Customized Integrated Circuits for Scientific and Medical Applications

Zur Erlangung des akademischen Grades eines

DOKTOR-INGENIEURS

an der KIT-Fakultät für
Elektrotechnik und Informationstechnik
des Karlsruher Instituts für Technologie (KIT)
genehmigte

DISSERTATION

von

M.Sc. Roberto Blanco

geb. in: Ludwigshafen am Rhein

Tag der mündlichen Prüfung: 29.01.2019
Hauptreferent: Prof. Dr. rer. nat. Ivan Perić
Korreferent: Prof. Dr.-Ing. Dr. h. c. Jürgen Becker

Customized Integrated Circuits for Scientific and Medical Applications

1. Auflage: 2019

To my parents
Angelo and Maria Blanco
and my loving wife
Andrea Blanco

Zusammenfassung

Intelligente Halbleitersensoren spielen eine wichtige Rolle in wissenschaftlichen und kommerziellen Anwendungen und haben in den letzten Jahren eine rasante Entwicklung erlebt. In Smartphones, Uhren, Autos, Kameras, Smart Home usw.; sie sind für uns heute überall gegenwärtig. Die größten Entdeckungen in der Wissenschaft basieren heute auf intelligenten und hochempfindlichen Sensoren. So ist der Nachweis von Elementarteilchen möglich, die in Teilchenbeschleunigern für die Hochenergiephysik erzeugt werden. Astronomen haben Exoplaneten oder Planeten außerhalb unseres Sonnensystems beobachtet. Mit Teleskopen und Spektrometern entdecken Astronomen jedes Jahr neue Exoplaneten. Teleskop- und Spektrometerdetektoren bestehen aus hochempfindlichen Halbleitersensoren. In der Wissenschaft wird eine Vielzahl von Halbleiterdetektoren eingesetzt, wie z.B. in der Elementarteilchenphysik, Optischen- und Röntgenastronomie, Kernphysik und Medizin. Die Zahl der Anwendungen nimmt heute rasant zu. Der Erfolg basiert auf der extrem hohen Zeitauflösung, geringes Rauschen, geringe Leistungsaufnahme, hohe Auslesegeschwindigkeit, der monolithischen Architektur von Pixeln, präzise Messung von Position und Energie.

Die Entwicklung intelligenter Halbleitersensoren in komplementärer Metalloxid-Halbleiter-Technologie (CMOS) ist eine der wichtigsten Entwicklungen unserer Gruppe. Monolithische Pixeldetektoren in Hochspannungs-CMOS-Technologie (HV-CMOS) ist eine Erfindung unserer Gruppe. Sie sind eine der fortschrittlichsten Detektoren für die Detektion von hochenergetischen Teilchen. Pixelsensor und komplexe Pixelelektronik (ladungsempfindliche Verstärker, Integrator, Diskriminator und Speicherzellen) werden in derselben Wanne (N-Well) integriert. Das KIT ASIC and Detector Laboratory (KIT-ADL) ist auch an der Entwicklung von Einzelphotonen-Detektoren und Photonensignalzählern beteiligt, welche in monolithische Chips integriert sind. Die Einzelphotonen-Detektion hat in der Wissenschaft eine große Bedeutung erlangt. Inzwischen werden diese Art von Detektoren in den verschiedensten Anwendungen und Bereichen eingesetzt, darunter in Raumfahrt, Quanteninformationsverarbeitung, Medizin, Meteorologie, Messtechnik und Biotechnologie. Die Einzelphotonen-Detektion wird mit den sogenannten Photomultipliern erreicht. Da diese Photomultiplier groß sind und hohe Spannungen benötigen, wurden Halbleiter Einzelphotonen-Lawinensensoren entwickelt. Diese können in einem Technologieprozess integriert werden und benötigen nicht so hohe Spannungen wie bei Photomultipliern. Im Rahmen dieser Arbeit wurde ein monolithischer Einzelphoton-Detektor-Auslese-Chip in HV-CMOS Technologie vollständig entwickelt. Der Einzelphoton-Detektor-Auslese-Chip wurde von mir mit der Unterstützung von Professor Ivan Perić entwickelt. Der Sensor besteht aus einer sogenannten Single-Photon Avalanche Diode (SPAD). Darüber hinaus wurde ein Chip zur Photonenzählung vollständig charakterisiert. Der Chip kann gleichzeitig Photonensignale zählen und die Gesamtladung (Gesamtenergie) der Photonen mit Hilfe eines Integrators messen. Der Chip wird im Rahmen dieser Arbeit jedoch nicht

betrachtet. Dieser Chip wurde von Professor Ivan Perić entwickelt. Die Entwicklung des Testsystems sowie die Messungen wurden von mir durchgeführt.

Darüber hinaus beteiligen wir uns an der Entwicklung von Auslesemikrochips für die sogenannten DEPFET-Sensoren. Die DEPFET-Technologie wurde von der Max-Planck-Gesellschaft Halbleiterlabor in München (MPG HLL) entwickelt. Für die DEPFET-Sensoren haben wir zwei Arten von Mikrochips entwickelt - den Drain Current Digitizer (DCD) und den Steuerungschip (Switcher) für die Zeilenselektierung der DEPFET Matrix. Beide ASICs wurden von Professor Ivan Perić entwickelt. Zur Anwendung kommt diese Sensortechnologie im Vertex-Detektor (VXD) für das Belle-II-Experiment im Forschungszentrum KEK in Japan. Im Rahmen dieser Arbeit wurde eine Chip Probe Station und das Testsystem für Serientests von DCD und Switcher von mir mit der Unterstützung von Richard Leys entwickelt. Es wurden mehr als 1000 Chips getestet.

Ein weiteres Forschungsgebiet ist die Entwicklung des Elektronendetektors für das KIT-Experiment TRISTAN. Ziel des Experiments ist es sterile Neutrinos zu finden. Der Detektor ist in der Lage, die Energie von sogenannten langsamen Elektronen genau zu messen. Die Existenz von sterilen Neutrinos wurde zwar in der Theorie vorhergesagt, aber experimentell bisher nicht bestätigt. Sterile Neutrinos sind geeignete Kandidaten für Dunkle Materie. Im Rahmen dieser Arbeit wurde ein ADC-Design von mir entwickelt. Das Testsystem wurde ebenfalls von mir entwickelt und zahlreiche Messungen durchgeführt. Die rauscharmen Verstärker wurden von Professor Ivan Perić und Richard Leys entwickelt.

Eines der wichtigen Projekte am IPE ist die "Ultraschall Computertomographie (USCT)". Es handelt sich hier um eine neue bildgebende Methode zur Früherkennung von Brustkrebs. Ziel ist die Erkennung von Tumoren mit einem durchschnittlichen Durchmesser von kleiner als 5 mm. Eine neue Version des Transducer Array System (TAS) befindet sich in der Entwicklung. Die nächste Stufe der Entwicklung ist der Einsatz eines Mikrochips und die Entwicklung einer kleinen Empfänger- und Sendeelektronik. Dieser Mikrochip dient als Vorverstärker für den Sender und als Verstärker für den Empfänger. Das Design des Hochspannungskanals im Mikrochip wurde von mir entwickelt. Der 3-stufige rauscharme Verstärker wurde von Professor Perić und Richard Leys entwickelt. Garima Singhal und Bhavana Imanani haben das Testsystem entwickelt und die Messungen mit der Unterstützung von Richard Leys und mir durchgeführt.

Die Entwicklung von HV-CMOS-Sensoren für das ATLAS-Experiment ist das letzte Projekt, das von mir durchgeführt wurde. Mein Anteil war die Schaltungs- und die Layoutentwicklung der digitalen Triggerlogik. Die digitale ungetriggerte Logik und das Pixeldesign wurden von Professor Perić und Hui Zhang entwickelt.

Acknowledgment

My special thanks are directed to my doctoral supervisor Prof. Dr. rer. nat. Ivan Perić for the trust he has placed in me and for his outstanding support during my work. Through his valuable experience I was able to extend and deepen my knowledge of circuit design.

Also special thanks to Prof. Dr.-Ing. Dr. h.c. Jürgen Becker from the Institute for Information Processing Technologies (ITiV) for the acceptance of the second supervision.

I would also like to thank Prof. Dr. rer. nat. Marc Weber for proofreading my working reports. His revisions have enabled me to improve my scientific writing.

I would like to thank my colleagues at the IPE in particular Richard Leys for its programming expertise. Alexander Bacher, Uwe Bauer for their valuable support in technical questions such as difficult soldering tasks. Christian Krämer, a student of the department of electrical engineering at KIT who supported many of my projects. I would like to express my thanks to other members of IPE who have not been named here.

I would like to thank Avril Lang for proofreading this work and making a lot of helpful remarks.

Also my thanks and love goes to my wife Andrea Blanco. She is always there for me and I would like to thank her for the love and patience. I am very grateful to my parents because they always believed in me. Everything I am and because I exist, I owe it to the Almighty God, Sovereign Lord Jehovah. Without the "power beyond what is normal" that comes from Him, this work would not have been possible.

Karlsruhe, March 8, 2019

Roberto Blanco

Do not be afraid, for I am with you.
Do not be anxious, for I am your God.
I will fortify you, yes, I will help you,
I will really hold on to you with
my right hand of righteousness

- Isaiah 41:10

Contents

1 Introduction

Intelligent semiconductor sensors play a major role in commercial applications and have experienced a rapid development in the last years. Whether in smartphones, watches, cars, cameras, smart homes etc. they can be found everywhere today. The greatest discoveries today are based on intelligent and highly sensitive sensors. The detection of cosmic particles, elementary particles generated in particle accelerators for high-energy physics, is possible. Astronomers have been observing exoplanets, or planets, outside our solar system. Every year, astronomers are using telescopes and spectrometers to discover exoplanets. A large variety of semiconductor detectors are used in science such as elementary particle physics, optical and X-ray astronomy, nuclear physics, medicine, etc. These applications demand extremely high time resolution, low-noise, low power, high readout speed as well as the precise measurement of energy and position of particles.

The development of intelligent semiconductor sensors in commercial complementary metal oxide semiconductor (CMOS) technologies is the focus of our research group. Monolithic pixel detectors in commercial high-voltage CMOS (HV-CMOS) technology are one of our inventions. They are one of the most advanced detectors for detecting high-energy particles. Pixel sensor and complex pixel electronics including charge sensitive amplifier, integrator, discriminator, memory are embedded inside the same substrate.

The KIT-ADL is also involved in the development of single-photon detectors and photon signal counters integrated in a monolithic chip architecture. Single-photon detection has become very important in modern physics research. In the meantime these detectors are used in a large variety of applications and fields including space, quantum information processing, medical physics, meteorology, metrology and biotechnology, starting in 1934 with a solid rock technology as photo-multiplier to the age of semiconductor single-photon avalanche sensors. Within the scope of this work, a single-photon monolithic readout chip using HV-CMOS technology has been developed. The sensor consists of a single-photon avalanche diode and integrated fast readout electronics. The circuit and layout design was developed by myself with the support of Professor Ivan Perić.

A microchip for photon counting and a test system has also been developed. The ASIC was developed by Professor Ivan Perić and I carried out the test system and the measurements. The chip can simultaneously count photon signals and measure the total charge (total energy) of the photons using an integrator. The microchip is not described in this thesis.

Additionally we are contributing to the development of readout microchips for DEPFET sensors. The DEPFET technology was developed at the Max-Planck Society Semiconductor Laboratory in Munich (MPG HLL). We are developing two types of microchips for the DEPFET sensors - the current digitizer (DCD) and a row control microchip (Switcher). Both ASICs were developed by Professor Ivan Perić. The application of this sensor technology

is the Vertex detector (VXD) for the Belle-II experiment at the KEK research center in Japan. Within the scope of this work, I, supported by Richard Leys, developed a chip probe station and the system for serial testing of the DCD and Switcher; we tested more than 1000 chips.

Another research field is the development of the electron detector for the KIT experiment TRISTAN. The goal of the experiment is to find "sterile" neutrinos. The detector is able to measure the energy of slow electrons accurately. The existence of sterile neutrinos has been predicted but not yet confirmed experimentally. Sterile neutrinos are appropriate candidates for dark matter. Within the scope of this work, an ADC has been designed by me. I developed the test system and carried out the measurements of ADC, and the low-noise amplifier was designed by Professor Ivan Perić and Richard Leys.

One of the important projects at IPE is the "Ultrasound Computer Tomography" (USCT), a new imaging methodology for early breast cancer detection. The goal is the detection of tumors with an average diameter of less than 5 mm to improve the survival probability of the patients. A new version of the Transducer Array System (TAS) is under development. Next stage of improvement is the use of a microchip and the development of small receiver and transmitter electronics. This microchip acts as a pre-amplifier for the transmitter and an amplifier for the receiver. The 3-stage low-noise amplifier was designed by Professor Ivan Perić. The digital interface was mainly designed by Richard Leys. Garima Singhal and Bhavana Imanani developed the test system (software) and carried out the measurements with the support of Richard Leys and me. The design of the microchip (high-voltage channel) and measurements were done within this work.

The development of HV-CMOS sensors for ATLAS experiment is the last project done by me. My contribution was the circuit and layout design of the digital triggered readout cell. The untriggered and pixel circuit and layout design were developed by Professor Ivan Perić and Hui Zhang. The ASIC is not described in this thesis.

2 Introduction to CMOS Design

The principle of metal-oxide-silicon field-effect transistors (MOSFETs) was invented by J.E. Lilienfeld [58] in the early 1930s - before the invention of the bipolar transistor. Due to fabrication limitations, MOS technologies became practical not before 1960s. In 1960s the complementary MOS (CMOS) devices (n-type and p-type) were for the first time built, thereby a revolution in integrated circuits [76].

Digital integrated circuits are dominated by CMOS technologies. CMOS technologies can be found in computers, smartphones, digital cameras, television (satellite and terrestrial). This only became possible because the dimensions of MOS transistors could be scaled down more easily than those of other types of transistors. The next step was to utilize CMOS technology for analog design. The possibility of placing analog and digital circuits on the same chip reducing the cost of both development and packaging as well as the possibility of making chips with a large number of transistors made CMOS technology attractive. At the beginning, MOSFETs were slower and noisier than bipolar transistors [76]. CMOS technology has become dominant in analog integrated circuits. The main

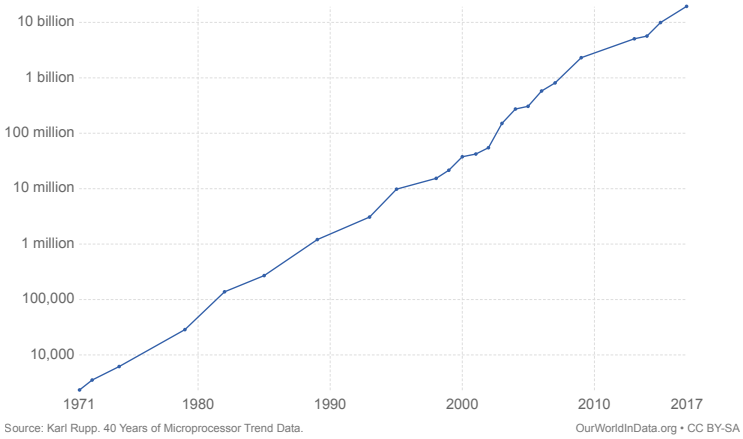

Figure 2.1: Moore's Law: The number of transistors on integrated circuit chips [6].

reason was the device scaling because it brings several benefits. Firstly, it improves the speed of the MOSFETs. The intrinsic speed of MOS has increased by orders of magnitude in the past decades, so that MOSFETs have exceeded the bipolar transistors. Another significant advantage over bipolar transistors is the operation at lower supply voltages between 1V and 2V. These supply voltages allowed low-power consumption for high-integrated circuits [76].

In 1965, Gordon Moore found that plotting the number of transistors that can be most economically manufactured on a chip gives a straight line on a logarithmic scale [65]. Moore predicted the transistor count doubles every 18 months. This prediction has been called Moore's Law. Figure 2.1 shows that the number of transistors doubles approximately every two years. In 2016, the largest number of transistors counted in a commercial processor is 7.2 billion (Intel Broadwell-EP Xeon). There are different classification levels of integration like small-scale, medium-scale, large-scale, and very large scale. Small-scale integration (SSI) circuits have less than 10 gates, with roughly 6 transistors per gate. Medium-scale integration (MSI) circuits have up to 1000 gates. Large-scale integration (LSI) circuits have up to 10000 gates [88]. The very large-scale integration (VLSI) holds for the most integrated circuit since 1980s. The clock frequencies of microprocessors have doubled every 34 months. In the year 2004 the power wall was hit with frequencies around 3 GHz. Luckily, not only the clock speed is important but also the ability to run an application in parallel. This performance is driven by the number of cores on the chip. The processors available on the market usually consist of several processor cores on one die [88]. The size

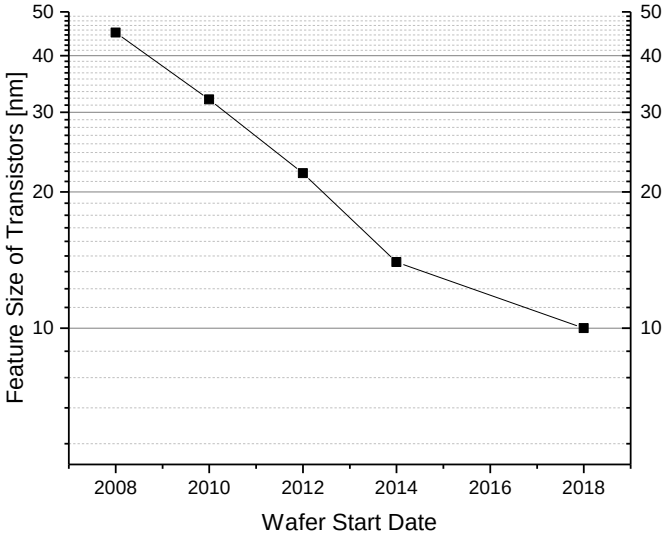

Figure 2.2: Transistor Feature Size has decreased by about a quarter when compared to 10 years ago.

of CMOS transistors was 10 µm in 1971. Microprocessors in 2008 had a feature size of 45 nm. Every 2 to 3 years the feature size decreases by 30 %. The next figure 2.2 shows the evolution of feature size since 2008. Moore's Law has proved to be a self-fulfilling prophecy until now the down scaling it cannot go forever [88].

2.1 Basics of MOSFET Transistors

Silicon (Si) forms the basis for highly integrated circuits. A silicon atom contains four valence electrons. A valence electron, an outer shell electron, contributes to the formation of chemical bond. Silicon is a Group IV element; it can compose a crystal when surrounded by four other neighboring atoms. A part of the periodic table containing a number of elements, each with three to five valence electrons [75], is shown in figure 2.3. Each atom

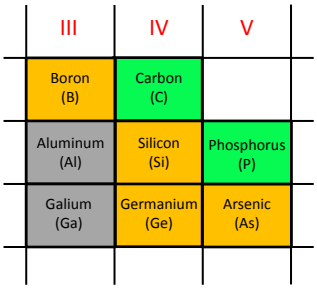

Figure 2.3: Periodic Table of Group III to V.

has one valence electron with its neighbor, thereby completing its own shell and those of the neighbor atoms. The so-called "covalent bond" emphasizes the sharing of valence electrons [75]. The formation of covalent bonds with adjacent atoms is shown in figure 2.4a.

The lattice is two-dimensional for easier drawing but actually forms a three-dimensional cubic crystal. Pure silicon is a poor conductor because its valence electrons are involved in chemical bonds. The conductivity can be improved by implanting small impurities, so-called dopants, in the silicon lattice [75].

2.1.1 Intrinsic and Extrinsic Semiconductors

Pure silicon is an example of an intrinsic semiconductor. Intrinsically the silicon has a high-resistance. It is possible to manipulate the resistivity of silicon by implanting atoms of another material, the extrinsic semiconductor [75].

Arsenic, a dopant from the Group V (see figure 2.3) has five valence electrons. In this case, the excess electron (fifth electron) is responsible for the improved conductivity. One silicon atom is replaced within the lattice with an arsenic atom. The fifth electron is freely bound to the arsenic atom as shown in figure 2.4b. At room temperature it is enough to move the free electron leaving a positively charged arsenic ion and a free electron. This electron contributes to the electrical current so that the conductivity is improved. This type of doping is n-type doping or n-type semiconductor. Boron, a Group III dopant, has three valence electrons as shown in figure 2.4c. Boron can form only three complete covalent bonds in silicon lattice [75]. The result is that the fourth bond contains a hole. The hole is ready to absorb an electron because the hole is positive charged and the electron

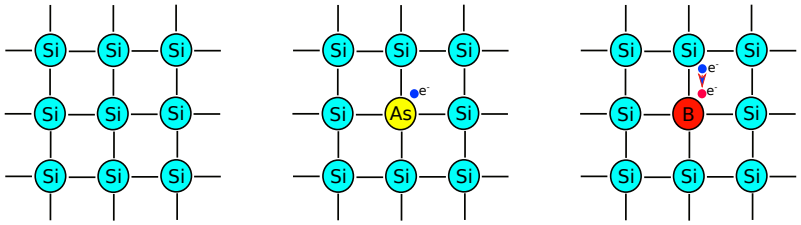

Figure 2.4: (a) Silicon Lattice without dopants. (b) One silicon atom replaced with arsenic atom. (c) One silicon atom replaced with boron atom.

negative charged. The hole acts as a positive carrier. This type of doping is p-type or p-type semiconductor. In summary it can be said that an n-type semiconductor has electrons as majority carriers and holes as minority carriers. For p-type semiconductors the holes are majority carriers and electrons minority carriers. The arsenic atom is donator and boron atom acceptor dopant [88].

In an intrinsic semiconductor, the electron density, n, is equal to the hole density, p. The densities in a doped material can be proved as follows [75],

$$np = n_i^2 \tag{2.1}$$

where n represents electron densities and p hole densities in the extrinsic semiconductor. The number of n_i represents the densities in the intrinsic semiconductor and independent of the doping level for silicon. The density for intrinsic silicon can be described as [75]

$$n_i = 5.2 \cdot 10^{15} \cdot \sqrt{T^3} \cdot e^{\frac{-E_g}{2kT}} \tag{2.2}$$

where E_g is the minimum bandgap energy (silicon: 1.12 eV), $k = 1.38 * 10^{-23} J/K$ the Boltzmann constant. If an intrinsic semiconductor is doped with a density of N_D ($\gg n_i$) donor atoms per cubic centimeter, then the mobile charge densities are given by [75]

$$\text{Majority Carriers: } n \approx N_D \tag{2.3}$$

$$\text{Minority Carriers: } p \approx \frac{n_i^2}{N_D} \tag{2.4}$$

For the acceptor atoms (per cubic centimeter) this can be described as follows

$$\text{Majority Carriers: } p \approx N_A \tag{2.5}$$

$$\text{Minority Carriers: } n \approx \frac{n_i^2}{N_A} \tag{2.6}$$

The exponential behavior of n_i reveals the effect of the band gap energy (E_g) on the conductivity of the material. Insulators have high band gap energy ($E_g = 2.5$ eV diamond).

Conductors have a small band gap energy. Semiconductors have an E_g between 1 eV to 1.5 eV (figure 2.5).

Silicon atoms arranged in a diamond structure forming energy bands. The valence band is the highest filled energy band with its Energy E_v and the conduction band E_c this energy band is empty. Both energy bands are separated by the band gap E_g:

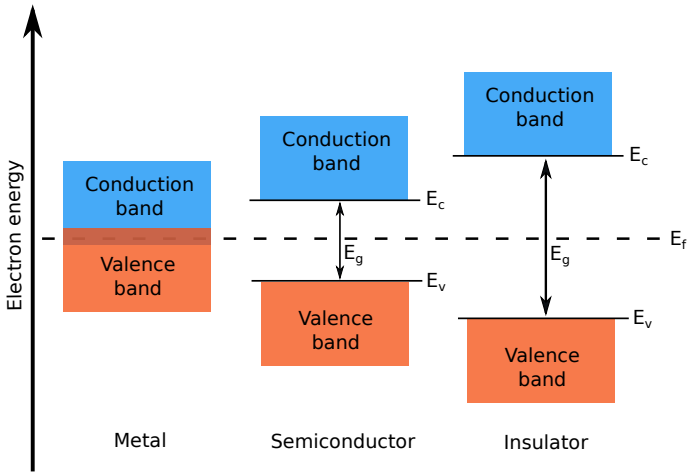

Figure 2.5: Simplified energy band structure of metal, semiconductor and insulator. Metals have the Fermi level within the conduction band, the Fermi level for semiconductors and insulators lies between conduction and valence band.

$$E_g = E_c - E_v \tag{2.7}$$

At low temperatures the valence band is completely filled while the conduction band is empty. When the temperature rises valence electrons can lift electrons from valence band to conduction band [59]. The formation of energy bands of a semiconductor is not described in this work.

2.1.2 Direct and Indirect Semiconductors

Direct semiconductors have a direct band gap (shown in figure 2.6). This means, the minimum of the conduction band and the maximum of the valence band are at the same wave number vector. An electron at the conduction band can recombine with a hole at the valence band maximum any time. In an indirect semiconductor (indirect band gap), on the other hand, an additional momentum must be transferred at the transition from conductance band minimum to valence band maximum. A third particle, usually a phonon, must be involved in the recombination process. Therefore, the probability of recombination is lower compared to a direct semiconductor because such a phonon is not always available. Direct semiconductors are particularly suitable for optical components like semiconductor

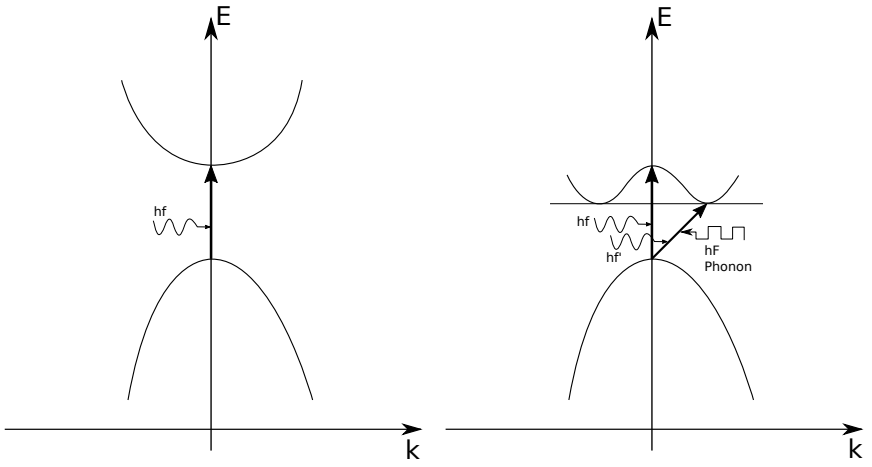

Figure 2.6: Direct and indirect band transitions. The optical transition is practically vertical in the E(k) diagram, the photon takes less momentum. In indirect band transition, electron and holes can only recombine if phonons reach the momentum balance.

light amplification by stimulated emission of radiation (LASER), light emitting diodes (LED).

2.1.3 Charge Movement of Semiconductors

Current can flow in a resistance, if a potential difference or an electrical field is applied. The electrical field accelerates charge carriers from one end to the other. Charge carrier movement caused by an electrical field is called drift. Semiconductors have a similar behavior. The accelerated charge carriers collide with atoms from the lattice (shown in figure 2.7). Acceleration and collision lead to the following expression of constant velocity

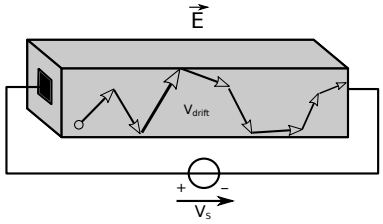

Figure 2.7: Charge carrier drift in a semiconductor.

for the carriers [75]

$$v = \mu \cdot E \tag{2.8}$$

where v (velocity) is proportional to the electric field E. The mobility μ is expressed in terms of cm^2/Vs. The mobility of electrons of the intrinsic silicon is μ_n = 1350 cm^2/Vs and that of holes μ_p = 480 cm^2/Vs. Physically accurate interpretation of the velocity for electrons and holes would be [75]

$$\vec{v_e} = -\mu_n \cdot \vec{E} \tag{2.9}$$

$$\vec{v_h} = \mu_p \cdot \vec{E} \tag{2.10}$$

For electrons, the equation 2.8 must have a minus sign because electrons move in the opposite direction to the electrical field. The current density is generally defined as electric current flowing through surface A or as product of number charge carrier per volume unit (n, p), the elementary charge e and the average velocity v [75]

$$J = \frac{I}{A} = n \cdot e \cdot v \tag{2.11}$$

where e is the elementary charge ($e = 1.602 \cdot 10^{-19}$ C). Equations 2.9 and 2.10 for holes and electrons can be inserted for the current density [75]

$$J_n = \mu_n \cdot E \cdot n \cdot e \tag{2.12}$$

$$J_p = \mu_p \cdot E \cdot p \cdot e \tag{2.13}$$

A total drift current density can be defined in the presence of electrons and holes in response to an electric field E [75]

$$J_{tot} = \mu_n E n e + \mu_p E p e \tag{2.14}$$

$$J_{tot} = eE \cdot (\mu_n n + \mu_p p) \tag{2.15}$$

In a semiconductor the mobility of carriers is independent of the applied electrical field and the velocity rises linearly with E. If the electric field rises for higher values, there is no longer linear dependence between velocity and electrical field. Due to the collision of carriers with the lattice, very often the carriers cannot accelerate anymore. A saturation effect occurs for high-electric fields. It can be defined with a velocity saturation shown in figure 2.8. This effect cause limits the performance of MOS transistors [75]. The slope is the mobility, μ_{eff}. At fields of critical level, E_c, the velocity saturates 10^7 cm/s for electrons and 8x10^6 cm/s for holes [68]. The expression $v = \mu \cdot E$ can be approximated reasonably with the following expression [84]:

$$v = \begin{cases} v = \frac{\mu_{eff} E}{1 + E/E_c} & E < E_c \\ v_{sat} & E \geq E_c \end{cases} \tag{2.16}$$

the critical electric field is

$$E_c = \frac{2 v_{sat}}{\mu_{eff}} \tag{2.17}$$

In addition to the drift current phenomenon, another effect causes current flow. If charge carriers are injected into a semiconductor they will create a nonuniform intensity. In the

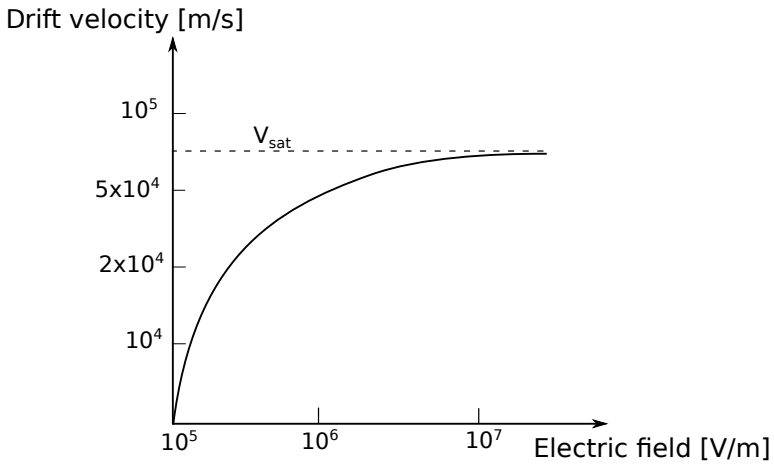

Figure 2.8: Velocity saturation of charge carriers in semiconductors.

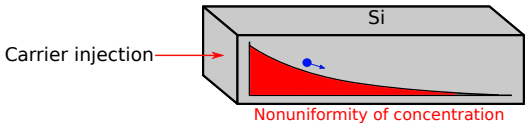

Figure 2.9: Drift effect in a piece of silicon.

absence of an electrical field the charge carriers tend to move from regions of high concentration to regions of low concentration. This kind of effect is called diffusion. Therefore diffusion is different from drift because electrical current is carried as long as the nonuniformity is maintained. A nonuniform charge profile is created by continuing injecting charge carriers. Figure 2.9 shows a simplified illustration of the diffusion phenomenon. The diffusion current for electrons and holes can be defined as [59]

$$\vec{j}_{n,Drift} = -eD_n\vec{\nabla}n \tag{2.18}$$

$$\vec{j}_{p,Drift} = -eD_p\vec{\nabla}p \tag{2.19}$$

$D_{n,p}$ are the diffusion constants for electrons and holes which are specific to a particular semiconductor. For instance, in intrinsic silicon, $D_n = 34$ cm^2/s and $D_p = 12$ cm^2/s. ∇n and ∇p are the concentration gradients. In summary, the current densities (for electron and holes) in a semiconductor can be written as follows

$$\vec{j}_n = \vec{j}_{n,Drift} + \vec{j}_{n,Diff} = -e\mu_n n\vec{E} - eD_n\vec{\nabla}n \tag{2.20}$$

$$\vec{j}_p = \vec{j}_{p,Drift} + \vec{j}_{p,Diff} = e\mu_p p\vec{E} - eD_p\vec{\nabla}p \tag{2.21}$$

Thus, the quotient of diffusion and mobility for electrons and holes can be expressed by fundamental constants such as elementary charge and Boltzmann constant. This expression is called "Einstein Relation"[59].

$$\frac{D_n}{\mu_n} = \frac{kT}{e} \tag{2.22}$$

$$\frac{D_p}{\mu_p} = \frac{kT}{e} \tag{2.23}$$

2.1.4 A Brief Introduction to P-N Junction

The p-n junction [36] is the simplest semiconductor device and also serves as part of such complex structures as transistors. If an n-type and p-type silicon are brought into contact with each other, it forms a p-n junction. This structure forms a diode and plays an important role in many semiconductor devices. The p-type part is called anode and n-type part of the silicon is the cathode. If an auxiliary voltage is applied on the p-type part and raised above the n-type the diode is forward biased. Reverse biasing occurs if the anode voltage is less than, or equal to, the cathode voltage.

A depleted region is formed at the p-n interface and this area is depleted of holes or electrons. In the p-doped region the holes form the majority carriers and majority carriers in the n-doped region are electrons. Because of the strong concentration gradient of the two different charge carrier types, this results in a diffusion current, as described in section 2.1.3. Electrons of the n-doped part will diffuse into the p-doped part of the crystal, and the holes from the p-doped part to the n-doped part. A recombination of the charge carriers occurs and an area of no charge carriers forms in the depletion zone. The fixed, ionized atoms are positively charged in the n-region and negatively charged in the p-region. A space charge is formed leading to an electric field. The field leads to a drift of electrons and holes. An build-in voltage (V_{bi}) exists for zero applied auxiliary voltage. The built-in voltage depends logarithmically on doping levels [83].

$$V_{bi} = \frac{kT}{e} \ln\left(\frac{N_A N_D}{n_i^2}\right) \tag{2.24}$$

For typical doping concentrations, this voltage is a few of hundred mV. A p-n junction with no external connections (no applied voltage) can be defined as a junction in equilibrium. Once the junction reaches the equilibrium, the electric field is strong enough to completely stop the diffusion currents. Therefore, drift currents, resulting from the electric field and diffusion currents cancel each other out and the equilibrium state is established. Figure 2.10 conceptually shows how the junction evolves with time [75]. Figure 2.11 shows the development of charge density, electric field and potential for an abrupt transition. The electric field increases linearly with a maximum electrical field at the boundary between p

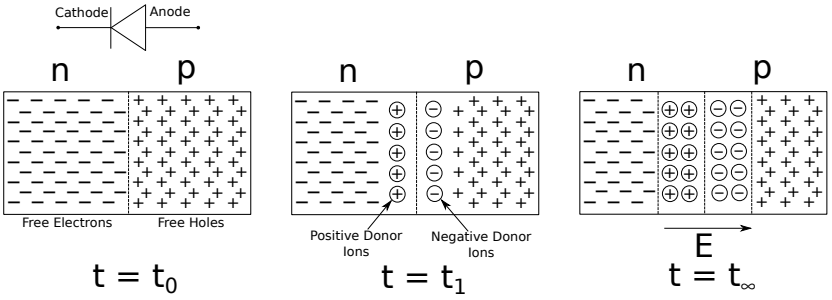

Figure 2.10: p-n junction evolution: Adjoining p- and n-type silicon form a p-n junction. The charge of valence electrons and holes is balanced by the charge of the atomic cores. The charge neutrality is maintained (equilibrium state) [75].

and n conductors caused by the space-charge zone. The space-charge zone in equilibrium depends only on the doping semiconductors. The charge density is given by [49]

$$\rho(x) = \begin{cases} -eN_A & -x_p < x < 0 \\ +eN_D & 0 < x < x_n \end{cases} \tag{2.25}$$

The electrical field $E(x)$ and the potential difference V_{bi} can be calculated for the charge density. Charge neutrality is given in the regions outside the space-charge zone. The required zero E-field can be written as [49]

$$E(-x_p) = E(x_n) = 0 \tag{2.26}$$

where x_p and x_n represent the penetration depths of the space-charge zone in the p- or in the n-doped part of the semiconductor. In order to maintain the neutrality of the semiconductor, the carrier numbers in both parts of the space-charge zone must be the same [49]

$$N_A x_p = N_D x_n \tag{2.27}$$

Using the one-dimensional Maxwell equation for the E-field as well as the boundary conditions (using equation 2.25 and 2.27) [49]:

$$\frac{dE}{dx} = \frac{1}{\epsilon\epsilon_0}\rho(x) \rightarrow E(x) = \begin{cases} \frac{-eN_A}{\epsilon\epsilon_0}(x + x_p) & -x_p < x < 0 \\ \frac{+eN_D}{\epsilon\epsilon_0}(x - x_n) & 0 < x < x_n \end{cases} \tag{2.28}$$

The maximum E-field can be calculated at x=0:

$$E_{max} = -\frac{eN_A}{\epsilon\epsilon_0}x_p = \frac{eN_D}{\epsilon\epsilon_0}(-x_n) \tag{2.29}$$

The voltage drop across the depletion zone, the built-in voltage V_{bi}, can be described as the difference of potentials in the n or p area outside the space-charge zone. The built-in voltage

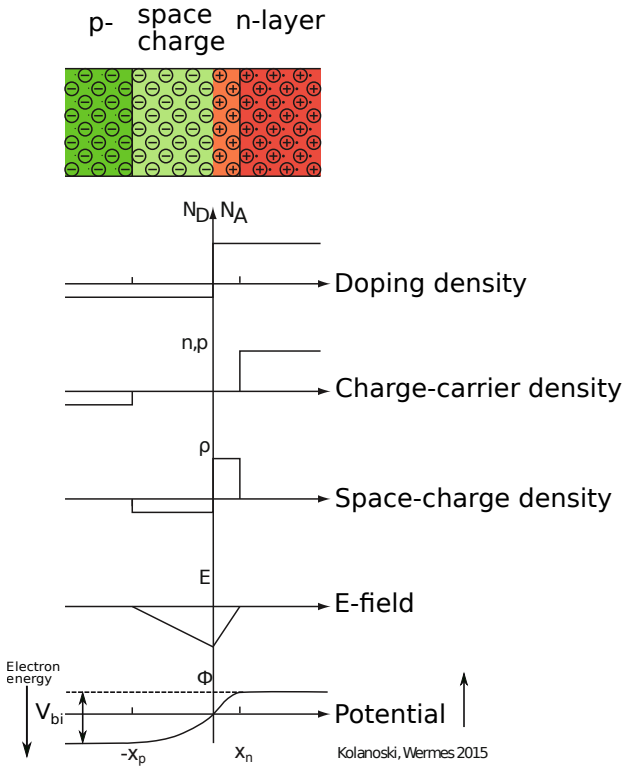

Figure 2.11: Graphical derivation of doping and space-charge densities, electrical field and potential profile in an abrupt p-n transition [49].

can be described as $V_{bi} = \phi_p - \phi_n$ where ϕ_p and ϕ_n are the potentials at the boundaries of the depletion zone [49]. The built-in voltage depends on the difference between the intrinsic (E_f) and extrinsic (E_F) fermi energies in the respective n- and p- regions. Since $E_f - E_F^p = -e\phi_p = kTln\frac{N_A}{n_i}$ and $E_F^n - E_f = -e\phi_n = kTln\frac{N_D}{n_i}$ follows [49]

$$V_{bi} = \phi(x_n) - \phi(x_p) = E_F^n - E_F^p = \frac{kT}{e} ln \frac{N_A N_D}{n_i^2}. \tag{2.30}$$

Thus, the x_n of the depletion region can be defined as the width w_0 if $N_A >> N_D$ and $x_n >> x_p$:

$$w_0 \approx x_n \approx \sqrt{\frac{2\epsilon_{Si}}{eN_D} V_{bi}}. \tag{2.31}$$

To determine the width of x_p, the following equations can be written: $N_D >> N_A$ and $x_p >> x_n$ therefore [49]

$$w_0 \approx x_p \approx \sqrt{\frac{2\epsilon_{Si}}{eN_A} V_{bi}}. \tag{2.32}$$

When forward bias is applied, the equilibrium state no longer holds. With a positive potential on the p-side and negative on the n-side, the flow of valence electrons and holes across the junction increases. When negative potential is applied on the p-side and positive on the n-side the width of the depletion grows. The width can be determined as follows:

$$w(V_{ext}) = \sqrt{\frac{2\epsilon_{Si}}{e} \frac{N_A + N_D}{N_A \cdot N_D} (V_{bi} + |V_{ext}|)} \tag{2.33}$$

where V_{ext} is the external voltage applied on the p-n diode. For asymmetric junction e.g. p^+n - diode it follows $N_A >> N_D$ with depth

$$w(V_{ext}) \approx x_n \approx \sqrt{\frac{2\epsilon_{Si}}{eN_D} (V_{bi} + |V_{ext}|)}. \tag{2.34}$$

In order to detect particles, semiconductor detectors are operated in reverse bias. The positive potential is on the n-doped part and the negative potential on the p-doped part of the diode. A large depletion zone is needed which represents the volume for particle detection [49].

A diode current with respect to applied voltage can be described by the "Shockley equation" [83]:

$$I_D = I_0 \left(e^{V_{ext}/kT/e} - 1 \right) \tag{2.35}$$

For forward bias voltages the exponential behavior dominates and the current increases.

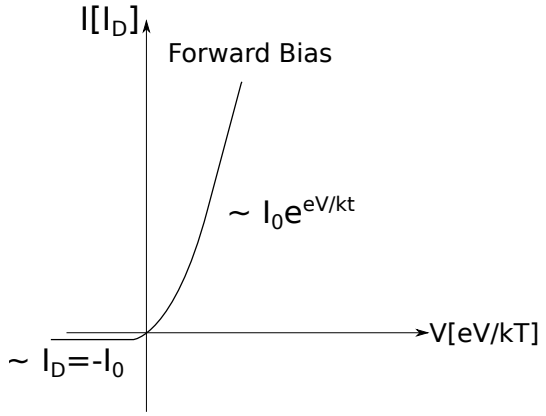

Figure 2.12: I/V characteristic of a semiconductor diode.

For reverse bias voltages the exponential term becomes insignificant and the current would be $I = -I_0$ (figure 2.12).

2.2 MOSFET Fundamental Device Characteristics

As already described in the introduction (chapter 2) the MOSFET is the most important device for high integrated circuits such as for memories like static random access memory (SRAM), dynamic random access memory (DRAM), read-only memory (ROM) and microprocessors, field programmable gate arrays (FPGA), application specific integrated circuits (ASIC).

A MOS consists of superimposed layers of conducting and insulating materials. These materials are placed in a pizza-like structure. A series of chemical processes are involved in the manufacture of these structures. Some of the chemical processes involve silicon oxidation, deposition and etching of contacts and metal layers, ion implantation. MOS transistors are developed on highly pure single silicon crystals which are produced as thin wafers (circular). Two types of transistors are provided by complementary MOS (CMOS) technology: n-type transistor (NMOS) shown in figure 2.13 and a p-type the so-called PMOS transistor [83], based on the field effect. NMOS and PMOS consist of a pizza-like structure of the conducting gate (heavily doped polysilicon), an insulating layer of silicon dioxide, $Si0_2$ and the silicon wafer also called substrate or bulk. An NMOS transistor is developed with a p-type body has n-type regions called the source and drain. A PMOS transistor is similar to an NMOS but has an n-type body and p-type regions for drain and source. NMOS and PMOS transistors are usually located in "local" substrates called wells (figure2.14). The transistors are fully symmetrical. That means, it doesn't matter how

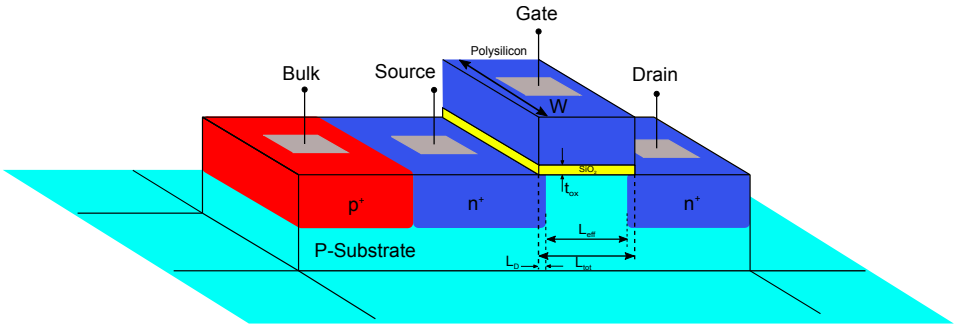

Figure 2.13: Structure of a MOS device.

source and drain are arranged. Thus, a MOSFET transistor has four electrodes: source, drain, gate and substrate (bulk). Free charge carriers (NMOS: electrons, PMOS: holes) are generated on the source contact and the drain collects them. The gate is used as control input. Source and drain are in the p-substrate. The substrate has its own contact - called bulk contact. Roughly explained an NMOS transistor works as follows: the substrate is grounded so the p-n diodes of the source and drain to substrate are reverse biased. If the gate has 0 V or ground no current can flow through the reverse biased p-n diodes. Thus, the transistor operation is off. When a positive voltage is applied to the gate, the moving holes are repelled from the substrate. The negative of acceptor atoms remains. The substrate is depleted. In chapter 2.1.4 it was explained that a space charge is accomplished in the

depleted area. This charge generates an electric field. If an electric field exists, the potential also changes. A conducting path of electron carriers is formed from source to drain and current can flow. Now, the transistor is on. With a PMOS transistor the conditions are reversed. The substrate is set to positive potential. If the gate is also at a positive voltage, the source and drain p-n diodes are reverse biased and no current will flow from source to drain. The transistor is off. When the gate potential is lowered, positive charges are attracted. If the voltage is lowered enough, this inverts the channel, forming a conducting path of positive carriers from source to drain. The transistor is on [88]. When drain and

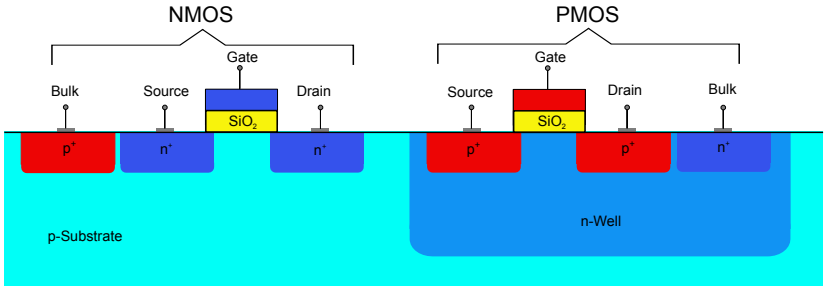

Figure 2.14: Complementary MOS (CMOS) device structures. PMOS inside an n-well and NMOS inside the local substrate (p-Substrate).

source regions are implanted, some of the n or p implant diffuses underneath the gate poly. The total length (L_{tot}) of the gate along source-drain is the lateral dimension (see figure 2.13). Where L_D is the length of the lateral diffusion [76]. An effective length can be defined

$$L_{eff} = L_{tot} - 2L_D \tag{2.36}$$

perpendicular to the length is called width (W). Since the SiO_2 has a certain thickness a t_{ox} is denoted. Both the gate oxide thickness and effective length play an important role in the speed of MOS circuits (some values for $L_{eff} \approx 25$ nm and $t_{ox} \approx 1.5$ nm) [76].

The circuit symbols used to represent NMOS and PMOS transistors are depicted in figure 2.15. The circuit symbols contain four terminal contacts denoted by "S" (source), "D" (drain), "G" (gate) and "B" (bulk) for the substrate contact.

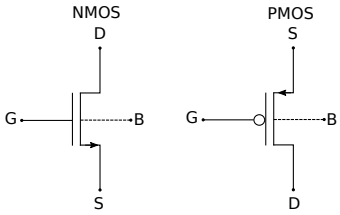

Figure 2.15: NMOS and PMOS symbols.

