt bedanken, die mich während meiner gesamten Studienzeit begleitet haben und mich dabei immer wieder angestachelt haben, auch andere Themengebiete kennenzulernen. Ähnliches gilt für Sinan Kepezkaya, mit dem ich seit dem Abitur in einem freundschaftlichen Konkurrenzkampf stehe. Natürlich soll auch mein Freundeskreis nicht unerwähnt bleiben, der immer wieder für Ablenkung sorgte und vor allem für den sportlichen Ausgleich. Dazu gehört natürlich auch meine Freundin Natalie Ortega, die leider in den letzten Wochen etwas vernachlässigt wurde. Zu guter Letzt möchte ich mich bei meinen Eltern bedanken, auf deren Unterstützung ich immer zählen kann.

Erklärung:

Ich versichere, dass ich diese Arbeit selbstständig verfasst habe und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.

Heidelberg, den (Datum) 03.06.2020 B. Weinläder

.....