2.2.1 Derivation of I/V Characteristics

Before considering the I/V behavior of the NMOS transistor it is of great importance to understand the threshold voltage. Since the gate (dielectric) and the substrate form a capacitor, as V_G is raised, the holes in the p-substrate are repelled from the gate, leaving negative ions (figure 2.16b). As discussed in this chapter a depletion region is formed. When V_G increases, the depletion zone width and gate oxide also increase. This structure forms a capacitive voltage divider (figure 2.16a): the gate oxide capacitor and the depletion zone capacitor. If V_G is positive enough, electrons flow from source to the gate-oxide and to the drain. Charge carriers are formed under the gate oxide and a channel is established (figure 2.16c). Now the transistor is turned on. The gate-oxide interface (also called the inversion layer) is inverted. A threshold voltage, V_{th} can be defined. This threshold indicates when the transistor becomes conductive or the inversion layer is formed. Depletion region remains constant even when V_G increases and a greater current is provided between source and drain. It can be proved that [76]

$$V_{th} = \Phi_{MS} + 2\Phi_F + \frac{Q_{dep}}{C_{ox}} \tag{2.37}$$

the potential Φ_F is given by

$$\Phi_F = \frac{kT}{e} ln\left(\frac{N_{sub}}{n_i}\right) \tag{2.38}$$

where Q_{dep} is the charge in the depletion region. From the p-n junction theory it can be shown that [76]

$$Q_{dep} = \sqrt{(4e\varepsilon_{si}|\Phi_F|N_{sub})} \tag{2.39}$$

ε_{si} is the dielectric constant of silicon, N_{sub} the doping density of the substrate, n_i density of electrons of undoped silicon. Φ_{MS} is the difference in the working functions of the polysilicon gate and the substrate. The thermal energy at room temperature corresponds to a voltage of $V_T = 25$ mV (kT/e). For an oxide thickness of 2 nm the value of C_{ox} can be scaled proportionally, the value of the oxide thickness in this case is $C_{ox} \approx 17$ fF/um^2. The depletion capacitance is determined by the depletion zone depth. Normally, the doping and oxide thickness are chosen in such a way that C_{ox} is about twice as large as C_{dep}. Therefore, following the voltage divider formula $\frac{V_{ox}}{V_{dep}} = \frac{1}{2}$.

Three regions of operation have a MOSFET: subthreshold region, linear (or triode) region, saturation region. To obtain the relationship between drain current and the terminal voltages, it is important to derive the first order model of an NMOS transistor. This model is based on the Shockley equations [81]. The Shockley model assumes that the current through a transistor is zero if the transistor is set to off. When gate-source voltage ($V_{gs} > V_{th}$) is raised, the gate carriers attract carriers (electrons) to form a channel. Electrons drift from source to drain proportional to the electric field. If the amount of charge and the rate is known, it is possible to determine the currents. The charge on the capacitor plates is defined as [88]

$$Q_{channel} = C_g\left(V_{gc} - V_{th}\right) \tag{2.40}$$

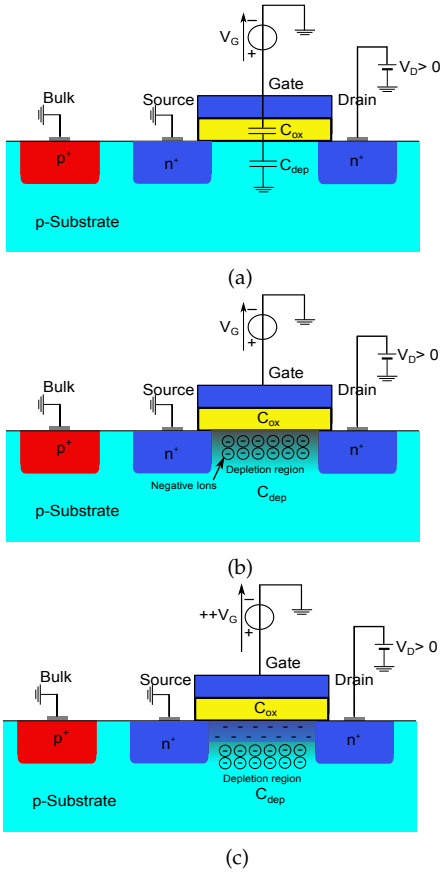

Figure 2.16: (a) Capacitive voltage divider of oxide and dynamic depletion capacitance. (b) Depletion region formation. (c) Inversion layer formation [76].

where $(V_{gc} - V_{th})$ is the amount of voltage attracting charge to the channel beyond the minimum required to invert from p to n. An Average gate to channel potential can be written as [88]

$$V_{gc} = \frac{V_{gs} - V_{ds}}{2} \tag{2.41}$$

The gate capacitance is proportional to area over thickness. If the gate has the length L and width W and a oxide thickness (t_{ox}), the capacitance can be defined as [88]

$$C_g = \varepsilon_{ox} \frac{WL}{t_{ox}} = C_{ox} WL \tag{2.42}$$

where ε_{ox} is permittivity (ε_0) times the permittivity of SiO_2. The equation 2.42 shows that the relationship (ε_{ox}/t_{ox}) corresponds exactly to the oxide capacitance. To cross the channel the required time of the carriers is length divided by the velocity. The constant velocity for carriers is described in 2.1.3. Between source and drain the total amount of charge in the channel is defined by the following current [88]

$$\begin{aligned} I_{ds} &= \frac{Q_{channel}}{L/v} \\ &= \frac{Q_{channel}}{\frac{L}{\mu_n E}} \\ &= \frac{\mu_n C_{ox} W (V_{gs} - V_{ds}/2 - V_{th})}{\frac{L}{\mu_n E}} \\ &= \frac{\mu_n C_{ox} W (V_{gs} - V_{ds}/2 - V_{th})}{\frac{L}{\mu_n V_{ds}/L}} \\ &= \mu_n C_{ox} \frac{W}{L} \left(V_{gs} - V_{th} - \frac{V_{ds}}{2} \right) V_{ds} \end{aligned} \tag{2.43}$$

$$= \mu_n C_{ox} \frac{W}{L} \left(V_{gs} - V_{th} - \frac{V_{ds}}{2} \right) V_{ds} \tag{2.44}$$

For $V_{gs} > V_{th}$ the transistor is in the linear region (see figure 2.17a) of operation but V_{ds} is small. $V_{ds} << (V_{gs} - V_{th})$ is called linear or resistive (also known as triode mode) because I_{ds} increases almost linearly with V_{ds} the same as a resistor. If $V_{ds} > V_{dsat} = (V_{gs} - V_{th})$ in the vicinity the channel is pinched off. After this point, the drain saturation voltage, has no further effect on current changing with the increasing drain voltage (see figure 2.17b). In equation 2.44 V_{ds} can be substituted with V_{dsat}. A new expression of drain source current can be expressed that is independent of V_{ds} [76]:

$$I_{ds} = \frac{\mu_n C_{ox} W/L}{2} \left(V_{gs} - V_{th} \right)^2 \tag{2.45}$$

Equation 2.45 is only valid for $V_{gs} > V_{th}$ and $V_{ds} > V_{dsat}$. An additional factor involves λ, the channel-length modulation parameter. This parameter models current dependence on drain voltage due to the Early effect (λ) or channel length modulation. In section 2.1.4, it was shown that a p-n junction forms a depletion region. This depletion between drain (or source) and substrate has a certain width and increases with drain substrate (or body) voltage. This effect shortens the channel length. As the channel length gets even shorter, the effect of the channel length becomes more important. The channel length modulation must be considered because this reduces the gain of amplifiers. The drain source current can be approximately modeled as:

$$I_{ds} = \frac{\mu_n C_{ox} W/L}{2} \left(V_{gs} - V_{th} \right)^2 \left[1 + \lambda (V_{ds} - V_{dsat}) \right] \tag{2.46}$$

The electrical state of the transistor is described by two voltages (V_{gs}, V_{ds}) and two currents (I_{gs}, I_{ds}). For DC signals I_{gs} is zero. As the gate previously mentioned represents

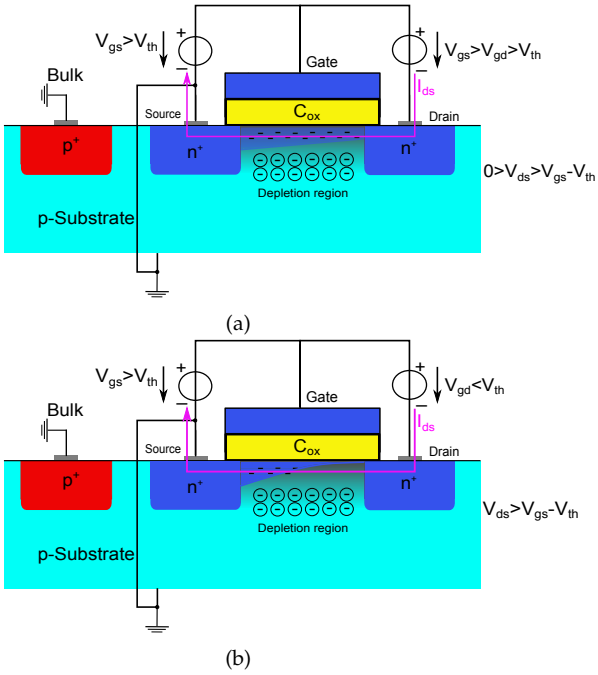

Figure 2.17: (a) Linear Region: channel formed I_{ds} increases with V_{ds}. (b) Saturation Region: channel pinched off I_{ds} independent of V_{ds} [88].

only a capacitance (neglecting the tunnel effect). Transistor behavior can be described with following characteristic curves:

- I_{ds} as function of V_{ds} for different V_{gs} (output characteristics)
- I_{ds} as a function of V_{gs} for different V_{ds} (input characteristic)

In the area on the left (see figure 2.18) the current sinks as V_{ds} (resistive, triode region decreases). For small V_{ds}, the current-voltage dependence is approximately linear. In the area on the right, the current of V_{ds} is practically independent (saturation region). Ideally $I_{ds} = I_{dsat}$ for all $V_{ds} > V_{gs} - V_{th}$. Figure 2.18 shows the NMOS I-V characteristics. The behavior of PMOS transistors is similar, but the signs of all voltages and currents are reversed. The I-V characteristics are on the third quadrant. As mentioned in section 2.1.3 the mobility of holes in silicon is lower than that of electrons. It follows then that PMOS transistors provide less current than NMOS transistors hence they are slower. Intersection points form a parabola (I_{dssat} curve) for different V_{gs} voltages. This curve indicates the boundary between triode region and saturation. The parabola can be derived from the formula for saturation current and the condition $V_{ds} = V_{gs} - V_{th}$. In the saturation region, the transistor behaves like a voltage-controlled current source (figure 2.19). For the triode region (small V_{ds}) the transistor behaves as a variable resistor (figure 2.20) or as an

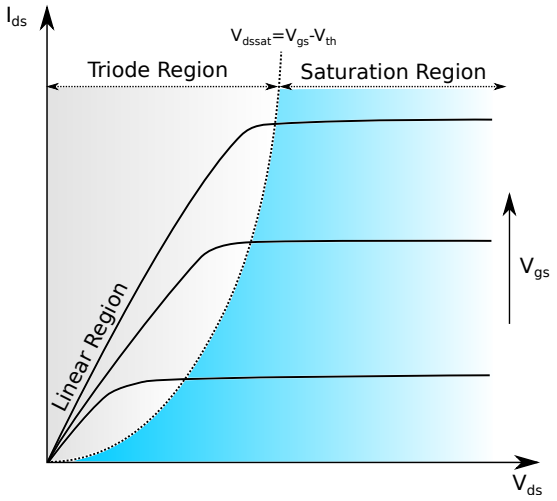

Figure 2.18: I-V characteristics for NMOS and PMOS.

electronic switch (linear region). The linear relationship implies that the path from drain to source can be defined as linear resistor

$$R_{on} = \frac{1}{\mu_n \cdot C_{ox} \frac{W}{L}(V_{gs} - V_{th})}. \tag{2.47}$$

The NMOS or PMOS transistor operates as variable resistor under the condition given $V_{ds} << 2(V_{gs} - V_{th})$. A representation of several I_{dssat} curves for different V_{ds} is possible. If only the I_{dssat} curve is displayed, I_{dssat} is a function of V_{gs}. Parabolic behavior rises from the value V_{th} upwards. The input characteristic curve (I_{dssat}) is usually linearized in the range around the operating point - the gradient is called the transconductance. The transconductance can be written as

$$g_m = \frac{dI_{dssat}}{dV_{gs}} \tag{2.48}$$

Linearization at the operating point results a small-signal model of the transistor. The small-signal model allows arbitrary positive and negative V_{gs} and I_{ds} values (small signals). However, the negative small signal current must not exceed the DC current, otherwise the total current I_{ds} would be negative. A negative I_{ds} cannot be achieved with a positive

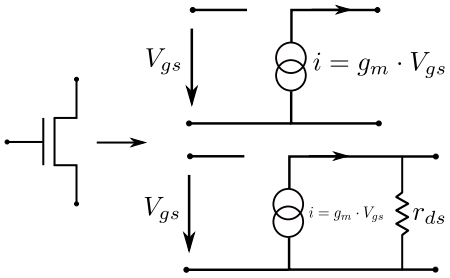

Figure 2.19: Small signal model: In saturation region the transistor acts as a voltage-controlled current source (channel-length modulation represented as a resistor in parallel r_{ds}).

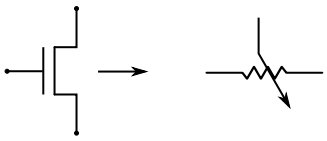

Figure 2.20: In triode region the transistor acts as a variable resistance.

V_{ds}. Looking closely at the saturation region the line is not ideally flat. Channel length modulation forms a slope in the saturation range and it can be defined as output resistance

$$g_{ds} = \frac{\partial I_{ds}}{\partial V_{ds}}$$
$$r_{ds} = \frac{1}{g_{ds}} \tag{2.49}$$

Figure 2.21 shows the output resistance r_{ds} in saturation region. As the drain voltage increases, the pinch-off region becomes larger. Due to the channel length modulation, the drain current continues to increase in saturation. The output resistance is not infinite but is higher the longer (length of transistor) the transistor is and the smaller the current. A small g_{ds} is usually preferable because the transistor then behaves more like a current source. Small g_{ds} values (or large r_{ds} resistance) results for "long" transistor (it means increasing the length of the gate) and for small currents.

In figure 2.22 the transconductance is depicted as linearization around the operating point of the input characteristic. This g_m serves as the strength of the transistor [76]. A higher value corresponds to a greater change in the drain current for a given V_{gs}. In a larger sense the transconductance represents a kind of sensitivity of the transistor. The transconductance is expressed in Siemens (S) and is the inverse of the resistance (Ω). For an analog designer the MOSFET operates as voltage to current converter (transconductor), converting the voltage change to a current change. Calculating the derivative of g_m at the

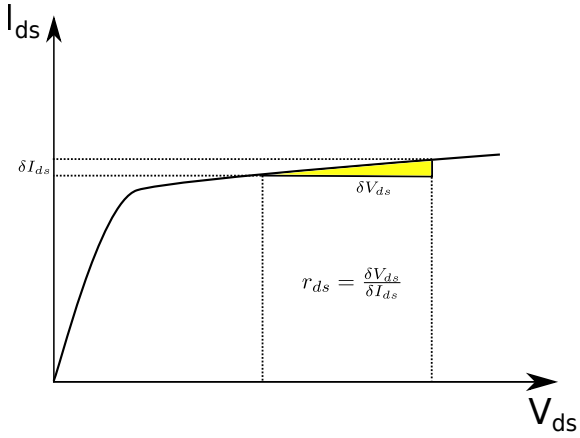

Figure 2.21: Output resistance caused by channel length modulation.

operating point produces the following expression:

$$g_m = \frac{\partial I_{ds}}{\partial V_{gs}}\bigg|_{V_{ds}=const} = \mu_n C_{ox} \frac{W}{L}(V_{gs} - V_{th}) \tag{2.50}$$

In order to find another expression that highlights the dependence of g_m on the DC drain current:

$$g_m = \sqrt{2\mu_n C_{ox}(\frac{W}{L})I_{ds}} = \frac{2I_{ds}}{V_{gs} - V_{th}} \tag{2.51}$$

2.2.2 Small-Signal Model

The large-signal model of MOSFETs describes the quadratic characteristics (Eq. 2.45) including voltage-dependent capacitances. For large-signal model the complete component equations are used for the calculation. The large-signal analysis therefore represents the true physical behavior of switching. In small-signal analysis, the circuit is first linearized at the operating point and then all types of analyses are performed with this linearized circuit. As a result, any changes in the operating point due to an analysis parameter are not taken into consideration (e.g. if the operating point of a circuit is frequency dependent). In many analog circuits the MOSFETs are biased in saturation to work as an amplifier, it is important to derive the small-signal model. To derive the small-signal model an approximation of the large-signal model around the operating point can be employed [76]. The procedure is used specifically is to apply a certain bias voltage to the terminals of the transistor. When the potential difference between two of the terminals is incremented, other terminal voltages remain constant. All terminal currents will change and can be measured. Specifically if the gate-source voltage changes, $\delta V = V_{gs}$, (here V_{gs} is the small-signal voltage) the result is that the drain current changes by $g_m V_{gs}$ (see figure 2.19). The model refers to voltage-dependent current source. Regarding channel-length modulation, drain-source

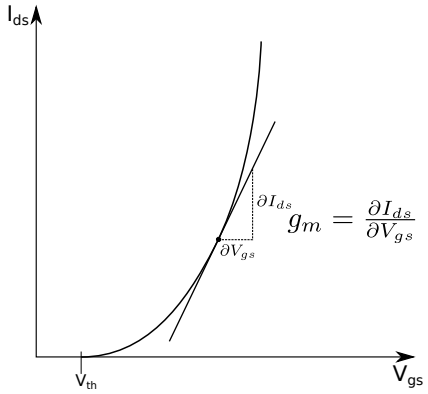

Figure 2.22: Transconductance determined by the slope of the input characteristic.

voltage varies with the drain current. This effect can be modeled by voltage-dependent current source. This current source is linearly dependent on the voltage across it and is equivalent to a resistor (figure 2.23). This resistor is applied between drain and source. The resistor can be calculated as follows

$$\begin{aligned} r_{ds} &= \frac{\delta V_{ds}}{\delta I_{ds}} \\ &= \frac{1}{\frac{1}{2} \cdot \mu_n C_{ox} \frac{W}{L} (V_{gs} - V_{th})^2 \cdot \lambda} \qquad (2.52) \\ &\approx \frac{1 + \lambda V_{ds}}{\lambda I_d} \\ &\approx \frac{1}{\lambda I_d} \qquad (2.53) \end{aligned}$$

The output resistance, r_{ds} affects the performance of the analog circuits limiting the voltage

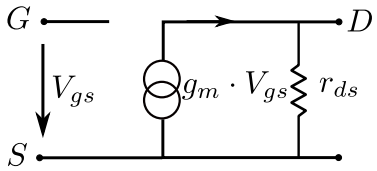

Figure 2.23: Channel-length modulation represented by a resistor.

gain amplifiers. The bulk potential influences the threshold voltage and hence the gate-source overdrive [76]. The bulk behaves as a second gate. The model dependencies change

by a current source connected between drain and source. A bulk transconductance can be defined and can be expressed as follows in the saturation region [76]

$$g_{mb} = \frac{\delta I_{ds}}{\delta V_{bs}} \tag{2.54}$$

$$= \mu_n C_{ox} \frac{W}{L}(V_{gs} - V_{th}) \left(-\frac{\delta V_{th}}{\delta V_{bs}}\right) \tag{2.55}$$

$$g_{mb} = g_m \frac{\gamma}{2 \cdot \sqrt{2\Phi_F + V_{SB}}} \tag{2.56}$$

$$\gamma = \sqrt{2e \cdot \epsilon_{si} N_{sub}} / C_{ox}$$

$$g_{mb} = (\eta - 1) \cdot g_m \tag{2.57}$$

The parameter η is defined as body factor. The bulk effect is taken into consideration adding a second current source. A total small-signal current can be calculated

$$i_d = g_m V_{gs} + g_{mb} V_{bs} + g_{ds} V_{ds} \tag{2.58}$$

As expected, g_{mb} is proportional to γ. For $g_m V_{gs}$ and $g_{mb} V_{bs}$ have the same voltage

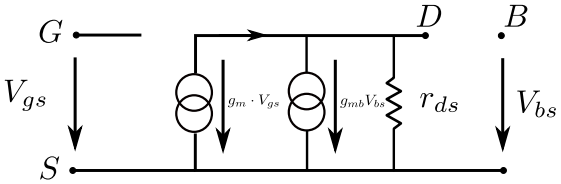

Figure 2.24: The body effect shown by adding a new dependent current source.

polarity. That means, raising the gate voltage has the same effect as raising the bulk voltage. In order to show the complete small-signal model all MOSFET device capacitances must be included. For low-frequencies the model in figure 2.24 can be used. For higher-frequencies, this model must be expanded accordingly due to parasitic capacitances. These are unwanted (parasitic) capacitances, but still part of the transistor. With the resistances in the circuit, they put an upper limit on the speed of the transistor. All these capacitances can be introduced into the low-frequency small-signal model to make it into a high-frequency small-signal model depicted in figure 2.25. Most of the capacitors in the small-signal model are related to the physical transistor. The largest capacitor is C_{GS}. This capacitance is primarily due to the change in channel charge resulting from a change in V_{gs}. It can be shown that C_{GS} is given by [86]

$$C_{GS} \approx \frac{2}{3} W L C_{ox} \tag{2.59}$$

The next largest capacitor is C_{SB} this is the capacitance between the source and the substrate. This capacitor occurs because of the depletion capacitance of the reverse-biased source junction, and it includes the channel-to-bulk capacitance. The size of the capacitance is given by [21]

$$C_{SB} = (A_s + A_{ch}) C_{js} \tag{2.60}$$

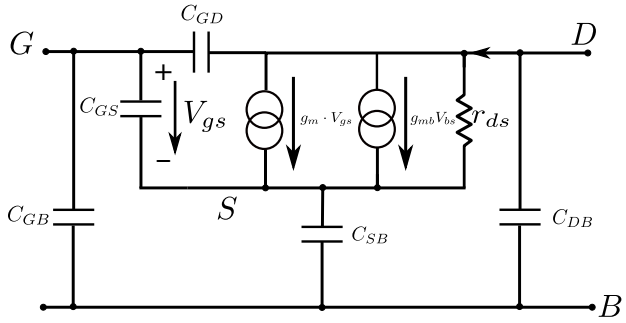

Figure 2.25: High-frequency small-signal model including all device capacitances.

where A_s is the area of the source junction, A_{ch} is the area of the channel ($A_{ch} = WL$) and C_{js} is the depletion capacitance defined as

$$C_{js} = \frac{C_{j0}}{\sqrt{1 + \frac{V_{SB}}{V_{bi}}}} \tag{2.61}$$

where C_{j0} is the depletion capacitance per unit area. The capacitance C_{GD} also called Miller capacitance. The Miller capacitance is essential when there is large voltage gain between gate and drain and can be defined as [21]

$$C_{GD} = C_{ox} W L_{ov} \tag{2.62}$$

The Miller capacitance (C_{GD}) is due to physical overlap of the gate and drain regions as well as fringing fields [21].

2.3 Processing Technology of CMOS Devices

In order to design integrated circuits, it is essential to understand the basic processing steps for manufacturing an integrated circuit in CMOS technologies. It is essential for an ASIC designer to know the limits of the process because during the layout certain rules must be considered. Today's semiconductor technology requires that ASIC designers have a good knowledge of each of these process steps. This chapter explains briefly the processing technology of CMOS devices. The aim is to have an understanding of the processing or fabrications steps and the importance for the design and layout of the integrated circuits. Modern CMOS processes have more than 200 processing steps. For the purpose of this work a brief overview of the following processes will be given: photolithography, oxidation, ion implantation.

Photolithography

Photolithography is a process used to pattern structures on the substrate (wafer). Light is used to transfer a geometric pattern or circuit layout information to a light-sensitive photoresist on the wafer. The layout consists of polygons showing different layers, e.g. n-well, source/drain regions, polysilicon, contacts etc. Each layer must be created with a very high precision. The active or diffusion serves as the source/drain regions of the transistors. Concerning n-well pattern, this must be transferred to a transparent glass mask using a very high accuracy electron beam. The wafer is covered by a thin photoresist material which engraves the pattern inside, or allows a new material to be deposited in the pattern on the material underneath the photoresist. The light pattern is projected over the mask or on top of the wafer. After the selective exposure of the photoresist to ultraviolet (UV) light the wafer is placed in an etchant that "removes" the part under the opaque geometric form (see figure 2.26). An n-well can be created in the exposed area. The light used for this process is UV light. Projection of a clear image of small size onto the wafer is limited by the wavelength. State-of-the-art photolithography uses deep ultraviolet light with wavelengths 248 nm to 193 nm which allow small feature sizes down to a few ten nanometers [76].

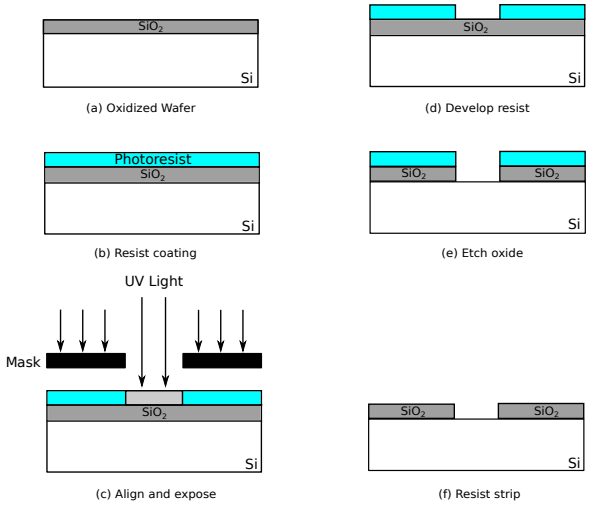

Figure 2.26: Simplified procedure of the most important steps for the implementation of photolithography [16].

Oxidation

When exposed in an oxidizing atmosphere at elevated temperatures (900-1200 °C), silicon forms a thin layer of oxide on the surface. The oxide of silicon formed is silicon dioxide (SiO_2). Figure 2.26 shows the important steps for the implementation of photolithography.

The SiO_2 serves as a dielectric device structures, such as gate oxide. In addition to serving as the gate dielectric, SiO_2 can act as a protective coating. In areas between devices, a thick layer of SiO_2, called "field oxide"(FOX), forms. The rate of growth depends on the type and pressure of the atmosphere, temperature and doping level of the silicon. Two types of oxidation can occur: oxidation in O_2 is called dry oxidation, whereas in H_2O it is called wet oxidation. In general, dry oxidation yields a denser and higher quality oxide than a wet oxidation. In a modern fabrication facility, oxidation occurs in a tube furnace or rapid thermal processing (RTP) tool [16].

Ion Implantation

The introduction of dopant impurities into silicon is necessary to affect majority carrier type, carrier concentration, carrier mobility, carrier lifetime and internal electrical field. The most common technology is ion implantation. This is the preferred method in modern CMOS fabrication processes. Dopant atoms are ionized and then accelerated as a high-energy focused beam, hitting the surface of the wafer and penetrating in the material.[76] Intensity and duration controls the doping level of the implantation. The energy of the beam sets the depth of the doped region. With a high energy the peak of the doping concentration occurs below the surface, creating a retrograde profile. As shown in figure 2.27 an n-well is formed by entering dopants into the exposed area [76]. A negative effect

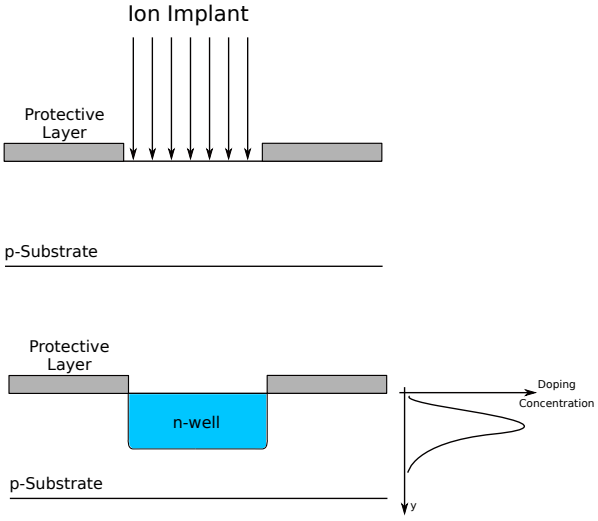

Figure 2.27: Ion implantation creating a retrograde profile [76].

is that ion implantation damages the silicon lattice. For this reason, the silicon or wafer is heated to roughly 1,000 °C for a certain period of time allowing the lattice bonds to reform [76]. During the creation of source and drain regions of a transistor, the annealing effect

leads to diffusion of dopants and broadening the structure in all directions. Annealing results an inside diffusion creating overlap with the gate area [76].

The fabrication of devices like NMOS, PMOS, resistors and capacitors requires deposition of different materials for example metal layers, dielectric materials, polysilicon. A deposition technique called chemical vapor deposition (CVD) is a common method used for CMOS processing. Heated gases react in the vicinity of the wafer and a chemical reaction produces the desired material. Contact structures with small dimensions and large depths must be etched with a very high precision. There are different types of etching. Wet or dry etching can be used to remove undesired metal. The use of fluorine and chlorine gases for metallization steps is called plasma etching (dry etching) [88]. Placing the wafer in a chemical solution is a wet etching process and works where low precision is sufficient. The plasma etching mentioned can be described as a bombardment of plasma gas on the wafer. Sharp profiles can be achieved with a high precision [76].

3 NeutrinomassV3 ASIC for the TRISTAN Experiment

Neutrino (also called "ghost particle of the Universe") is probably the most fascinating type of elementary particle. Neutrinos are the lightest particles in the Universe. In the Standard Model of elementary particle physics, the known neutrino types ν_e, ν_μ, ν_τ were initially assumed to be mass-less. Various experiments - atmospheric, solar and reactor neutrinos - indicate that the neutrino rest mass is different from zero. All these experiments show neutrino oscillations. Experiments such as Karlsruhe Tritium Neutrino Experiment (KATRIN) allow determination of the absolute masses. The components for this big experiment were assembled here at KIT. Furthermore, the Tritium Beta Decay to Search for Sterile Neutrinos (TRISTAN) experiment is planned at the KATRIN facility. The aim of the experiment is to search for heavy ("sterile") neutrinos by measuring the electron energy spectrum in β-decays. These neutrinos are candidates for dark matter particles; their mass in the range of a few keV.

Within the scope of this work, three prototypes of ASICs were developed for the TRISTAN experiment in standard 350 nm CMOS technology. The ASIC has low-noise amplifiers and shapers. Furthermore a 9-bit pipelined ADC with a sampling rate of 50 MS/s was developed to sample the waveform produced by the amplifiers. The general function of the KATRIN and TRISTAN experiment, the full ASIC functionality (amplifiers, shaper and pipelined ADC) and the readout system for the characterization of the ASIC and measurements will be discussed in this chapter. A silicon detector is used to detect the electrons. Silicon drift detectors will be produced in collaboration with the semiconductor laboratory of the Max-Planck Society Munich (HLL). Other detectors are being produced with a thin entrance window at Fondazione Bruno Kessler (FBK). All measurements were performed with FBK sensors at KIT.

3.1 KATRIN and TRISTAN Experiments

The KATRIN experiment is designed to measure the kinetic energy of electrons emitted in a tritium β-decay. The beta decay of gaseous tritium shows that an electron and electron-antineutrino are emitted. Tritium decays to ^{3}Helium, an electron and an electron anti-neutrino (see equation 3.1).

$$^3_1H \rightarrow ^3_2 He^+ e^- + \overline{\nu}_e \tag{3.1}$$

In nuclear physics a beta decay is a radioactive decay in which fast energetic electron or positron and a neutrino are emitted from the atomic nucleus. Two types of beta decay are known as beta minus and beta plus. In beta minus (β^-) decay, a neutron is converted to a proton, and the process creates an electron and an electron antineutrino. The beta

plus (β^+) decay a proton is converted to neutron and the process generates a positron and an electron neutrino (positron emission). Beta decay is the result of the weak force, which is characterized by relatively lengthy decay times. The nucleons consist of up quarks and down quarks. The weak force allows a quark to change type by exchanging a W boson (elementary particle for force carriers of weak interaction) and creating an electron/antineutrino or positron/neutrino pair production. KATRIN in particular emits electrons from high luminous tritium source. The released decay energy results in a continuous β-spectrum (figure 3.1). The energy region is near the energy endpoint $E - E_0$ where E_0 is 18.575 keV. This corresponds to the mass difference between neutron and proton multiplied by c^2. A non-zero mass of the electron antineutrino leads to a shift in the electron energy spectrum by the corresponding mass value. KATRIN can measure the spectrum around this endpoint (figure 3.1) with high precision [50]. Only electrons that leave the

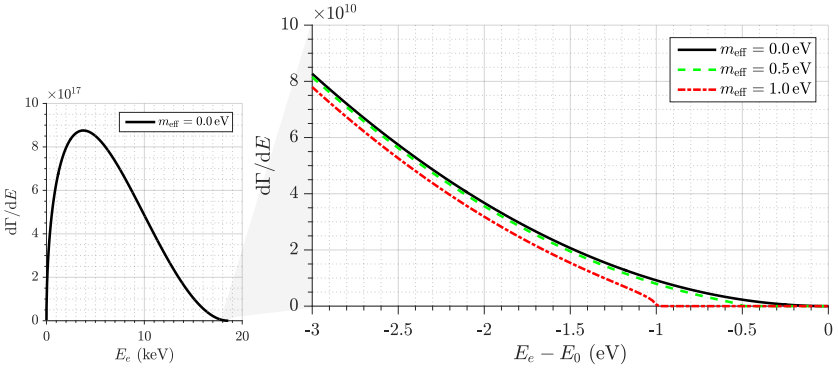

Figure 3.1: Energy spectrum of the tritium β-decay. Three different neutrino masses are shown (different graph color). These differ only in the range near the high-energetic endpoint [50].

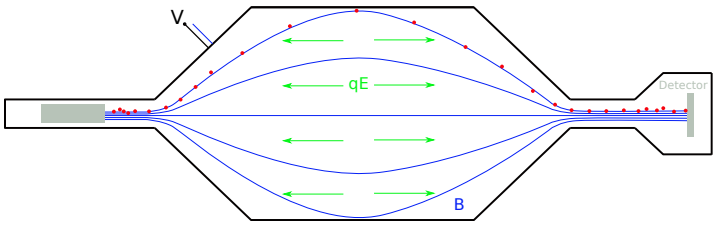

Figure 3.2: A simplified block diagram of the MAC-E filter. Electrons are guided through the spectrometer by the magnetic field. A negative potential is applied to the main spectrometer, such that only electrons with a higher kinetic energy than the potential can pass and follow the path to the detector.

tritium with almost the full decay energy are interesting for the experiment. Their energies of about $E_0 = 18.575$ keV must be measured and differentiated with a precision of 1 eV. This

can be achieved with simple particle detectors. Therefore two spectrometers, the so-called MAC-E filters (Magnetic Adiabatic Collimation in combination with Electrostatic Filter) are used. The MAC-E (see figure 3.2) filters out all electrons below the corresponding energy using an adjustable counter-voltage to produce an energy-dependent spatial resolution for the remaining electrons. Electrons with an energy of about 18.575 keV hit the detector (this energy is defined by the counter-voltage). In a pre-spectrometer a counter-voltage of about -18 kV reduces the electron flow and the same voltage is applied to guide back slower electrons that pass the filter.

TRISTAN investigates the technical realization of the keV-scale sterile neutrino search with KATRIN. It allows investigation of a mass range 0 keV to 18.6 keV. This would require a measurement of the whole β-decay energy spectrum and an increased electron rate at the detector by a factor of 10^8. At keV-scale the tritium β-decay spectrum is similar to eV-scale sterile neutrino. The tritium β-decay spectrum with no sterile neutrino is superimposed with an additional energy spectrum of keV sterile neutrino. The sterile neutrino spectrum is shifted by the mass of the sterile neutrino generating a distortion of the spectrum (kink) at $E_0 - 10keV$ (figure 3.3). The sterile neutrino is defined by its mass and the $sin^2(\theta)$ that allows us to calculate the probability that β decay generates sterile neutrino.

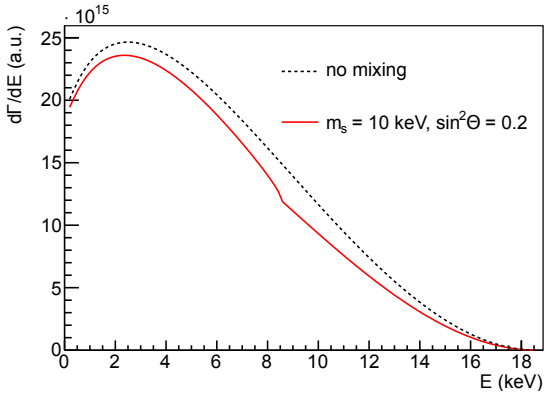

Figure 3.3: A tritium β-decay with no mixing (no sterile neutrino as dashed line) compared with a keV sterile tritium β-decay spectrum (red line). A distortion (kink) is manifested in the spectrum. Figure reproduced from ref. [64].

3.2 NeutrinomassV3 ASIC Architecture

The development of the Neutrino ASIC started in 2015. We developed all three versions in standard 350 nm HV-CMOS technology. A lot of improvements were made during three years of development. The NeutrinomassV1 ASIC contains an 8 channel two-stage low-noise amplifier. A special sensor diode (BARC diode) is implemented on the ASIC. This sensor diode is used to test the low-noise amplifiers. The input signal from Fe-55 source is about 1660 e, and leads to a signal at the output of the first amplifier of the

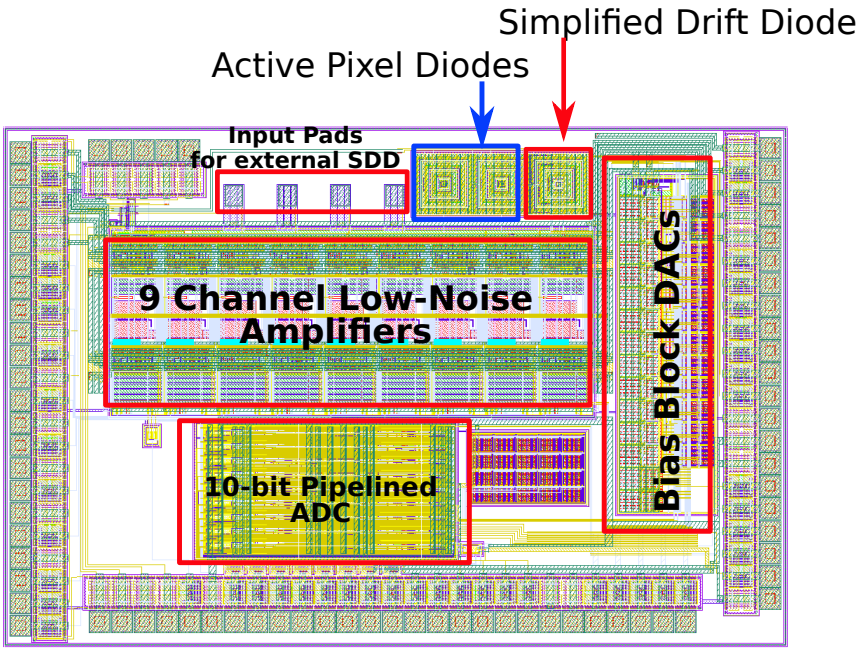

Figure 3.4: Layout of the NeutrinoV3 ASIC.

order of 10 mV. Noise at the output of the second amplifier corresponds to an input signal of ≈ 30 e. In the TRISTAN experiment an existing Data Acquisition (DAQ) system is used. Our idea is to digitize the amplified signal on the chip. A pipelined 8-bit (25 MS/s) analog-to-digital converter (ADC) was designed fully differential. Each stage performs a 1.5 bit conversion. Investigations of linearity and noise of the ADC will be shown in this chapter. The ADC works with a digital linear correction to improve the linearity. The linearity of the pipelined ADC is mainly limited by the capacitor mismatch and finite gain in the multiplying converter stage. The full ASIC layout is shown in figure 3.4. The NeutrinomassV3 has a simplified drift diode and two active pixel diodes. In addition, an active reset is implemented in one channel, as well as PMOS feedback channels, injection and a channel for JFET readout. The amplifier can operate in continuous and pulsed reset [20] mode. A 10-bit and 25 MS/s fully differential pipelined ADC is also implemented to digitize the amplitude. This version is tested with the FBK dodecagonal drift diodes. The preamplifiers are specifically designed to be used with a silicon drift detectors (SDD). Furthermore, the ASIC has a channel that can be used to read out an external SDD with integrated junction gate field-effect transistor (JFET). The ASIC can be used with the FBK diodes or the JFET inputs. Table 3.1 show the features of NeutrinomassV3 ASIC.

Different currents and thresholds are generated in a bias DAC block implemented on the periphery of the ASIC. These thresholds and bias currents are needed for the preamplifiers and for the fully differential pipelined ADC. The bias voltages for the current sources are generated by 12 on-chip 6-bit DACs. Each DAC is written by using a shift register.

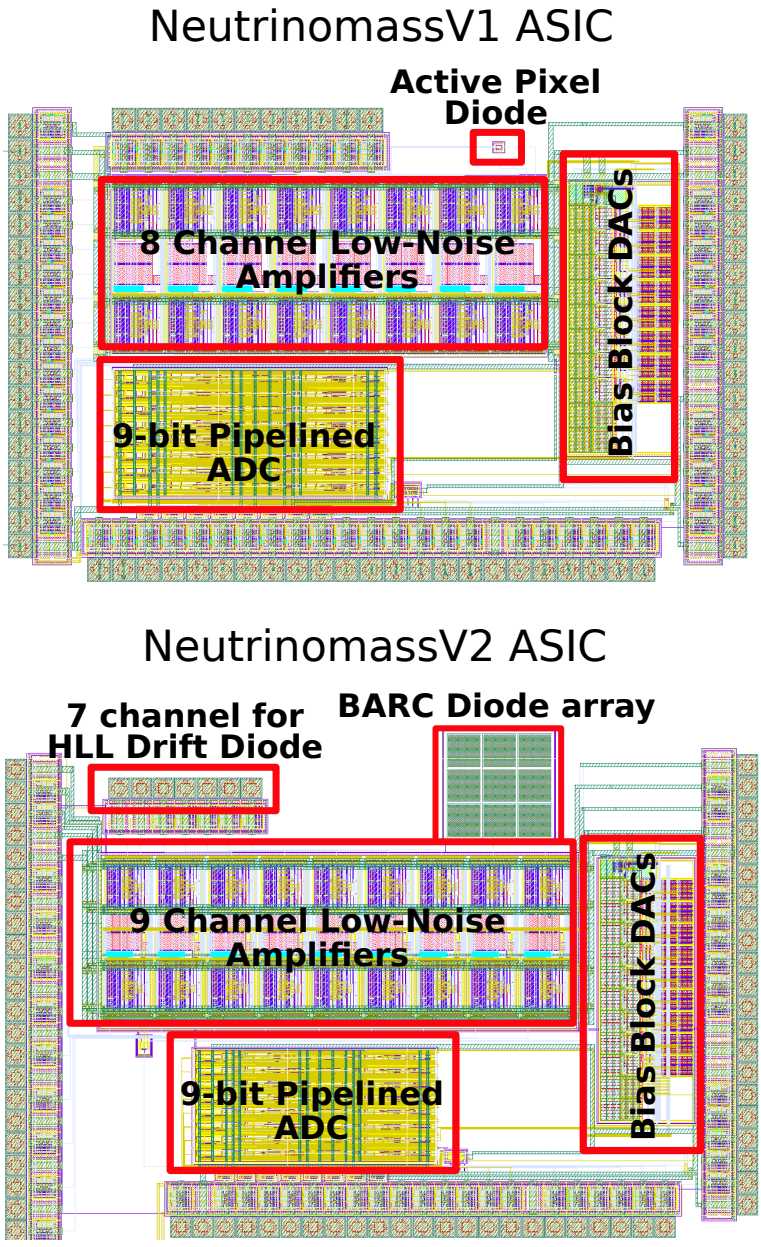

Figure 3.5: Top Layout of Neutrinomass ASIC V1 (top) and V2 (bottom).

Technology	AMS 350 nm HV-CMOS
Chip size	2.29 mm x 3.46 mm
ADC architecture	pipelined fully differential ADC
ADC resolution	10 bit
Amplifier	CSA + CR-RC shaper
FBK SDD entrance window	10 mm^2
FBK SDD J_{leak}@ 1.5 V_{dep}	285 pA/cm^3
Energy resolution	300 eV @ 5.85 keV

Table 3.1: NeutrinoV3 ASIC and FBK SDD main characteristics.

Figure 3.5 shows the NeutrinomassV1 with the active pixel diode and 8 channel low-noise amplifiers, 9-bit fully differential pipelined ADC and bias block. NeutrinomassV2 (figure 3.5) has a BARC diode array, 7 channel input for HLL drift diodes, 9 channel low-noise amplifiers, 9-bit fully differential pipelined ADC and bias block. With regard to the BARC diode array, BARC stands for bottom anti reflective coating, which is a special AMS 350 nm process to remove process layers above the active diode area. The diode consists of a simple N+ diode in a deep n-well. The collection time is constant. 6 µm of metal filling layers are removed, leaving a layer of 30-90 nm. A high-voltage pixel (active pixel diode) is used for the NeutrinomassV1 ASIC.

3.3 Low-Noise Amplifier Stage for TRISTAN Experiment

Low amplitudes in the range of 90-100 µV have to be amplified by the TRISTAN two-stage amplifier. Detector system requirements are: 10^8 counts/s with the capability of handling these high rates and with an excellent energy resolution of 300eV at 20keV. A low energy threshold of < 1keV is also required. Using the SDD it is possible to detect electrons with such energies. An SDD consists of a pn-junction. When a reverse voltage is applied to the sensor the detector will deplete. This depleted region corresponds to the sensitive volume. When the electron travels through the detector, it loses a certain amount of energy. Electron scattering occurs and its energy is deposited in the detector. This energy produces small amplitudes which have to be amplified. We designed a two-stage amplifier that consists of a charge sensitive amplifier (CSA) stage (first stage) and CR-RC shaper (second stage). The stages use two types of amplifying modes: continuous and pulsed reset mode. Pulsed reset is used to further reduce the noise level. For the in-house measurements the FBK SDD sensor and a FE-55 source was used. A test setup was developed including readout FPGA board and interface board. The SDD sensor and NeutrinomassV3 ASIC were bonded to the bottom and top of the interface board. As explained in chapter 3.4 the amplifiers are fully configurable with an application written in C/C++. A Qt graphical user interface (GUI) displays all adjusting bias voltages. Output signals such as unshaped and shaped signals are analyzed and histograms created to display the spectrum and calculate the noise level. Histograms can be calculated offline, but we used the oscilloscope histogram function to characterize the spectrum. The idea of using the oscilloscope histogram function is to find the best setting during the tuning process in order to keep the noise as low as possible and

produce a low-noise spectrum. Another measurement variant is to operate the interface board in vacuum. These conditions simulate the inside of the Tristan detector and the sensor can be fired with electrons using an e-gun.

3.3.1 The Charge Sensitive Amplifier and CR-RC Shaper

In our multi-stage front-end ASIC uses the CSA. A CSA is a form of transimpedance amplifier in which the feedback resistor is necessary to set the DC operating point. If it is large enough it gives a negligible contribution. In other words the feedback resistor (R_f) is infinite. An ideal charge sensitive amplifier is shown in figure 3.6 [78]. The CSA behaves

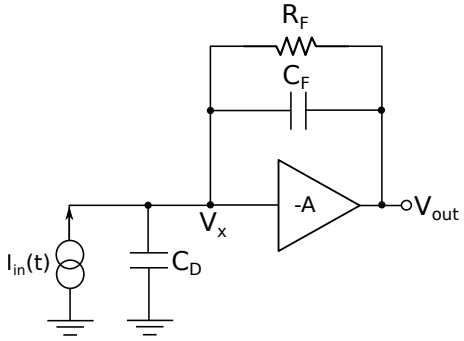

Figure 3.6: Ideal Charge Sensitive Amplifier.

as an ideal integrator and the output voltage can be calculated by

$$V_{out} = \frac{1}{C_f} \int i(t)dt \tag{3.2}$$

The input signal can be approximated with a δ-like pulse

$$I_{in}(t) = Q_{in}\delta(t) \tag{3.3}$$

Studying the transfer function of the CSA we obtain

$$H(s) = \frac{V_{out}}{I_{in}} = -\frac{R_f}{1 + sC_fR_f} = -\frac{R_f}{1 + s\tau_f} \tag{3.4}$$

When $s\tau_f \gg 1$ and $\mathscr{L}\{I_{in}(t)\} = I_{in}(s)$ (Laplace transform) becomes

$$V_{out}(s) = -\frac{Q_{in}}{sC_f} \tag{3.5}$$

For large open loop gain (A) the detector capacitance C_D can be neglected as follows

$$V_{out} = -AV_x = \frac{-AQ_{in}}{sC_D + sC_f(1+A)} \approx -\frac{Q_{in}}{sC_f} \tag{3.6}$$

The boundary between the two frequency regimes $s\tau_f \gg 1$ and $s\tau_f \ll 1$ is for $sR_fC_f = 1$. For lower frequency values, the impedance of C_f is greater than R_f and they are in parallel, the current flows through R_f. For high frequencies, the impedance becomes smaller, and the current flows through C_f. Taking the Inverse Laplace Transform ($\mathscr{L}^-1$) the following temporal domain is obtained for the output voltage

$$V_{out} = \frac{Q_{in}}{C_f} e^{-(t/R_fC_f)} \tag{3.7}$$

The output reaches the peak value given by the Q_{in}/C_f. The CSA transforms the charge pulse into voltage. The output of the preamplifier is processed by CR-RC shaper to improve the signal to noise ratio (S/N) [22]. This kind of shaper or filter consists of high-pass filter followed by a low-pass filter (see figure 3.7). For decoupling the two time constants, the filters are isolated by a voltage buffer. The derived CSA amplitude is multiplied by the next stage. In the Laplace domain the total transfer function can be written as:

$$V_{out} = \frac{Q_{in}}{C_f} \frac{\tau_1}{(1+s\tau_1)(1+s\tau_2)} \tag{3.8}$$

where $\tau_1 = R_1C_1$ and $\tau_2 = R_2C_2$. We can observe that the output of the low-pass filter

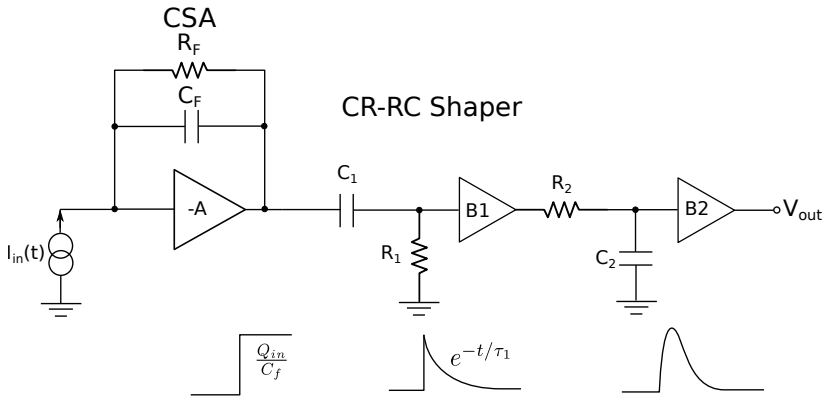

Figure 3.7: CSA and CR-RC shaper as high-pass and low-pass filter.

reaches the maximum value of the CSA (Q_{in}/C_f) and then decays exponentially by the time constant τ_1. The role of the high-pass filter is to slowly cut the varying parameter of the CSA output. This creates a signal that goes back to the baseline before the next pulse arrives. The output starts fading immediately after the peak value has been reached. The peak must be captured very fast. To filter out the high frequency components, a low-pass

filter is used at the second stage. The low-pass filter acts as an integrator (high-pass filter as a derivator). Calculating the $\mathscr{L}^{-}1$ of 3.8 the time domain signal is obtained[78]

$$V_{out}(t) = \frac{Q_{in}}{C_f} \frac{\tau_1}{\tau_1 - \tau_2} (e^{-t/\tau_1} - e^{-t/\tau_2}) \tag{3.9}$$

In the case where the time constants are equal $\tau_1 = \tau_2 = \tau$ in the Laplace domain

$$V_{out}(s) = \frac{Q_{in}}{C_f} \frac{\tau}{(1+s\tau)^2} \tag{3.10}$$

which results in

$$V_{out}(t) = \frac{Q_{in}}{C_f} \left(\frac{t}{\tau}\right) e^{-t/\tau} \tag{3.11}$$

By taking the derivative of equation 3.9 and setting the equation to zero the time can

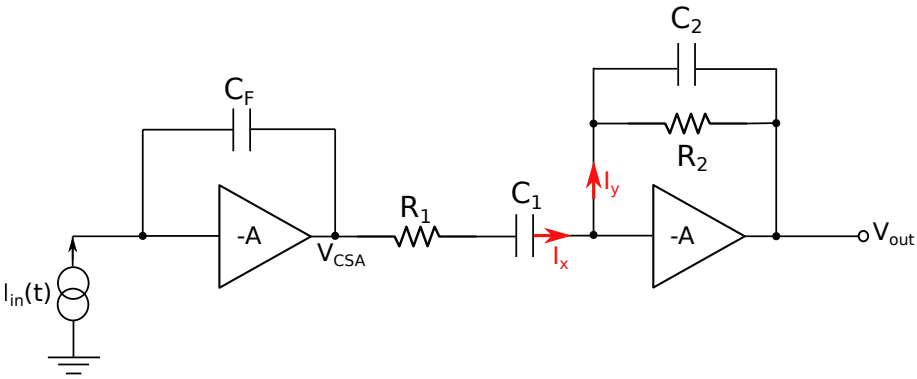

Figure 3.8: Real implementation of TRISTAN amplifier.

be calculated when the signal reaches its peak or the so-called peaking time. In the real implementation there are a couple of differences with respect to the CR-RC shaper in figure 3.7. No buffers are inserted to decouple the stages. Instead of buffer stages amplifiers are used which are implemented as a folded cascode amplifier topology. The output of the CSA is connected to high-pass filtering capacitance. Figure 3.8 shows the real implementation

structure of the TRISTAN low-noise two-stage amplifier. Calculating the transfer function for the real shaper implementation we obtain (using KCL):

$$
\begin{aligned}
I_x &= \frac{V_{CSA}}{\underline{Z}_1} \\
\underline{Z}_1 &= \frac{sR_1C_1 + 1}{sC_1} \\
\underline{Z}_2 &= \frac{V_{out}}{I_y} = \frac{R_2}{sR_2C_2 + 1} \\
I_x - I_y &= \rightarrow I_x = I_y \\
V_{CSA}\frac{sC_1R_2}{1 + sR_1C_1} &= V_{out}\frac{1 + sR_2C_2}{R_2} \\
H(s) = \frac{V_{out}}{V_{CSA}} &= \frac{sC_1R_2}{(1 + sR_1C_1)(1 + sR_2C_2)} \\
&= \frac{sC_1R_2}{(1 + s\tau_1)(1 + s\tau_2)}
\end{aligned}
\tag{3.12}
$$

The transfer function has two poles (τ_1, τ_2) which can be calculated from the denominator.

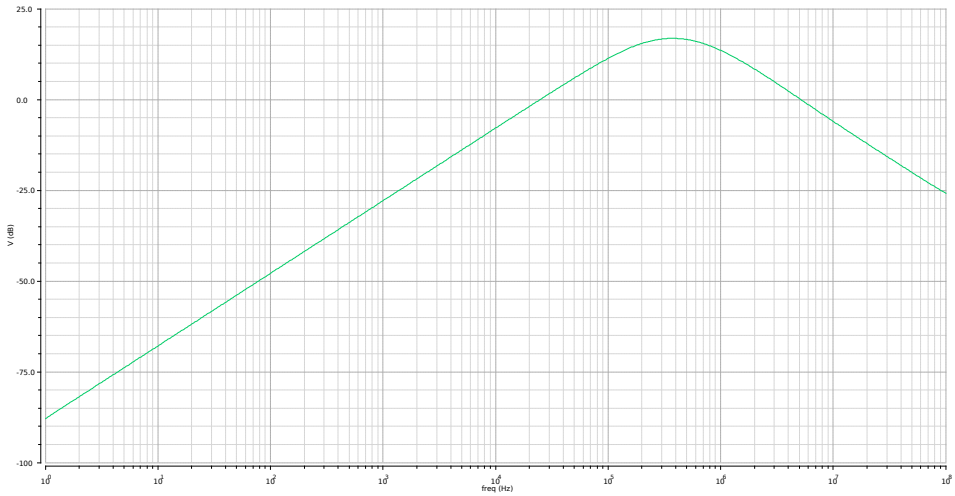

Figure 3.9: Bode plot of the shaper in closed loop configuration.

By solving the equation $N(s) = 0$ (nominator) we find the zeros of the transfer function. Zeros and poles can be located on the left or right half of the S plane or complex plane. If the poles are right half plane then the system oscillates because its time-domain grows exponentially. Even when the poles are at zero in the complex plane the system sustains oscillations. On the other hand, if the poles are on the left half plane the system is stable and the oscillation decays to zero due to the exponential behavior. Observing the AC response of the entire system we obtain Bode plots for magnitude and phase.

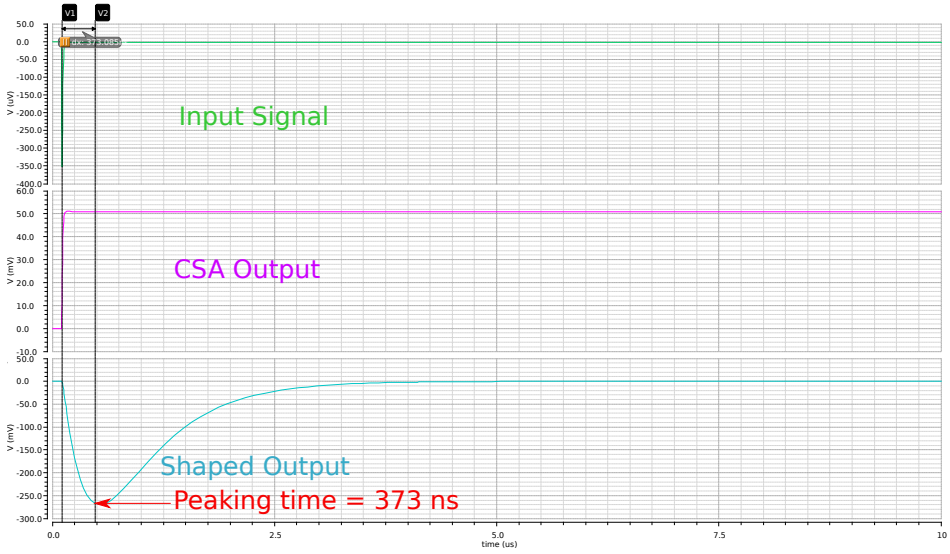

Figure 3.10: Transient response of the generated input signal, CSA and shaper output.

The calculated poles at $\tau_1 = R_1C_1$ and $\tau_2 = R_2C_2$ are shown in the simulation in figure 3.9. As expected, the nominator sC_1R_2 produces a slope +20db/dec. A low-pass function is realized with the two other poles. The first compensates the +20dB/dec and the second pole produces -20dB/dec. Thus, a characteristic transfer function for the real shaper implementation is generated.

Furthermore, studying the transient response (figure 3.10) of the amplifiers we obtain an unshaped signal with an exponential behavior that leads the signal to zero. The shaped output has a peaking time of 373 ns.

3.4 Measurement Results of the NeutrinomassV2 Pipelined ADC

In addition to ASIC development, application and firmware implementation is an essential part of testing the ASIC and analyzing the data. Before testing an ASIC a test setup consisting of a test board (interface board) and FPGA readout board is required. The interface board was designed in an Altium Designer Environment [11]. The development board consists of two additional smaller voltage boards providing auxiliary voltages for the pipelined ADC and TRISTAN amplifier. A Digilent Nexys ATLYS [31] development board containing a Xilinx Spartan 6 FPGA is used to read out the data. Read out of the ASIC can also be done by including firmware and software implementation configuration.

The communication between client PC and FPGA is implemented via asynchronous EPP Protocol. The Digilent asynchronous parallel interface is patterned after the EPP mode. The interface consists of an 8-bit bi-directional data bus and four control signals. The data

transfer speed that can be achieved depends on the particular communication subsystem and firmware version being used. The asynchronous parallel interface is made up of an address register and a set of 8-bit data registers. The address register can be up to eight bits wide. It holds the data register currently being accessed. Access to the registers is accomplished via bus cycles. The three type of bus cycle are Address Write, Data Read and Data Write. An address write cycle is used to write to the address register. Data Read or Data Write cycles read or write the data register whose address is currently held in the address register. Once an address has been written to the address register, any number of data read or write cycles can be performed on the selected data register. The EPP interface protocol was implemented as Verilog module acting as a finite state machine (FSM). An ATLYS general user constraint file (UCF) allows all pins between ASIC and FPGA to be defined [30]. In addition, in the UCF-files, logical networks (input and output port in the top level design) are connected to physical pins. Timing conditions can also be mapped.

The test application (software) for NeuntrinomassV3 ASIC is mainly based on Scala software. The Scala language is a functional language which complies to Java Byte Code. Hence, the applications run on the standard Java Virtual Machine and any Java library can be used. The software interacts with the register file generator (RFG) [57]. The RFG offers a tool command language (TCL) interface to describe registers present in FPGA or ASIC, and generate the required hardware/software interfaces. As result of the RFG, a Verilog output file is generated which can be connected to an existing firmware project. Different modules build up the firmware: a top level module, RFG interface file, read/write FIFO. This interface includes intercommunication between top level design and FPGA, EPP main control unit. The main control unit uses an FSM that controls the four EPP control signals as well as Address read/write and Data read/write. Asynchronous design is implemented for the EPP protocol. Further, the deep shift registers are implemented in the design to delay the data received from the ASIC. As previously mentioned, a software application written in Scala offers a Webserver interface which has the functionality to send a specific configuration to the interface board and configure the ASIC. Readout of the ASIC is possible and the configuration of the auxiliary voltage is controlled by voltage boards. In order to configure the ASIC, the bias block write sequence must be used. The sequence uses two-phase clocking (Ck1, Ck2), SIn and finally the generation of the Load signal to transmit the data (see chapter 3.6).

Different kinds of low level function were implemented to send and receive data. To define the registers for bias voltages and auxiliary voltages, a TCL script generates a predefined register output file for the RFG. Figure 3.11 shows a photograph of the hardware test setup. FPGA board (left) is connected via a VHDCI connector to the NeutrinomassV2 ASIC interface board (right).

The results which were carried out with the test setup described above are considered here. As previously mentioned, to control the pipelined ADC, an exact control sequence of amplify and reset must be implemented in the FPGA. Shifting the produced data in the deep shift registers must also be precisely controlled. This sequence was realized as follows: the application sets two coarse sampling points and one fine sampling point. Two internal counters are used. If the value of the first counter matches sampling points 1 and 2 and the second counter matches the fine sampling point, then the data is shifted to the deep shift registers. Thus, b1 and b0 (residual digital outputs) are generated. To

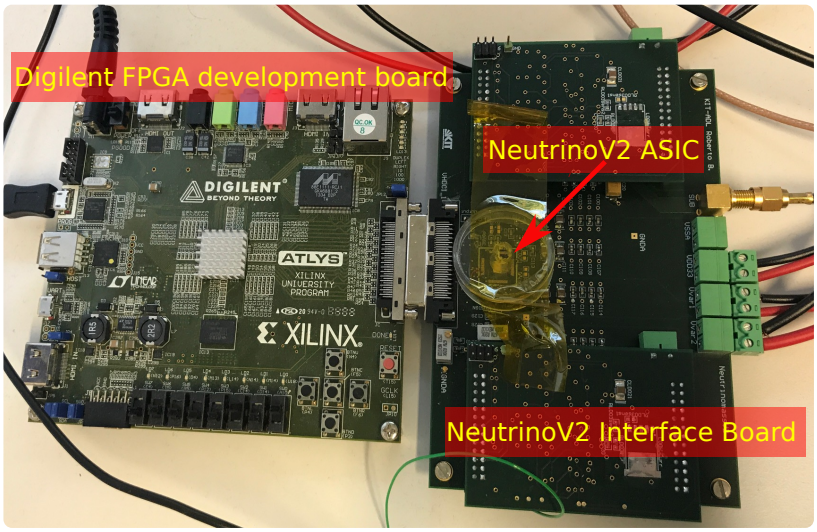

Figure 3.11: Hardware test setup consisting of FPGA and interface board.

control the amplify and reset sequence, the second counter must be compared with an ADC change time. This value corresponds to a sampling point. If this point is wrongly set, the ADC data received will become worse. This creates ampR and the inverted ampR control signal. All other signals like ampBR, ampBL (amplify), resL, resR (reset) and sel (select for multiplexer stage) are derived from these two generated signals. The L and R stand for left and right due to the fully differential architecture. The produced data are loaded in an internal FPGA FIFO and can then be read out by the application.

The expected transfer characteristic was measured. Due to the imperfections of the pipelined ADC architecture, we will observe nonlinearities in the transfer characteristics (see figure 3.12). For that reason, a software linear correction was prepared. A linear correction is calculated by introducing an alpha coefficient (α) which corresponds to the gain as follows

$$\begin{aligned} result = (Q_1(0) - Q_2(0)) + \alpha(Q_1(1) - Q_2(1)) + \alpha^2(Q_1(2) - Q_2(2)) \\ +.. + \alpha^7(Q_1(7) - Q_2(7)) = \sum_{j=0}^{7}[Q_1(j) - Q_2(j)]\alpha^j \end{aligned} \tag{3.13}$$

Alpha should be a value of 2.0 (because of the multiplication factor of the multiply-by-two stage). Due to the mismatch of the multiply-by-two stage (a noninverting feedback amplifier using an input and a feedback capacitance) the alpha value should be corrected between 1.95-2.02. The multiply-by-two gain circuitry depends on the capacitor mismatch. To measure the error in the pipelined ADC, a software integral nonlinearity (INL) error calculation is used. The INL specifies the deviation of pipelined ADC transfer characteristic from its ideal value. We used the endpoint method to calculate the INL error. Another

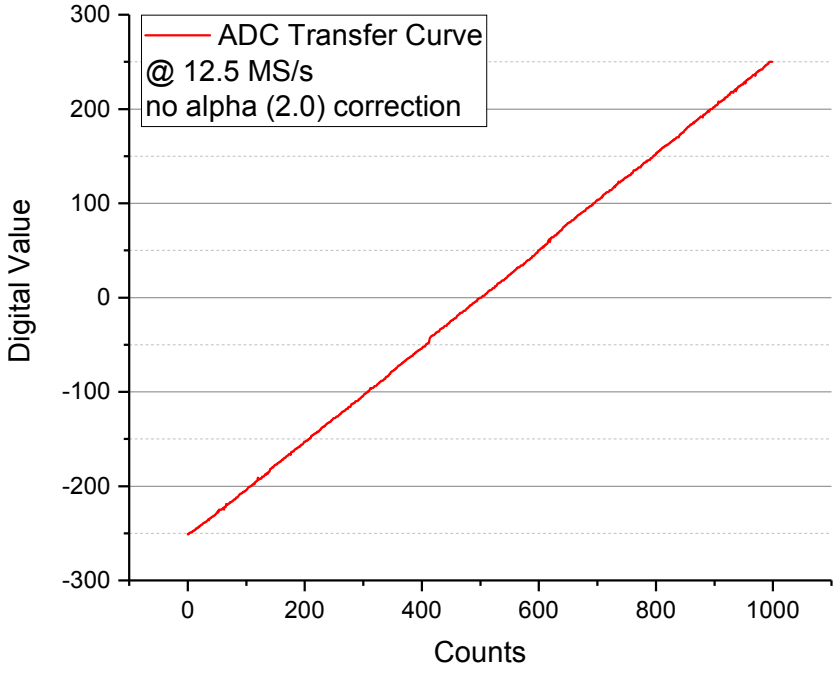

Figure 3.12: Uncorrected transfer characteristic.

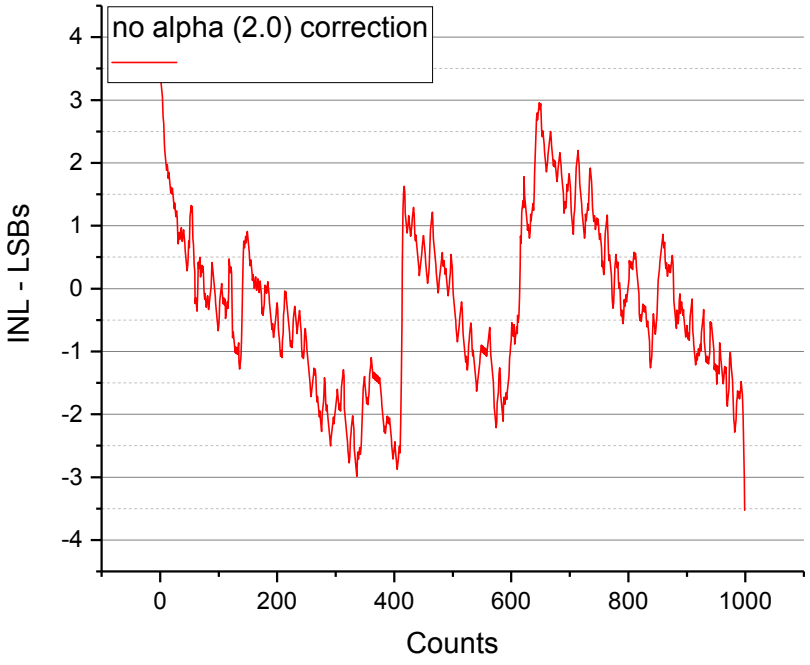

Figure 3.13: Uncorrected INL.

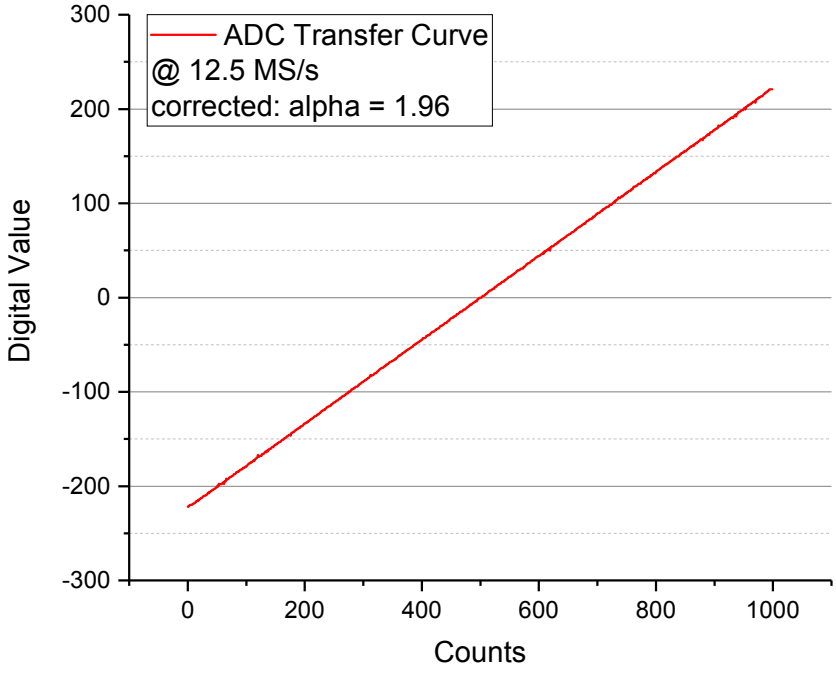

Figure 3.14: Linear correction applied to transfer characteristic.

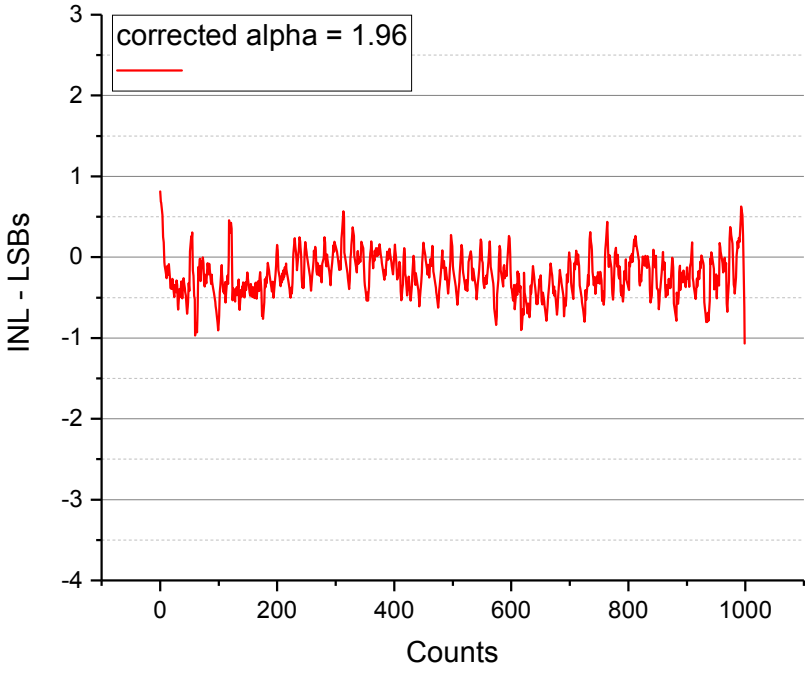

Figure 3.15: Linear correction applied to INL.

measurement would be to find the best-fit straight line so that the maximum difference is minimized [21]. In our measurement we swept the input voltage over its full scale range. A first measurement was performed without using the linear correction (see figure 3.12). A second measurement shows that the applied linear correction improves the INL error significantly (see transfer characteristics in figure 3.14). In NeutrinomassV2 the pipelined ADC is running at 12.5 MS/s and has a resolution of 9-bit. The full speed is 25 MS/s. Because of a software application bug, the speed of 25 MS/s could not be reached. Alpha coefficient is set to 2.0 for uncorrected measurement result.

For the corrected measurement result the alpha coefficient is set to 1.96. This alpha value gives the best result with the smallest INL error but slightly smaller gain (slope of the transfer characteristic). A standard deviation of $\sigma = 1.28$ (uncorrected) and $\sigma = 0.26$ (corrected) was measured. In terms of LSB measurement a fluctuation between -3.5 to 3.47 LSB (figure 3.13) and -1.06 to 0.81 LSB (figure 3.15) is calculated. Due to the significant improvement with linear correction, the next step would be to integrate a digital linear correction on-chip. Further improvements for the next generations are the integration of more parallel pipelined cells to achieve a higher resolution and an interleaved structure of the pipelined cells to obtain higher sampling speed.

3.5 Measurement Results of the NeutrinomassV3 Two-Stage Amplifier

The test setup developed in our group consists of interface boards and an FPGA board which is responsible for the ASIC configuration. Instead of Digilent ATLYS FPGA board and a Scala application (see chapter 3.4) the setup is implemented with a new Digilent FPGA board and a C/C++ software. Based on our experience, a C/C++ application is more versatile than a Scala application and reduces development time. Configuring the Nexys board is very similar because the interface used to create the registers was generated by the RFG. An Fe-55 source is used for the measurement. The Fe-55 decays and the SDD (FBK drift diode) generates a signal that is then amplified by the two-stage amplifier. Output signals are sensed by fast oscilloscope with the capability to histogram the maximum values of all the amplified output amplitudes. An output spectrum of Fe-55 source is measured. As shown in the simulation (see figure 3.10), the output amplitude (≈ 275 mV) is large enough to be measured without using an external amplifier. The normal settings of the oscilloscope are used as well as 1MΩ input impedance and AC coupling to remove the DC part of the signal. A DAQ system that will later be used for the TRISTAN experiment requires output amplitudes of at least 150 mV for the ADCs. As mentioned before our test setup is based on Nexys board. There are preparations for the DAQ system to be developed by another group at KIT IPE. It is also possible to store the data and use a trapezoid filter that has been developed with the C/C++ software. An offline analysis is also possible and remote connection can be established. The remote connection is established by a Raspberry Pi Linux computer. The C/C++ application also runs on Raspberry Pi. Finally, two measurements were carried out: using Fe-55 source (outside measurement in room temperature environment) and interface board used in the vacuum chamber. The sensor is then hit with electrons generated by an e-gun. It was also necessary to design the interface

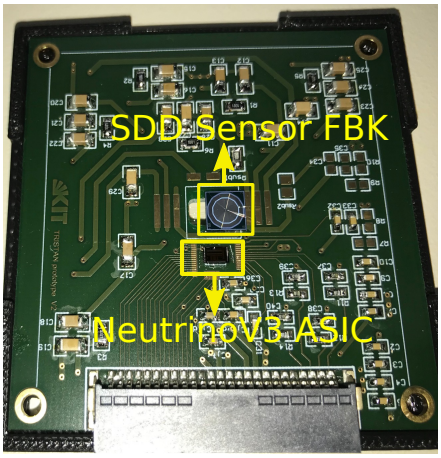

Figure 3.16: Interface board of the NeutrinomassV3 ASIC.

board in such a way that it fits with the vacuum flange. The SDD sensor is positioned exactly in the middle of the electron beam line. Photographs of the test setup are shown in figures 3.16 and 3.17. The measurements show promising results with the Fe-55 source. The signal output amplitudes produced for unshaped output is about 38 mV (figure 3.18) and for the shaped output is, as expected, 200 mV (3.19) and peaking time roughly corresponds to the simulated value. 50k analog signals were histogrammed at room temperature to determine the Fe-55 spectrum. The resulting spectrum shows (see figure 3.20) a full width at half maximum (FWHM) of about 28 electrons (240 eV). Electronic noise is given as input referred noise and expressed by the equivalent noise charge (ENC) in electrons (e). The first peak corresponds to 5.9 keV and the second smaller peak to 6.5 keV. All measurements were done in continuous reset mode. To reduce noise level we implemented a pulsed reset mode. These measurements will be presented in a separate paper.

3.6 Bias Block Design

This chapter briefly describes the design of the bias block. It generates the bias voltages for the current sources for the amplifiers and the pipelined ADC. The bias voltages are generated by a current flow through a diode-connected MOSFET. Every bias voltage can be adjusted using a current-mode 6-bit DAC [70] (Digital-to-Analog Converter). Each DAC register segment consists of 7 bits (one spare bit). The DACs are written using a shift register. The write sequence (see figure 3.21) uses the signals **SIn**, **Ck1**, **Ck2** and **Load**. The clocks are running in the two-phase clocking mode. A non-overlapping two-phase clocking scheme must be used in order to obtain a robust clocking. After the **Load** signal is issued, the content of the shift register is stored in the latches. There is a latch memory cell attached to each bit of the shift register. It is possible to implement a read back functionality. An additional signal **ShiftEnB** is used. The basic element for biasing

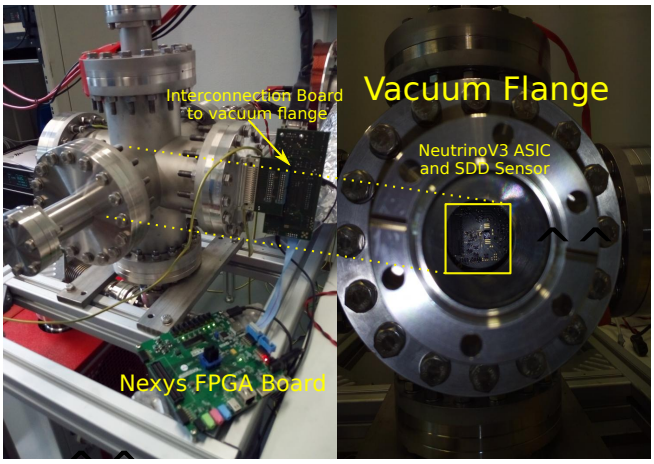

Figure 3.17: Vacuum measurement setup: Nexys board and NeutrinomassV3 ASIC inside the vacuum flange. Photo courtesy of M. Korzeczek.

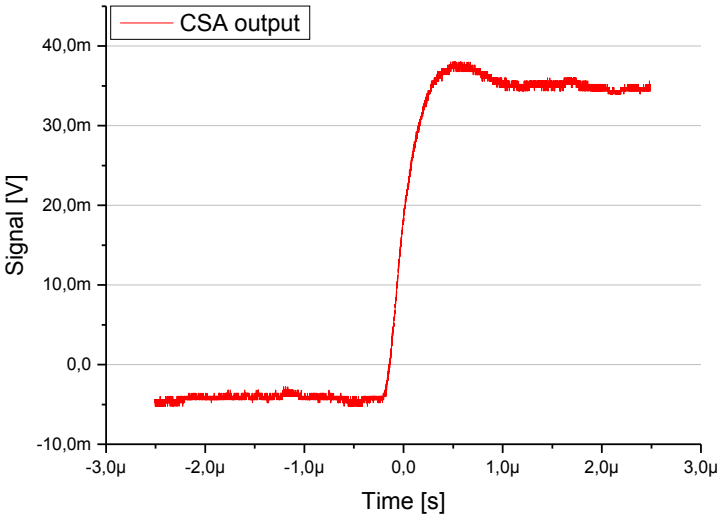

Figure 3.18: CSA output response to Fe-55 source (at room temperature).

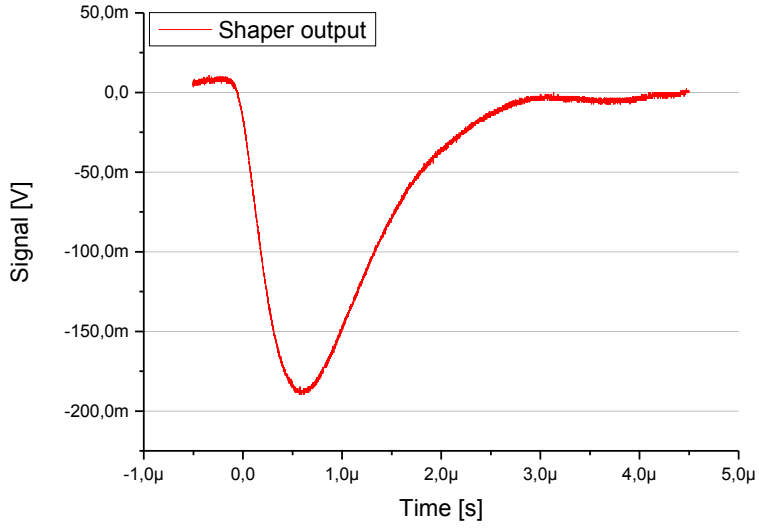

Figure 3.19: Shaper output response to Fe-55 source (at room temperature).

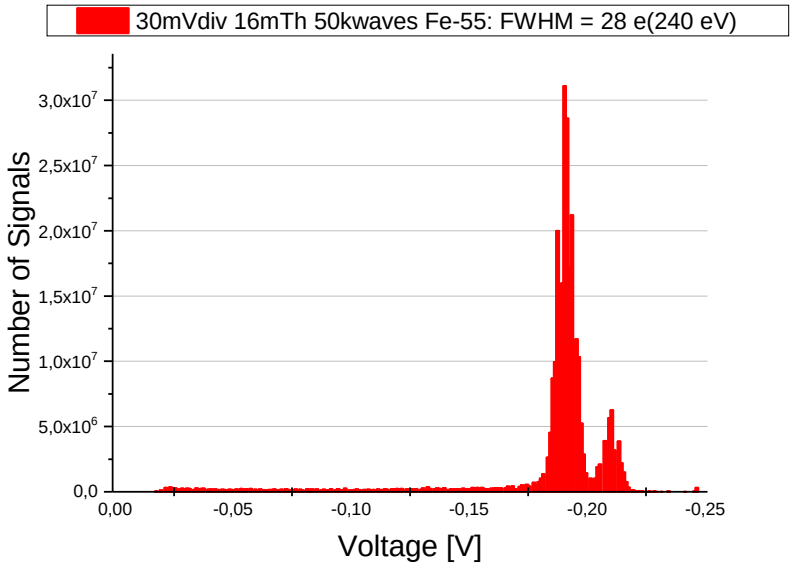

Figure 3.20: Fe-55 spectrum produced with 50k analog outputs. Noise of about 28 e (240 eV) was calculated from measured Fe-55 spectrum.

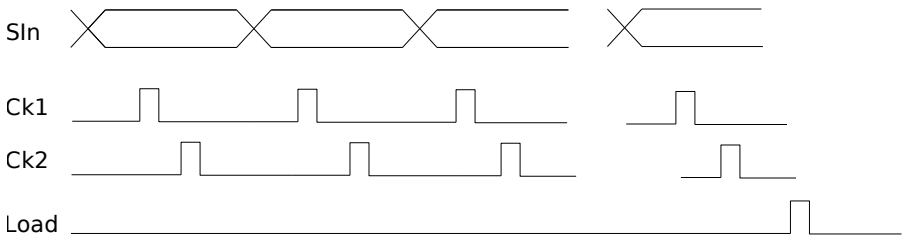

Figure 3.21: Configuration sequence.

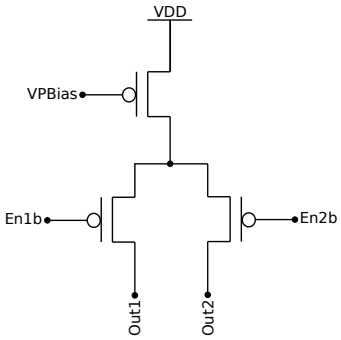

Figure 3.22: Differential PMOS current source (MDAC cell).

is the current mirror. A current mirror replicates an injected current. In an ASIC, current mirrors are always implemented as MOS transistors. Furthermore, another key component is the DAC to allow adjustable bias currents. A DAC generally, converts a digital input value into an analog signal or, more general a physical quantity. This physical quantity is usually a current or voltage. The current-mode 6-bit DAC consists of the shift register and multiplying DAC (mDAC) cells [70]. The mDACs are designed as differential current source topology. It is a core-matrix with 63 identical PMOS current sources. The current sources are grouped in six groups consisting of 1, 2, 4, 8, 16, 32 sources (see figure 3.23). The drains are connected together. The current generated by the PMOS transistors can be directed either into the output **Out1** or **Out2**. The output **Out1** is connected in this case to ground. Using the switch transistors (PMOS) are controlled by the differential logic signals **EN1b** and **EN2b**. The switches are arranged using the binary weighted scheme which implemented for its simplicity. There are other possibilities to implement a current mode DAC converter. The generated current from the differential PMOS current source is then distributed to current mirrors that are connected as a current divider topology. The current divider consists of distributed current mirrors with multiple outputs. One of the outputs is connected to diode-connected transistor to generate the desired bias voltage for the ASIC. A simplified block schematic of the ASIC bias block is shown in figure 3.23.

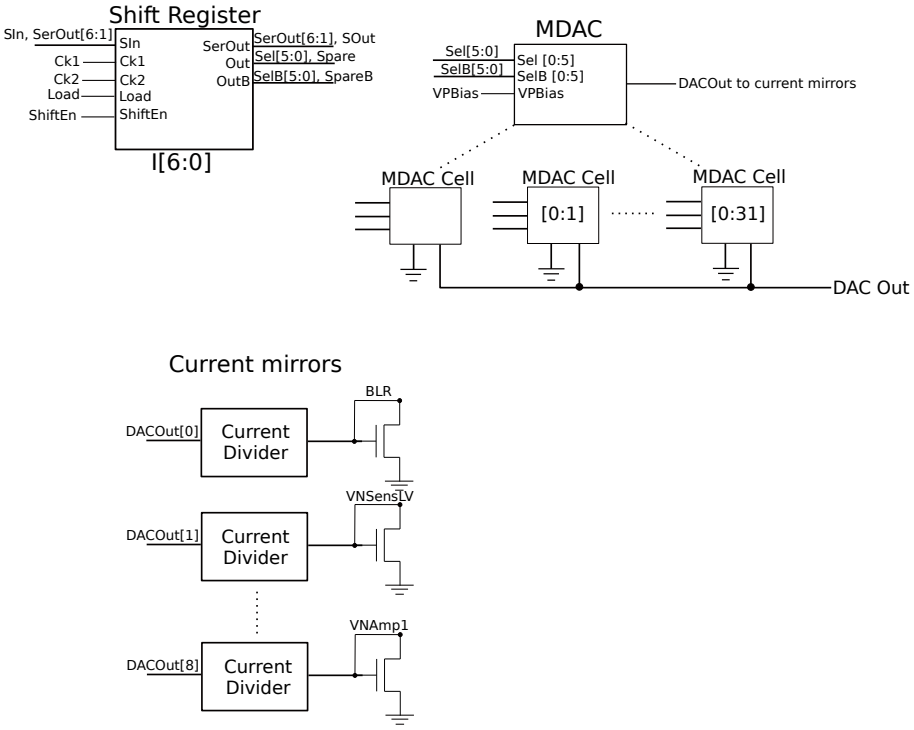

Figure 3.23: Simplified block schematic of the bias block.

4 Single Photon Avalanche Detector in Standard CMOS Technology

In chapter 3 we described the design of the NeutrinomassV3 ASIC as an amplifier and ADC ASIC for TRISTAN. For the TRISTAN experiment an external sensor is used to detect electrons. In this chapter we will present a different kind of ASIC. It consists of a single-photon sensitive sensor and readout electronics built as monolithic detector. This monolithic avalanche diode array is designed in a commercial AMS 350 nm high-voltage CMOS (HV-CMOS) process. It comprises a single-photon avalanche diode (SPAD), active quenching circuit [26] and readout electronics. The SPAD consists of a p+ diffusion/n-well junction surrounded by a shallow p-well acting as guard ring to prevent edge breakdown. This monolithic detector has a matrix of 20 x 15 pixels. The ASIC was developed at KIT-ADL.

A growing interest in single-photon detection and imaging can be observed in various scientific and industrial applications such as astroparticle physics (e.g. cherenkov light detection), meteorology, medicine (e.g. positron emissions tomography), biology (e.g. bioluminescence). Light detection at the single-photon level was invented more than 80 years ago. The photomultiplier tube (PMT) was the first device used for single-photon detection. In 1887 Heinrich Hertz discovered the photoelectric effect [45]. Albert Einstein later found that this effect is definitely quantum, and photons of light transfer their energy to electrons or to a single electron [34]. The photoelectric effect is used to translate photons to an electrical signal that can be amplified. However, the signal generated by only one electron is too small to be detected. In 1902 L. Austin and H. Starke discovered the mechanism of secondary emission [14]. A secondary emission arises when an incoming primary electron has enough energy to release the electrons in the target material. The electrons emitted in this process are called secondary electrons. The number (on average) of generated secondary electrons increases rapidly with the energy of the primary electron; then saturation occurs. Saturation depends on the target device but most surfaces saturate at primary electron energies of ≈ 1 keV [10]. Leonid Kubetsky invented a multistage secondary emission device where the output produced is transmitted to a secondary emission stage [51]. With this technique, he was able to overcome the gain limitation. Kubetsky measured an amplification gain of about 1000. PMTs are designed with an evacuated glass housing containing a photocathode, several dynode stages and an anode. A PMT has high gain, low noise and an ultrafast response. The tubes also have several disadvantages such as limited detection efficiency, especially for longer wavelengths and are unsuitable for compact design of integrated systems. Semiconductor detectors are a valuable alternative to PMTs due to the well-known advantages of solid state detectors such as small size, low-power dissipation, low-power supply voltage, low cost.

Application ideas for our monolithic detector could be positron emission tomography (PET) or the distance measurement of objects (e.g. cars). PET is a type of imaging system used in cancer detection. A positron emitting substance is injected into a patient. When a molecule is emitted, a positron will react with an electron creating two anti-parallel gamma-rays (annihilation), each with 511 keV of energy. The gamma-rays (gamma photons) can be detected by converting with a scintillator to optical photons. State of the art systems have a scintillator coupled with a photo-multiplier. The 511 keV gamma-rays are converted to electrical signals. If the detectors are placed in a specific geometry such as a ring, gamma-rays simultaneously incident on two of the detectors (see figure 4.1). It is possible to localize their source along a straight line of coincidence or the so-called the line of response (LOR). The image reconstruction is a coincidence event representing the detection of annihilation photons. If the detectors are able to determine arrival time with an accuracy of < 500 ps a technique called time of flight (TOF) is used. The time information can help to determine the position of annihilation of the photons [38]. The TOF-PET gives a significant improvement in image quality (improvement of signal to noise ratio and time resolution).

Scintillators are placed on top of the sensitive SPAD area of the monolithic detector. The incoming gamma-rays are converted into photons and are detected by the single photon array (SPADs operating in Geiger-mode). A principle application idea for a PET system is shown in figure 4.2. A simplified PET system consists of the single photon detector ASIC and scintillator crystals. The detected photon signal triggers the time stamp circuitry. Two ASICs are able to measure the coincidence time of the photons. In the self-trigger or auto-trigger mode the photons generate a hit information and the hit is compared to a threshold. The time stamp runs continuously. In this mode measurements must be taken to determine which pixels have similar time stamps.

Another application of the ASIC would be to use it for distance measurement of objects (e.g. distance measurement of cars see figure 4.3). A laser is pointed at an object and a time stamp ramp is generated for the ASIC. The reflection pulse (reflection caused by the object) can be detected. The time stamp can be measured by the delay of the ramp (when the ramp starts) and the delay of the light reflection pulse. The time stamp is proportional to the difference of the delays (delay of the ramp and the delay of the light reflection pulse).

4.1 The Semiconductor Photon Sensor: SPAD

A semiconductor detector, such as a SPAD, or avalanche photodiode (APD), is a semiconductor photon sensor. The fundamental difference between SPADs and APDs is that SPADs are designed to operate in reverse bias voltage well above the breakdown voltage. The SPAD detector is operated in Geiger-mode. When a photon hits a SPAD, the charge carriers generated trigger an avalanche multiplication. It has an internal amplifying coefficient. The avalanche can produce 10^5-10^6 carriers. The SPAD is sufficiently biased above the diode breakdown voltage. Since the avalanche in Geiger-mode is self-sustained, a quenching circuit must be used to prevent overheating and the destruction of the diode. The purpose of the quenching circuit is to stop the avalanche current by lowering the bias voltage. There are three conditions in which the SPAD operates: reset (where the diode is "armed" with a bias voltage higher than breakdown), avalanche and quenching. A silicon photomultiplier

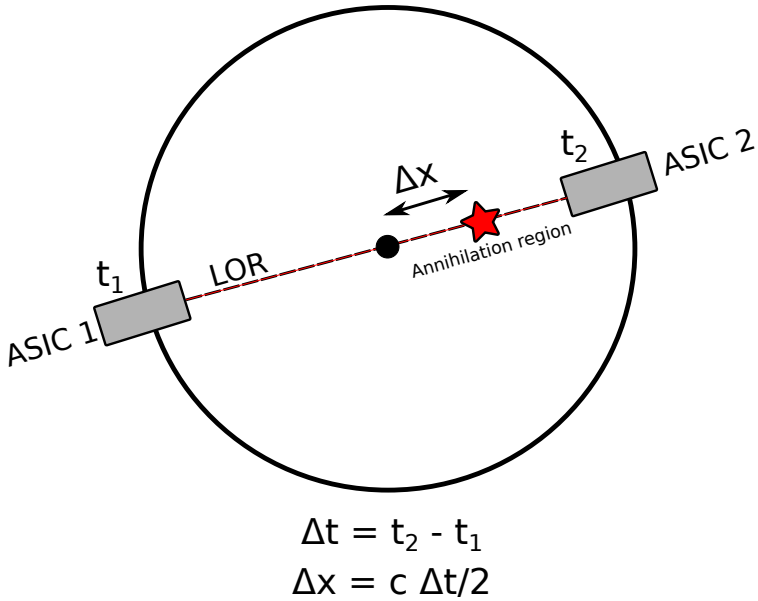

Figure 4.1: Application idea for the monolithic single photon detector: TOF-PET. Measuring of t_1 and t_2 with the time stamp circuitry is possible. TOF-PET allows the localization of the annihilation point on the LOR. In TOF-PET, the distance to the origin of the detector (Δx), where c is the speed of light, t_1 is the arrival time on the first detector and t_2 is the arrival time on the second detector.

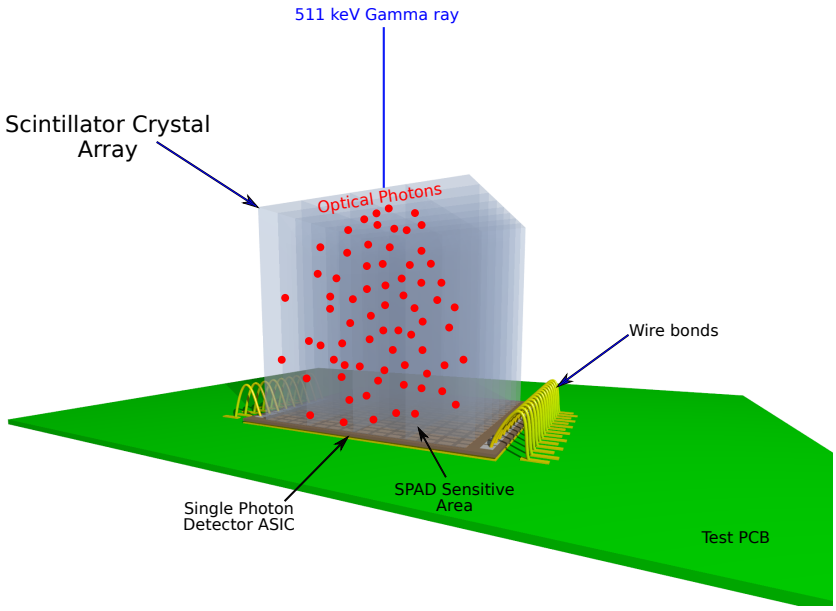

Figure 4.2: Principle application idea: the PET system consists of single photon detector ASIC and scintillator crystal array which convert the gamma ray into optical photons.

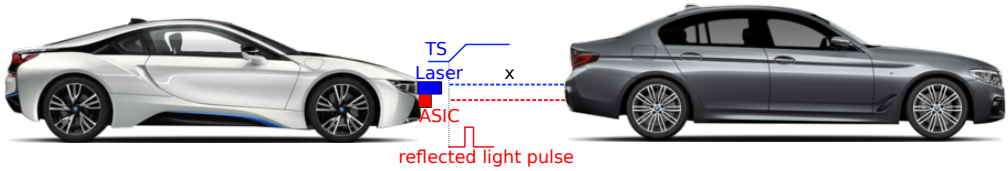

Figure 4.3: Application idea for the monolithic single photon detector: distance measurement of objects. Here the distance (x) between two cars is measured. The laser points (blue dashed line) to another car and starting a ramp (TS) for the time stamp circuitry. The light reflection pulse (red dashed line) is received by the ASIC. The time stamp is calculated by the delays of the reflected light pulse ramp.

(SiPM) in its simplest form consists of SPADs defining a microcell. In other words, the SiPM is equivalent to an array of SPAD microcells with a quenching circuit [18].

4.1.1 Principle of Operation

In a SPAD the avalanche effect plays the most important role. The avalanche creates a cascade of moving carrier pairs for each photon detected. The SPAD is a strongly reverse-biased photodiode. When a p-n junction (see 2.1.4) is reverse biased above its breakdown voltage and a quenching circuit is connected, this constitutes a SPAD. A p-n junction will operate in one of these modes: forward, reverse and breakdown (figure 4.4). If we apply

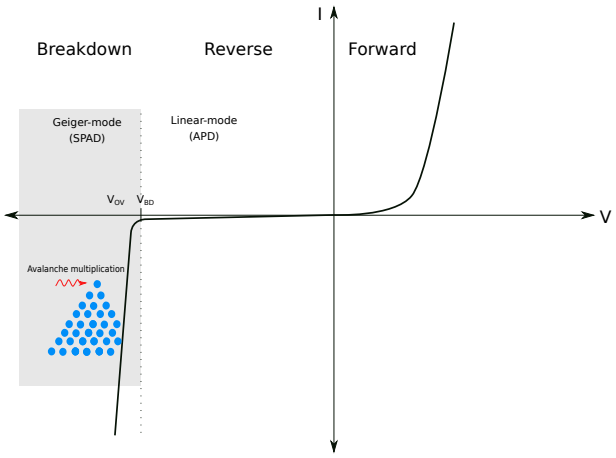

Figure 4.4: I-V characteristic of diode (p-n junction). The voltage applied at the diode defines the operating mode (forward, reverse and breakdown).

positive voltage (forward mode) the current can be described by the ideal diode Shockley equation (see equation 2.35). A p-n junction in reverse bias is partially or fully depleted. The width of the depletion can be calculated according to equation 2.34. The depletion zone represents the volume for particle detection. In this volume a photon is absorbed creating an electron-hole pair (see figure 4.5 a). The average energy needed for producing 1 eh-pair in silicon is $E_{eh} = 3.61$ eV (e.g. X-ray photons). For silicon the energy band gap is $\Delta E = 1.12$ eV. For photoconductivity the energy of a photon must be greater than the energy band gap of silicon $h \cdot f > \Delta E$. The rest of the energy is produced in phonons (heat).

Figure 4.5 shows a typical e-h pair in the depletion region of a SPAD. A photon is absorbed (a) creating an e-h pair (an electron in the E_c and a hole in the E_v). The electron accelerates under the effect of the strong electrical field. A saturation velocity is reached by random collisions due to the interrupted acceleration process [15]. Due to the high electrical field the electron can impact, ionize and generate a secondary electron-hole pair (b). The holes generated at (a) and (b) are also accelerated, while moving to the left. These holes may generate hole-initiated electron-hole pairs (c) [15]. Figure 4.6 shows the energy

band diagram of linear and Geiger-mode. In linear mode the avalanche triggered by electrons. In Geiger-mode both carriers contribute to the avalanche process. The avalanche process is also called self-sustained.

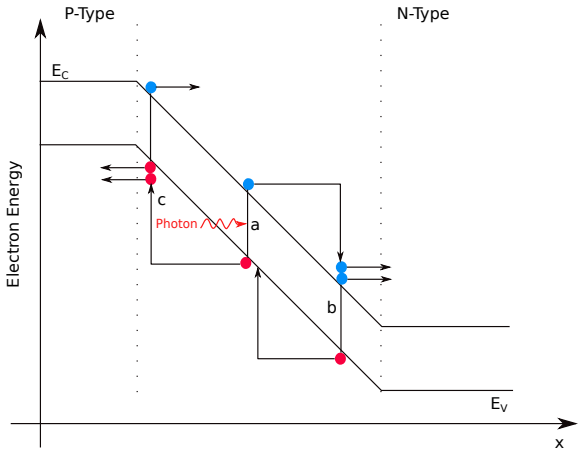

Figure 4.5: Energy-band diagram of a strongly reversed biased pn junction. Multiplication process of an APD. Blue dots represent electrons and red dots are holes.

As previously mentioned the SPAD is operated in Geiger-mode. The breakdown region can be divided into the linear region and Geiger-mode region. Diodes operated in the linear regions are APDs or linear mode APDs. In the linear region the diode is biased slightly below the breakdown voltage. The electrical field is strong enough to cause significant ionization through free electrons [37]. Due to the higher ionization coefficient of electrons in comparison with holes the impact ionization is dominated by electrons in linear mode. The linear mode has a moderate gain that makes detection of single photons unsuitable. APDs in linear mode are used to detect clusters of photons. To detect single photons, we have to bias the APD in Geiger-mode transforming the APD into a SPAD. In Geiger-mode, another process is added, the process of secondary avalanches triggered by holes and secondary photons. At this point a self sustaining avalanche multiplication process is started. Both carriers are now involved: electrons and holes creating a positive feedback effect that makes the multiplication process [33] (figure 4.7). The current then grows exponentially. As previously mentioned, reset, avalanche and quenching are the key operations in Geiger-mode (figure 4.8). If the sensor is biased above V_{BD} and no current flows, the sensor is in the OFF state. The sensor is set to ON state when a macroscopic current flows. Even when a single-photon is absorbed and an electron-hole pair is created, the avalanche multiplication starts and remains self-sustained. The sensor remains in the ON state until the avalanche is quenched (or the current is limited) by a quenching circuit. This circuit may have one component only, e.g. a resistance (figure 4.9). It can also be quenched with an active quenching circuit. This circuit drives the applied voltage down to the breakdown voltage V_{BD} or lower than breakdown. The reset phase is initiated by the

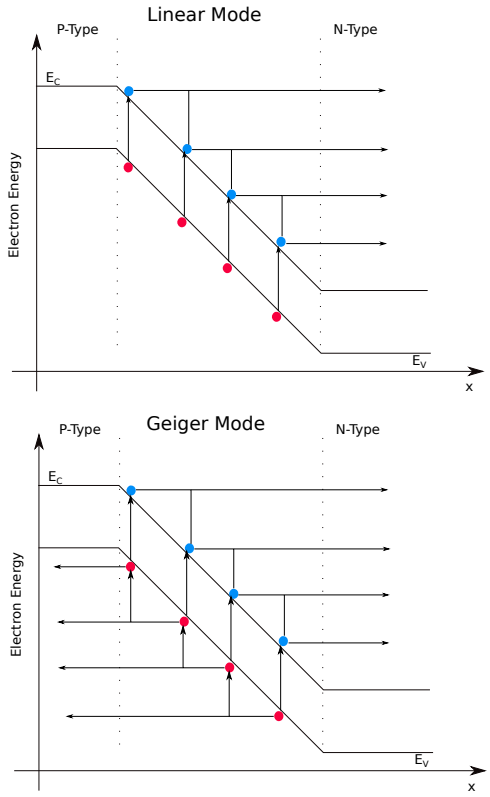

Figure 4.6: Energy-band diagram of linear (top) and Geiger-mode (bottom). In the linear mode the avalanche is triggered by electrons (blue dots). In Geiger-mode both carriers (holes are red dots) contribute to the avalanche process (self-sustained).

quenching circuitry by resetting the voltage to its original point above breakdown. We can define a voltage above breakdown as

$$V_{HV} = V_{BD} + V_{OV} \tag{4.1}$$

V_{BD} being the breakdown voltage and V_{OV} the overvoltage to operate in Geiger-mode. Between avalanche to reset state the sensor is insensitive to an incoming photon. The sensor has the so-called dead time. The charge generated in an avalanche is proportional to its gain (A) given by

$$A \sim Q = C_{det} \cdot V_{OV} \tag{4.2}$$

where C_{det} is the intrinsic detector capacitance. In this operation mode the sensor works as a digital flip-flop. In contrast to linear mode APD in which the gain is proportional to the impinging radiation flux, this proportionality is lost for a SPAD.

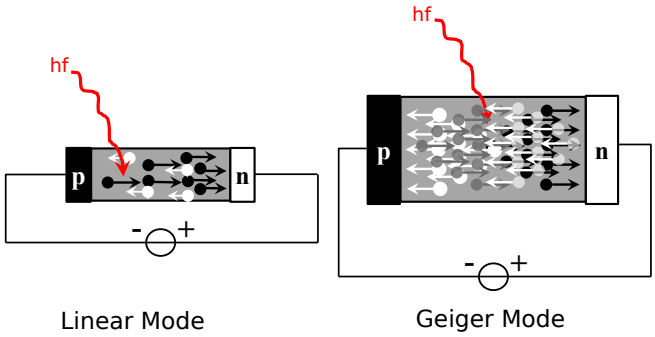

Figure 4.7: A p-n junction for Linear and Geiger-mode is shown. In the linear mode avalanche is triggered by electrons due to the high ionization coefficient (black dots are electrons and white dots are holes). In Geiger-mode both carriers (electron and holes) contribute to the avalanche process starting a secondary avalanche multiplication.

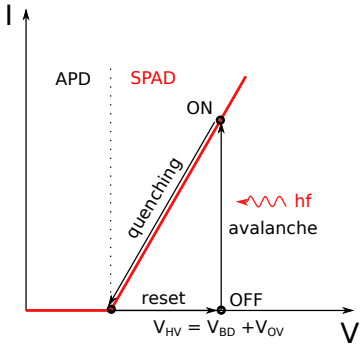

Figure 4.8: Reverse I-V characteristic for SPAD operation.

4.1.2 Quenching Circuits for SPAD Operation

There are two types of quenching methods: passive and active quenching. A passive quenching circuit (PQC) consisting of a series resistance and a SPAD sensor forms the simplest quenching circuit (figure 4.9). The series resistance is in the order of $R_Q = 100$ kΩ to 1 MΩ. By using a high value for the series resistance (R_Q), the passive quenching leads to small quenching time but with a high resetting time. The avalanche current develops a voltage drop across the series resistance R_Q. The voltage across the sensor decreases toward V_{BD} and the avalanche current also decreases. If the sensor voltage reaches V_{BD} the decrease rate slows down. The avalanche current flows through the series resistance and can be written as

$$I_d = \frac{V_{HV} - V_{BD}}{R_Q} = \frac{V_{OV}}{R_Q} \tag{4.3}$$

Due to the fact that the series resistance has a high value it reduces the current to less than a few tens of 1 µA [10] to quench the avalanche. After this quenching process the voltage V_{det} starts to recover or recharge slowly toward V_{HV} (reset) with the time constant $\tau = R_Q \cdot C_{det}$. One disadvantage of PQC is that the timing of the sensor is set by the internal time constant and cannot be adjusted. This time constant is long because R_Q is not small and $C_{det} = 2$ pF. Calculating the reset time for a typical sensor with PQC configuration R_Q = 0.5 MΩ, $C_{det} < 2$ pF results in

$$\tau_{reset} = R_Q \cdot C_{det} = 0.5\,M\Omega \cdot 2\,pF = 1\,\mu s \tag{4.4}$$

In an ASIC, the series resistance can be realized with an NMOS or PMOS transistor whose

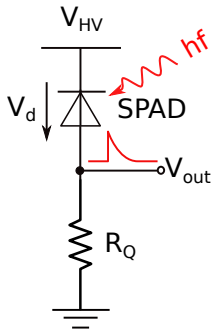

Figure 4.9: Passive quenching technique realized with a series resistance and SPAD sensor.

gate voltage can be set externally. It is possible to implement a quench resistance as poly resistance inside the ASIC. A quench resistor can also be used as a poly resistor in the layout design. Poly resistors take up a relatively large amount of space in the ASIC.

To completely avoid the disadvantages of PQC an active quenching circuit (AQC) must be used. An AQC has for example a comparator, monostable circuit and a switch as a feedback. The principle of AQC is to avoid long quenches and reset transitions. A feedback switch generates a short quench and reset pulse. Timing is then dominated by the external AQC and not by internal RC time constant. As depicted in figure 4.10 a comparator detects the SPAD signal and converts it into a full CMOS level (e.g. 3.3 V). The comparator output signal is fed to a Monoflop circuit which produces a hold-off pulse which is fed to a driver or feedback switch producing a short quench pulse in this loop. The hold-off time can be set with the monoflop circuit. The width of the hold-off pulse can be adjusted by a bias current. By adjusting the bias current using an NMOS acting as a current source, it can produce faster rising and falling edge for the hold-off pulse. The output of the monoflop can be latched or disabled to exclude noisy pixels. The quenching pulse should be as fast as possible to reduce overheating of the sensor, light emission from hot carriers [10]. A disadvantage relates to the feedback loop in an AQC. This feedback produces a delay in cases when the sensor is not located near the quenching circuit. This feedback should be as short as possible to avoid a delay. To obtain high count rates, the hold-off time must be minimized and an adjustable hold-off time reduces afterpulsing [10]. Another point is the reset transition. During this time the sensor is insensitive to incoming photons. If the reset

pulse is very short, the sensor is then quickly reset to its initial efficiency and is ready to detect a next photon. If a fast reset pulse is produced it can cause overshoots, slow recovery to base line voltage level. A fast and precise reset ensures accurate photon counting and timing. Dead times of 50 ns depending on the process technology and active quenching circuitry [91] can be reached. Further, achieving accurate photon-timing involves other requirements for fast quenching. In particular with SPAD devices with an active area greater than 10 µm a clear dependence of the time resolution on the quenching circuit for timing the pulses can be observed [10]. This can be investigated by the physical processes that describe the rise of the avalanche current [62, 10, 52, 53, 54].

A mixed passive-active quenching is an approach where a passive load is used providing a drastic reduction of the avalanche current. With this mixed topology, the advantages of PQC and AQC can be exploited. A feedback loop forces a fast reset transition making it possible to introduce an adjustable hold-off time. In the mixed passive-active quenching circuit the resistance can be smaller than the value required for a classic passive quenching circuitry. There are other possibilities when mixing these topologies. An active-quenching circuit with passive reset consists of bipolar n-p-n transistors, comparator, fast switching diodes and passive load resistance. Passive quenching with active reset - this type of topology has fast voltage switch implemented as MOSFET switch and a load resistance. A mixed active-passive reset is used to avoid overshoots.

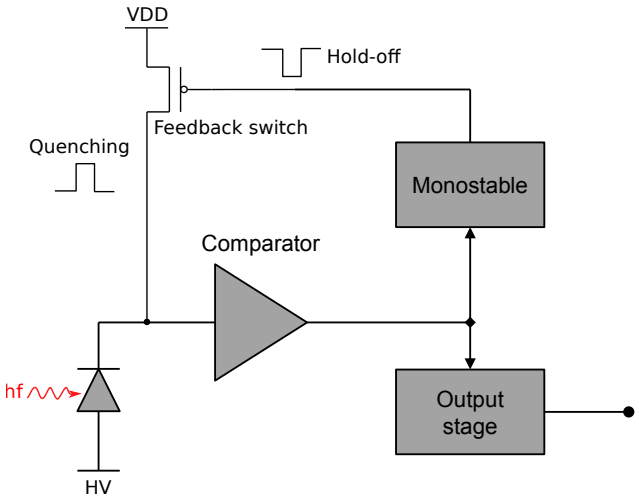

Figure 4.10: A simplified active quenching circuit consists of comparator, monoflop circuit, and feedback switch.

4.1.3 Performance of SPAD Devices

In this section performance parameters of SPAD sensors are briefly explained. The parameters are: photon detection efficiency (PDE), dark count rate (DCR), breakdown voltage,

afterpulsing, crosstalk and fill-factor. These parameters are important when characterizing the SPAD sensor. An ideal single-photon sensor should have

- Detection Efficiency = 100%
- Broad spectral range
- Timing jitter = 0
- Dead Time = 0
- Afterpulsing = 0
- Photon number resolving
- Room temperature operation
- Large active area
- Integration in a ASIC as an array of SPADs

Unfortunately, SPAD sensors developed in deep-submicron CMOS, HVCMOS process technology have far from ideal values. For example, avalanche multiplication can be started by a primary carrier but not started by an incident photon. There are false detection events or uncorrelated avalanche multiplication events or the so-called dark counts. These limit the sensor in its detection efficiency.

Photon Detection Efficiency When a photon generates an electron-hole pair it is important that the primary electron-hole pair is able to trigger an avalanche. The probability of avalanche triggering increases with the overvoltage. In other words, the PDE is the probability that an absorbed photon ignites an avalanche. This ignition starts a real event. The PDE is dependent on the intrinsic detection efficiency, absorption (is a function of the wavelength), geometry or the fill-factor, avalanche probability and overvoltage. PDE is depicted in figure 4.11.

Dark Count Rate Dark counts are generated internally in the sensor. The dark count increases with temperature. This effect is due to thermally generated carriers inside the sensor. The dark count has Poisson distribution as in dark current of photodiodes. Another dark count source could be background radiation. When the overvoltage increases, the DCR also increases (figure 4.11). In silicon the thermal generation of charge carriers takes place through local energy levels that are located deep in the band gap. Traps which are impurities are located between conduction and valence band [83]. These levels are closer to the midgap. Poole-Frenkel and trap-assisted tunneling occur at higher fields ($> 10^5$ V/cm) and can intensify deep energy levels (field-enhanced generation)[61, 87]. At higher fields a direct band-to-band tunneling occurs. The tunnel-assisted generation is independent of temperature. That means, it is not reduced by lowering the temperature. The electric field within the SPAD sensor must be simulated. It is important to avoid band-to-band tunneling and field-enhancement generation of charge carriers [10].

Breakdown Voltage (V_{BD}) Another important parameter is the measurement of the breakdown voltage (figure 4.12). This breakdown point can be determined by increasing

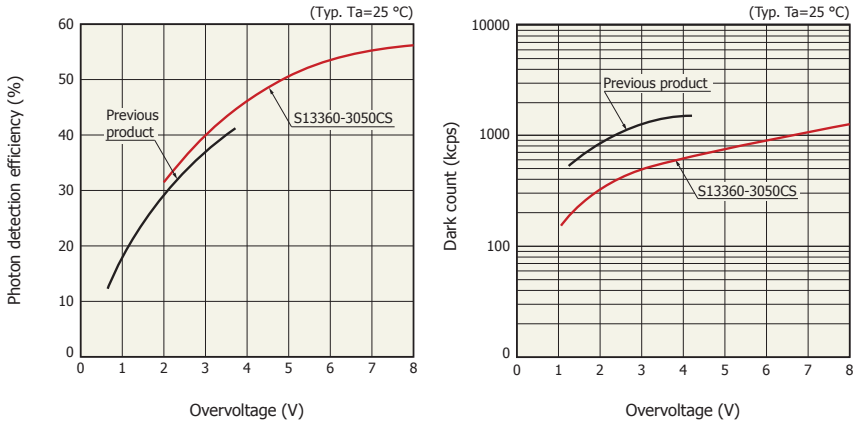

Figure 4.11: PDE and DCR as a function of overvoltage. Both curves show a rise for higher overvoltage values. Graphs produced by Hamamatsu company ref. [44].

the voltage on the SPAD sensor and measuring the current through the PQC. This method is called the I-V method. Once the breakdown is reached, the current rises exponentially. With this kind of circuit the voltage can be determined precisely. For measuring such small currents (a few hundred picoamperes) a source measurement unit (SMU) is needed. Furthermore, it also important to darken the SPAD sensor completely before reaching the breakdown.

Crosstalk For SPAD sensors used as arrays or pixel arrays there is a probability that the absorbed photon in the incident pixel, can also be detected in the neighboring pixels. This effect is called crosstalk. Two types of crosstalk are known: the optical crosstalk and electrical crosstalk. Optical crosstalk occurs when secondary photons (radiative emission from the hot carriers) [55] are emitted in the p-n junction while the sensor is in avalanche. The emitted secondary photons can trigger an avalanche in the neighboring SPAD sensors causing other avalanches. The crosstalk is increased when the distance between the neighboring SPAD sensors is reduced. To reduce the optical crosstalk and avoid the avalanche, trenches are placed (a kind of optical shield) in between two adjacent SPAD sensors. These trenches are coated with metal [77]. Electrical crosstalk occurs during the avalanche process when the generated carriers can overcome the junction and propagate into and trigger the neighboring SPAD sensors. The penetration depth increases with wavelength e.g. photons in red ad NIR ranges induce more electrical crosstalk than short wavelength photons.

Fill Factor In an array of SPAD sensor a geometric fill factor can be defined. The fill factor is defined as the ratio of detection area to pitch. The pitch is defined as the addition of gap length and detection area. A large filling factor has a high pixel density and therefore a high PDE. For monolithic SPAD pixel ASIC is the goal to increase the fill factor up to 90% to reach high detection efficiency.

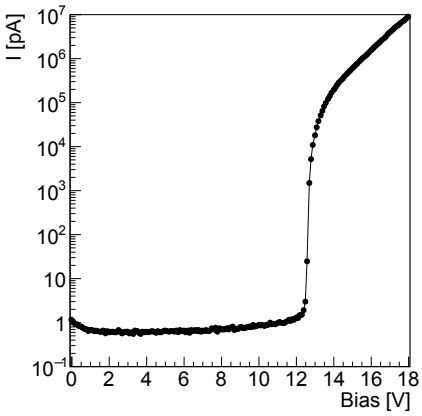

Figure 4.12: I-V breakdown in dark condition. Before breakdown a dark current of a few picoamperes is observed. When the breakdown is reached, the current rises abruptly to a few microamperes. Figure reproduced from ref. [28].

Afterpulsing Trapped electrons are located in deep-level trap (located in the midgap). During an avalanche process such electrons may be trapped and then released. These electrons can retrigger an avalanche producing the so-called afterpulses. Afterpulses are correlated in time or noise. The probability of triggering an afterpulse depends on the number of deep-levels and the amount of charge carriers produced in these traps. The deep-level lifetimes are longer than the produced dead time in PQC [37]. The afterpulsing causes a higher dark count in the sensor. A quenching procedure can be found for reducing the afterpulses, for example by tuning the hold-off time to be extended. Lowering the temperature decreases the dark count rate but makes the afterpulsing worse. The trap release process becomes slower at low temperatures [27] requiring a longer hold-off time but this limits the dynamic range for counting photons.

4.2 H35-SPADV1 ASIC Architecture in Standard CMOS Technology

We present a fully integrated monolithic avalanche diode array (SPAD array) in a commercial AMS 350 nm HV-CMOS process. The ASIC has a monolithic architecture that comprises a SPAD, AQC and readout electronics. The size of one SPAD pixel (subpixel) is 38 x 92 μm^2. One pixel defined as sub-pixel and readout electronics has a total size of 115 x 111 μm^2 with a fill factor of 24.3% and a chip size of about 2.8 x 2.6 mm^2. The ASIC has 15 x 20 SPAD pixels. Table 4.1 summarizes the main H35-SPADV1 features. We developed the layout of the SPAD sensor with Cadence Virtuoso software. The signal generated by the avalanche effect is DC coupled to a fast CMOS comparator. A configurable bias block was included in the chip architecture for bias setting. The designed bias block in the NeutrinomassV3 ASIC is also used for H35-SPADV1 ASIC, but was slightly modified because

Technology	AMS 350 nm HV-CMOS
Chip size	2.8 mm x 2.6 mm
SPAD pixel size (sub-pixel)	3496 μm^2
Pixel size (sub-pixel+front-end)	115 µm x 111 µm
Quenching Circuit	Active
SPAD Design	p+/shallow n-well/deep n-well
Breakdown Voltage	11.7 V
Fill Factor	24.3%
Dark Count Rate (DCR)	174 Hz/μm^2
Pulse Detection Efficiency	30 % (λ= 420 nm)

Table 4.1: H35-SPADV1 ASIC and SPAD main characteristics.

of the different bias currents. Bias blocks for example are multipurpose and designed to speed up and simplify the ASIC development [18].

Each pixel has the following features: fast comparator, active quenching, edge detection, monoflop circuit, 1-bit hit flag, analog time stamp memory, Hit-OR bus and data output. Single pixel and full matrix readout are possible. A row and column control is implemented. The row column control consists of shift registers to select a row and a column (single pixel select). After the comparator stage a monoflop circuit defines precisely an inhibition pulse or hold-off pulse to ignite the inhibition transistor and the same signal is fed to a delay circuit to set a reset pulse by an NMOS transistor. The quench was realized in an NMOS transistor acting as a current source with an adjustable bias setting. The SPAD signal is transformed to CMOS level over the inhibition transistor that keeps the signal to 3.3 V during hold-off phase. The generated SPAD signal is large enough to trigger a two-transistor comparator. The threshold of the comparator is set by an auxiliary voltage supply [18]. Figure 4.13 shows the full layout of H35-SPADV1. To improve the fill factor the SPAD sub-pixel was designed as large as possible and also has a large photon sensitive area. The front-end electronics has a minimum distance of 15 µm between the deep n-wells.

4.2.1 SPAD Sensor Architecture

We chose industrial HV-CMOS (AMS 350 nm) technology because it guarantees low cost production and is used for a variety of applications such as automotive, power electronics etc. Long-term availability is also guaranteed [13]. The CMOS process allows high-voltages up to 120 V. High-voltage NMOS and PMOS operates up to 120 V and can be combined on the same substrate. On our ASIC a SPAD sensor operates at high-voltage regime (breakdown voltage) in a deep n-well and the electronics (front-end) is supplied with 3.3 V in a separate deep-nwell but on the same p-substrate [18].

The layout of the HV-CMOS SPAD was designed with Cadence Virtuoso. This structure (figure 4.14) is implemented in a deep n-well (DN), builds the cathode and insulates the photosensitive area of the SPAD. The deep n-well is biased at positive $V_{HV} = V_{BD} + V_{OV}$, V_{BD} being the breakdown voltage and V_{OV} the overvoltage to operate in Geiger-mode. A p+ implant, acting as the anode, and a shallow n-well (SN) define the high-field region in

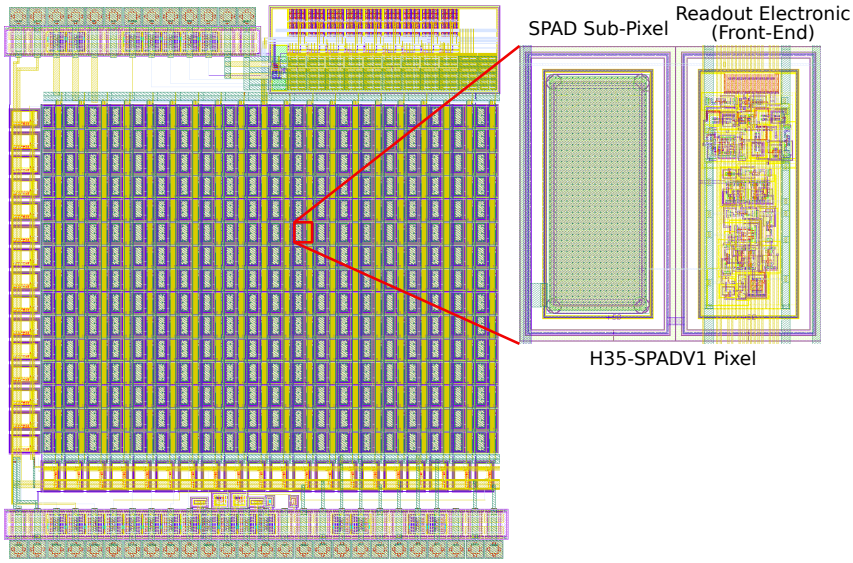

Figure 4.13: Layout of the monolithic H35-SPADV1 ASIC (left) and magnified view of SPAD sub-pixel and readout electronics (right).

the active area. The junction is surrounded by shallow p-well (SP) implantation acting as a guard ring to prevent edge breakdown of the SPAD sensor. The p-substrate is shared with readout electronics. The SPAD signal is DC-coupled to the front-end. Since the SPAD n-well is at high voltage and the n-well for the readout electronics at 3.3 V, a minimum spacing of 15 µm must be allowed between these wells. The corners of the SPAD are round to avoid high electric field peaks. We designed a large SPAD with an active area of 38 x 92 μm^2 to increase the fill factor. A fill factor of 24.3% was achieved. The design of a SPAD pixel is shown in figure 4.14. In this first version no technology computer aided design (TCAD) simulations were performed for the SPAD sensor. TCAD simulations show, for example, the distribution of the electrical field for breakdown condition and help to design a suitable shape to avoid avalanches at the edges of the sensor. Based on our experience with the AMS 350 nm HV-CMOS technology, we have designed the layout of the SPAD sensor. The breakdown voltage could only be determined from the measurements. The breakdown voltage of the SPAD diode depends on the p-tub guard ring and of the design of the SPAD shape. Transient simulations were carried out with an equivalent circuit of the SPAD sensor. To exhibit a moderate DCR a SPAD diameter between 8-20 µm is used [85]. A small pixel size is preferable, because the DCR decreases with the pixel size. The size of the active area influences DCR and breakdown voltage. Larger area SPADs show a higher DCR as can be seen from [69] and [2]. Table 4.2 shows different pixel diameters, DCR and breakdown voltage for 350 nm technology node. Our SPAD sensor shows high DCR due to large active area of 3496 μm^2 [18] and possible crosstalk between neighboring pixels. Measurement results will be shown later in this chapter.

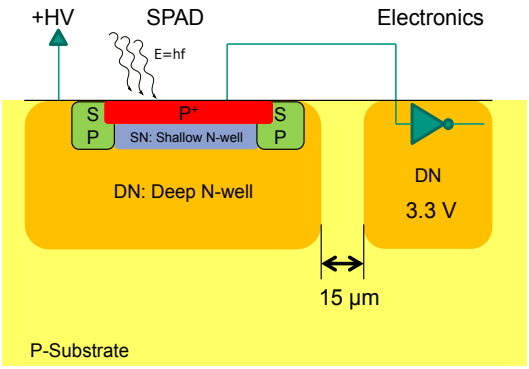

Figure 4.14: SPAD sensor in HV-CMOS technology. Front-end electronics and SPAD sensor are separated between two n-wells with a minimum distance of 15 µm.

Pixel Diameter (µm)	Breakdown Voltage (V_{BD}[V])	DCR (Counts/s)
200	11.7	overflow
100	11.7	overflow
50	11.7	650000
20	11.8	21000
10	12.1	2020
7	12.9	1900
6	13.9	1900
5	16.2	1900

Table 4.2: DCR and Breakdown voltage dependent on the pixel diameter. At 22.5 °C the dark count rates at 800 mV overvoltage were measured. This table is reproduced from ref. [69].

4.2.2 Front-End Electronics

Our design implements an AQC. A simplified schematic diagram of the electronics is shown in figure 4.16. In order to achieve a better pixel integration and adjustable quench timing an AQC design is used to fit these requirements. AQC avoids a fixed quench timing by an internal time constant (passive quenching). The H35-SPADV1 ASIC has the same configuration scheme as the NeutrinomassV3 ASIC. A two phase clocking (*Ck1*, *Ck2*), serial (*SIn*) and load (*Ld*) signal configure the internal DACs and can therefore select row and columns.

The external electronics consist of an active inhibition and reset switches (a PMOS and NMOS), fast comparator, edge detection, monoflop circuit, 1-bit hit flag, analog time stamp memory, Hit-OR bus and data output [18]. If a photon hits the SPAD sensor (SPAD is working in Geiger-mode) the SPAD signal will rise above the comparator threshold voltage (V_{thr}). The comparator threshold can be generated in bias DAC block implemented on the periphery of the ASIC. This threshold is also used to suppress the DCR during photon

signal generation. As soon as the signal exceeds the threshold, the signal is pulled to 3.3 V CMOS level (inhibition switch a PMOS). At this point the hold-off time starts. The hold-off (*delLoHold*) can be adjusted with a delay circuit that is connected to the reset input of the RS-flipflop. Delays in the circuit are implemented as MOSFET capacitances and current sources. MOSFET capacitances are either connected to a component input or a slow input of the edge detectors. A slow input is foreseen to delay the edge at the output of the edge detector. The edge detector has a gated inverter at the input and a NAND gate to generate the output pulse. One input of the NAND gates is slowed by the MOSFET capacitance. The output of the RS-flipflop controls the inhibition switch and the input of the edge detector. This edge detector has a defined edge time (*delHiEdge*, *delLoEdge*) which triggers the reset switch. After the comparator stage, the same *delHiRo*, *delLoRo* time, is fed to the edge detector. This edge detector is used as monoflop to produce the hit signal. Return to base line speed of the SPAD diode is defined as relaxation speed which is dependent on the edge time setting (*delLoEdge*/*delHiEdge*) and the *VNBias*. High *VNBias* setting and low *delLoEdge*/*delHiEdge* lead to short relaxation time. An NMOS transistor is used to quench SPAD sensor. This transistor has a gate voltage *VNBias*. Thus, setting an operating point to saturation effectively limits the avalanche current. The saturation region is defined as

$$V_{dssat} \geq VNBias - V_{th} \tag{4.5}$$

The current is limited by the size of saturation current. The resistance of the NMOS transistor in triode region is inversely proportional to *VNBias*. The relaxation time is

$$\tau = R_{on} * C_{SPAD} \tag{4.6}$$

Figure 4.15 shows the intersection of NMOS transfer characteristic and SPAD I-V characteristic. As shown in the previous chapter the NMOS has a triode region (where R_{ON} is defined) and the saturation region. The SPAD sensor has an on-off branch. The hit monoflop signal is connected to the D input of a D latch. The load signal for this latch is controlled externally. The D latch is used as the hit flag to store the hit. The output of the hit flag is used to store the analog time stamp. A row can be selected so that the hit flags of the selected row can be read out. The *Hit-OR* and *Hit* signals are controlled by the monoflop output. The amplitude of *Hit* is proportional to the number of hit pixels and can be used as trigger. Upon receiving monoflop out, the external circuit can generate a *Store* signal that puts the hit latch into store mode. The *Hit-OR* is provided for self-trigger mode. Single pixel, full matrix and self triggered readout are possible. A voltage ramp must be generated for the time stamp ramp. The ramp can be either generated externally, starting from a moment when we expect a signal or *Hit-OR* can be used for auto generation. The voltage level at the time stamp line (the ramp) will be stored in a small capacitor within a pixel when the hit flag output goes high. The stored voltages can be read out in the same way as the hit flags, by selecting the rows from one shift register and multiplexing different columns to a common output using another shift register. The switches are implemented as transmissions gates. Figure 4.16 shows the simplified front-end circuit of a pixel. An adjustable edge time is fed to the reset transistor to precisely adjust the reset time. After the quenching process the reset should be initiated using a fixed reset time to set the SPAD diode back to avalanche mode again. Thus, the dead time is programmable.

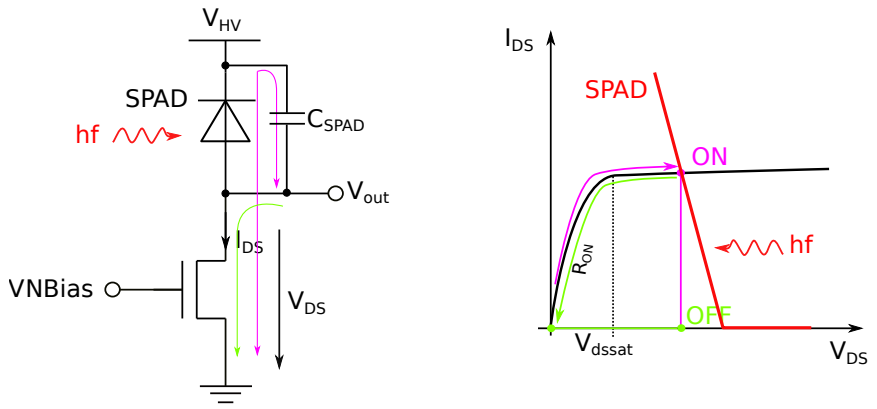

Figure 4.15: Intersection of NMOS transfer characteristic and SPAD I-V characteristic. In darkness the V_{out} will be ≈ 0 and the transistor will always be in triode region no matter how large *VNBias* is. During avalanche the current (ON branch) rises and is limited by the size of saturation current. Quenching will occur and the SPAD will return to OFF branch.

To detect the avalanche, a threshold is set on the comparator. It has been shown that when the DCR is low the circuit is not working properly and has a low detection efficiency. As previously shown, the DCR is also dependent on the environment such as light and temperature. One weak point is the *VNBias* voltage. This voltage should be generated on the chip but there is some distortion on the power supply line during bias generation (bias block), which will influence the *VNBias* and the operating point. The NMOS quench was designed too weak and it has been found that it is sensitive to such distortions. To keep the transistor more stable, the *VNBias* voltage is generated by a linear regulated power supply. The *ReadRow_i* lines are depicted in the schematic 4.16. There is one row control block consisting of 15 shift register blocks (or 15 row lines). To select *ReadRow_i* the FPGA generates (set a row by the application) in the same way as for the bias generation, **SInRow**, **Ck1Row**, **Ck2Row** (two phase clocking), the **Ld** (load signal). When selecting a column **SinCol**, **Ck1Col**, **Ck2Col** (two phase clocking) and **Ld** signal is also generated by the FPGA. The same control block is used for selecting the columns (20 columns). Figure 4.18 shows a simplified block schematic of the architecture of the row and column control block. In this way each pixel can be chosen in matrix of 15 (rows) x 20 (columns) pixels by an application. This functionality makes a picture of the entire SPAD matrix and reads out the entire matrix or only one pixel. Furthermore, the latch functionality can reduce the dark counts in such a way, that a defined laser pulse is generated shortly before the rising edge of *Store* signal. With a short laser pulse as well as a small monoflop window setting the probability of dark counts within this time interval is very small. A detailed transient SPAD signal during avalanche and quench operation is shown in figure 4.17. If a photon ignites an avalanche, the signal rises to the same amount as V_{OV}. The inhibition PMOS transistor pulls the signal to 3.3 V and holds it for the defined hold-off time. After the hold-off time the *delHiEdge*, *delLoEdge* set the discharge time. High values mean that

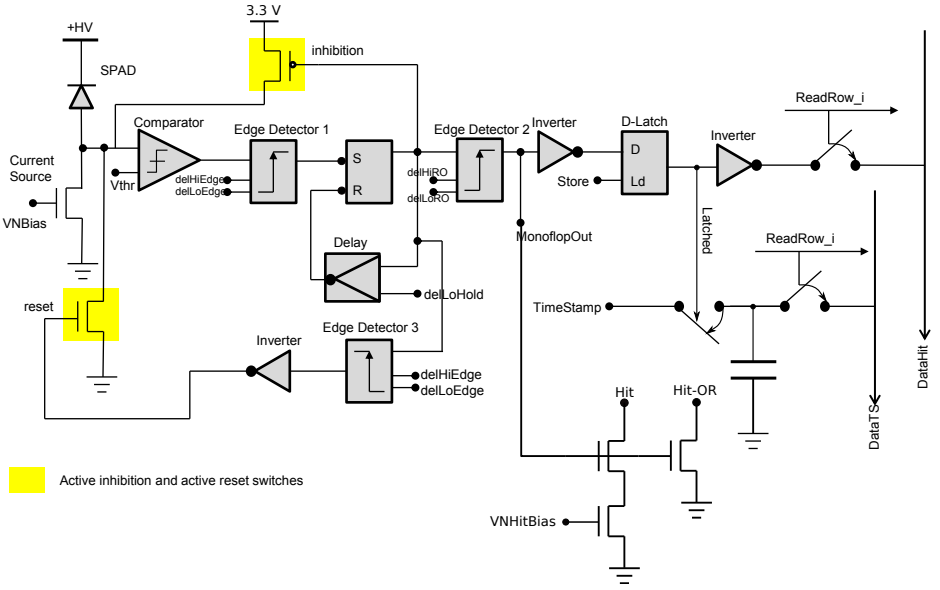

Figure 4.16: Simplified front-end electronics. Active quenching is driven by the inhibition and reset transistors. The monoflop output is fed to both transistors.

the reset transistor is on for shorter time, this leads to an incomplete discharge. The full discharge is then performed by *VNBias*. This all leads to longer discharge times. The comparator switches to high as soon as the signal exceeds the threshold and switches to low when the falling edge of the SPAD signal is below threshold. The speed can be adjusted with *delLoEdge*, *delHiEdge*, *delLoHold* and *VNBias* (see figure 4.16).

A detailed timing diagram from SPAD signal to monoflop output is shown in figure 4.19. When the comparator becomes active (SPAD signal exceeds threshold), the edge detector is set. The width of this edge can be adjusted with the signal *delLoEdge*/*delHiEdge*. A fast pulse is generated by the edge detector. This pulse sets the RS flipflop. The output of the RS flipflop initiates its reset. Reset is delayed by the time that depends on setting *delLoHold*. Therefore the time PMOS is on depends on *delLoHold*. When the SR latch is reset, edge detector 3 (see figure 4.16) produces the reset signal. The width of this signal depends on *delHiEdge*, as already explained. After avalanche, quenching and reset phase, the SPAD sensor is in the initial state and waiting for the next photon. An adjustable monoflop window width is started by the rising edge of the flipflop output. There are *DataHit*, *Hit bus* and *HitOR* outputs. *DataHit* is formed by the output of the D-latch and a transmission gate which is controlled by *ReadRow_i* and used for matrix readout. *DataHit* is a digital hit information. The *Hit* signal is generated by two NMOS transistors (see figure 4.16). The first transistor is a current source (current amplitude can be tuned) and the second NMOS is connected in series to the current source and used as a switch. The monoflop output is fed to the second NMOS transistor. The generated *Hit* signal is active low and its

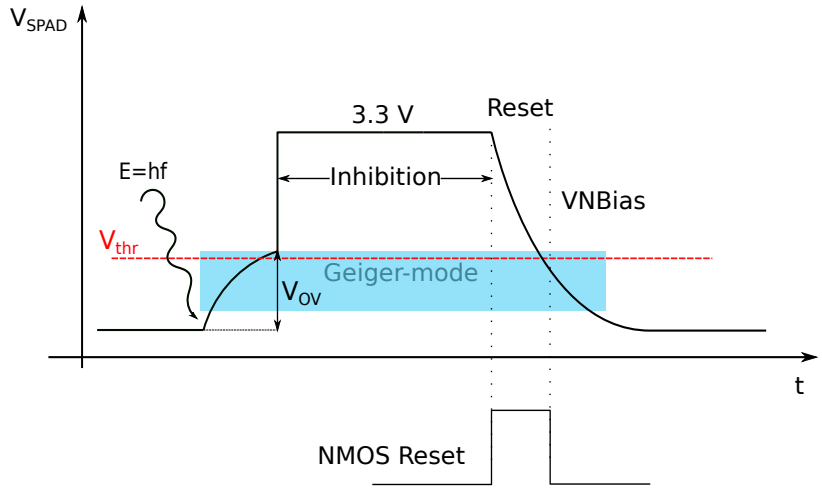

Figure 4.17: Transient SPAD Signal to CMOS level conversion. Quenching and hold-off time.

amplitude is proportional to number of hit pixels. *Hit-OR* is a digital signal and it shows the hit OR function of the entire matrix. *HitOR* consists of one NMOS transistor [18].

For time stamp (*DataTS*) information two transmission gates, a capacitance and PMOS buffer are used (see figure 4.20). The first transmission gate stores the latched information to the capacitance (C_S). A 90 fF capacitance stores this information. A PMOS T1 is used for the bias of the source follower T2. The source follower buffer is used to drive the *DataTS* Bus of the entire matrix. The second transmission gate switch *ReadRow_i* is for matrix readout.

4.2.3 Simulation of the Front-End Electronics

A full transient simulation was performed starting from the SPAD sensor to the digital ASIC outputs. An accurate SPAD transient signal representing the breakdown speed of rising edge was taken into account when estimating the detector capacitance. All bias voltages based on the implemented bias block are used for the simulation. A *Store* signal and a time stamp ramp are generated to the drive the readout for one pixel or the entire matrix readout. A timing diagram of a simulation is shown in figure 4.21. The depicted red line shows the SPAD signal which is clamped to 3.3 V (*SPADSig*). The comparator reacts (*CompP*) when the comparator threshold is passed. The set signal for the SR-flipflop is generated by the edge detector (*SetB*). A inhibition signal (*PullUpB*) is fed to the PMOS transistor. The monoflop circuit has generated a monoflop output (*MonoflopOut*). The hit is now latched (*Store* = 1) by the *Store* signal. For *Store* = 0 the D-latch is transparent. A reset is now executed which switches the NMOS (reset branch) transistor (*PullDn*) and pulls the SPAD signal below the comparator threshold. A reset signal (*ResetB*) is set and

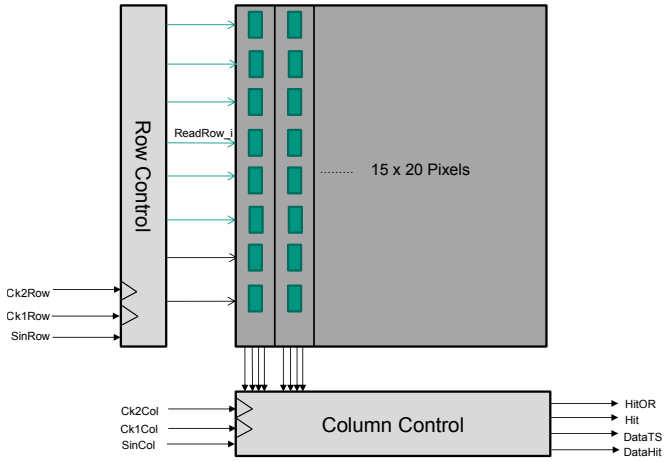

Figure 4.18: Architecture of H35-SPADV1 ASIC. Row and column control blocks can choose a pixel in the SPAD matrix or readout the full matrix.

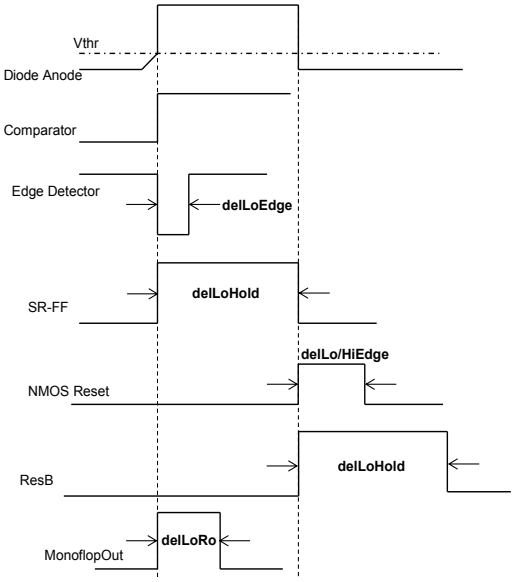

Figure 4.19: Timing diagram of the front-end electronic. The monoflop window can be adjusted with *delHiRo* and *delLoRo*.

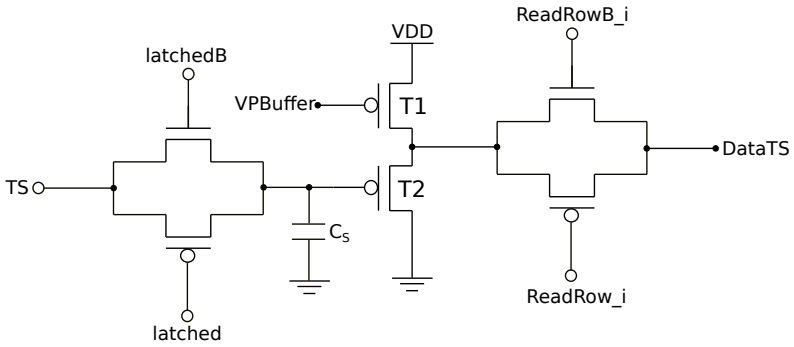

Figure 4.20: Time stamp circuitry. The circuit consists of two transmission gates, capacitance for storing the latched information, and a source follower (transistors T1 and T2) to drive the *DataTS* of the entire matrix.

the SPAD sensor is ready for the next photon. As mentioned in the previous chapter, the hold-off time is defined between the *PullUpB* and *PullDn*. A closer look at the SPAD signals reveals that the discharging returns relatively quickly ($\approx$ 13 ns) below the comparator threshold, but takes some time to return to the base line. This can be influenced by tuning *delLoEdge/delHiEdge* and *VNBias*.

For dark count simulation a noise source is needed that generates poisson distributed pulses in time. This type of simulation was not performed. It is possible to simulate several SPAD pulses that are below the comparator threshold. It has been shown that the front-end electronics are accurate enough to detect both photons and dark counts respectively.

In our simulations we used exactly one photon pulse. To show how to capture a few hundred photons it is essential to construct this scenario. The next timing diagram shows (figure 4.22) when a fast laser diode pulse (e.g. a pulse containing a few hundred photons) occurs. The *HitOR* line shows that, before the diode laser pulse occurs, some dark counts are present. For a dark count the hit (Hit line) amplitude is smaller than for a real photon pulse (LED pulse is generated) because the *Hit* gives the number of pixels firing. The more pixels that are hit, the greater the amplitude. The time stamp ramp is triggered by the dark count. With the falling edge of the diode pulse (figure 4.22) the *DataTS* goes high. The analog time stamp (*DataTS*) is proportional to $\approx Delay2 - Delay1$. Delay1 can be measured between a trigger signal and time stamp ramp start. Delay2 is also measured between trigger signal and the rising edge of the laser diode signal. The *Store* signal is generated later by an FPGA (Store is inverted in the FPGA) where a Delay3 is defined. The *Store* = 0 (D-latch is latched) and therefore the *DataHit* goes to low when *Store* returns to high again (*Store* = 0 (latched), *Store* = 1 (SPAD sensor is sensitive for next photon)). In principle the simulations and timing diagrams show that the front-end electronic fulfills the expected behavior for AQC. Hold-off time, quench time, reset time and monoflop window are fully programmable. The bias setting and DC operating points of all transistors in the front-end are also programmable.

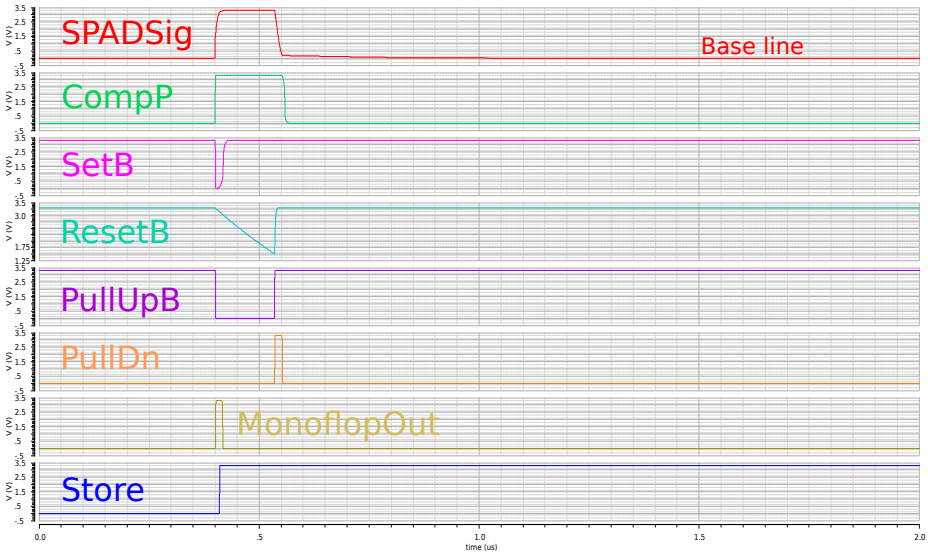

Figure 4.21: Full transient simulation of the internal front-end electronics. SPAD signal is simulated with the estimated breakdown voltage and detector capacitance (C_{SPAD}).

4.2.4 Layout of the H35-SPADV1 ASIC

After all simulations and calculations were completed, the layout with the simplest cell was started. The layout was divided into front-end electronics and SPAD sensor. Starting with the front-end which is divided into sub-cells. Bias Block layout was slightly modified but all the layout design is shared. This layout consists of two sub-cells:

- Readout cell: AQC, monoflop circuit
- Readout digital cell: Hit, Hit-OR bus, 1-bit latch, switches as transmission gates and *DataHit* and *DataTS* output

After the layout of the sub-cells was completed, a cell was made to connect both sub-cells. These two cells are connected by vertical bias and signal lines. The design follows vertical signal flow. After this point the design of the SPAD cell was introduced. A p+ implant with shallow n-well (active area where avalanche multiplying occurs) is surrounded by a guard ring. A rectangular shape with round edges was chosen to avoid avalanche at the edges of the SPAD sensor. As mentioned before, to increase the fill factor a large active SPAD area was designed. A pixel now consists of two sub-cells building a readout cell and a SPAD sensor. Both layouts are separated by guard rings and have a minimal deep n-well to deep n-well distance of 15 µm. A column is built out from a pixel. One column consists of 15 pixels. This column shares all bias, signal and power supply lines that are routed from the ASIC periphery. After column design, the SPAD pixel matrix was started. The pixel matrix has 20 columns that form the entire pixel array. The design of vertical and

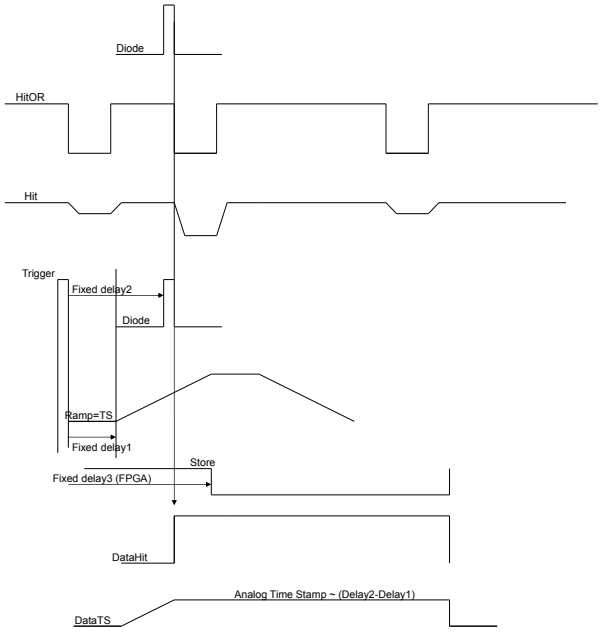

Figure 4.22: Timing diagram of laser pulse capture and time stamp calculation.

horizontal control logic is placed vertical to every row and horizontal for every column. The bias block is also placed in the ASIC periphery. All bias block connections from the bias block fan out to all columns and some pads. The whole ASIC layout is shown in figure 4.13. SPAD pixel and readout electronics is illustrated in figure 4.23. Bias, signal, power supply lines and a SPAD pixel row are depicted in figure 4.24. Figure 4.25 shows the arrangement of the horizontal and vertical control logic of a pixel row.

4.2.5 Measurement Results of the H35-SPADV1 ASIC

As we have seen from chapter 3.4, the readout implementation to evaluate the ASIC plays an important role. A special interface board (test board) was developed. This interface board is attached to the FPGA board via an FMC connector. The H35-SPADV1 is mounted on a dedicated interface board and covered with a special 3D printed black case to shield the ASIC from ambient light (figure 4.26). The test setup includes a LED that illuminates the entire ASIC. We used another type of FPGA board for readout - the Digilent NEXYS Video development board featuring a Xilinx Artix-7 FPGA[32]. The communication between computer and FPGA is implemented via Future Technology Devices International (FTDI) [40] based on FT2232H [39] chip. It integrates two independent FIFO controllers. These control the FIFO controllers. A USB protocol engine controls and manages the FIFOs of the FT2232H chip. This device allows data to be sent and received over USB connection. The FPGA communicates with the FT2232H chip using an 8-bit data bus. It provides

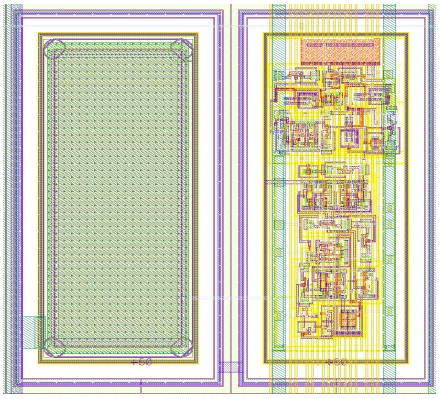

Figure 4.23: SPAD pixel: SPAD sensor and readout electronics consisting of two sub-cells.

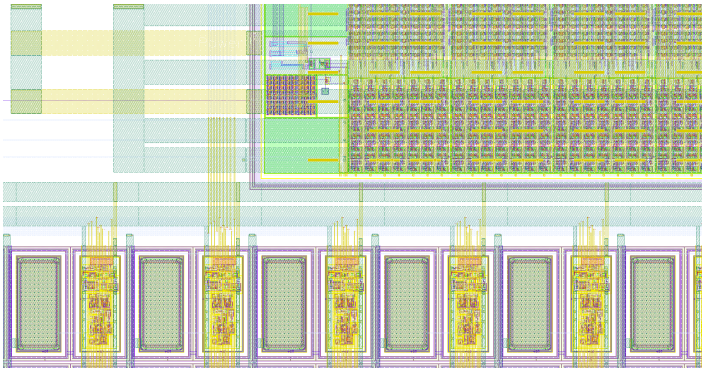

Figure 4.24: Bottom: a pixel row is sharing bias, signal and power supply lines. Top: Bias block with bias lines fan out.

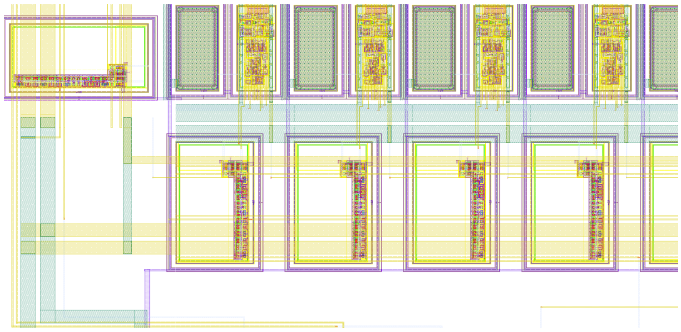

Figure 4.25: Layout of vertical and horizontal control logic and pixel in a SPAD pixel row.

two modes: synchronous and asynchronous FIFO mode. Communication in synchronous mode is established via 8-bit bus at 60 MHz with a clock generated by the device. For asynchronous FIFO mode it is possible for the device to deliver the clock to the FTDI chip. FTDI provides a libftdi library. This library allows the detection of FTDI chips, port configuration, as well as the receiving and sending of data. The library provides C/C++ functions like *ftdi_ usb_ open* or *ftdi_ set_ bitmode* etc. For our purpose we used the asynchronous FIFO mode.

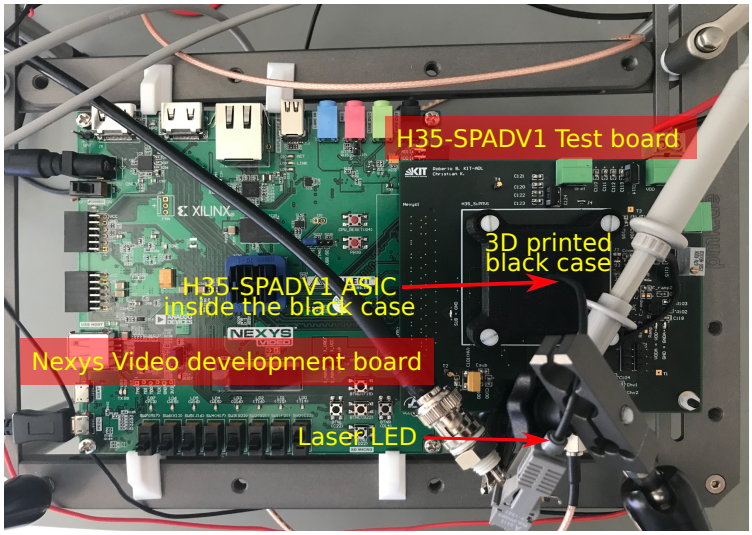

Figure 4.26: H35-SPADV1 test setup.

The software was developed in C/C++ Qt environment. The application handles the FTDI communication using libftdi. Low level bit controlling is also implemented in the application. A send bit function is used, e.g. to send a single bit using bit banging the slow

bits to generate the two phase clocking sequence. A TX (send) and RX (receive) buffer must be introduced to send the data from application over FT2232H chip. For the TX buffer it is important to send the register address and value first. Many other functions were implemented. For configuration of vertical and horizontal control, this function sends a single bit to the vertical configuration register also using bit banging config bits to generate two phase clocking. Single pixel and matrix readouts are controlled by the application and the FPGA. Interconnection between ASIC and FPGA is accomplished by defining the UCF file. In addition, the interface between FPGA and application is done via firmware main control unit. An 8-bit data in and FIFO output is defined for receiving and transmission of data. An FTDI FSM controls the transfer and receiving of the FIFO data. Registers are defined in the firmware and used in the application (address register of the ASIC) and a sequence handler writes data to, or reads data from, a specific address. A write toggle bit controls FIFO write. FIFO write is fed to the FTDI FSM. FIFO out is an 8-bit data output. The application can read/write dependent FIFO write signal generated with the write toggle bit. A simplified block schematic of the readout chain is shown in figure 4.27. A pulse counter was implemented in the FPGA for counting the dark counts.

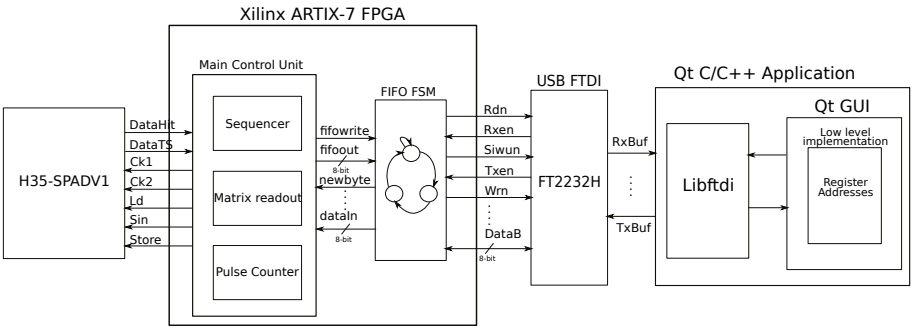

Figure 4.27: Simplified readout scheme of H35-SPADV1 test setup.

A GUI reports the DCR. When a GUI is used (implemented with Qt Creator) the ASIC can be configured and read out by counting the DCR sent by the ASIC to the FPGA in single pixel mode. An edge detector was also implemented in the FPGA for detecting the DCR or fast photon pulses. We observed that a maximum readout frame up to 25 MHz is possible. Furthermore, when generating fast pulses for laser light emitting diode (LED) a pulse generator produces a fast pulse as trigger signal for an external pulser. For full matrix readout it is essential to generate a *Store* signal and at the same time a triggered laser pulse. The laser pulse is then captured or latched by the Store pulse and read out. A picture is taken which can be read out as full matrix. Since the FPGA counts the number of pulses, the representation is displayed as the number of counted pulses (DCR or real photon pulses) for each pixel. A histogram or 2D matrix can be developed as heatmap to analyze the data. The heatmap represents the ASIC matrix and is color coded.

We present the initial experimental results for breakdown voltage and DCR. The breakdown voltage values have been determined from the I-V characteristics, using a precision source measurement unit (Keysight B2901A). To carry out this measurement, the sensor

must be set in passive quenching mode. Specific settings of *VNBias*, *delLoEdge*, *delHiEdge*, *delLoRo*, *delHiRo* etc. set the sensor to passive quenching mode and the sensor behaves as a PQC. Thus, the measured current flows over the SPAD sensor and the transistor which is working in linear region (resistance of few hundred kΩ). Using this technique, we can measure the current and determine when the current starts to increase exponentially. Figure 4.28 shows a breakdown voltage of 11.7 V [18] that was measured at 24.0 °C in dark conditions for a SPAD pixel size of 3496 µm^2. The y-coordinate is plotted in logarithmic

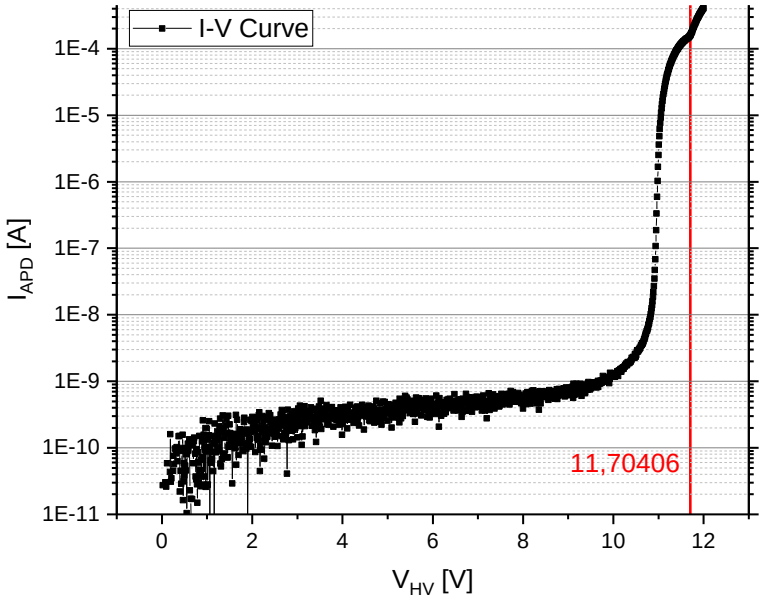

Figure 4.28: I-V curve measured in dark condition and room temperature [18].

scale to visualize the exponential rise of the current. A small kink can be observed after breakdown probably because of the comparator threshold is exceeded, which causes an additional switching current. The I-V curve of the SPAD sensor is divided in the following section: *no avalanche* occurs between 0 V and 9 V. The current start rising for values above 10 V. At this point the SPAD sensor is in the *linear mode*. The quench transistor acts as a resistor. The current is limited by the resistor. At the breakdown voltage (V_{BD}) the SPAD sensor is in *Geiger-mode*. Figure 4.29 shows the I-V curve and the sections mentioned above.

In figure 4.30, we measured the temperature sensitivity. The temperature has a strong influence on the breakdown voltage. We started the measurement at 5 °C to -25 °C and calculated a temperature sensitivity of 8.4 mV/°C.

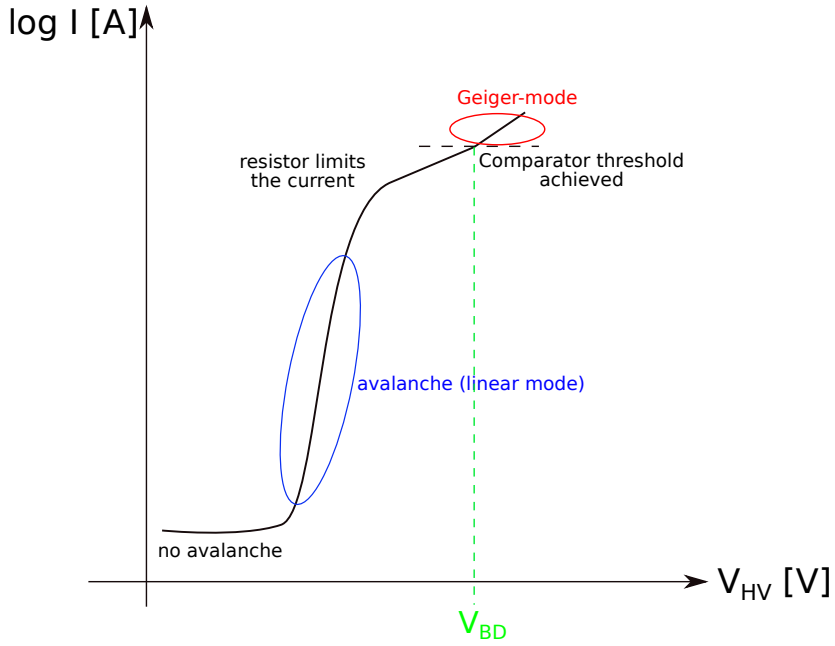

Figure 4.29: I-V curve and different sections of the SPAD sensor.

There are two possibilities of measuring the DCR: via HitOR signal or via DataHit bus. The amplitude of *Hit* is proportional to the number of hit pixels. In the full matrix or single pixel readout mode a readout window is generated. A *Store* signal is produced by the FPGA within the readout window (a timing diagram is shown in figure 4.31). The readout window is the time in which all the pixels are read out. The DCR produced is then counted via the DataHit bus. The DCR of one pixel was measured in darkness. At 12.0 V ($V_{OV} = 300$ mV) a dark count rate of 600 kHz corresponding to 174 Hz/µm^2 was measured [18]. *VNBias* strongly influences the DCR measurement. When *VNBias* is set to larger values the DCR and detection efficiency increases. For *VNBias* values between 0.65 V and 0.85 V the highest PDE and DCR over 600 kHz/pixel was measured (figure 4.32). When VNBias is varied, the time constant as well as the RC window will be changed. It has been shown that the RC window is dependent on *delLoHold*, *delLoEdge*/*delHiEdge* and comparator threshold. Another measurement reveals that, when the overvoltage (V_{OV}) was increased, the DCR drops to zero. This can be explained by the fact, that the inhibition transistor pulls the anode of the diode to 3.3 V. After the hold-off time, the anode potential decreases because it is pulled down by the NMOS transistor. If the V_{OV} is so high that the SPAD sensor is sensitive during the reset phase another avalanche may occur before the anode potential falls below the comparator threshold and no pulse or dark count can then be detected by the electronics. As a consequence the DCR drops to zero at a certain V_{OV} value. In the region where the DCR is zero, a current was measured in the SPAD sensor using the source measurement (SMU) unit.

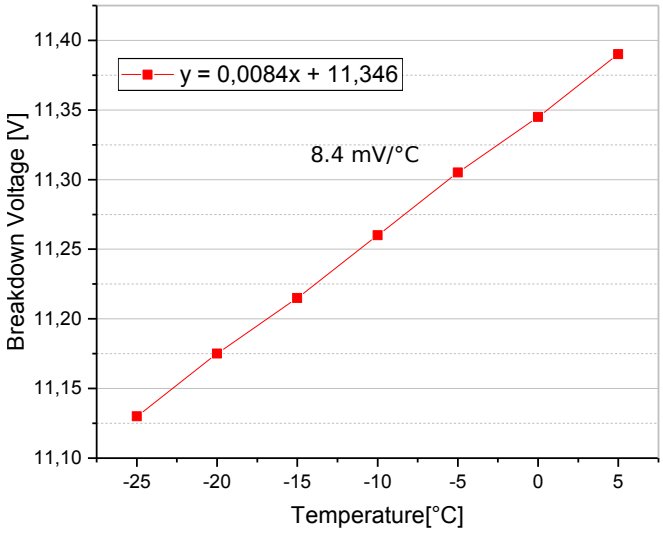

Figure 4.30: Temperature sensitivity measurement for breakdown voltage.

It was also determined that the DCR is weakly dependent on temperature (figure 4.33). This is a surprising result, since thermal generation and leakage currents usually depend exponentially on temperature. An explanation could be that the trap-assisted band-to-band tunneling is responsible for the generation of free charge carriers. Another indication of band-to-band tunneling is the low breakdown voltage. At such low breakdown voltages, the band-to-band tunneling plays an important role and significantly deteriorates the DCR. Despite a high DCR resulting from band-to-band tunneling, the sensor is sensitive to a few hundred e-h pairs. When the comparator threshold, *VNBias*, hold-off time, LED pulse timing, *Store* signal sequence are precisely set, the sensor can be sensitive to the pulsed photons. The idea behind this is to capture the pulsed photons within the Store sequence or Store pulse. This could be in nanosecond scale. The FPGA generates a trigger signal for an external fast pulser with the pulse generator.

Now we will describe the measurements of detection efficiency with a blue LED (here Store sequence is used for full matrix readout). The pulser produces a pulse of a few nanoseconds for a blue LED (the LED pulse contains a certain number of photons). Light intensity can be regulated by adjusting the pulse amplitude. On the scope it is possible to display the FPGA generated Store pulse for matrix readout and LED pulse triggered by the matrix readout sequence. It is difficult to calculate how many photons generate N electrons. One possible approximative formula is:

$$N_{eh} = \frac{I_{SMU}}{N_{pix} \cdot e \cdot f} \tag{4.7}$$

N_{eh} being the number of e-h pairs, I_{SMU} measured current of the SPAD sensor, f frequency of the pulses in continuous mode, N_{pix} number of pixels and e is the elementary charge.

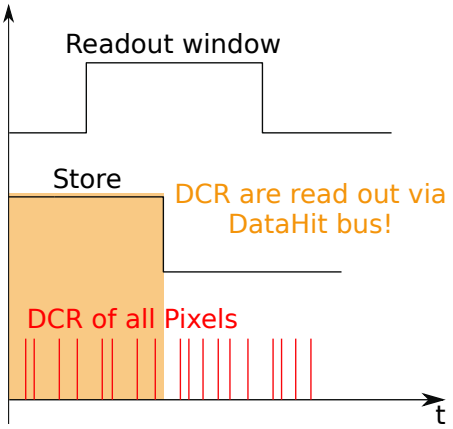

Figure 4.31: DCR measurement via DataHit bus: a readout window is generated. A *Store* signal is generated within the readout window. For *Store* = 0 the D-latch is latched, no additional dark counts are accepted after *Store* goes 0.

Equation 4.7 gives the relation between the measured current (I_{SMU}) and the number of e-h pairs. For measuring I_{SMU} it is important to operate the SPAD with a reverse voltage of 5 V. Keeping the LED in continuous pulsed mode the intensity can be adjusted and I_{SMU} can be measured. When the LED amplitude is varied, the I_{SMU} changes. By keeping the pulser settings unchanged (pulse width and amplitude) we have performed the matrix measurement. I_{SMU} is the measured current of all SPAD sensors. The LED illuminates the entire ASIC area. This means that all SPAD sensors are illuminated (figure 4.34). Notice that the total number of electrons consists of the primary electrons generated in the high-field region (avalanche) and electrons generated in the low-field region of the SPAD sensor. The produced electrons in the low-field region do not contribute to the avalanche effect. Therefore, the electrons produced in the areas of low-field region must be subtracted by the total number of produced electrons. The number of remaining electrons corresponds to the primary electrons produced in the high-field region. For instance, the equation 4.7 gives us 100 electrons. The width of the low-field region is 6 um (see figure 4.35), the number of the generated primary electrons are $\approx$ 40 and 60 electrons in the low-field region.

A LED pulse with a defined N_{eh} is generated for every matrix readout. With this measurement it can be shown whether the SPAD sensors can detect the calculated number of e-h pairs. The produced light is distributed over the whole chip surface. Notice that the Equation 4.7 gives the number of e-h pairs generated in the sensor. In order to calculate the number of photons, we will assume that to generate one e-h pair an energy of E_g = 3.61 eV is required in the silicon SPAD sensor. The incident photon total energy must be greater than E_g to generate an e-h pair. When the number of e-h pairs are multiplied by 3.61 eV the total energy required to generate N_{eh} (e.g. 6 e-h) pairs is obtained

$$E_{eh} = N_{eh} \cdot E_g = 6\,eh \cdot 3.61\,eV = 21.6\,eV \tag{4.8}$$

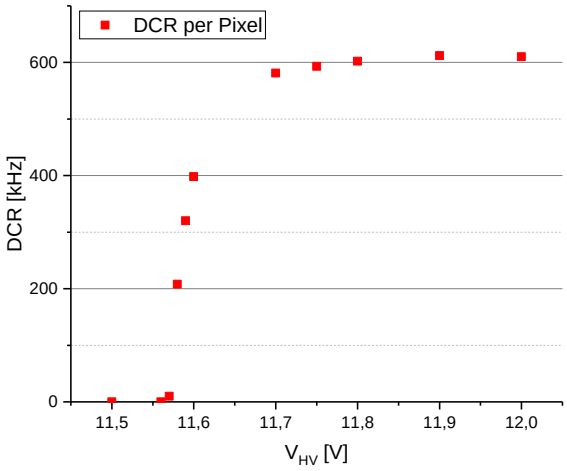

Figure 4.32: DCR/pixel as a function of breakdown voltage.

However, we are aware that equation 4.8 is very approximative. For this reason we will express the results in the following text in detected primary electrons. For blue light (λ= 420 nm) for instance, the average photon energy can be calculated as follows

$$E = h \cdot f = 4.136 \cdot 10^{-15}\, eVs \cdot 7.14 \cdot 10^{14}\, Hz = 2.95\, eV \tag{4.9}$$

To calculate the number of primary electrons we must divide the total energy for 6 e-h pairs by the average photon energy

$$N_{pe} = \frac{E_{eh}}{E} = \frac{21.6\, eV}{2.95\, eV} \approx 7 \tag{4.10}$$

Equation 4.10 is an example of calculation of the average number of photons. To detect a few hundred primary e-h pairs an overlaid hit map or image was produced. Overlaid means that we record the sum of hits for every pixel per matrix readout. The sensor was masked in two places. A SPAD silicon sensor PDE is dependent on the light wavelength. Different semiconductor materials have different absorption coefficients [4]. The absorption coefficient determines how far light can penetrate the material at a particular wavelength. For instance in silicon, blue light (420 nm) can penetrate ≈ 300 nm and red light (637 nm) 3 µm [4]. The active SPAD has a certain depth where an avalanche process is produced. If the light is absorbed too deep it will produce less primary e-h pairs and the PDE is reduced. Blue light has a higher PDE than red light [85]. For that reason we used both blue and red light for measurements. Figure 4.36 shows one overlaid image with LED pulse illumination (figure 4.36 right) and one without LED illumination (4.36 left). Without illumination the figure 4.36 shows the dark count rate (poisson distributed) of the SPAD

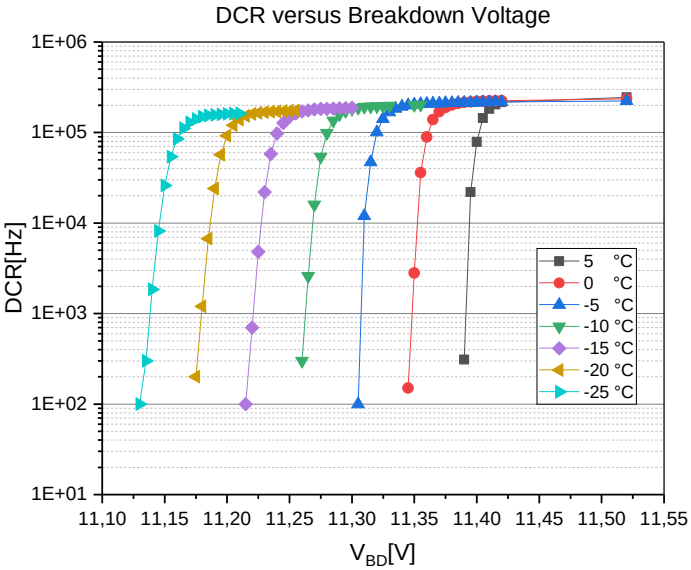

Figure 4.33: DCR/pixel as function of breakdown voltage for different temperatures [18].

sensors for matrix readout. In order to obtain the detection efficiency, the DCR of the whole SPAD matrix can be subtracted by the image with LED illumination. In this way, a pulse detection efficiency of 15.6% was calculated.

A pulse initially generates 360 e-h pairs. This means on average ≈ 360 primary electrons were generated in the sensor. In this measurement setup the masked areas are clearly visible and can be distinguished from the areas without the mask. This mask is dark in comparison to the area without mask. It can be observed that there are less hits produced on these two masked positions. Figure 4.37 shows 10k events overlaid. In the overlaid image on the left two dark areas can be seen (360 e-h pairs). The right image shows the same measurement with reduced laser intensity. 17 e-h (≈ 21 primary electrons generated) pairs are generated in one pulse. The dark masked areas become blurred due to the lower pulse detection efficiency. For 360 e-h pairs the pulse detection efficiency is 15% and for 17 e-h pairs it is 6.5 %. The PDE can be calculated as follows:

$$N_{hit} = (1 - (1 - \alpha_{PDE})^{N_{pe}}) \cdot N_{meas} \tag{4.11}$$

where N_{pe} is the detected number of primary electrons, N_{meas} number of measurements and α_{PDE} the PDE. For instance, detecting 6 primary electrons the PDE is 3.6% using the values from the measurement shown in figure 4.36. N_{hit} is the number of detected photons subtracted from the dark counts. It will be shown later that the efficiency can be improved by setting *VNBias* to higher values. To improve the PDE the *VNBias* should

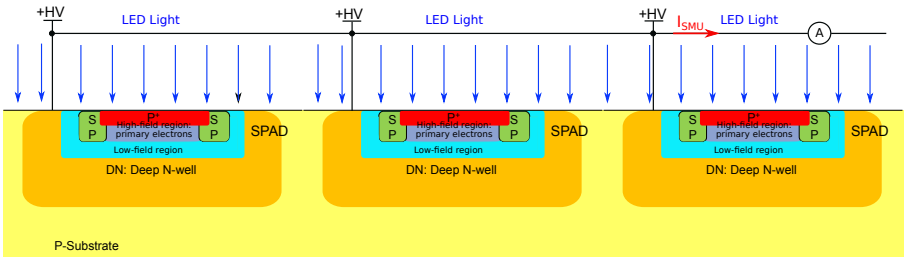

Figure 4.34: SPAD array is illuminated by a blue LED. I_{SMU} is measured for all SPAD sensors. In the high-field region primary electrons are generated. In the low-field region electrons are also generated but do not contribute to the avalanche process.

be tuned to higher values, because this speeds up the return to baseline and reduces the dead time. Additionally, increasing the overvoltage and shortening the LED pulse leads to improvement. Table 4.3 shows how the PDE can be improved by tuning *VNBias* and by increasing of the overvoltage setting. After adjusting these parameters a matrix readout

VNBias (V)	**PDE (%)**
0.4	0.66
0.45	2.82
0.5	7.6
0.55	14.24
0.6	24.9
0.65	35.0
0.7	52.8
0.75	54.5
0.8	61.7

Table 4.3: PDE improved for higher VNBias values. All measurements done with blue LED ($\lambda = 420$nm).

was taken with 100 overlaid events. Using new laser setup (photograph in figure 4.38) developed here at KIT, it is possible to spot only one pixel in the matrix (figure 4.39). The whole setup is in a light-tight box to prevent ambient light. Precise step motors move the laser to the target position. High accuracy in x, y, z position is guaranteed. The control software (written in C/C++) is implemented in the test application so that measurements can be configured with the same ASIC test application. The light intensity can be attenuated to obtain single-photons. The laser setup is able to generate picoseconds of pulses that trigger a red laser (637 nm) with beam spots smaller than a pixel size of the H35-SPADV1 ASIC. For the next measurement the emitted photons have not been attenuated, so that a maximum pulse detection efficiency of ≈ 30 % was measured for a single pixel. When a pixel is illuminated with laser source (see figure 4.39), it generates secondary photons that hit neighboring pixels (optical crosstalk). Another reason for light getting into the

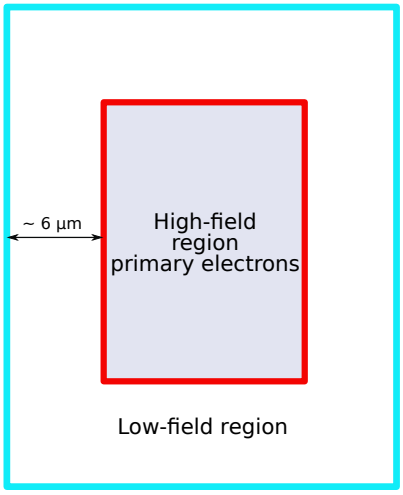

Figure 4.35: Top view: High-field and low-field regions of the SPAD sensor. The primary electrons are generated in the high-field region.

neighboring pixels can arise because the laser beam spot is not a perfect circle but has a more Gaussian shape. Some photons also get into the neighboring pixels.

To demonstrate optical crosstalk, one possibility could be pseudo-crosstalk measurements [89]. Two different SPADs are activated at the same time forming an emitter and detector pair. An emitting SPAD generates photons during the avalanche process. Other SPADs are triggered by the secondary photons. To observe the crosstalk, the SPAD with the highest dark hit counts is considered as emitter and the neighboring SPADs are the detectors and are activated at the same time. Two crosstalk measurements could be used: one with the emitter on, and one with the emitter off. These dark counts are subtracted to obtain the crosstalk without DCR contribution. This functionality to enable/disable the pixels (kill pixel) is not implemented. The kill pixel functionality will be integrated in the next ASIC version. To make the photons generated by crosstalk visible, one possibility could be observation using a light sensitive camera placed on top of the H35-SPADV1 ASIC. An emission microscope (Phemos 1000 developed by the company Hamamatsu) is an option to localize the emitted photons during avalanche processes. These secondary photons trigger new avalanche processes in the neighboring pixels. This kind of measurement was not performed in this work.

Response of the sensor to a defined laser pulse is briefly considered here. A laser pulse with a fixed width of 100 ns was generated and the response of *DataHit* and *Hit bus* was measured (see figure 4.40). *DataHit* is a digital output that reaches a 3.3 V CMOS level while detecting photons. The measurement was performed in the following way: one pixel was selected and targeted by the laser. Full matrix readout is disabled during this measurement. The laser injection occurs for the chosen pixel. The next version of the ASIC should be able to select one pixel and disable all other pixels during single pixel readout.

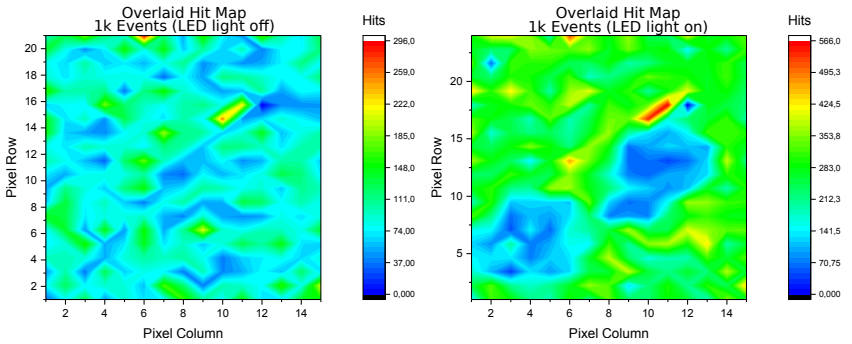

Figure 4.36: Hit map of 1k overlaid events without LED pulse (left) and LED pulse illumination (right).

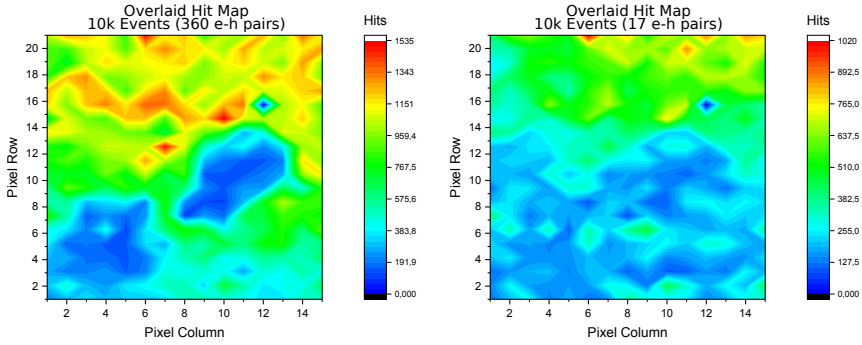

Figure 4.37: Overlaid image (10k events) of pulsed blue LED with 360 e-h pairs (left) and 17 e-h pairs (right).

The analog signal output - *Hit bus* - behaves differently. Depending on the *VPBias* setting or current setting of the PMOS transistor the *Hit bus* shows a slightly different shape. In figure 4.40 the *DataHit* and *Hit bus* bus are shown. As expected, the *DataHit* and *Hit bus* bus responds to a laser pulse. To measure the pixel time resolution the *VPBias* can be tuned in such way as to put more current in the *Hit* bus and to increase the speed of the rising edge of the *Hit* bus. A time resolution (and jitter) is measured between laser pulse and Hit analog output response.

To conclude the H35-SPADV1 measurements, the pixel front-end electronics is fully operational and the AQC principle works well. Development of a new monolithic ASIC including SPAD sensor and pixel readout electronics offers many advantages because digitizing is integrated in-pixel without the use of an amplifier. AMS 350 nm technology has shown that the SPAD sensor has relatively high DCR, and it is possible to detect several hundred photons. Noise caused by band-to-band tunneling deteriorates the sensor. Furthermore, optimizing the SPAD sensor with TCAD simulations would be the next

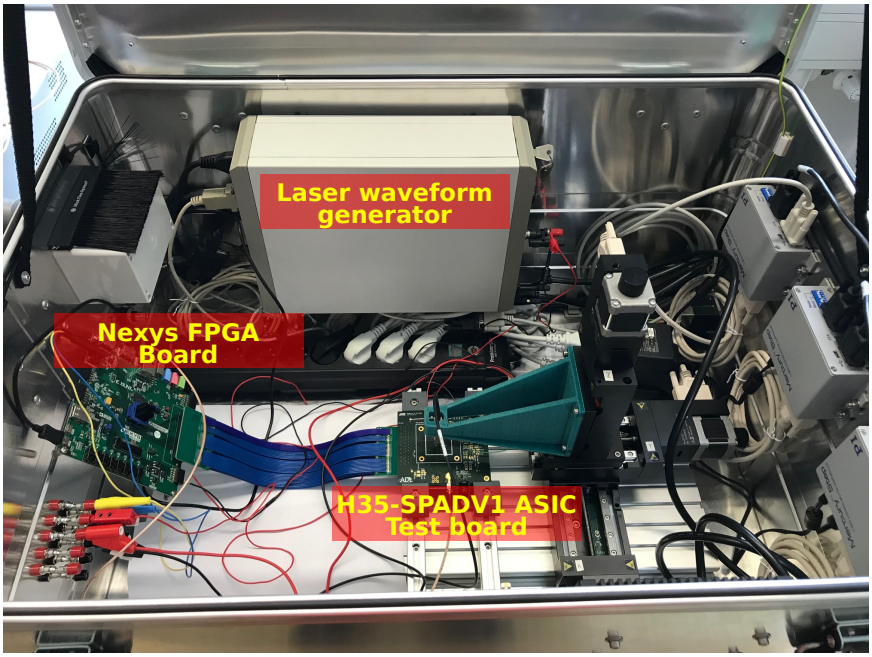

Figure 4.38: Laser setup with picosecond pulser that drives a 637nm laser.

step to improve the SPAD sensor. Another idea is to import the same architecture into a different HV-CMOS technology. A second version of the ASIC is under development. Two technologies were selected for this purpose: TSI 180 nm and LFoundry 150 nm. The ASIC based on HV-CMOS TSI 180 nm has a simpler architecture with smaller diodes to reduce the DCR. In the LFoundry 150 nm HV-CMOS technology the same architecture as H35-SPADV1 will be used. This new version includes some new features that are missing in first version. A new SPAD developed by the company FBK will be used to reduce the DCR. Hence, much smaller DCR values are expected. In the next version it should be possible to detect single-photons which unfortunately is not possible with H35-SPADV1 ASIC.

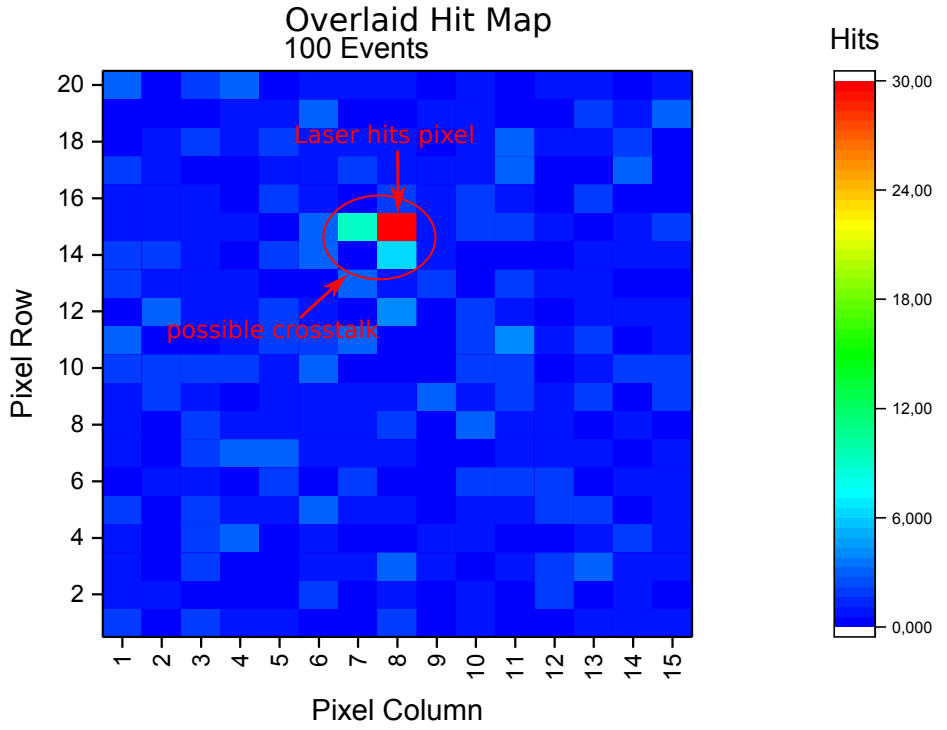

Figure 4.39: Single pixel target for a non-attenuated laser pulse. Overlaid image for 100 events. New measurement setup uses a fast laser pulser for picosecond pulse generation and a possible single-photon output.

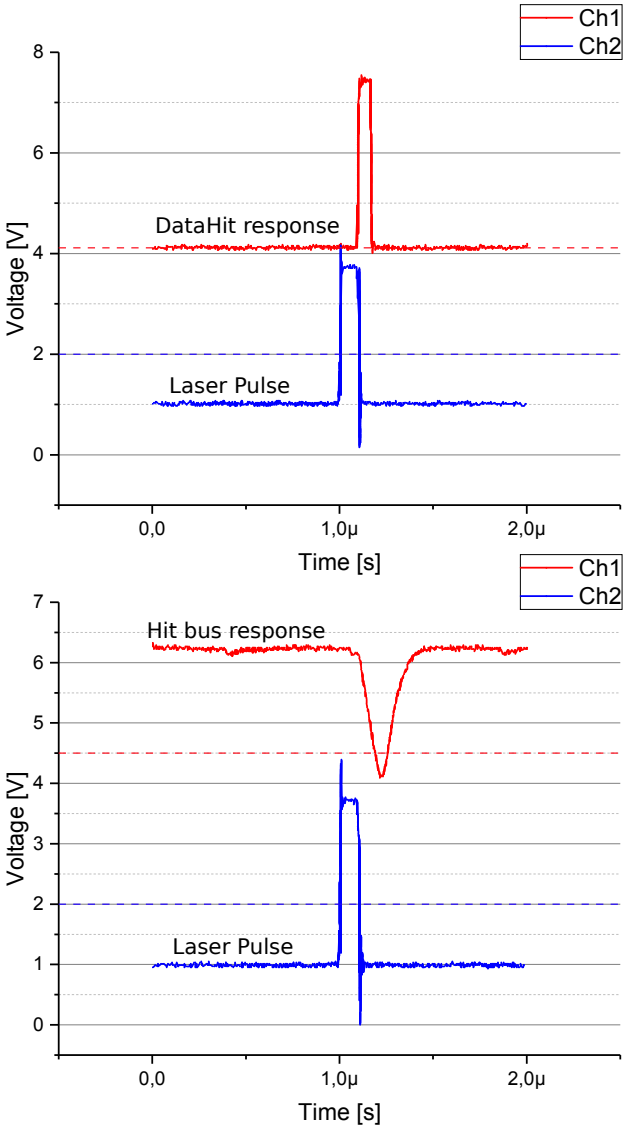

Figure 4.40: *DataHit* and *Hit bus* response of a laser pulse. A single pixel was selected and single pixel readout activated.

5 Integrated Readout Electronics for Belle II Pixel Detector

Chapter 4 was dedicated to a new single photon avalanche detector in standard CMOS technology. Interest in single photon detection is growing in various applications. In particle physics, on the other hand, the inner structure of matter is studied by means of the construction of particle accelerators. To reveal secrets of the structure of inner matter enormous particle accelerators have been built in the past decades. The Large Hadron Collider (LHC) [7] at CERN is, in terms of energy and frequency of particle collisions, the most powerful particle accelerator in the world. A great discovery of the LHC was the Higgs particle. These accelerators prove the Standard Model of Physics. The knowledge of elementary particles and their interactions are summarized in the Standard Model of Elementary Particle Physics. A description of the strong, weak and electromagnetic interactions in the form of quantum field theories allows the Standard Model. The Standard Model contains twelve particles and twelve antiparticles which are divided into leptons and quarks.

The goal of the colliders is the search for new physics at high energies. Other particle detectors such as the Belle II [5] detector proves the Standard Model with high statistics at the precision regime. The Belle experiment is located at the Japanese Research Center for Particle Physics KEK and focuses on the so-called B-physics. B-physics is the field of particle physics that investigates hadrons containing a bottom quark. Hadrons are subatomic particles that are held together by a strong force. The best-known hadrons are the nucleons: neutrons and protons. Protons and neutrons form the atomic nuclei. Depending on the spin the hadrons are divided into two types: Meson and Baryon. Mesons are made of one quark and one antiquark while Baryons are made of three quarks. Protons and neutrons are baryons. Of particular interest for Belle II are decays of B mesons to determine more precisely parameters of the Standard Model. Belle II is an upgrade to enlighten unknown sources of Charge Parity (CP) violations [24] beyond the Standard Model. CP violation (C stands for the charge or charge conjugation and P for parity) is the violation of CP in-variance. The latter states that physical laws in a system should not change if all particles are replaced by their antiparticles and all space coordinates are simultaneously mirrored.

The Belle II is an electron-positron collider which produces B meson and anti-B meson. These mesons decay into other particles. The particle decay is studied at Belle experiment. The SuperKEKB accelerator is an upgrade of the KEKB accelerator with the goal of 40 times higher collision performance. To achieve this high collision performance the beam sizes are squeezed to the nanometer level. 50 times more data than Belle experiment will reveal violations of the symmetry between particles and antiparticles. This asymmetry between particles and antiparticles occurred in the early stage of the universe. The Belle II

experiment may lead to new physics and explain why matter has gained the upper hand in the universe.

This chapter introduces the ASICs that have been developed and tested at KIT. A dedicated chip probe station was also developed at KIT to test the full functionality of the ASICs. Entire wafers and single chip tests are possible. The probe station has a special probe card with dedicated needle ring board to contact the ASIC for measurements and tests. After the test procedures the ASICs are delivered and mounted in the Belle II detector. Important for the test phase is that the tests are carried out carefully. During the tests a high yield was achieved and the ASICs were delivered on time. The KIT developed ASICs are: Drain Current Digitizer (DCD) and Switcher. They are essential for Belle II experiment and play an important role in Vertex detector **(VXD)**.

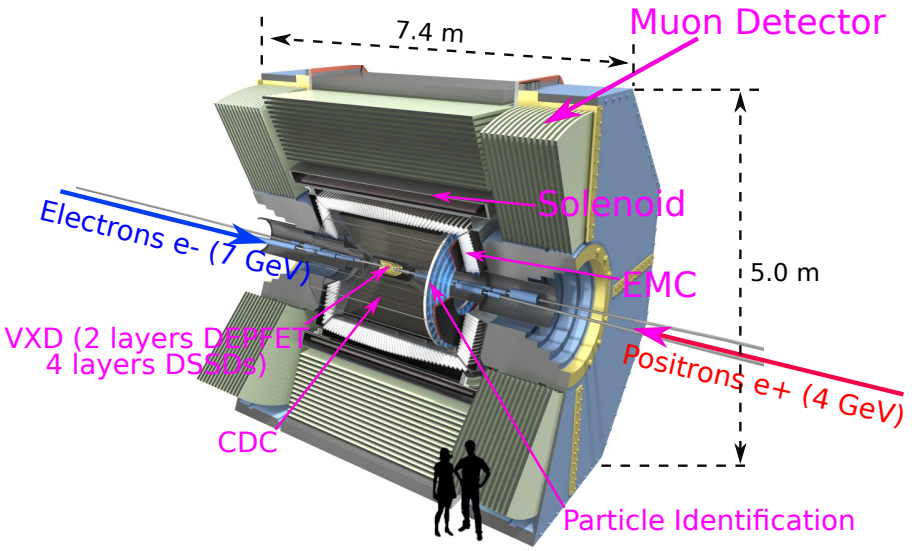

Figure 5.1: Belle II detector comprising sub-detectors.

5.1 Belle II Detector

The Belle II consists of sub-detectors [67] and is located around the beam pipe at the interaction point. Particle trajectory reconstruction emerging from the particle collisions (electrons and positrons) is the main functionality of the Belle II. Modern detectors in high energy physics are arranged in an onion shape formed by sub-detectors. To cover almost 2 π solid angle they are placed cylindrically. All sub-detectors have a special task. The Belle II can be divided into inner and outer part (figure 5.1). The outer part detectors are muon detectors **(KLM)**, super conducting **solenoid** and the electromagnetic calorimeter **(ECL)**. The ECL consists of 6624 (CsI:TI) scintillators to detect photons of the decay products

of neutral particle (neutral pions). Other particles can be identified: electrons, hadrons [67]. The super conduction solenoid generates 1.5 T magnetic field [60]. Charged particles are deflected by the strong magnetic field, which allows particle reconstruction using the tracking information from the central drift chamber **(CDC)**. The CDC is cylindrical and the chamber is filled with a helium and ethane gas mixture. Momenta and track reconstruction of charged particles are measured and the energy loss dE/dx within the gas volume is measured to provide particle identification [8, 23]. The principle of the CDC follows: charged particles cross within the gas mixture and collide with gas atoms, and electrons are generated. The electrons flow to sensor wires, creating an electrical signal that can be sensed by electronics such as amplifiers. The above mentioned VXD detector is composed of a silicon vertex detector **(SVD)** and pixel detector **(PXD)**. The VXD has six detector layers. The SVD in Belle II experiment consists of four layers of double-sided silicon detectors (DSSDs). There is a single layer of strips on the top and bottom. A photograph of the combination of SVD and PXD is shown in figure 5.2 (commissioning at KEK Japan 2018). The main task of the SVD is to reconstruct the tracks of passing particles and to determine their momenta. The PXD has two layers of Depleted Active Pixel detectors based on p-channel field effect transistor (DEPFET) sensors forming the innermost Belle II detector. The inner layer consists of eight sensors, while the outer is equipped with twelve sensors. It comprises eight megapixels which are read out every 20 µs [67].

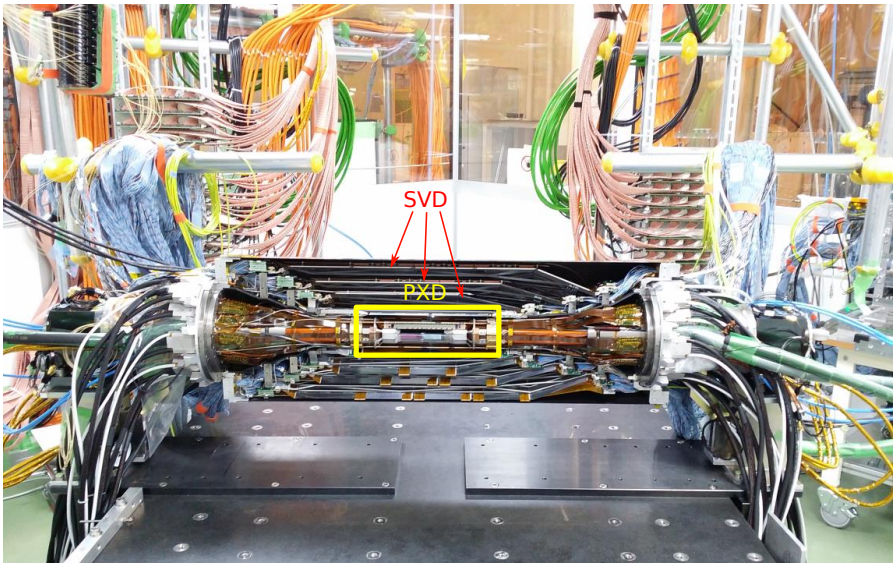

Figure 5.2: Combination of SVD and PXD at KEK Japan (commissioning 2018). Photo courtesy of Felix Müller and Hua Ye [35].

5.1.1 The Pixel Detector (PXD)

To cope with the designed luminosity of $8 \cdot 10^{35}$ cm^{-2}/s and considerably increased background, the PXD has been developed based on DEPFET sensor technology. The main requirements for the PXD is the small amount of material used and reduced multiple scattering. The total amount of material should be less than 0.3 % of the radiation length per layer [66]. In order to achieve this requirement, the standard DEPFET sensor technology has been modified in such a way that the sensor wafer is bonded to a handle wafer. The handle wafer is then structured so that a thin active region (down to 50 µm) and thicker frame provide mechanical stability.

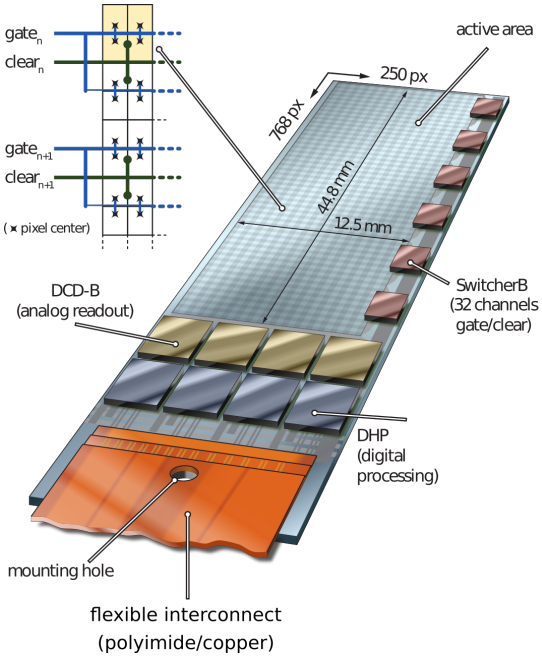

Figure 5.3: A PXD module. DCD and Switcher ASICs are arranged on the balcony. This figure is reproduced from ref. [48].

The PXD consists of two layers of DEPFET modules located at 14 mm and 22 mm radii from the interaction point. The inner layer consists of eight and the outer of 12 ladders; each ladder is made up of two modules (see figure 5.3 and 5.4). The length of the outer layer of the module is 8.5 cm and the width 1.2 cm. Each module contains a DEPFET sensor matrix with 250 columns and 768 rows. The pixel size in the inner layer is from 50 µm x 55 µm to 50 µm x 60 µm depending on the pixel position. The pixel dimensions in the second layer are 50 µm x 70 µm and 50 µm x 85 µm, respectively [66]. The matrix is steered and read out using three types of ASICs: 1) Switchers, 2) DCDs and 3) Data Handling Processors (DHPs). The Switcher chips are used to select the pixel segments (rows) for readout and

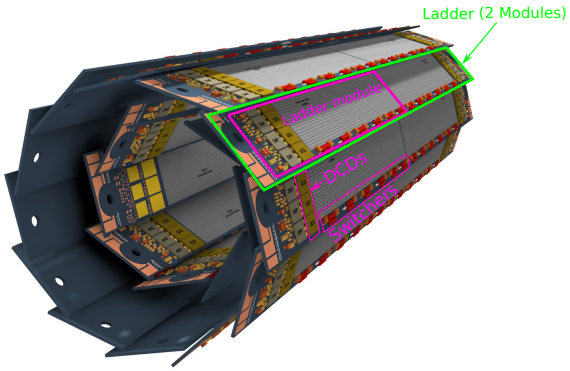

Figure 5.4: A ladder PXD modules with DEPFET matrix and Switcher and DCD ASICs for readout. Figure reproduced from ref. [9].

clear the stored signal charge. Switcher generates fast signals with high-voltage amplitudes (amplitude up to 20 V, offset up to 50 V). DCD digitizes the currents produced in the pixels. DHP carries out the digital processing of the data provided by DCD and controls the timing of the Switchers. The ASICs are directly bump-bonded to the sensor substrate. The DCDs and DHPs are arranged at the bottom of each half-module matrix and the Switchers are on the longer edge rim, the so-called balcony. Switcher and DCD have been designed at KIT and the University of Heidelberg, DHP at the University of Bonn [19]. Table 5.1 summarizes the most important requirements for DCD and Switcher.

DCD	**Switcher**
ADC resolution: 8 bit	Number of Switcher channels: 2 x 32
Radiation tolerance: 20 MRad	Radiation tolerance: 20 MRad
ADC offset correction: 2 bits	Voltage range: 20 V
Power consumption: < 2 W	DC current consumption: < 50 mA
Analog CMC	Speed: 15 ns rise time (150 pF load)
Noise (sigma): 80 nA	
ADC LSB: 80 nA	

Table 5.1: Requirements for DCD and Switcher.

5.1.2 DEPFET Sensor

The DEPFET is the first successful monolithic pixel detector technology on depleted substrate. It was invented by G. Lutz and J. Kemmer [47]. This invention paved the way for modern depleted active detectors. The DEPFET pixel contains a p-channel metal oxide semiconductor (PMOS) transistor. The whole n-type bulk of the sensor is depleted. This is achieved by a back side p+ implant, that is set to a negative high-voltage (see figure 5.5). The signal is generated in the depleted bulk. The charge in the internal gate modulates the

PMOS current which can be switched on or off using an external gate. The charge from the internal gate must be cleared after readout. This is done by applying a high-voltage on a clear contact. Because the signal charge is collected in the depleted potential well and there is an amplifying device in the pixel, there may be a low noise in the DEPFET detector. When the readout electronics are optimized, noise levels below one electron can be achieved. DEPFET sensors are now mostly used for X-ray spectroscopy [19]. The DEPFET readout has three steps: 1.) Sensing 2.) Readout and 3.) Clear. In the sensing phase, ionizing particles passing the DEPFET create e-h pairs (figure 5.5). The holes drift to the p+ contact while the electrons are collected by the internal gate. The current flows between drain and source is modulated by these electrons. Therefore, a voltage controlled current source is required where the current is modulated by the gate voltage and by the electrons in the internal gate. Applying a negative voltage at the external gate forms a conductive channel between drain and source. During the readout a negative voltage is applied on the MOSFET that is switched on. On the other hand, applying a positive voltage switches the MOSFET off. In the clear phase the accumulated charges (electrons) are cleared or removed from the internal gate by a clear voltage pulse on the clear contact. An n+ implant close to the MOSFET forms a clear region. A conductive channel must be created between clear n+ implant and internal gate. A positive voltage is applied between clear implant and internal gate, causing electrons to flow to the clear region. The electrons moved to the most positive potential and the clear region are large enough to remove the electrons from the internal gate. The DEPFET sensor is switched on by applying a low gate voltage or (gate pulse) and the pixel is cleared by applying a high clear voltage (clear pulse). An equivalent circuit of the DEPFET is shown in figure 5.6. This circuit comprises MOSFET with external and internal gate and coupling capacitance between clear gate and clear. The current of the DEPFET is modulated or controlled by the voltages of external and internal gate. There are parasitic components such as coupling capacitances forming low-passes and leading to a determined time constant. Capacitance also leads to charge injections. Like any other semiconductor sensor the DEPFET is affected by noise. For

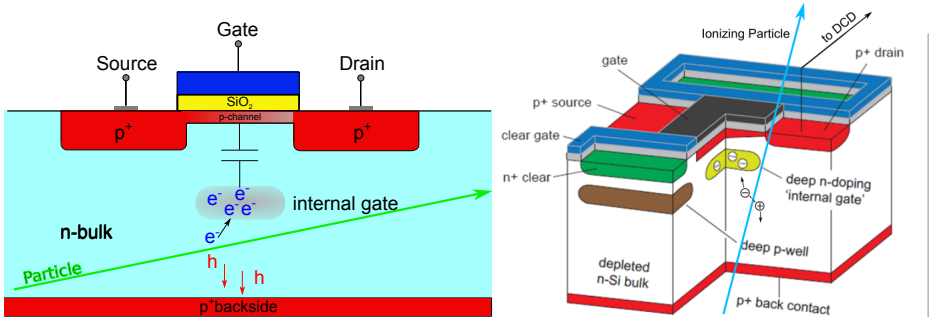

Figure 5.5: Simplified cross-section of the DEPFET sensor and DEPFET sensor reproduced from ref. [43].

high-bandwidth readout was better than 50 e noise presented in [79]. For low-bandwidth

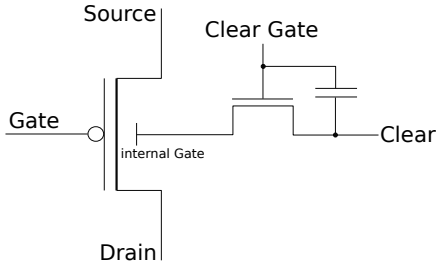

Figure 5.6: Equivalent circuit of the DEPFET sensor.

readout the noise of the device can be reduced to a few of electrons. Measurements with Fe-55 at 5.9 keV give a spectral resolution of 143 eV that corresponds to 2.3 e [12].

5.2 DCD ASIC Architecture

Although the DEPFET sensor has amplification in pixels, there are several factors that make the readout and signal processing with the required S/N ratio challenging:

(1) required readout time of 20 µs

(2) large fixed pattern noise (offset variation)

(3) large capacitance of connections between the pixels and DCDs (column lines)

(4) relatively small signal amplitudes

PXD modules will be read out in the rolling shutter mode. Four-row matrix segments are selected consecutively by setting the gate voltages of the DEPFETs transistors. This is performed by the Switcher ASICs. The signal current flow to the long connections to the DCDs placed on the periphery. At the end of the cycle the signal charge from the internal gate is cleared by Switchers. A clear pulse of about 20 V is required. The gate voltage must be kept low during the clear phase. Since the PXD module contains 768 rows, and four rows are read out in one cycle, 192 steps are needed to read out a complete matrix. The frame time plays an important role because a frame time larger than 20 µs would lead to a very large number of background noises or hits and make it hard to find the relevant tracks. Hence, a great effort has been made to achieve a segment readout time in the order of 100 ns. In this case the matrix readout time is 19.2 ms (192 steps x 100ns for a segment readout time) [19].

Each DCD is composed of 256 analog channels. Since one row-group consists of 1000 DEPFET pixels (four rows each containing 250 pixels), four DCD are required to read out all the columns in parallel. The DCD is implemented in UMC 180 nm CMOS process technology using a radiation hard design technique. The chip occupies an area of 5 mm x 3.2 mm. Deposition of SnAgCu (SAC) bumps is implemented by the vendor. SAC technology is commonly used for electronic soldering. Figure 5.7 shows a photograph of the DCD III ASIC with SAC bumps. Every DCD channel is composed of one input

signal amplifier and one ADC. The amplifier keeps the column line potential constant which is necessary to achieve fast readout. It compensates for the DEPFET offset current variation, amplifies the signal and provides shaping for noise reduction. The amplifier has a programmable gain and bandwidth (by switching on resistors in the feedback path), a 2-bit DEPFET offset current compensation (using digital data from the DHP ASIC) and analog common-mode correction. The analog signal is digitized using current-mode pipelined ADCs. This pipelined ADC has a similar architecture to the pipelined ADC used for the NeutrinomassV3 ASIC, the only difference being that the NeutrinomassV3 pipelined ADC has fully differential architecture. A synthesized digital block (Verilog code synthesized on chip) decodes and de-randomizes the ADC data which is transmitted to the DHP using 256 MBit/s digital outputs. The channel architecture has been described in [72]. The current-receiver, analog common-mode correction (ACMC) and ADC will be described in this chapter.

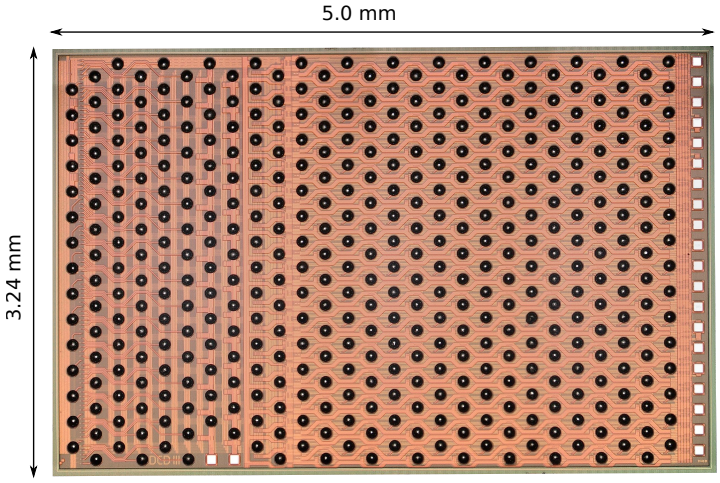

Figure 5.7: DCD III photograph with SAC solder bumps and metal signal traces can be seen. The technology used is UMC 180 nm process.

5.2.1 The Current-receiver

The following main circuit exists: a current-receiver based on a transimpedance amplifier with a resistor on its output that amplifies the current generated by the DEPFET sensor. A current-mode pipelined ADC based on current memory cells (CMC), memory cells (CM) and comparators (CMP). A 2-bit DAC for pedestal (DAC) correction, a calibration circuit for the ADCs (Cal), a decoder for generating the control signals for the pipelined ADC and a pixel configuration register (Config). A simplified analog channel is depicted in figure 5.8.

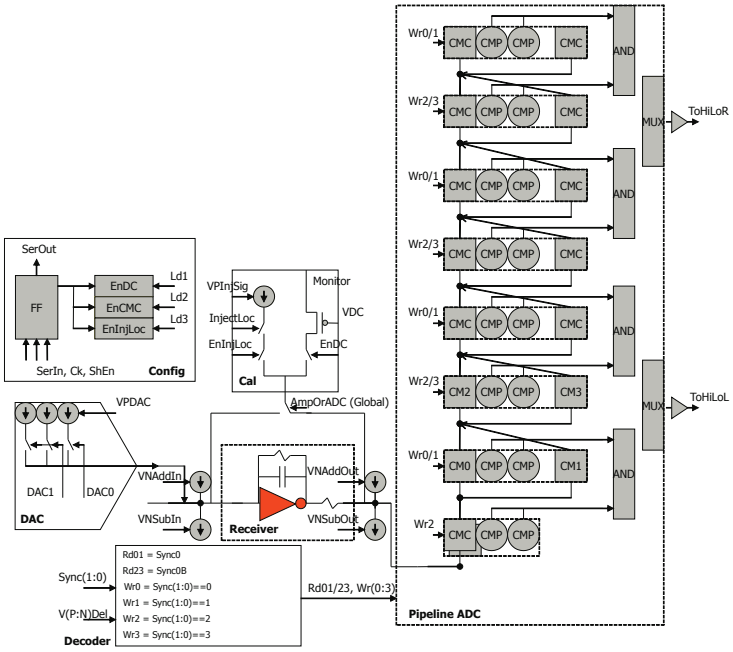

Figure 5.8: Simplified block diagram of the DCD analog channel.

The heart of the current-receiver block is a transimpedance amplifier (amplifier with a feedback resistor R_{fb}) that receives the DEPFET current (I_{in}) and converts it into a voltage V_{out}. This voltage can be described

$$V_{out} = V_{in} - I_{in} \cdot R_{fb} \tag{5.1}$$

Resistor R_s is connected between V_{out} and the input of the ADC (ADCin). Since the ADC holds its input at a constant potential, the current flowing into the ADC is defined as $I_{out} = (V_{out} - V_{ADCin})/R_s$ or substituting the previous equation

$$I_{out} = I_{offset} - G \cdot I_{in} \tag{5.2}$$

where G is the current gain $G \equiv R_{fb}/R_s$ and offset current $I_{offset} \equiv (V_{in} - V_{ADCin})/R_s$. So the current gain G would be higher than 1 if the resistance R_s is smaller than R_{fb} due to the ratio of R_{fb}/R_s. Current sources SubIn and AddIn, SubOut and AddOut and the 2-bit DAC can be used to compensate I_{offset} (figure 5.8). The settling time of the current-receiver and the current gain G can be adjusted by changing the R_{fb}, R_s and C_{fb}. The schematic of the current-receiver is shown in figure 5.9.

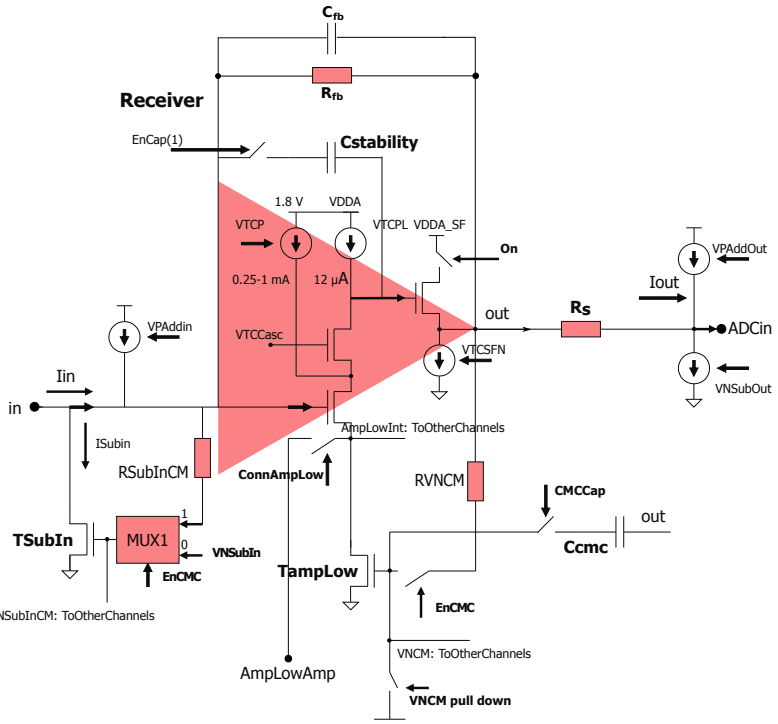

Figure 5.9: Schematic of the current-receiver. A cascoded transistor and the input transistor form the amplifier in this simplified schematic.

5.2.2 Analog Common-Mode Compensation

There are several noise sources in the PXD module: (1) transimpedance amplifier introduces thermal noise, (2) offset variation is caused by the DEPFET threshold mismatch, (3) fluctuations of DEPFET matrix and bias voltages cause common-mode noise. Common-mode noise is the correlated signal fluctuation that effects all the channels in the same way. The readout system should cope with all noise sources explained. The thermal noise can be reduced by using a single input transistor amplifier and maximizing the bias current of the amplifier. This amplifier has less noise than the differential amplifier using the same bias current. Shaping of the signal (RC filtering) is not a feasible option because the readout time is relatively short and the filtering time constant is limited [19].

A variation of the offset is partially compensated at the input of the amplifier in order to increase the available dynamic range. This is achieved by using a 2-bit current-mode DAC. A typical current-mode DAC makes up a cascode current source where PMOS transistors are biased at VPDAC and switching transistors where controlled by DAC0 and DAC1. The digital values for the DAC are stored in the DHP memory and transmitted to the DCD. The remaining offset variation will be compensated in the DHP. The common-mode

fluctuations may exceed the amplitude of other noise sources. If they are not compensated in the analog channel, a large part of the ADC range will be occupied by the common-mode noise, leading to a reduced signal range. The common-mode noise is compensated by the analog feedback loop. This operation is described as ACMC-mode. In the ACMC-mode all 256 amplifiers behave as distributed differential amplifiers with 256 inputs and 256 outputs. 256 analog channels is a new approach in the readout of pixelated sensors. There are two common-mode feedbacks in the channel. One serves to stabilize the inputs and the other stabilizes the outputs of the amplifiers [19].

The amplifier input is used to determine the bias voltage VNSubInCM. VNSubInCM is therefore the arithmetic mean of all input voltages *in* (see figure 5.9). Summing of the voltages is executed using resistors RSubInCM. Similarly, the amplifier output is used to determine the bias voltage VNCM. VNCM is also the arithmetic mean of all output voltages *out*. Summing of the voltages is also executed using resistors RVNCM. Note that VNSubInCM and VNCM are gate voltages of the transistors TSubIn and TampLow. These voltages are distributed by all the channels [19].

5.2.3 Pipelined-Current-Mode ADC

A channel contains a pipeline-current-mode ADC in addition to the current-receiver. The principle of cyclic ADC is explained here in this section. The ADC uses analog memory cells to implement the redundant signed-digit conversion. The algorithm starts with the comparison of the input signal with two thresholds, one positive and one negative. If the input signal exceeds the positive threshold, the output is set to 10 (+1), then a reference current is subtracted. The output code is set to 01 (-1) when the input signal is lower than the negative threshold voltage and the reference current is also added. The bits are set to 00 (0) when the input signal is in between the thresholds. The output signal is multiplied by factor two and the result undergoes the same operation for the next iterations. A separate electric block is used (for the pipelined ADC) for every iteration. For 9-bit resolution eight pipelined ADC blocks (1-bit redundancy) are required. The pipelined block structure with more schematic details is shown in figure 5.8. These pipelined channels are composed of CMC, CMP, CM cells, AND gatter and multiplexers [19].

Current-Memory Cell Current-mode memory cells (CMC) are used to implement the conversion algorithm described above. The CMC cell is reported in figure 5.10. The capacitor C_{fb} serves to store the analog information. A transconductor TC converts the voltage across C_{fb} into a current. One of the most critical requirements is the linearity of the TC. A high-linearity TC is not required for this application due to 9-bit resolution. For high-resolution ADCs a high linearity TC is certainly an advantage for high resolution applications. Switch *Sw1* samples the voltage across C_{fb}. Switch *Sw2* is used to disconnect the TC from the input when necessary. Switch *Sw3* connects the TC current to the next cell during *read* state. In the following consideration we assume that the amplifier has an infinite gain and no offset (assuming an ideal amplifier). In the *write* state; *Sw1* and *Sw2* are closed. There is a negative feedback generated by amplifier A and TC. The voltage on **node X** takes the value at which the TC can drain the input current completely, so that no current flows into C_{fb}. Opening *Sw1*, the voltage across C_{fb} remains stored. In the

read state, *Sw1* and *Sw2* are opened and *Sw3* is closed. The input current I_{in} flows out to the previous cell (see figure 5.10). Further considerations of the memory cell have been explained in Ref. [71].

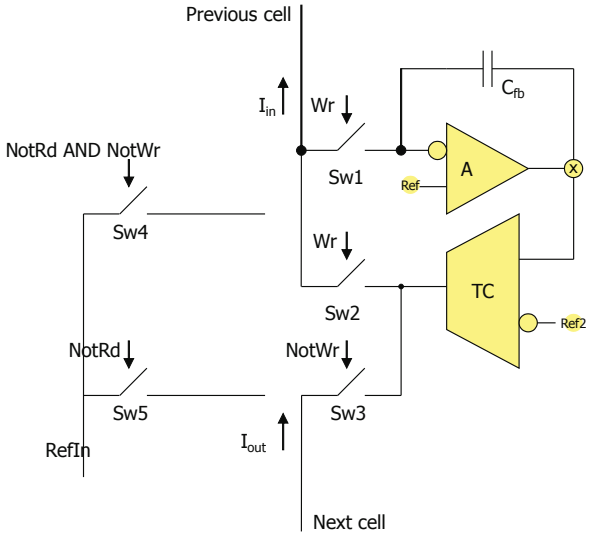

Figure 5.10: The current-mode memory cell with amplifier A and transconductor TC.

5.2.4 Current-mode realization of the Pipelined ADC

The ADC is composed of 8 double cell blocks (figure 5.11). For simplicity we show only the first three blocks. All eight blocks are shown in figure 5.8 with less details. One double cell block contains two memory cells (*MemCell1*, *MemCell2*) and two comparators (*Comp Lo*, *Comp Hi*). The comparators are connected to the voltage output (**node X** in figure 5.10) of the first memory cell. *MemCell2* does not drive the comparators. All memory cells have additional switchable current sources that can be used to add or subtract the reference current. The analog to digital conversion is performed in the following way [19]:

- State 1: The signals from the previous cells are written in cell 1. There is a difference in the first double block, which is explained later in this section. The comparators connected to these cells are in *reset* state.
- State 2: Signals from the previous cells are written in cell 2. The comparators connected to cell 1 are in a *compare* state.
- State 3: Comparators connected to cell 1 are latched; cell 1 and 2 are read out. Comparison results are used to decide whether to add, subtract or disable the switchable reference sources in cell 1 and 2, in accordance with the conversion algorithm. If the thresholds are chosen correctly, the residue currents flowing out of the cells have half of the full signal range for an input signal $\pm 2R$ (R is the reference

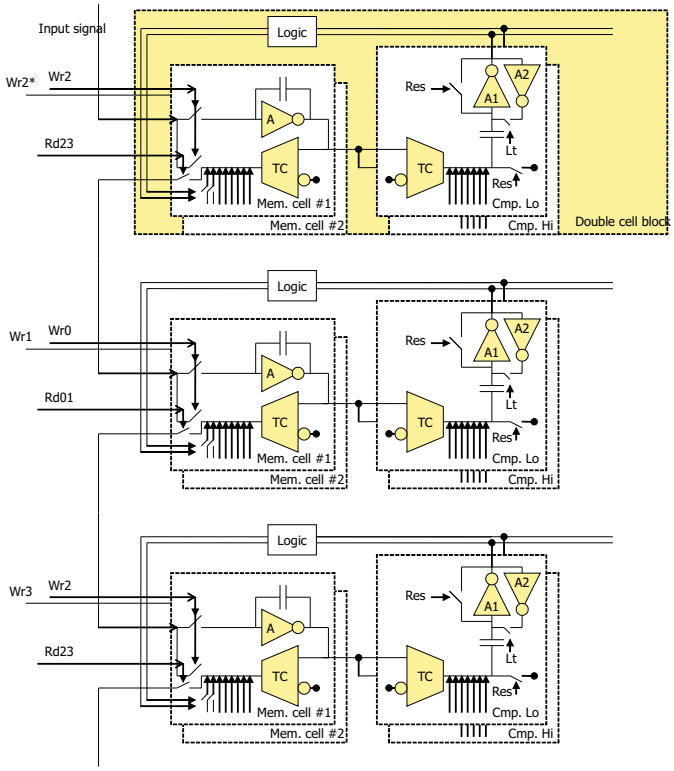

Figure 5.11: Block diagram of the pipelined ADC cell.

current). The residuals can be summed and the sum is fed into cell 3. Since the residue currents are equal, the summing is equivalent to multiplication by two. Cell 3 is in *write* state, the comparators connected to this cell are in *reset* state.

- State 4: The summed residue currents are stored in cell 4. The comparators connected to cell 3 are in *compare* state [19].

The states 1-4 are repeated for every cycle. Note that the first double block is controlled in a slightly different way to the remaining blocks. For example, in State 1 both cells are written at the same time. Thus nothing happens in State 2. This means that the ADC takes only one sample within one state. Since the input current is stored in two cells at the same time, it has twice the ADC range. Transferring the signal from two memory cells into one has the same effect as doubling the current [19].

A digital output pair (h_i, l_i) generated by the comparators when the corresponding double cell block is in *read* state. When one block generates valid outputs the next block generates invalid output (00). Therefore, the digital outputs of the two subsequent blocks can be merged by the AND function (see figure 5.8). Further, a time multiplex is performed by the multiplexers, also shown in figure 5.8. The idea behind this is to reduce the number

of digital lines in total to four (2 x *CompHi*, *CompLo* outputs) per ADC block. The generated bit stream (h_i, l_i) is the redundant binary representation of the input signal. To evaluate the final conversion result (D), bit streams h_i and l_i must be subtracted and the subsequent conversion result can be written as

$$D = \sum_{i=0}^{7} 2^i (h_i - l_i) \tag{5.3}$$

This is accomplished in the digital readout block placed in the ASIC periphery. An 8-cycle quantization of the input signal results in a signed discrete number

$$in = D \cdot \frac{R}{128} \tag{5.4}$$

where D takes all integer values between -255 and 255; D corresponds to 8+1-bit binary representation of the input current I_{in}. The digital data transfer is simplified by ignoring the LSB of D and only the 8-bit digital codes are transferred. The simplified code can be represented within the range -127 and 127 [19].

5.3 Switcher ASIC Architecture

The Switcher (see photograph in figure 5.12) steering ASICs will be mounted on the rim of the longer edge of the DEPFET module, the so-called balcony. These chips can generate fast (15 ns for 120 pF load) voltage pulses of up to 20 V amplitudes to activate DEPFET rows and clear charges deposited in the DEPFET sensor. The gated mode operation, which will be explained below, is also supported. The final version of the Switcher ASICs has been implemented in the AMS 180 nm HV-CMOS process technology. The ASIC size is 3.62 mm x 1.5 mm. Bumps deposition is accomplished by the Fraunhofer IZM Institute in Berlin. Six Switchers are required to readout one DEPFET module with 768 pixel rows, each with 32 channels. The Switcher has following fundamental blocks (figure 5.13) [19]:

- 32 high-voltage channels each generating two high-voltage signals. Fast high-voltage drivers are implemented.
- A low-voltage control block that is composed of shift registers. This low-voltage control block is used to select the high-voltage channels.
- Voltage regulators are used to generate auxiliary high-voltage supplies
- A digital block or slow control block where a Joint Test Action Group (JTAG) FSM is implemented

The outputs of the high-voltage channels switch between upper (*CHI*, *GHI*) and lower voltages (*CLO*, *GLO*), respectively. The channels are enabled one after the other by a shift register. The ASICs can be daisy chained by connecting the serial output to the serial input of the next ASIC.

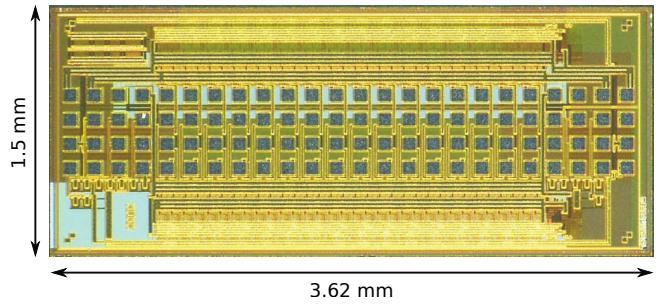

Figure 5.12: Photograph of Switcher ASIC in AMS 180 nm high-voltage technology.

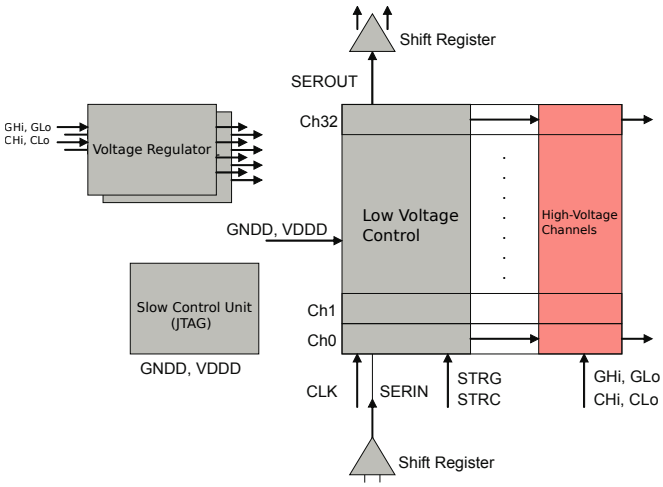

Figure 5.13: Simplified block diagram of the Switcher ASIC with the main blocks.

5.3.1 High-Voltage Channel

The main block - high-voltage channel of the Switcher is described in this section. Each channel contains a digital low-voltage control block (*LV ctrl.*), a CMOS-to-differential-current converter (CDC), two high-voltage MOSFETs (T_{nHV}, T_{pHV}) with their gate drivers (see figure 5.14). The CDC uses the floating 1.8 V supply and receives two input signals, *Sleep* and *LVIn* from the low-voltage control. The high-voltage power transistors T_{nHV} and T_{pHV} can sustain high drain-gate and drain-source voltages. The gate-source voltage must be within 1.8 V. The ON-resistance of both devices is about 20 Ω for gate-source voltage (V_{gs}) of 1.8 V. The gate control block uses four high-voltage supplies; in the case of the block connected to T_{pHV} they are *CHI/GHI* and *C/GHI0*, when the block is connected to T_{nHV} they are *C/GLO* and *C/GLO0*. The Supplies *C/CHI* and *C/GLO* are generated auxiliary: *C/GHI0* and *C/GLO0* on the ASIC use two voltage regulator blocks. *C/GHI0* is set to 1.8 V lower than *C/GHI* and *C/GLO0* is set to 1.8 V higher than *C/GLO* [19].

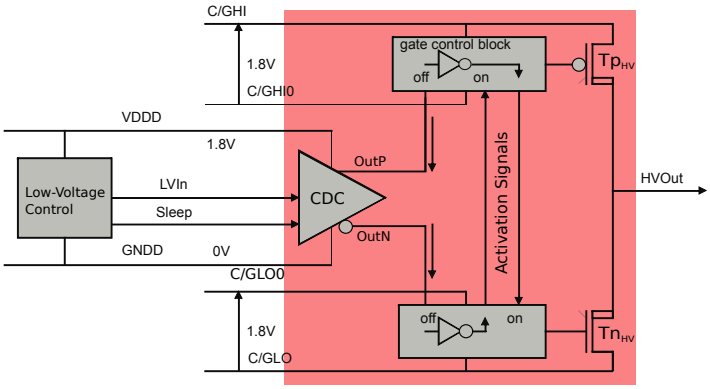

Figure 5.14: Photograph of Switcher ASIC in AMS 180 nm high-voltage technology.

If the CDC receives a positive input *LVIn*, it generates a current through the *OutP* path. The gate control block which is connected to T_{pHV} is turned off. An activation signal which is produced by the gate control block, is set to the gate control block connected to T_{nHV} which then sets its output to *C/GLO0* causing T_{nHV} to turn on. *HVOut* has the same potential as *C/GLO*. If the polarity is switched from high to low, the CDC generates a current in the negative path *OutN*. The current flows in the gate control block connected to T_{nHV} generating low voltage at the output and turning T_{nHV} off. An activation signal is produced again and sent to the gate control block connected to T_{pHV}, turning T_{pHV} on. The *HVOut* has the same potential as *C/GHI* [19].

5.3.2 The Gated Mode Operation

The SuperKEKB experiment uses the so-called continuous injection scheme [66]. Every 20 ms bunches of electrons and positrons are injected into the accelerator. These bunches generate a large background for about several milliseconds. The noisy bunches pass the detector region every 10 microseconds. Since the DEPFET has 20 microseconds integration time, every frame would be flooded by the signals from the noisy bunches and the data would be unusable for 10 % of the time. A gating mechanism has been developed to solve this kind of problem. During the gate phase, the clear voltages for all rows are set high while the transistors are off. In this way the "old" electrons are kept in the internal gates while the next generated electrons (or "new" electrons) are not collected. These electrons flow into the clear region and are removed. With this technique the sensors old information is protected [19].

5.4 Data Handling Processor

The DHP ASIC is also used for the DEPFET PXD sensor readout. The ASIC is used to reduce the amount of data before sending them off module via multi Gb/s differential

signals. It receives the data from the DCD via 64 single ended digital inputs. This data stream is corrected for CM and has a fixed pattern noise. Hit clusters are identified by thresholding operation and the hit data are sent out serially. The data are reduced by zero-suppression and with a maximum frame rate triggering of 30 kHz. Pixel data with signals exceeding a defined threshold are sent out for the DAQ. Signals below a certain threshold are not processed by the DAQ. The DHP is capable to transmit data with a rate of 1.6 Gbps using 8b/10b encoding [56]. A more detailed description of the DHP is not part of this work [19].

5.5 Development of a Chip Probe Station for DCD and Switcher Tests

About 1000 chips are required for the construction of the whole detector. Before mounting them on the detector, extensive tests must be carried out on the ASICs. The KIT developed a chip probe station for testing the main functions. The following tests must be performed for the DCD: Digital readout test, ADC curve measurement, INL, differential nonlinearity (DNL) noise measurement for every channel. For Switcher ASIC the following tests are essential: Digital configuration test and the test of all high-voltage outputs [19].

The chip probe station consists of a mechanical mounting and support (made at KIT), step motors (for x-y-z and phi axis movement), single and multiple vacuum chip carriers, wafer vacuum chuck, optics (high-resolution camera and laser) and probe cards. We have designed three probe cards containing a needle ring, FPGA, voltage regulators, multiplexers for readout. For testing the DCDs a PC135 and PC22 have been produced at PTSL. The PC135 can contact all the digital outputs which is dedicated for DCD with bumps only. PC22 contacts the test pads using standard needles and tests DCDs at full speed. PC135 and PC22 are able to read out an ID via JTAG, ADC clock programming, ADC values readout via JTAG and send a digital test pattern. The Switcher probe card (PCSw) is able to read ID via JTAG and test all high-voltage channels with a programmable external multiplexers. Figure 5.15 shows different probe cards used for the tests. The system has a monitoring computer for video processing and motor control. The monitoring software control environment is based on a web server application. The application runs with a windows PC and is based on Java/Scala implementation. The step motor controllers provide C/C++ DLLs. To use all motor control functions for Java/Scala a wrapper was developed to import all important functions. With these Java drivers the motors can be controlled via the Java/Scala application. Storage of position data of the motors and alignment between needles and ASIC is possible with the application. Automatic load, target and capture functionalities are available. Furthermore, a second PC is available which is used to program the FPGA on the probe card and to send the data between PC and probe card. The interface used for this communication is FTDI. The PC receives the data which is processed by C/C++ application. For both ASICs special firmwares were developed to control the ASIC data and send them to the PC. The firmware interacts with the C/C++ application where ADC clock programming, bias setting etc. of DCD and Switcher is possible. After the bit file has been uploaded to the FPGA, it is possible to select the software to be used. Two different GUIs for DCD and Switcher were implemented. The

Figure 5.15: Photograph of the DCD and Switcher probe cards.

DCD application can calculate INL, DNL, noise and full characteristic curve measurement. For the Switcher the application can set the gate and clear signals in the high-voltage channels. An internal counter, implemented in the Switcher firmware, can test every high-voltage channel inside the ASIC which is an essential part of the testing procedure for the Switcher. DAC configuration register must be implemented where is an essential part to set their operating point. Figure 5.16 shows a photograph of the chip probe station [19]. The photograph shows all the main blocks such as optics, step motors, wafer chuck, vacuum pump etc. A DCD test procedure is carried out as follows:

1 Contact the ASIC with needles (probe card)

2 The voltage used are VDDD = 1.8 V, VDDA = 1.8V, RefIn = 1.1 V, AmpLow = 0.3 V

3 In order to check the contact between needles and ASIC, JTAG ID readout is started. When JTAG ID is detected, the chuck raising is stopped

4 Send test pattern and read out test pattern

5 Analog power consumption after DAC configuration

6 It happens that RefIn current is not in the correct range after configuration

7 In this case the chuck is additionally lifted by less than 100 µm

8 ADC readout is started

9 ADCs are read out via JTAG; 256 are read out in parallel

10 The current source IP_Signal (internal current source) is used

11 If there is a contact problem, the ADC linearity gets worse. In this case the RefIn current usually oscillates. The chuck must be lowered a bit until RefIn current stabilizes. After adjustment the ADC linearity improves and fulfills the specification

12 The application can measure INL, DNL and ADC noise for all 256 ADCs

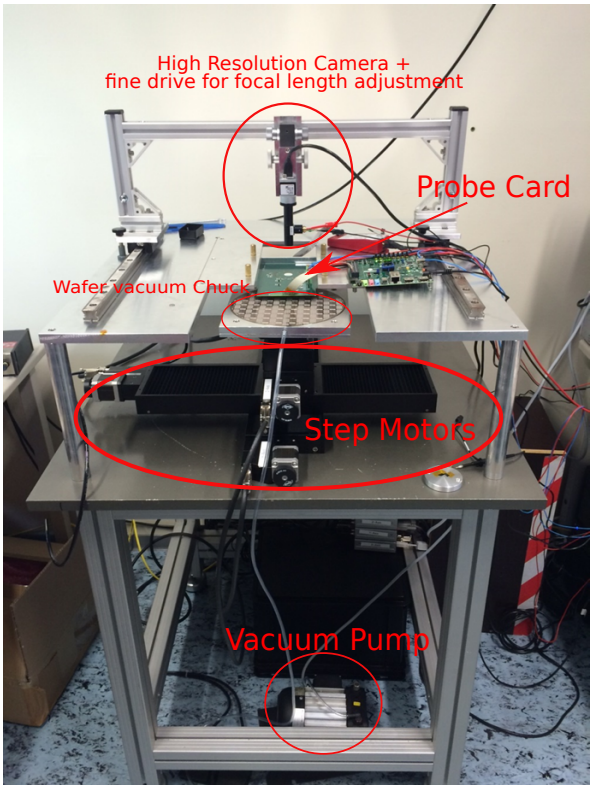

Figure 5.16: Photograph of the DCD and Switcher chip probe station. The chip probe station can test all ASIC functionality like ID readout over JTAG, ADC characteristic readout, high-voltage channel test.

This procedure can be automatized or carried out in manual mode. One single, one column and all 256 ADCs can be tested simultaneously. The next important test procedure for the Switcher ASIC has the following steps:

1 Contact Switcher with the needles

2 Used voltages: VDD = 1.8 V, $V_{gateHigh}$=3.3 V, $V_{clearHigh}$ = 3.3 V other voltages set to zero

3 The probe card contacts all clear and gate outputs - in total 64

4 The outputs are connected to a one 64 to 1 high-voltage multiplexer.

5 The output of the multiplexer is connected to a voltage divider and to a comparator, which has a variable threshold. The threshold should be 300 - 400 mV

6 The comparator output is measured by scope probe

7 JTAG ID test

8 Check VDD current consumption versus DAC value

9 Scope measurements of all 64 outputs. A counter in the FPGA is used to increment the channel number, all waveforms are checked

If a short occurs during the test procedure of all channels, then the test would fail.

5.5.1 Measurement Results

We have tested up to 1000 Switcher and DCDs comprising full functionality tests with the chip probe station. The measurements include full wafer and single chip tests. The DCD tests were performed characterizing ADC linearity, INL and noise. The application shows the characteristic curve and can calculate INL and noise. Figure 5.17 shows a chip probe station measurement of a single ADC of a DCD ASIC. The first graph on the top left shows the linearity or the ADC characteristic curve. The middle graph shows the ADC noise and the graph on the right illustrates the INL measurement. At the bottom of figure 5.17 the graphs show the mean of the noise measurement and the peak-to-peak mean of the INL measurement. The noise is 0.5 and the INL about 2 LSBs (figure 5.17). An eight ADC double column measurement is also shown in figure 5.18. Here all ADCs in one column are measured simultaneously. We have observed that the results are slightly worse for all 8 double columns. The reason is that all *VPSignal* currents are swept at the same time, therefore the current consumption changes and there is a variable voltage drop. This leads to instabilities during the measurement. The DCD readout speed was set to 150 MHz and all measurements are based on this frequency. The Switcher probe card can contact all clear and gate outputs - 64 in total. The outputs are connected to one 64 to 1 high-voltage multiplexer. A voltage divider senses the multiplexer output which is connected to a comparator. It has a variable threshold. An oscilloscope monitors the output. Measuring all 64 high-voltage outputs (all waveforms) requires a counter to increment the channel number. Figure 5.19 shows the clear and gate high-voltage output (high-voltage output was set to 3.3 V, so that the Switcher ASICs are not stressed too much during the tests).

Figure 5.20 shows an INL measurement for four DCDs in a working module (mass testing for the PXD module). Every cell in a column has its characteristic ADC curve and noise behavior. The INL is measured in every cell. Many cells show different INL values. Since all cells are not exactly the same, some can be nosier than others. A specific JTAG ID sent by the ASIC must match the reference ID: If the JTAG ID is missing the whole measurement will fail because the DCDs are configured over the JTAG chain.

Figure 5.21 and 5.22 show a DCD and switcher report of the measurements done with the chip probe station. This report summarizes all important functionality tests. The tests show a high yield for the UMC as well as the AMS process. Rarely have more than three ASIC failed through all tests. Difficulties were also experienced during contacting the needles with the ASIC. It was observed that the surface has oxidized. To remove the oxidized layer on the bottom of the needles, a special procedure guided by the application was executed. This procedure includes a short touchdown with a piece of test substrate and execution of a back and forth movement. After removing the oxide layer, it was possible to establish an ohmic contact between ASICs and needles. Slight deformation of the DCD

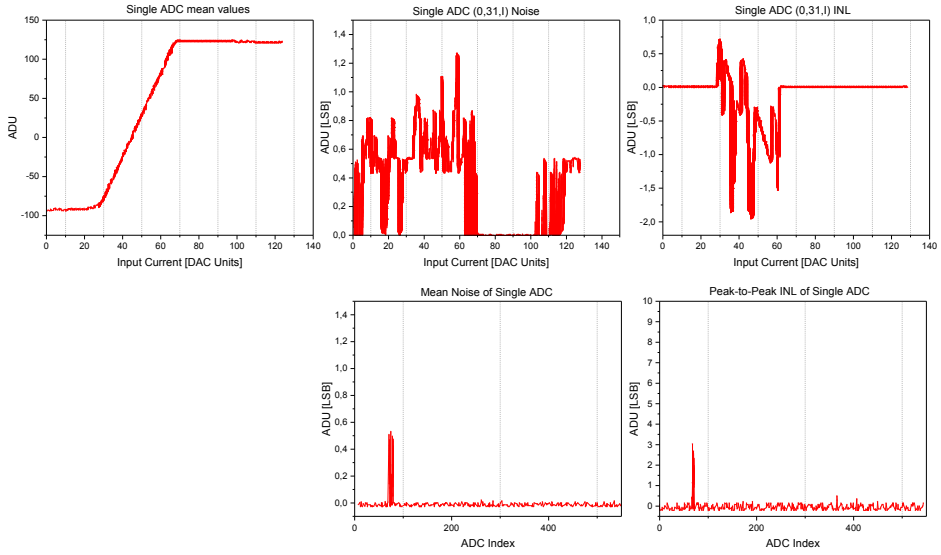

Figure 5.17: Probe station measurements: Single ADC characterization of linearity (top left), noise and INL (top right). Mean noise (bottom left) and the peak-to-peak mean (bottom right).

bumps was also observed by a small excessive pressure of the needles. These ASICs were removed as a precaution. For the Belle II detector, only those ASICs that have passed all tests successfully are used. After the test procedures, a microscope test was carried out to check damages on pads or bumps.

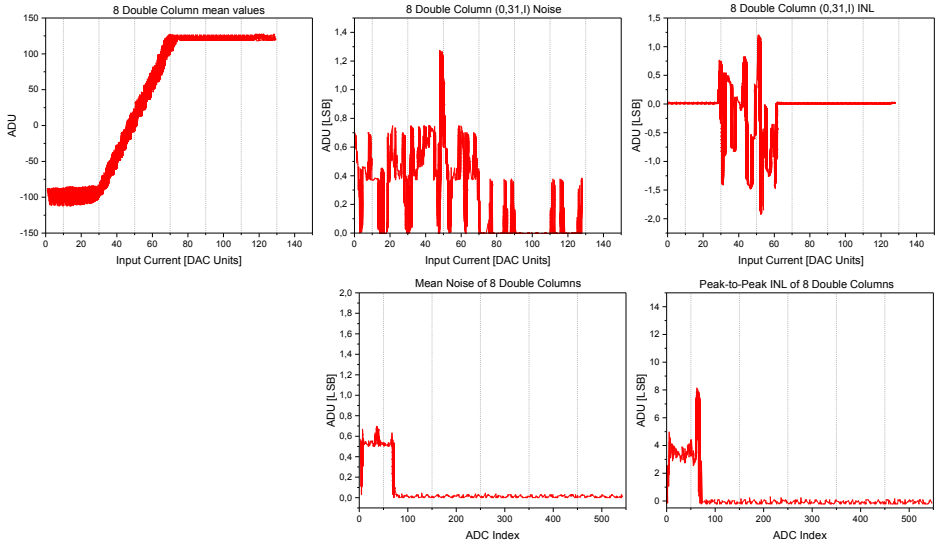

Figure 5.18: Probe station measurements: 8 double columns ADC characterization of linearity (top left), noise and INL (top right). Mean noise (bottom left) and the peak-to-peak mean (bottom right).

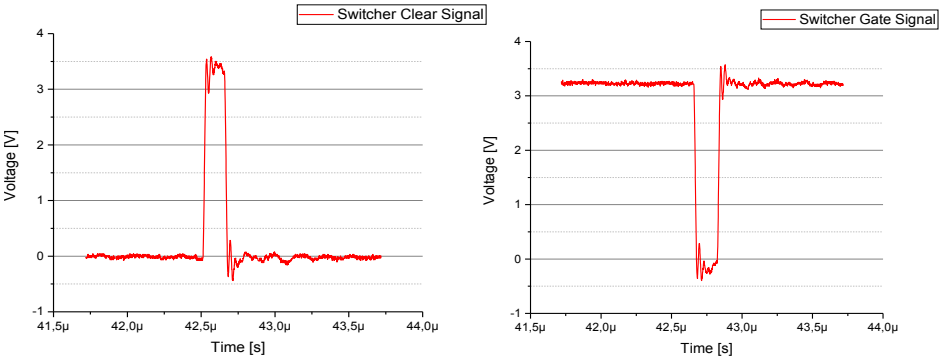

Figure 5.19: Clear and gate output of Switcher ASIC. High-voltage was set to 3.3 V to prevent electrical stress during the tests.

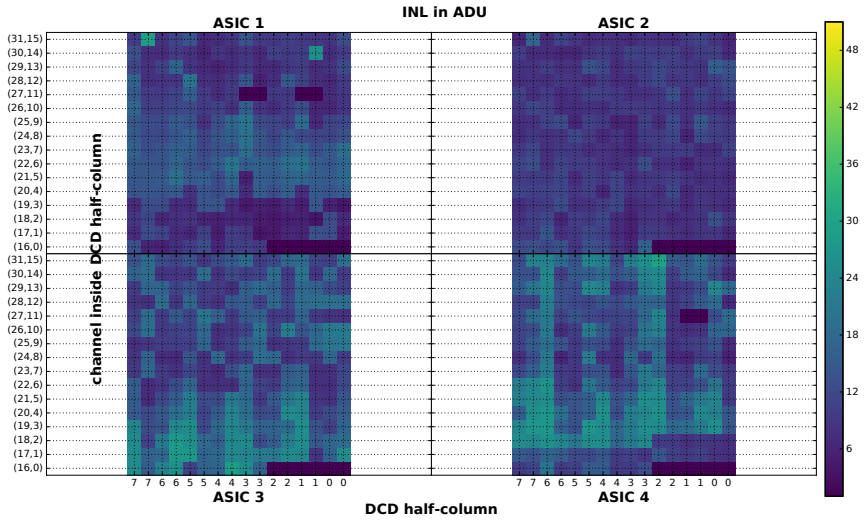

Figure 5.20: DCD INL characterization per cell: measurement for four DCDs (mass testing for PXD module construction). The y-direction shows the channels inside the DCD in a half-column and the x-direction shows the DCD half-columns [25].

Wafer D9KLK04-C4 121 DCD III Chips	Read JTAG ID	Write/read Test Pattern (ADCs)	Current Measurement	Noise Test (ADCs)
113	✓	✓	✓	✓
6	✓	✗	✓	✓
2	✗	✗	✗	✓

Wafer D9KLK05-E7 121 DCD III Chips	Read JTAG ID	Write/read Test Pattern (ADCs)	Current Measurement	Noise Test (ADCs)
111	✓	✓	✓	✓
9	✓	✗	✓	✓
1	✗	✗	✗	✓

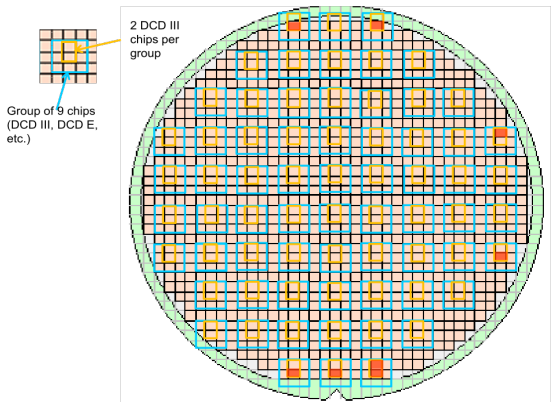

Figure 5.21: DCD wafer report. An entire wafer is composed of a group of 9 ASICs (DCDIII, DCDC etc.).

Switcher Chips	Read JTAG ID	Current Consumption (1.8V)	Current Consumption (HV)	64 HV-Channels	Bias Current	Boost Current
96	✓	✓	✓	✓	✓	✓
4	✓	✓	✓	✗*	✓	✓
0	✓	✗	✗	✗	✓	✓

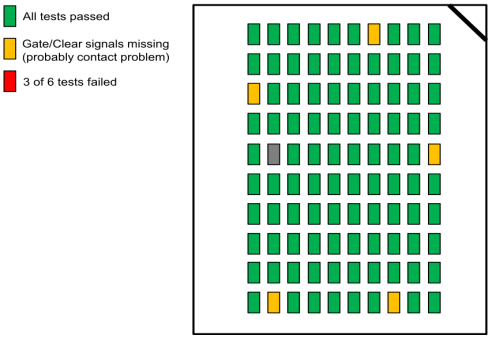

Figure 5.22: Switcher single chip test report. Rarely more than 3 Switchers failed 3 of 6 tests. All tested ASICs were provided in a chip gel pack.

6 USCT9C ASIC for Medical Application: 3D Ultrasound Computer Tomography

This chapter describes an ASIC that was developed specially for the 3D Ultrasound Computer Tomography (3D USCT) medical application. The architecture is similar to that of the NeutrinomassV3 ASIC described in chapter 3. It is based on a multi-channel ASIC with a receiver and transmitter amplifiers. We developed the receiver and transmitter ASIC using AMS 350 nm HV-CMOS process technology. The transmitter is composed of a high-voltage amplifier and the receiver is a three stage low-noise amplifier. It has a digital interface for communicating with a micro-controller. The bias block in this ASIC is similar and is a shared layout of NeutrinomassV3, H35-SPADV1 ASIC. The architecture of the low-noise amplifier has a more complicated structure when compared to the NeutrinomassV3 amplifier stages. It has more amplifier stages and an advanced feedback circuit. The development started in 2015 and we present the third version of the ASIC used for the 3D USCT project. An essential part of the third generation USCT III requires an ASIC that serves as a transmitter and receiver system. For USCT the advantages of using an ASIC is to reduce costs for a very high volume design and a smaller form factor. Discrete electronics are used for the USCT II generation. A discrete system has high form factors due to their large spatial requirements. Developing an ASIC in AMS 350 nm HV-CMOS is cheaper than the discrete component design.

3D USCT is based on the complex interaction of ultrasound with human tissue. While one transmitter is sending, all system receivers receive the emitted signal from the transmitter. The recorded signal is a reflection pulse. The reflection pulse has a small amplitude (in the order of a few µV) that needs to be amplified. The 3D USCT is composed of a patient bed with an embedded ellipsoidal measuring device covered by a 157 transducer array system (TAS). The main goal of 3D USCT is to detect tumors in the breast much earlier and with the same image resolution as systems already in use in medicine. Other high resolution 3D methods for breast cancer detection are either very expensive or use ionizing radiation. Examples for breast cancer detection: mammography based on low-energy X-rays, magnetic resonance imaging (MRI) and conventional ultrasound hand scanner. MRI is based on a strong magnetic field because the resolution of the MRI depends mainly on the magnetic field strength. MRI devices currently used in clinical practice have a magnetic field strength of 1.5 Tesla, but are increasingly being replaced by higher magnetic field MRI up to 30 Tesla, for a conventional hand scanner, which is moved manually across the breast tissue. Depending on the frequency the resolution is very high. Furthermore, due to the high frequency the penetration depth of ultrasound into the breast is limited, as the attenuation into the tissue increases with increasing frequency [29].

The USCT9C architecture and important measurements are shown in this chapter. All measurements were done with the USCT9C ASIC. A new version of USCT9CV2 was

developed in 2018 and will also be presented in this work. Further measurements of the USCT9CV2 will be presented in a separate paper.

6.1 The Development of USCT at KIT

Breast cancer is one of the most common and fatal cancers for women. There are ≈ 1.6 million cases of breast cancer every year. Although the breast is not a vital organ for women, but metastases resulting from the primary tumor become mortal. The best method of treating breast cancer is while the tumor is small and has not developed metastasis. If the breast cancer is diagnosed in the early stages, patients can be cured and the cancer can be treated. The motivation of this project is to build a cancer detection system that is comparable to existing systems e.g. CT, MRI. In the project 3D USCT a new imaging method for the early detection of breast cancer is developed. It promises three-dimensional images of the breast with high spatial resolution. The aim is the detection of tumors with an average diameter of less than 5 mm to improve the survival probability of patients. The development of high-end electronics and software methods for USCT is a big challenge. The USCT idea is based on over two thousand small ultrasonic transducers mounted in a water-filled ellipsoidal reservoir. In total, 2041 transducer are mounted in a 157 transducer array. For each image of the breast, a large amount of raw data is stored and then used for the reconstruction of various 3D volumes.

The development of USCT at KIT began in year 2000. The first 3D USCT or 3D USCT I was put into operation in 2006. A that time, the electronics of USCT I were not designed for large amounts of data and a measurement took several hours. This was a big disadvantage. It occupied several GB of data and was evaluated on a client PC. Evaluations of many different phantoms have shown that the expectations have been achieved. A reflection image of a wire smaller than 0.1 mm could be reconstructed with USCT I. The 3D USCT is composed of three staggered segments, which are 16 TAS. Each TAS has ultrasonic emitters or transmitters organized in a vertical row and four ultrasonic receivers assigned to one transmitter - in total eight transmitters and 32 receivers. There are 16 equally distributed and vertically installed TAS [42]. In addition, there is a DC motor that allows the aperture to be rotated to six defined positions. The total achievable measurements are up to 3.6 million [29] .

Further development was carried out on the second prototype, 3D USCT II (see figure 6.1). The 3D USCT II was developed for use in clinical studies. An optimized aperture shape as well as an improved spatial arrangement of the TAS was investigated (see ref. [80]). Furthermore, the optimization of the region of interest (ROI) and improvement of contrast with respect to tomography. The resulting aperture has a semi-ellipsoidal shape. Each of the TAS has four transmitters and nine receivers arranged in a square-like structure. The semi-ellipsoid can be rotated around the vertical axis and DC step motors are used to move it to the required positions. This results in more measurements for USCT II in comparison to USCT I. The same measurement duration of a few hours was also used as for the USCT I. This also corresponds to data volume of few tens of gigabytes. Improvements in the discrete electronic hardware design and in data compression have shown that more data can be stored in the internal memory of the DAQ system.

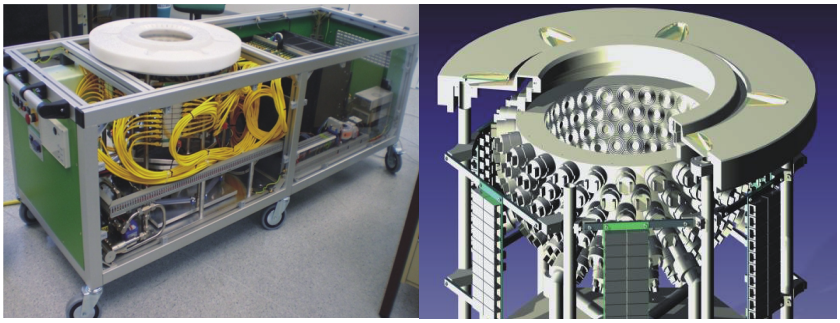

Figure 6.1: USCT II patient bed (left) and aperture (image reproduced from ref. [74]).

The measured USCT signals, so-called amplitude scans or A-scans, can be displayed in a transmission part and a reflection part. A transmitter emits an ultrasonic wave that propagates inside ellipsoidal water reservoir. The spherical wave hits on the way to spreader object, which in turn acts as point of source and emits a scattered ultrasound wave. In the receiver a pressure is produced and a pressure profile, called A-scan, is recorded over a certain period. In figure 6.2 two signals are highlighted (reflection signal

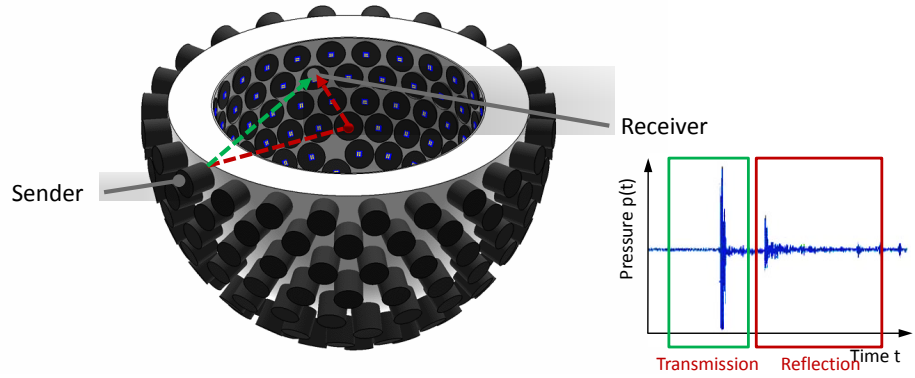

Figure 6.2: Image principle of 3D USCT. Left: TAS array and spreader (red ball). Right: transmission and reflection signals. These signals are a pressure (pressure profile) as function of time. Figure reproduced from ref. [3].

(red) and transmission signal (green). The first incoming pulse has taken the fastest path from transmitter to receiver, without to be reflected or scattered and is called transmission pulse. All delayed pulses are reflection pulses caused by the spreader object. The frequency range of ultrasound is between 20 kHz and 200 MHz. The basis of sending ultrasound and receiving ultrasound with a transducer is converting electrical energy into mechanical energy and back again. The transducers convert one form of energy to another. The ultrasound generated in a transducer is based on inverse piezoelectricity. The piezoelectric effect is used to transform the energy of an ultrasound wave into an electrical signal and

back again. Ultrasound generates a pressure which, in turn, generates electricity. This pressure is a mechanical quantity which deforms the piezo crystal. This was demonstrated in 1880 by French scientists [73]. The piezoelectric effect can be applied in two ways to generate an electrical signal or to produce ultrasound waves. Many materials, both natural and synthetic, exhibit piezoelectricity [90]. Figure 6.3 shows the inverse piezoelectric effect. The piezoelectric crystal vibrates to generate a sound wave when a voltage is applied.

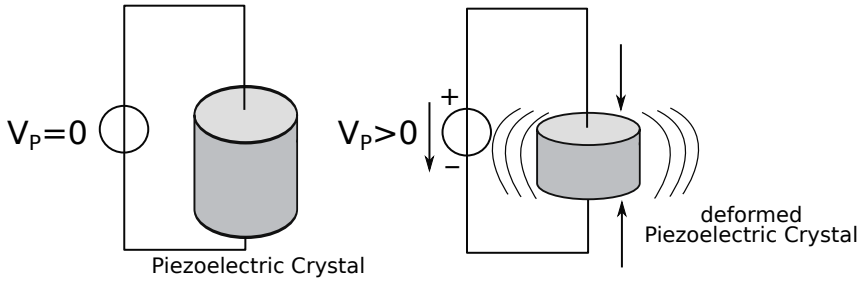

Figure 6.3: Left: No voltage applied on the piezoelectric crystal. Right: the same piezoelectric crystal after the application of the voltage V_P. A deformation results when voltage is applied and vibrates producing ultrasound waves (inverse piezoelectric effect).

6.1.1 Transducer Array System (TAS)

The TAS is composed of a matching layer with nine receiver transducers and four transmitter transducers with a backing layer (USCT II). A steel case or TAS casing contains PCBs and amplification electronics. For developing an imaging system with high spatial resolution the number of transducers has been increased (18 instead of 13 in USCT II) and the diameter has also been increased (figure 6.4). For USCT III the TAS system has a diameter of 43 mm (instead of 28 mm USCT II). In USCT II the emitter and receivers operate only in one direction. In USCT III the transducers are operated omnidirectional and the number of A-scans obtained per transducer quadruplicates [41]. In USCT II the existing TAS design includes a preamplifier as receiver and a step up transformer acting as an emitter. In addition, programmable voltage dividers, differential drivers are also used in the design. There are nine receiver transducers connected to a first stage of a receiver amplifier and a 3:1 multiplexer that has a second stage of amplifier. The first stage is a differential amplifier connected to a programmable voltage divider. The output of the programmable voltage divider is connected to the input of a current feedback amplifier used as a high-pass filter or as active filter. Further, the output is connected to ADC line driver (see figure 6.5). The emitter stage is composed of a step-up transformer, optocouplers producing a galvanic isolation to decouple the transducer after selecting one channel from the transformer stage. Figure 6.6 shows the schematic of the emitter stage. The challenge for the new TAS system is to ensure a high S/N ratio for the signal received and a high bandwidth. A large number of transducers are also required. The

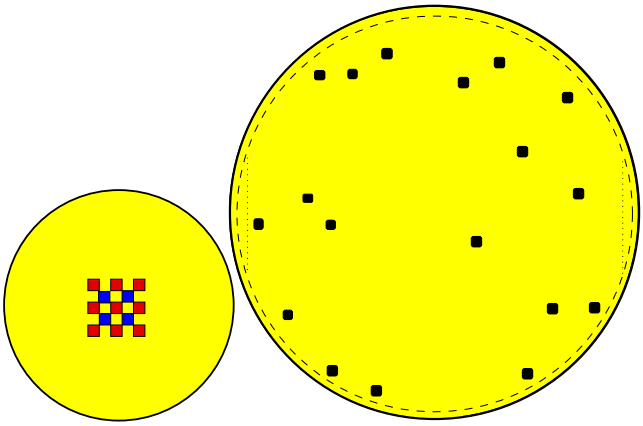

Figure 6.4: Distribution of transducer for USCT II and USCT III. The red dots are emitter and blue are the receivers working in one direction only. The TAS of USCT III has more transducers and diameter of 43 mm and the transducer are operated in omnidirectional mode (figure reproduced from ref. [41]).

idea is to integrate the discrete electronics into an ASIC which consumes low power, had high bandwidth and high S/N ratio and low crosstalk probability. In order to improve the contrast for USCT image reconstruction it was shown that the bandwidth plays an important role when using the synthetic aperture focusing technique (SAFT). The SAFT is an image reconstruction method. The contrast can be improved by choosing an appropriate frequency range between 0.25 MHz and 3.3 MHz. The emitted ultrasound has a center frequency of 2.5 MHz. The USCT9C ASIC expects a frequency bandwidth between 500 kHz and 5 MHz. The output bandwidth should be equal to the input signal bandwidth.

In this work, we will present an integrated solution using the USCT TAS electronics. Here two ASICs will be presented: USCT9C (figure 6.7) and USCT9CV2 ASICs. All ASICs were developed here at KIT-ADL. The USCTV1 ASIC (see figure 6.8) has only one channel consisting of one high-voltage channel (emitter stage replacing step up converter and optocouplers) and 2 stage low-noise amplifier (replacing the discrete amplifier electronics). USCT9C ASIC as second generation ASIC, features nine channels as required for the USCT III and using a 3:1 multiplexer. Furthermore, a digital block was implemented to program the ASIC over a serial peripheral interface (SPI) with existing micro-controller. Bias voltages and measurements can be executed with a micro-controller as interface between client PC and ASIC without the use of an FPGA board. Unfortunately during high-voltage measurements it was observed that the current strongly increased, thereby overheating the channels. We produced measurement results and ran the ASIC in save mode. The high-voltage channels were switched on and the input signal was applied for a short period of time. The high-voltage channels were then switched off again. The micro-controller controls the entire sequence. The USCT9CV2 will fix the bug in the analog part and a new software was developed to facilitate the measurement of the ASIC.

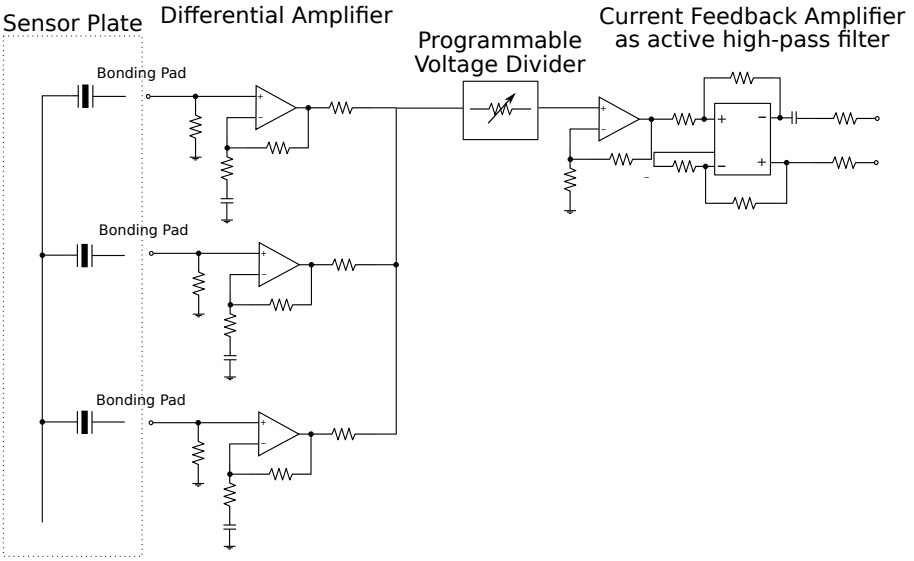

Figure 6.5: USCT II TAS receiver. This schematic shows only a part of the nine receivers. Sensor plate, differential amplifier, voltage divider and current feedback amplifier as active high-pass filter [63].

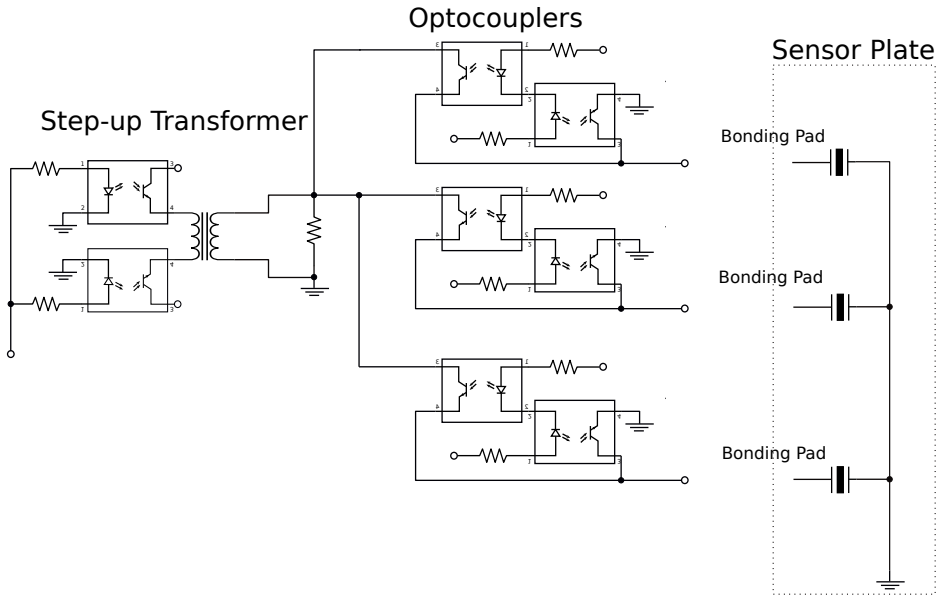

Figure 6.6: USCT II TAS emitter. This schematic shows the sensor plate, step up transformer, optocouplers [63].

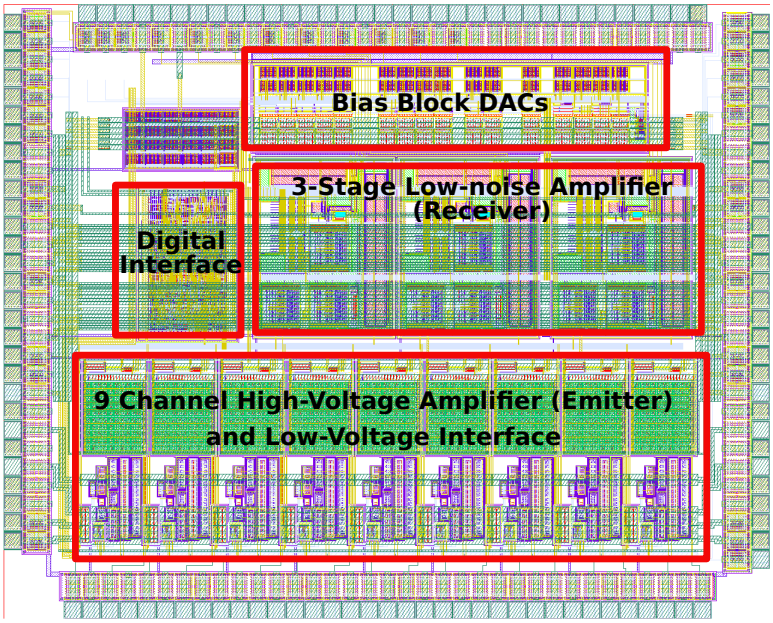

Figure 6.7: Layout of the USCT9C ASIC consists of bias block, 3-stage low-noise amplifiers, 9 channel high-voltage amplifier and low-voltage interface.

6.2 USCT9C ASIC Architecture

Development of the USCT ASIC started in 2015. We developed three generations of USCT ASIC in standard AMS 350 nm HV-CMOS technology. We carried out many experience with this technology. We gained the experience with further developments of NeutrinoMassV3 ASIC generations, H35-SPADV1 and USCTV1 ASIC. HV-CMOS transistors up to 120 V are provided by the HV-CMOS technology. The USCT9C ASIC (see figure 6.7) is composed of nine input channels with high-voltage emitter stage and three output channels consisting of a 3-stage low-noise amplifier receiver and a 3:1 multiplexer that is used to select one of the three input channels. The high-voltage amplifier is designed as an inverting amplifier and has a gain of 20. This high-voltage stage can generate signals up to 120 V amplitude. The rise time is about 120 ns. The gain produced by the low-noise amplifier first stage is around 40. The total gain is composed of all three stages ($G = G_1 \cdot G_2 \cdot G_3$). A bias current of about 100 µA flows between the $\pm$60 V and leads to a DC power consumption of 10 mW. Unused amplifiers are put into a sleep mode to reduce current consumption and crosstalk. In addition a digital interface is responsible for the configuration of the ASIC and controls the data stream between ASIC and micro-controller. The digital interface block was implemented in a prework. This block was separately implemented and later integrated in the top level of the ASIC layout. All measurements done with USCT9C will be presented as well as the USCT9CV2 simulations. As mentioned in the previous chapter, a problem in the channel of USCT9C was observed. In the second generation USCT9CV2

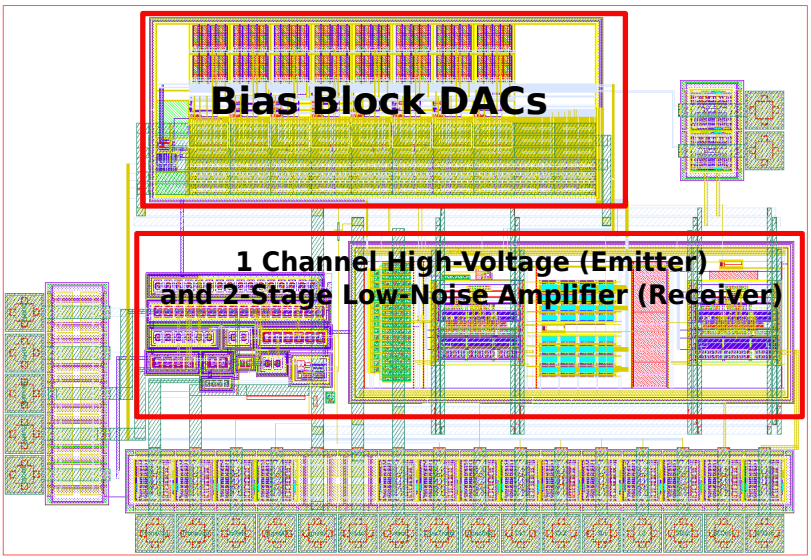

Figure 6.8: Layout of the USCTV1 ASIC consists of bias block, 1 channel of high-voltage and 2-stage low-noise amplifier.

the problem must be resolved with new dimensioning of the transistors. Furthermore, new simulations were done, and the bandwidth and S/N ratio were also improved. A peak detector was included as a new feature in the USCT9CV2 ASIC. The main characteristics of the USCT9C ASIC are reported in table 6.1. A simplified overview of the one input channel is depicted in figure 6.9. A piezoelectric sensor (transducer), emitter stage (high-voltage amplifier), tune capacitances and low-noise amplifier are connected to the output. Figure 6.10 shows the simplified multichannel block schematic with nine input channels and three output channels. Note, the following advantages can be described over a discrete component solution:

1 High bandwidth due to requirements of 3D USCT, i.e. transistors have low capacitance

2 Linear gain with negative feedback for gain control necessary requiring high signal accuracy and linear gain

3 Low power design to avoid heating of the transducer arrays

4 Small size design due to the integrated circuit

5 To reduce crosstalk the electronics is integrated on one chip

Technology	AMS 350 nm HV-CMOS
Chip size	3.3 mm x 4.1 mm
High-Voltage amplifier	Class AB amplifier (120 V transistors)
Low-noise amplifier	Inverting voltage amplifier, CR-RC shaper (3-stage)
Power consumption (high-voltage channel)	10 mW
Bias current	10 µA
Signal rise time	120 ns
3dB bandwidth	300 kHz - 6 MHz
Gain (high-voltage amplifier)	20
Gain (low-noise amplifier first stage)	40
Output related noise	121 µV
Input related noise	4 µV

Table 6.1: USCT9C ASIC main characteristics.

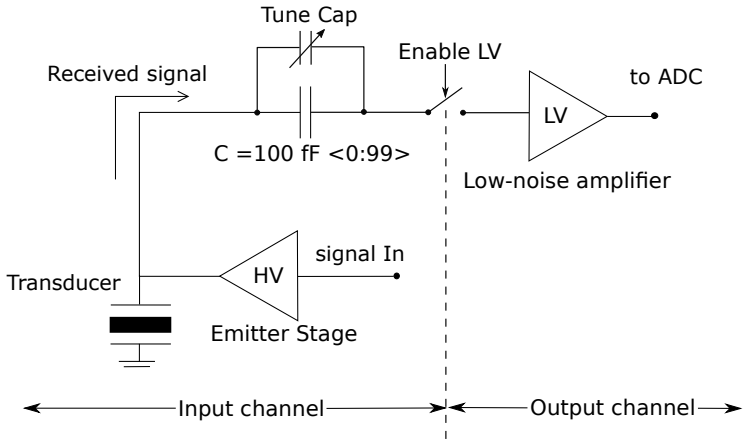

Figure 6.9: Simplified block schematic of the input channel composed of an emitter stage, low-noise interface and low-noise amplifier.

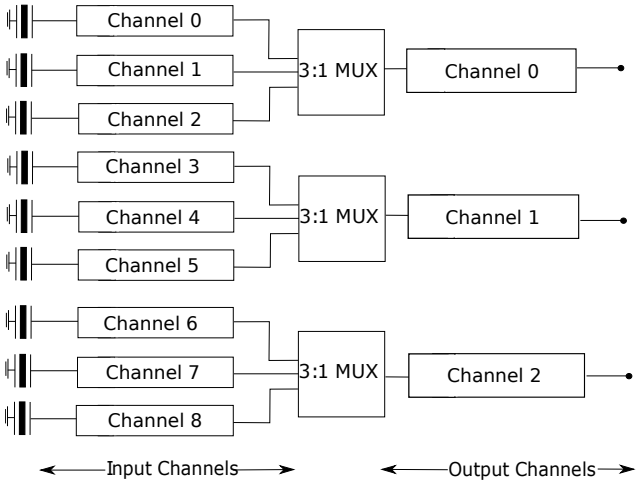

Figure 6.10: Simplified block schematic of the input channel.

6.2.1 High-Voltage Amplifier

The discrete step transformer was replaced by a transconductor and high-voltage class AB amplifier. The transistors used for the circuit design are 120 V NMOS and PMOS. The circuit has a differential transconductor as first stage. The first idea was to connect the transconductor to a basic push-pull stage (figure 6.11 b). A basic push-pull stage (figure 6.11 a) consists of NMOS transistor and current source. One disadvantage is that the circuit is slew rate limited. Transistor M5 charges the load capacitor (C_L) which allows the output voltage to track the input for rapidly increasing voltage. When the potential decreases the current flows out of C_L. Since the current source (I_0) is constant, the maximum output voltage change will be -I_0/C_L. It means that the fall time of the output signal depends on the slew rate of the capacitance C_L.

The second idea was to replace the current source with a PMOS transistor (class B output stage), a push-pull output stage can be implemented. The idea behind this is that M5 pushes the current to C_L and M6 pulls away from C_L (figure 6.11 c). Note that the load capacitance is connected to the source of the transistors; therefore, the transistors act as source followers. The class B output stage shows a distortion problem around Vin = 0. Between $|V_{thM6}| \leq V_{in} \leq V_{thM5}$ the class B stage turns off. In order to avoid the distortion an improved push-pull stage with two diode-connected transistors was introduced (third idea) (figure 6.11 d). The transistors M3 and M4 act as a voltage source. This shifts the operating points of the transistor M5 and M6 and the class B becomes a low distortion class AB amplifier. The class AB amplifier also improves the slew rate (shown in equation 6.1). The relaxation time can be defined approximately as:

$$\tau_{out} = 1/g_m \cdot C_L \tag{6.1}$$

where g_m is the transconductance of transistor M5 and M6.

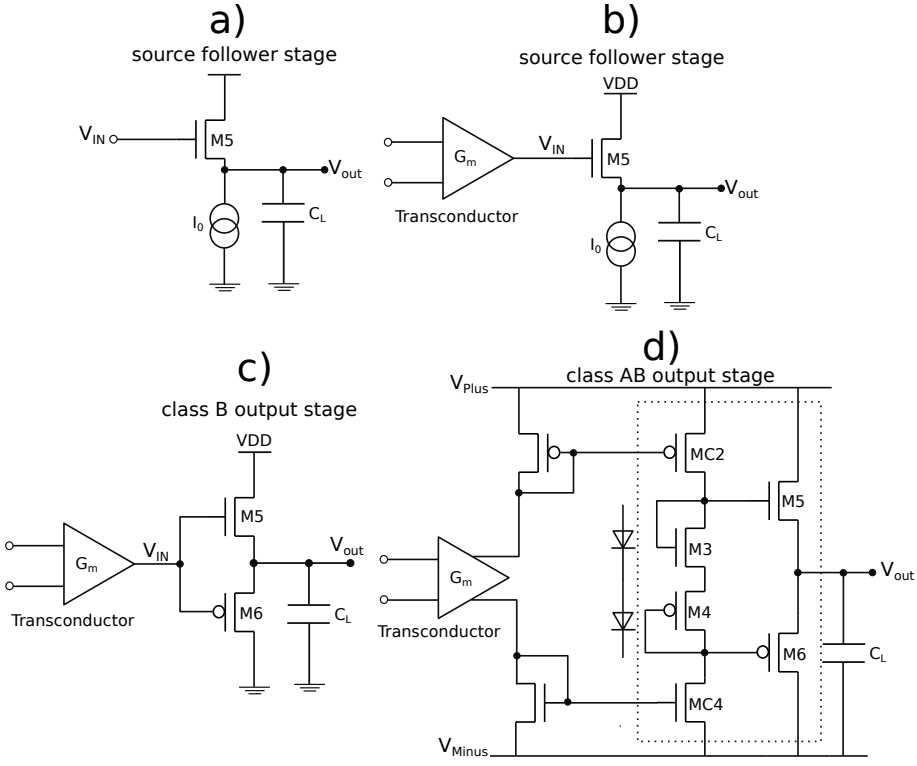

Figure 6.11: (a) Source follower circuit (slew rate limited), (b) source follower circuit (improved signal speed), (c) class B stage (push-pull circuit), (d) improved class B stage with two additional diode-connected transistors to prevent distortions. This circuit topology is also called class AB amplifier.

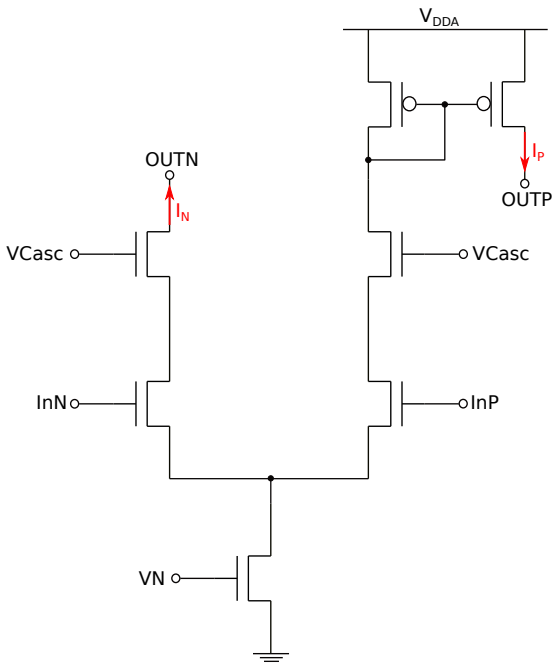

Figure 6.12: Simplified differential transconductor circuit.

The transducer is driven by this emitter or high-voltage amplifier. Generating a negative feedback allows a constant amplification (DC amplification), linear amplification and a linear response. The amplification is set in the order of 20. This amplification is set by two resistors - one in the feedback loop and one as input - resistor forming the ratio of R_{fb}/R_{in}. The high-voltage amplifier has three outputs. The output OUT is connected to R_{fb}. Thus, this allows a low voltage input signal leading to a high-voltage output. The differential transconductor does voltage-to-current conversion. It is implemented as a differential amplifier with two inputs I_{nN}, I_{nP} and two outputs OUTN, OUTP (see figure 6.12). The generated outputs of the differential amplifiers are currents; the difference in their values is proportional to the difference of the input voltages. Output current I_P flows into the transconductor amplifier and current I_N out of the transconductor block to the high-voltage transistor M2 (figure 6.13). Transistors M1 and M2 are connected to the transconductor outputs corresponding to P (OUTP) and N (OUTN) part. M1 is realized as NMOS and M2 as a PMOS transistor. The gate of M1 is connected to the low-voltage power supply (V_{DD}) and the gate of M2 is connected to ground (GND). For AMS 350 nm the low-voltage power is 3.3 V. M1 and is able to sustain a high-voltage between its drain and the sustained voltage up to 60 V. In the other direction, the PMOS transistor M2 is able to sustain negative high-voltage up to -60 V. Both drains of M1 and M2 are the so-called high-voltage drains. The high-voltage drain of M1 is fed to *current mirror 2*. It copies the current with a ratio 1:1 and the same applies for *current mirror 1*. *Current mirror*

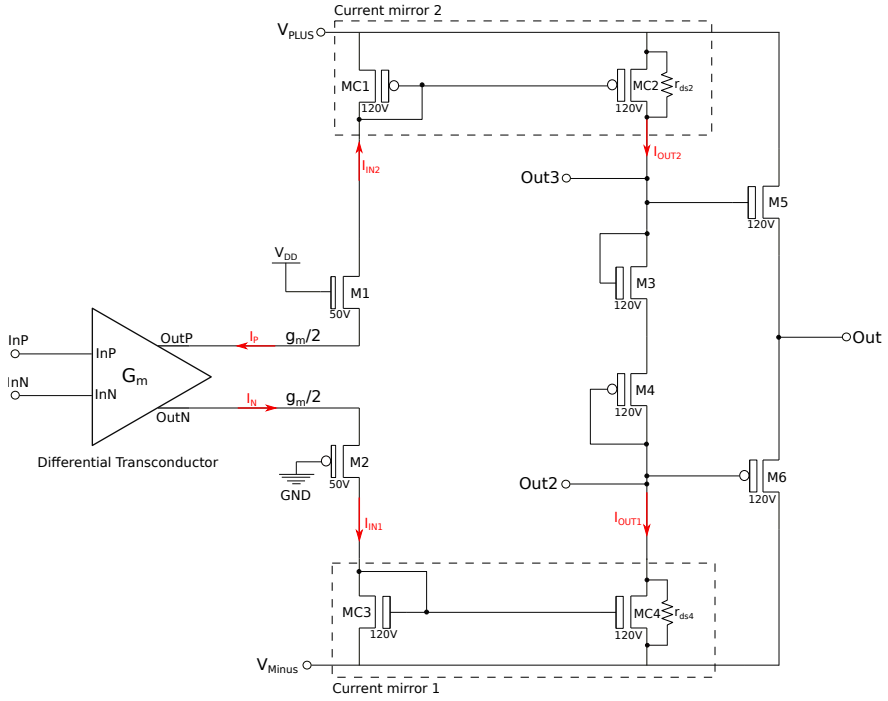

Figure 6.13: Circuit schematic of the high-voltage amplifier.

2 has an input I_{IN2}, output I_{OUT2} is supplied by V_{Plus}. V_{Plus} corresponds to the positive high-voltage power rail 60 V. Due to the 1:1 ratio of W/L geometry to the current, this can be calculated as following

$$I_{OUT2} = \frac{(W/L)_{MC1=MC2}}{(W/L)_{MC2}} \cdot I_{IN2} + r_{ds2}(V_{plus} - V_{out3}) \tag{6.2}$$

the current flowing out (I_{IN2}) is nearly equal to the current flowing out (I_{OUT2}). *Current mirror 2* is able to sustain a high-voltage of 120 V between V_{Plus} and Out. For the N part of the schematic M2 is connected to *current mirror 1*. This mirror is realized with NMOS transistors. We can also assume that a 1:1 ratio is realized for W/L geometry resulting in a copy of the current

$$I_{OUT1} = \frac{(W/L)_{MC3=MC4}}{(W/L)_{MC4}} \cdot I_{IN1} + r_{ds4}(V_{out2} - V_{minus}) \tag{6.3}$$

V_{Minus} corresponds to negative high-voltage -60 V. Also the *current mirror 1* can sustain high-voltage of 120 V between Out and V_{Minus}. Therefore, an overall output swing of 120 V can be reached. Outputs of the current mirror 1, 2 are connected in series with transistors M3 and M4. M3 is an NMOS and M4 a PMOS transistor. Both transistors are

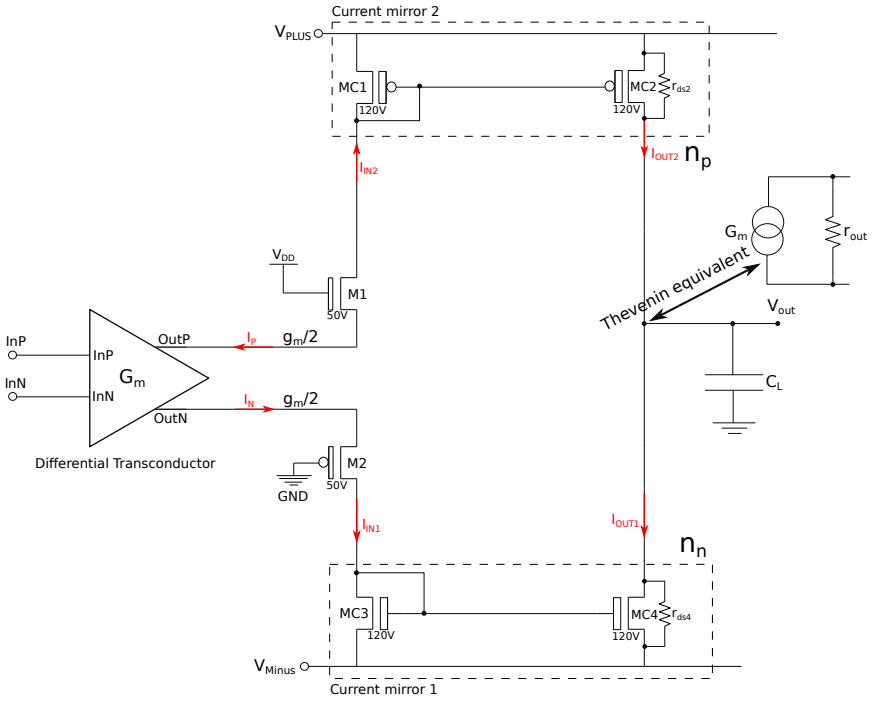

Figure 6.14: Equivalent circuit to calculate the open-loop gain.

diode-connected and saturated. The input resistance of these two transistors is given by $R_{in3} = 1/g_{m3}||r_{ds3}$ and $R_{in4} = 1/g_{m4}||r_{ds4}$ (M3 and M4 assure that M5 and M6 are always in strong inversion). Out2 and Out3 are connected to M5 and M6. The transistors M5 and M6 act as source followers, their purpose is to make the amplifier faster. They form the output power stage of the amplifier. In the case of unloaded output (Out) this output port is equal to the potential between the source of M3 and M4. Using the Thevenin theorem the open loop gain is defined as follows:

$$A_{OL} = G_m \cdot r_{out} \tag{6.4}$$

where G_m is the transconductance of the transconductor or the equivalent current source in the Thevenin theorem (figure 6.14). The G_m can be expressed as follows:

$$G_m = \frac{g_m}{2} \cdot (n_p + n_n) \tag{6.5}$$

where n_n and n_p is the current gain of current mirrors 1 and 2. The open loop gain can be calculated as follows:

$$A_{OL} = \frac{V_{out}}{(V_{inP} - V_{inN})} = (r_{ds2}||r_{ds4}) \cdot G_m = (r_{ds2}||r_{ds4}) \cdot \frac{g_m}{2} \cdot (n_p + n_n) \tag{6.6}$$

M5 and M6 can also be understood as an output stage. A load with a high capacitance can be connected without influencing the resulting circuit gain. If the potential of Out follows the rise of Out3, the V_{gs} of M5 will also increase and generate more current than in the unloaded state. This current serves to increase the change speed of Out. This results in fast change of node Out and the amplifier reaching rise times of up to 100 ns.

A piezoelectric sensor is connected as a load for this amplifier. The high-voltage produced by this amplifier drives the piezoelectric sensor to generate ultrasound waves in water. A feedback is applied to produce an inverting amplifier. Figure 6.15 shows an inverting amplifier with three outputs. Two out of them are fed to the feedback. Out3 and Out2 each have a feedback capacitance (C_{fb}) to increase the stability of the feedback.

6.2.2 Simulations of the High-Voltage Amplifier

From here the simulations of the high-voltage amplifier are considered. A lot of time was spent to tune all high-voltage transistors until the expected behavior was obtained. The bugs in USCT9C were removed and USCT9CV2 was redesigned and all simulations repeated. A chirp signal obtained from the USCT team was used to stimulate the high-voltage amplifier. The high-voltage signal input is set to receive analog high-voltage signals, which can be provided by an external high-voltage transmitter. The high-voltage signal used in the simulation can be specified as alternating voltage of varied amplitude, the frequency of which can periodically and continuously pass a specified range, like the chirp signal. The chirp signal varies whereby the frequency increase or decreases with time. A linear chirp means the frequency varies linearly with time, so the frequency response of this signal has a constant amplitude for the bandwidth of the chirp. The high-voltage

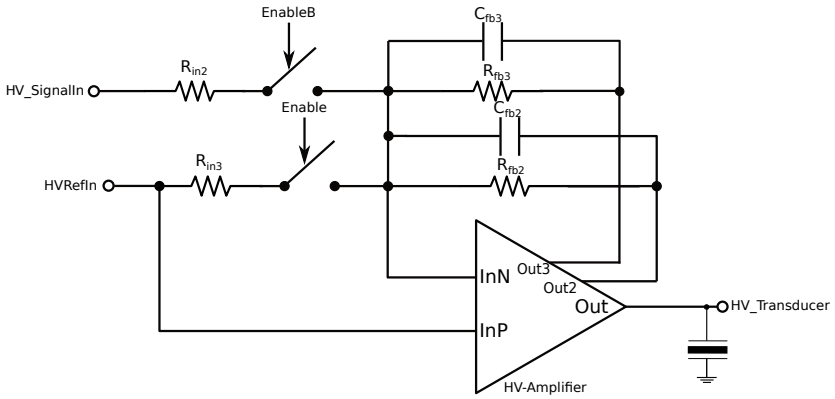

Figure 6.15: High-voltage amplifier implemented as an inverting amplifier with two feedbacks to improve the stability.

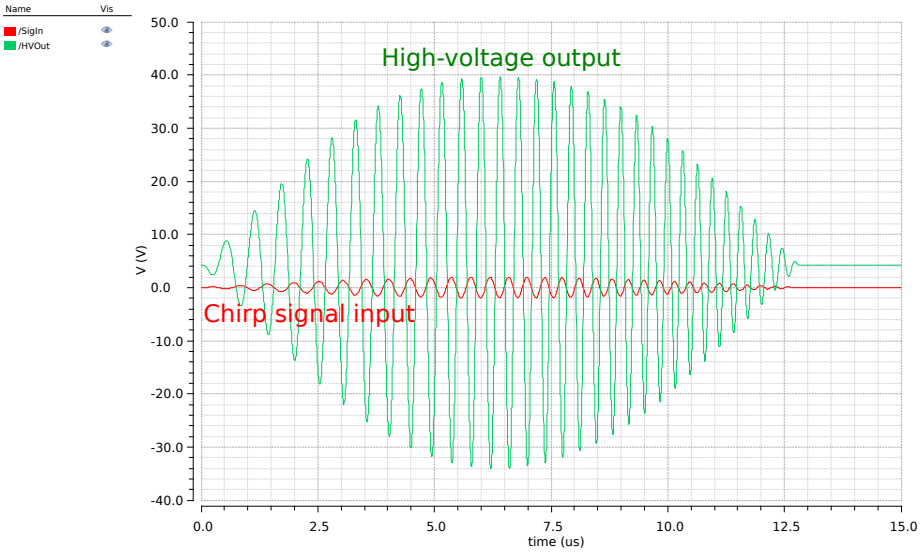

Figure 6.16: Simulation of high-voltage amplifier with chirp signal input (maximum peak-to-peak amplitude 3.8V). The high-voltage output swing reaches 71.5 V. A simulated gain of 18.8 was achieved.

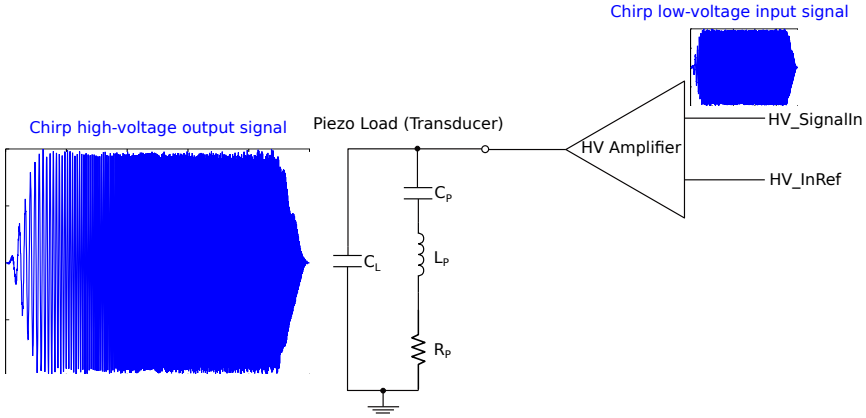

Figure 6.17: Simulation of the high-voltage amplifier with a piezoelectrical load forming LCR network.

block was simulated with the redesigned bias block. In the simulation (figure 6.16) a high-voltage of ±48 V was chosen. The total voltage output swing of 96 V can be reached. The maximum peak-to-peak amplitude of the input chirp signal is 3.8 V which, when amplified with the gain factor 20 (DC feedback of the resistors), gives a high-voltage of 76 V. In the simulation a total output voltage of 71.5 V for a certain bias configuration and voltage gain of 18.8 was achieved. Note that the bias setting for the high-voltage amplifier needs VNHV, TranCasc, HV_Offset setting. For measurements it is recommended that a parameter sweep be performed over all bias parameters to achieve maximum gain and highest linearity. Another signal type was also simulated which is based on rectangular pulse shape with a pulse period of 500 ns, a pulse width of 250 ns and an amplitude of ±2 V. This followed by a delay of 2.25 µs. After the delay time another rectangular pulse is generated with a 10 times smaller amplitude. This signal type was used to test the new peak detector circuit. The simulation of the peak detector is not part of this work.

An equivalent circuit of the piezoelectric sensor was used as load for the high-voltage amplifier. It contains an LCR series network and three parallel capacitances forming the capacitive load (C_L). It can be handled as a series resonant circuit. The piezo load is shown in figure 6.17. The LCR values of the piezo sensor were measured using a LCR meter. Following values were measured: L_P = 6 mH, C_P = 700 fF, R_P= 13 kΩ. In this simulation the high-voltage amplifier acts as the emitter.

6.2.3 Low-Voltage Amplifier Stage

In the USCT9C ASIC there are nine channels composed of high-voltage amplifiers and three 3-stage low-noise amplifiers. To increase the gain, the low-voltage amplifiers are connected in series. Each of these channels are connected to the input capacitors and a selection logic consisting of a NAND gatter and a switch. All the switches are designed as transmission gates. The ultrasonic signal received from the transducer (piezoelectric sensor) has low

signals in order of 1 µV to 100 µV. Only one of three input receivers can be selected. A gain 40 is produced in the first stage where a inverting voltage amplifier with capacitive loads was chosen to achieve low-noise. The second and third stages are composed of CR-RC shaper topology. To achieve the required bandwidth the amplifier feedback was designed with selectable resistors and tune capacitances. In the first stage a 100 fF capacitance was chosen and the second stage has a selectable 32 kΩ resistor. Furthermore, the gain of the second stage is dependent on the input capacitor. It can be tuned using a shaper tune capacitance. These tune capacitances can all be connected in parallel to increase the gain factor. In the third stage input capacitances can also be tuned. The output of stages 2 and 3 are fed into the output selection logic. It has a combined output pin with an option to bypass the third stage. The idea behind this is to measure the noise with and without the third stage. A test pattern can be injected for the purposes of channel identification.

As core amplifier for the inverting voltage amplifier and shaper a folded cascode topology was developed for the ASIC implementation. A folded cascode topology is shown in figure 6.18. This folded cascode structure is a single-ended topology with a PMOS

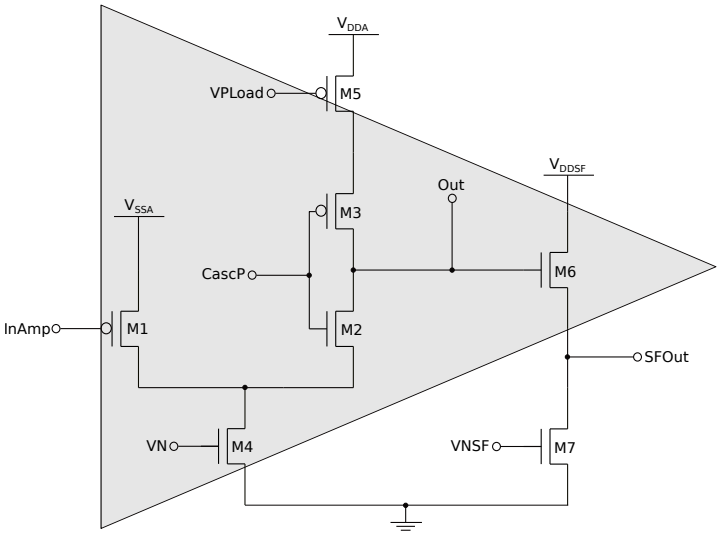

Figure 6.18: Folded cascode amplifier topology with PMOS input.

transistor (M1) as an input device. Two transistors form the cascode transistors with an NMOS-PMOS structure (M2 and M3). The current generated by transistor M4 is responsible for the biasing of M1. The transistor M5 is responsible for biasing of M2 and M3. A second bias current is generated by a PMOS transistor (M5) and acts as load resistance to increase the voltage gain. The amplifier circuit has two outputs (Out and SFOut). One output refers directly to the cascode output and the second output is a source follower output (output of folded cascode is fed to the source follower M6 and M7). In order to calculate the voltage gain and output resistance a simplified structure of the folded cascode is considered (see figure 6.19). Generally, the idea of a cascode structure is to convert the

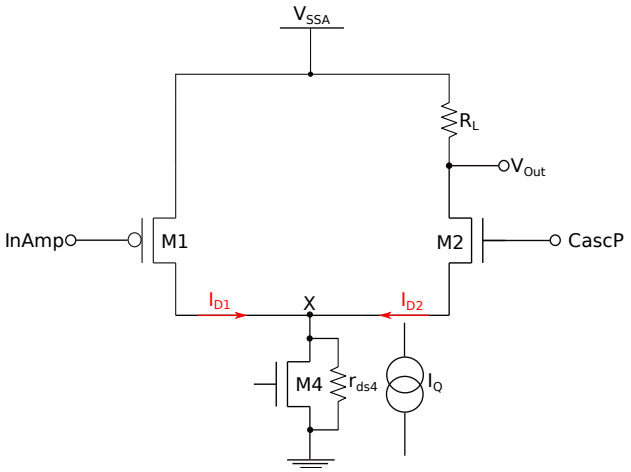

Figure 6.19: Folded cascode amplifier with proper biasing.

input voltage to a current and feed the result to a common-gate stage. In order to bias M1 and M2 a current must be added. The drain currents of M1 and M2 are at node X: $I_{D1} + I_{D2} = I_Q$. If InAmp becomes more positive, drain current of M1 decreases. I_{D1} forces I_{D2} to increase and Out drops. The stage in figure 6.19 is called folded cascode because the small-signal current is folded up. The large-signal characteristics of a folded cascode can be described as follows: if InAmp decreases from V_{ssa} to zero, InAmp > $V_{ssa} - |V_{th1}|$, M1 is off and M2 carries all the current of current source I_Q, where $V_{out} = V_{ssa} - I_Q \cdot R_L$. In case of InAmp < $V_{ssa} - |V_{th1}|$, M1 turns on and is in saturation region where the saturation current is given by

$$I_{D2} = I_Q - \frac{1}{2}\mu_p C_{ox}\left(\frac{W}{L}\right)_{M1}(V_{ssa} - InAmp - |V_{th1}|)^2 \tag{6.7}$$

Current I_{D2} continues falling to zero if $I_{D1} = I_Q$. This occurs at InAmp = V_{inM1} if

$$I_Q = \frac{1}{2}\mu_p C_{ox}\left(\frac{W}{L}\right)_{M1}(V_{ssa} - V_{in1} - V_{th1})^2 \tag{6.8}$$

Thus, V_{in1} can be resolved as follows

$$V_{inM1} = V_{ssa} - \sqrt{\frac{2I_Q}{\mu_p C_{ox}(W/L)_{M1}}} - |V_{th1}| \tag{6.9}$$

In figure 6.20 the large-signal analysis is plotted for a folded cascode. Using a telescopic cascode (NMOS-NMOS cascode) the voltage gain and output impedance can be obtained. The same analysis is obtained for a folded cascode with only minor modifications. For small-signal analysis all cascode voltages are set to ground. Figure 6.21 shows the telescopic

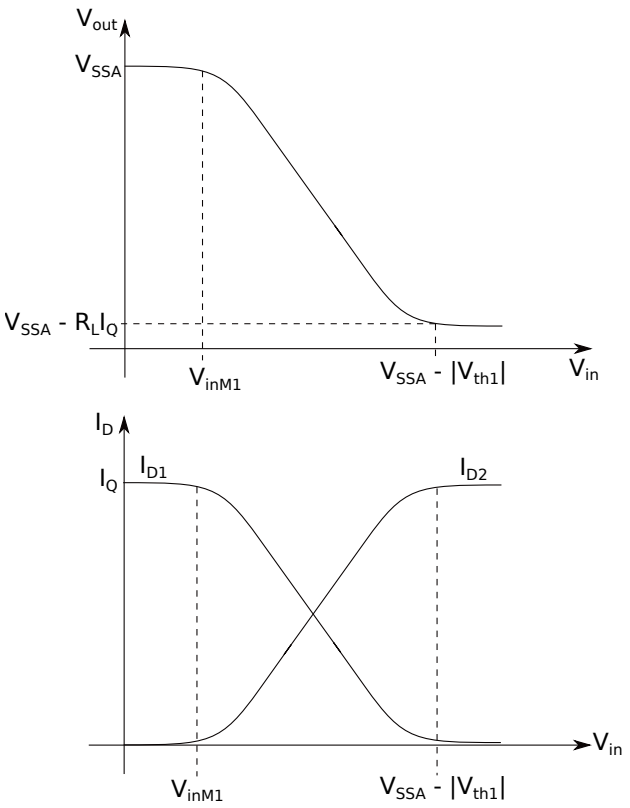

Figure 6.20: Large-signal characteristics of a folded cascode. If V_{in} falls below V_{inM1} the current I_{D1} enters in the triode region so that $I_{D1} = I_Q$. For $V_{in} > V_{SSA} - |V_{th1}|$ the current tends to the value of I_Q.

cascode topology and a small-signal equivalent circuit of the telescopic cascode amplifier. Looking into the output resistance of transistor M2 is approximately given by

$$r_{dM2} = g_{m2}(r_{ds1}||r_{ds4})r_{ds2} \tag{6.10}$$

The total output resistance can be calculated as

$$R_{out} = r_{dM2}||R_L \tag{6.11}$$

where R_L is the load of the folded cascode amplifier. Looking into the common-gate transistor M2 a input resistance (see figure 6.21) for low-frequencies can be described

$$r_{inM2} = \frac{r_{ds2} + R_L}{1 + g_{m2}r_{ds2}} \tag{6.12}$$

A common-source amplifier with a load resistance r_{inM2} is used from the input to the source of M2. The gain can be calculated as follows

$$A_1 = \frac{V_x}{V_{in}} = -g_{m1}(r_{ds1}||r_{inM2}) = -\frac{g_{m1}}{g_{ds1} + g_{inM2}} \tag{6.13}$$

The gain is derived for the common-gate stage from the source of M2 to the output

$$A_2 = \frac{V_{out}}{V_x} = g_{m2}(r_{ds2}||R_L) \tag{6.14}$$

A total gain of the telescopic cascode is the product of A_1 and A_2

$$A_t = A_1 \cdot A_2 = -g_{m1}g_{m2}(r_{ds1}||r_{inM2})(r_{ds2}||R_L) \tag{6.15}$$

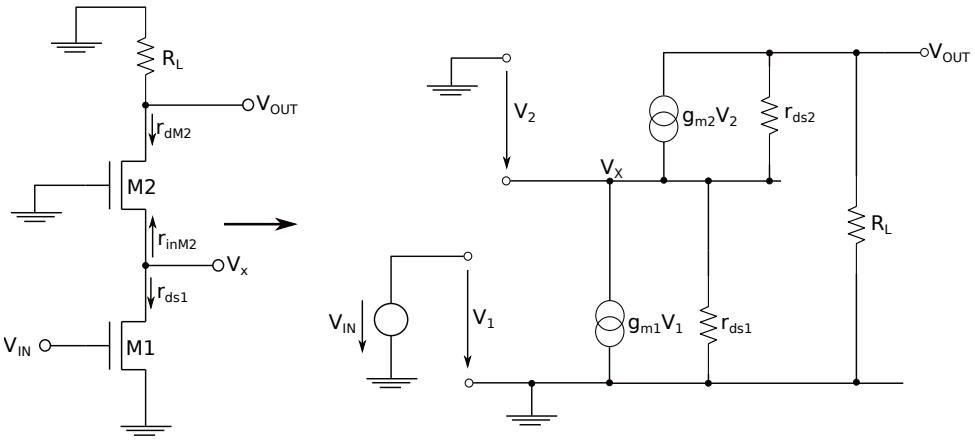

Figure 6.21: Telescopic cascode amplifier and small-signal equivalent circuit.

A simplified block schematic of the 3-stage low-noise amplifier composed of inverting voltage amplifier and CR-RC shapers is shown in figure 6.22. As already mentioned the folded cascode is used as core amplifier for all three stages. The output of the inverting voltage amplifier is connected to high-pass filtering capacitance. Calculating the transfer

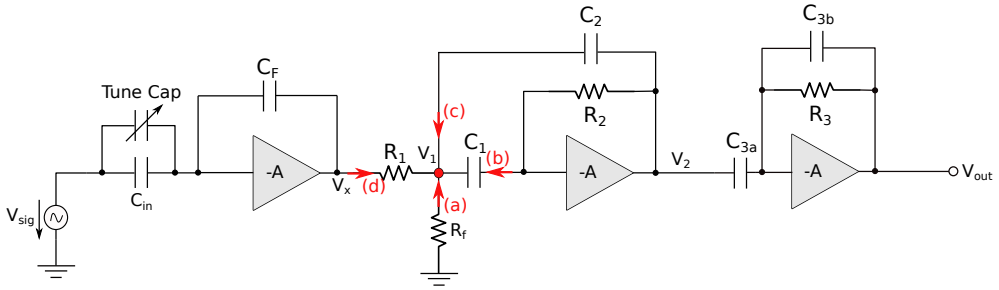

Figure 6.22: Simplified block schematic of USCT three stage low noise amplifier.

function for the real shaper implementation using KCL, we obtain:

$$0 = (a) : \frac{V_1}{R_f} + (b) : V_1 sC_1 + (c) : (V_1 sC_2 + V_1 s^2 C_1 C_2 R_2) + (d) : \frac{V_1 - V_x}{R_1}$$

$$\frac{V_x}{V_1} = s^2 C_1 C_2 R_1 R_2 + sR_1(C_1 + C_2) + \frac{R_1 + R_f}{R_f}$$

$$\frac{V_1}{V_x} = \frac{1}{s^2 C_1 C_2 R_1 R_2 + sR_1(C_1 + C_2) + \frac{R_1 + R_f}{R_f}}$$

$$\rightarrow \frac{R_f/(R_1 + R_f)}{1 + s^2 C_1 C_2 R_1 R_2 \frac{R_f}{R_1 + R_f} + sR_1(C_1 + C_2)\frac{R_f}{R_1 + R_f}}$$

$$\frac{V_1}{V_x} = \frac{R_f/(R_1 + R_f)}{1 + s^2 C_1 C_2 R_2 (R_1 || R_f) + s(R_1 || R_f)(C_1 + C_2))} \tag{6.16}$$

Now we can calculate the total transfer function of the two-stage amplifier

$$H(s) = \frac{V_2(s)}{V_1(s)} = sC_1 R_2 \tag{6.17}$$

$$\rightarrow \frac{V_2(s)}{V_x(s)} = \frac{sC_1 R_2 \frac{R_f}{R_1 + R_f}}{s^2 C_1 C_2 R_2 (R_1 || R_f) + s(R_1 || R_f)(C_1 + C_2)) + 1} \tag{6.18}$$

The transfer function of the third stage is simple because the series resistance is missing and is comparable to the transfer function of the second stage of TRISTANs amplifier (see 3.12). Thus, we get

$$G(s) = \frac{V_{out}}{V_2} = \frac{R_3 sC_{3a}}{(1 + sR_3 C_{3B})} \frac{sC_1 R_2 \frac{R_f}{R_1 + R_f}}{s^2 C_1 C_2 R_2 (R_1 || R_f) + s(R_1 || R_f)(C_1 + C_2)) + 1} \tag{6.19}$$

Calculating the poles of the denominator a second order equation is considered:

$$s^2 C_1 C_2 R_2 (R_1||R_f) + s(R_1||R_f)(C_1 + C_2)) + 1 = 0$$
$$s^2 + s \underbrace{\frac{C_1 + C_2}{R_2 C_1 C_2}}_{a} + \underbrace{\frac{1}{C_1 C_2 (R_1||R_f) R_2}}_{b} = 0 \tag{6.20}$$

Considering a second order transfer function this equation can be solved if $a^2 - 4b < 0$:

$$p_{1,2} = \mathrm{Re}(p) \pm j\,\mathrm{Im}(p)$$
$$\mathrm{Re}(p) = \frac{a}{2} \tag{6.21}$$
$$\mathrm{Im}(p) = \frac{1}{2}\sqrt{4b - a^2} \tag{6.22}$$

Inserting coefficients a and b of the transfer function the complex conjugate poles are written as:

$$p_{1,2} = \frac{1}{2}\left(-\frac{C_1 + C_2}{R_2 C_1 C_2} \pm j\sqrt{\frac{4}{C_1 C_2 (R_1||R_f)} - \frac{C_1 + C_2}{C_1 C_2 R_2}}\right) \tag{6.23}$$

Nominator zero of the transfer function is $s_z = 0$. Further, a third pole is defined at $p_3 = -\frac{1}{R_3 C_{3B}}$.

6.2.4 Simulations of the 3-Stage Low-Noise Amplifier

In this simulation part not only the 3-stage low-noise channel was simulated. The full transducer stage had to be simulated because the high-voltage channel affects the low-noise path with an additional resistance and therefore also has to be taken into consideration. The simulation schematic contains: bias block, high-voltage and low-noise amplifier blocks. The transducer is now switched to receive mode (low-voltage amplify, emitting mode off) whereby the high-voltage input is set to ground and the low-voltage path is activated through a select signal using the selection logic composed of NAND gatter and a switch which is controlled by the enable signal generated from the selection logic. All three low-noise stages are appropriately biased to achieve high gain and high linearity. Shaper tune resistances and capacitances are provided to tune bandwidth and noise. Depending on gain, speed, bandwidth a parameter set of the tune resistance and capacitance was chosen. The 3-stage low-noise amplifiers can be switched in the wide-band and narrow-band mode. The narrow-band mode has a maximum amplification of the resonance frequency of RLC transducer load. In addition, at a center frequency of 2.5 MHz a maximum amplification is expected. To achieve this, two NMOS transistors switch the amplifier to narrow-band mode. In the wide-band mode, the amplifier stages are designed to amplify signals starting from 100 kHz to 48 MHz at gain of 70 dB. The required frequency bandwidth between 500 kHz and 6 MHz is fulfilled.

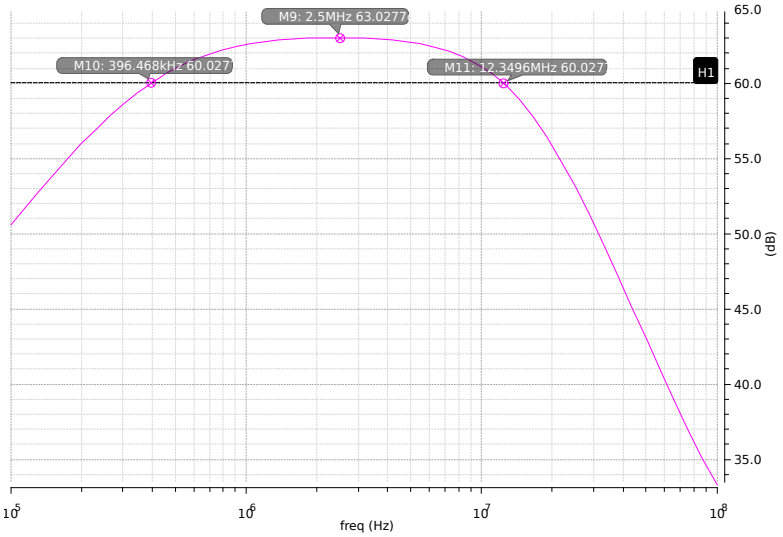

Figure 6.23: AC response of stage 2: the wide-band amplifier has up to 70 dB gain at frequency range of 100 kHz to 48 MHz. At 2.5 MHz a gain of 63 dB is obtained.

All resistance and capacitance tune values can be programmed (set the inputs of the tune capacitances and selectable resistances to VDD or GND in the simulation schematic). The simulation in figure 6.23 shows the wide-band amplifier AC response of stage 2. A 3 dB bandwidth of 11.95 MHz was measured in the simulation (see figure 6.23). At 2.5 MHz a gain of 63 dB is reached. Figure 6.24 shows the AC response of stage 3. A 3 dB bandwidth of 9 MHz was measured in the simulation. At 2.5 MHz a gain of 99 dB was reached for stage 3. As mentioned for the narrow-band simulation the amplifier should switch to narrow-band mode when a control bit is set to high. The resulting AC response is shown in figure 6.25. Approximately at a center frequency of 2.5 MHz a maximum gain value of 130 dB was measured in the simulation.

A transient response simulation was done to test the low-noise input channel with the transducer equivalent circuit. A current source with the transducer load generates an input signal (*Input Transducer* signal shown in figure 6.26) of a few hundred nano volts. The selected low-voltage output (Selected LV Output) which is selected by the selection logic, is fed to the 3-stage amplifier. After the second stage (Stage 2 Output) the signal is amplified to 550 µV. In this configuration the gain factor is $\approx$ 61 dB. After the stage 3 (Stage 3 Output), the amplified output signal reaches 31 mV which corresponds to a gain value of $\approx$ 96 dB compared to the input signal. A bypass circuit is used to bypass the signal output of stage 2 or stage 3. It is composed of several transmission gate switches acting as multiplexer (selected output).

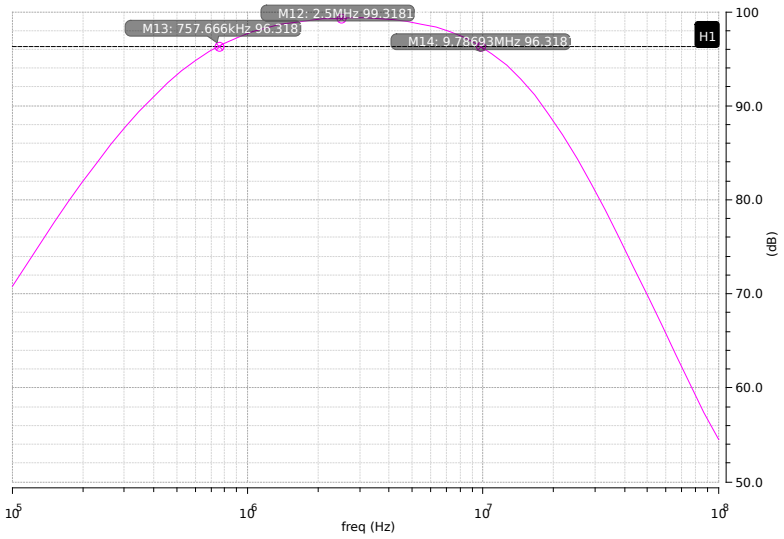

Figure 6.24: AC response of stage 3: 3 dB bandwidth from 757kHz to 9.7 MHz. At 2.5 MHz a gain of 99 dB is obtained.

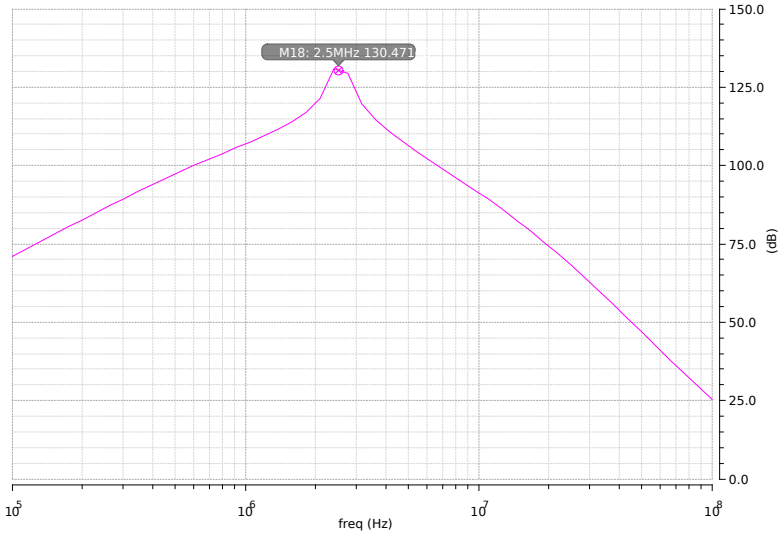

Figure 6.25: Narrow-band AC response simulation. At a center frequency of 2.5 MHz a gain of 130 dB is reached.

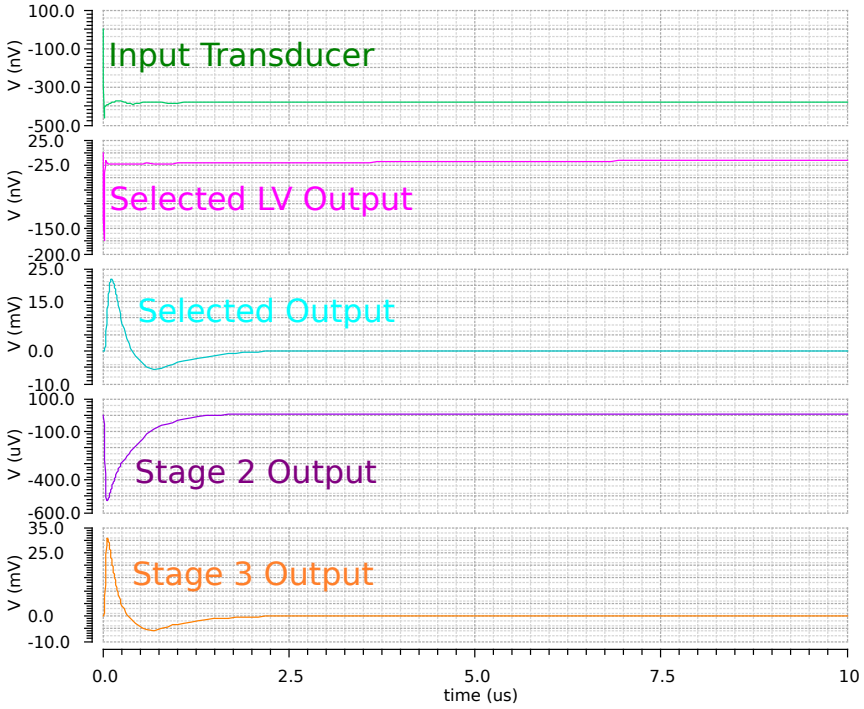

Figure 6.26: Transient simulation with transducer input signal and amplified output signals of stage 2 and 3. The amplitude of the selected output is slightly different to that of stage 3. It shows in principle that the bypass circuit works.

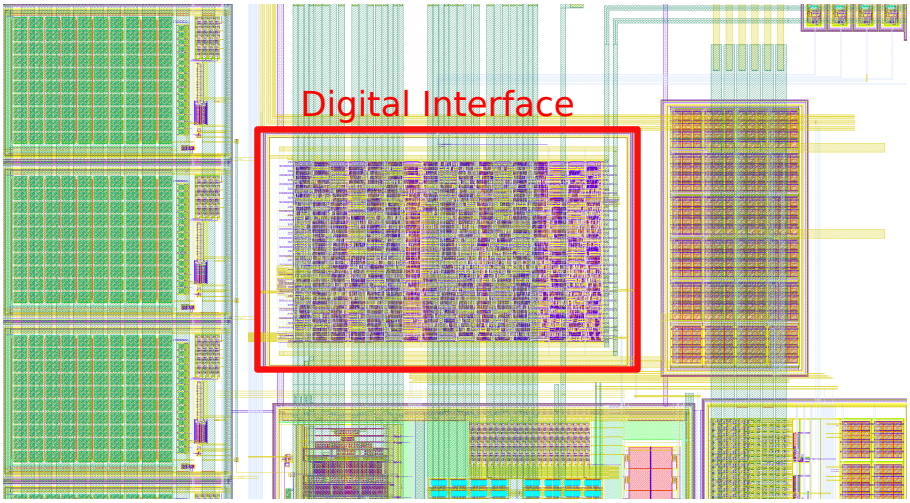

Figure 6.27: Integration of digital interface on the top level layout of the USCT9C ASIC.

6.2.5 Digital Interface

A micro-controller is used for the ASIC communication. The micro-controller handles the low-level communication with USCT9C ASIC using a simple read, write communication protocol. An SPI protocol was implemented as digital interface. SPI is a synchronous serial communication interface. This type of interface is commonly used to send data between a micro-controller and other actuators and sensors or, in our case, an ASIC (figure 6.28). There are two data lines called MOSI (Master Out Serial In) and MISO (Master In Serial Out). A chip select line (SS) or slave select line and SCLK is a clock generated from master to slave. SPI devices communicate using a master-slave architecture with a single master only. The master sends SCLK, MOSI and SS. SPI offers increased flexibility for extensions and is often considered a basic platform for custom protocol development for communication between devices. Following protocol was designed:

- The custom protocol is composed of a header byte, followed by some bytes or some dummy bytes to drive the ASIC output to write the bytes to be read [46]
- The two lower bits of the header control the read, write or burst operation. The six higher bits define the register address to be written or read [46]

The SPI can select the ASIC using the SS signal and acts as chip select signal. During send operation the data written to ASIC turns on the SPI clock and data from ASIC is also read back. The so-called Dummy bytes are used for dummy writes during which only data is read. The dummy bytes are written to ensure that the digital logic of the ASIC has enough clock cycles to complete the request [46].

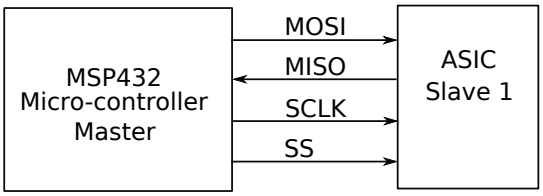

Figure 6.28: SPI interface showing a master (micro-controller) and slave (ASIC) configuration.

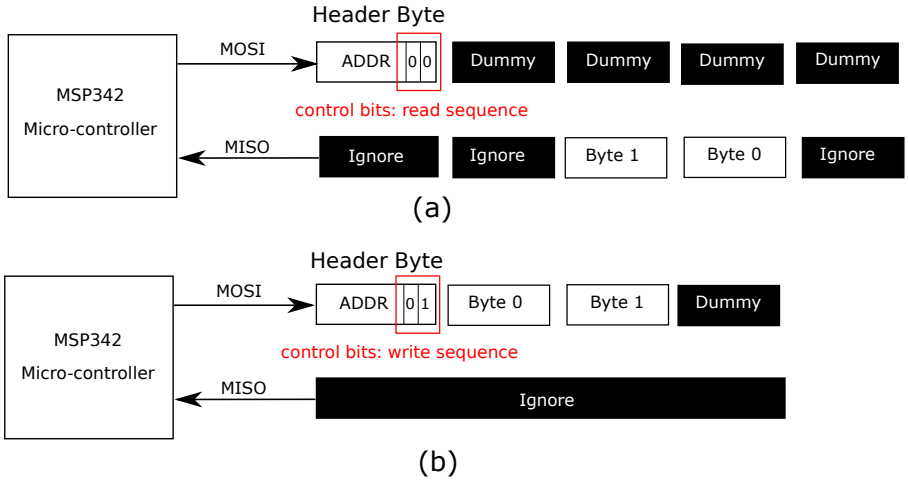

Figure 6.29: The implemented SPI protocol for USCT9C ASIC. Figure (a) shows the read operation. Figure (b) shows the write operation using the SPI interface. The dark boxes are data with garbage values. The dummy bytes are sent to generate extra SPI clock cycles. Figure is reproduced from ref. [46].

Implementation of the Digital Interface The digital design flow includes frontend processes such as register transfer level (RTL) design, functional verification and synthesis. Backend processes comprise floor planning, place and route, clock three synthesis, layout versus schematic (LVS) and generating a GDS file which is sent to the foundry for fabrication. The following steps are required to develop the digital interface on the ASIC (shown in figure 6.30):

1. The RTL design which corresponds to RTL coding (Verilog code) that describes the system specification. This Verilog code contains all definitions of address registers (used for USCT9C ASIC) and an SPI interface. The protocol described in figure 6.29 was implemented as SPI communication protocol. The registers described are 16-bits wide. SPI can send/receive only 8-bit data. A counter has been implemented for 8-bit data transfer and receive. Two further modules are defined in the Verilog

implementation: read register and write register. All modules are selected only when SS is low and data can be read or write. The MISO is the digital output from USCT9C ASIC to the micro-controller. The other micro-controller outputs such as MOSI, SCLK, SS are inputs to USCT9C ASIC (figure 6.28) [46].

2. The RTL design functional verification is the simulation of the RTL design. All modules contained in the Verilog code must be tested with the so-called testbench. The testbench is used to instantiate the digital interface and to produce a test pattern [46].
3. The next step is the logic synthesis and gate level simulation. The RTL is turned into design implementation in terms of logic gates by the synthesis tool. The design data, libraries, timing constraints are imported with the Verilog file and a gate level logic is generated using the RTL compiler. The gate level logic is verified by simulation. Further details of the gate level simulations are not described in this work [46].
4. The physical dimensions, input and output pins are placed during this step. This step is the so-called floor planning. In the current implementation the size of the digital interface is 500 x 800 μm^2 [46].
5. The place and route are used after the floor planning step. In the current design, power lines are placed in the core of the digital interface layout. During placement the so-called standard cells are placed in the design. Then the standard cells the clock tree synthesis (CTS) is carried out. Final routing is started after CTS optimization [46].
6. The last step is the generation of the graphic database system (GDS). The GDS represents the information of the layout in hierarchical form. This file is important for the LVS check. The LVS determines whether a particular circuit layout corresponds to the circuit schematic [46].

6.2.6 Measurement Results of the USCT9C

MSP432 Micro-Controller Development Board The entire measurement setup was implemented with Java/Scala language. All the main functionalities such as evaluation, acquisition, configuration have been developed in a single application. As stated in chapter 6.2.5 a micro-controller is used to allow the ASICs registers to be read/write and to communicate via SPI. An MSP432P401R (shown in figure 6.31) launch pad [1] used for the characterization of the USCT9C ASIC. The launch pad has an MSP432 ARM 32-bit Cortex-M4F micro-controller manufactured by Texas Instruments, which is used as the configuration device. In addition it has a special environment (developed by Texas Instruments) that enables the user to program and debug the projects without using additional tools. The MSP432P401R launch pad is connected to a client PC using Universal Asynchronous Receiver Transmitter (UART) communication via Universal Serial Bus (USB). The communication between micro-controller and USCT9C ASIC uses SPI. The micro-controller uses four EUSCI_A and EUSCI_B SPI modules. These modules can be used to implement a serial communication with USCT9C ASIC. The micro-controller code is written in C language. The hardware for the MSP432 development platform can be configured and controlled with the MSP432 driver library. An SPI master (micro-controller) is configured

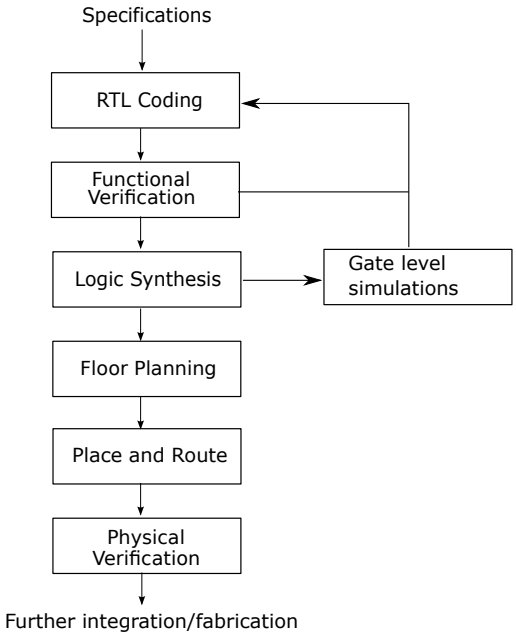

Figure 6.30: The design flow of the digital interface. The design flow is divided into front-end (specifications to gate level simulations) and back-end design (Floor planning to physical verification). Figure reproduced from ref. [46].

using three SPI lines SCLK, MOSI and MISO. For chip select an SS pin is also used. One single SPI write sends 8-bit data on the MOSI line and generates the SCLK. At the same time the data received through the MISO goes into an SPI buffer which is read before the next write sequence. The register write and read functions are implemented in accordance with the implemented protocol in the digital interface (6.29). Following steps must be fulfilled for a write sequence:

1. Chip select using SS set to low
2. A byte header (6-bit address and 2-bit for write select) is sent
3. Send lower byte (data to be written)
4. Send dummy byte to provide extra clock cycles (needed for slave)
5. SS set to high to deselect the ASIC

Following steps are necessary for the read sequence

1. A byte is sent (6-bit address and 2-bit for read)
2. Higher byte is read first, followed by a lower byte

A shift register interface was implemented for bias block programming. The bias shift registers are transmitted to the micro-controller in the form of a continuous string. This

string is split in the flash memory of the micro-controller. The SIn is shifted in using the shift register interface providing the two-phase clocking. The two-phase clocking is generated by using write operation in the 16-bit register (located in the micro-controller) of the shift register interface. After executing SIn and the two-phase clocking an Ld (load) signal is also generated from the micro-controller. As used in other ASICs in this work, the Ld signal completes the bias block programming procedure.

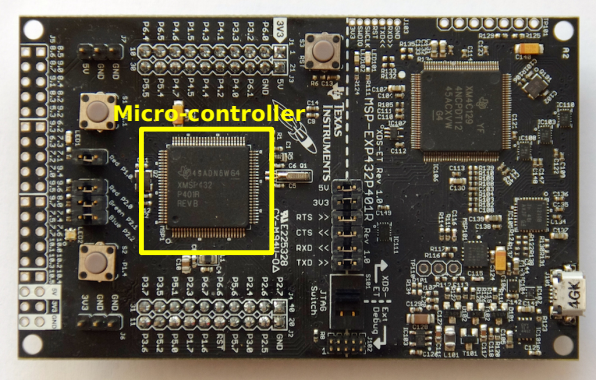

Figure 6.31: MSP432 development board with micro-controller for communication with ASIC and client PC. A three wire mode is used for the digital interface - MOSI, MISO and chip select.

Interface Test Board This interface test board provides power connections, input/output signal connector etc. The power supplies for the high-voltage part are:

- HV_Plus: +60 V
- HV_Minus: -60 V
- LEMO Connectors for high-voltage power supply
- Trans VDD: 3.3 V
- CascTrans (cascode voltage): 2.5 V
- psub!: HV_Minus

The two high-voltages are provided by the high-voltage Toellner power supplies that can each generate 64 V. The high-voltage supplies are provided using a phoenix connector. This special connector is capable of voltages up to 1 kV and therefore ensures the connection of such high-voltages. Electrolyte decoupling capacitors of 100 µF are used. The CascTrans is generated by a potentiometer. For fine tuning the potentiometer can be adjusted to produce a different voltage for the cascode stage. The potential of the substrate (psub!) should be connected to the lowest potential (HV_Minus). All the grounds are connected together to form a common ground. For the 3-stage low-noise amplifier following power supply voltages are used on the test board:

- VDDA: 3.3 V
- VSSA: 2.5 V

Beside the power supply connections there are:

- LEMO connectors for input and output channels
- Digital Interface Connector
- LEMO connectors for input excitation signal

Figure 6.32 shows the interface test board designed with Altium Designer Software [11].

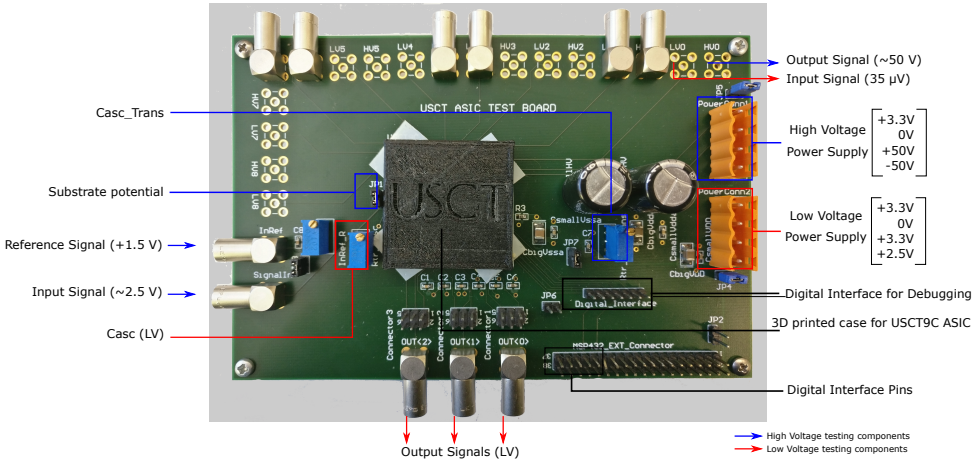

Figure 6.32: Interface test board for the USCT9C ASIC. The USCT9C ASIC is located beneath the 3D printed USCT casing. LEMO connectors are used for high-voltage input and low-voltage output. Figure is reproduced from ref. [82].

USCT9C ASIC Configuration A hex string contains the shift register bits that are arranged in bytes. This string is converted into an array of bytes in the micro-controller. The Scala application uses an XML file which contains all USCT9C registers. To generate a hex string, a USCT9C ASIC register is chosen and a value entered. A button in the application software then generates the hex string. This is sent to the micro-controller through serial port from the client PC. For instance, in the bias shift register the value VHVNAMP is to vary the current source of the high-voltage amplifier. There are also several other settings such as current biasing, tune capacitances and resistances that can be set with the application and sent to the USCT9C ASIC.

Furthermore, the Scala application automatically generates the input signal, sets the oscilloscope and obtains the waveform from it. The Scala application gathers the number of communication ports for COM port harvester (name of the implemented Scala function). Oscilloscope and other devices like waveform generator are carried out using the VISA library provided by National Instruments. Oscilloscope setting is done for channel selection

using VISA library functions. A location for saving the waveform files is defined. A folder will be made containing all the waveform files. All waveforms are saved as separate CSV files. Octave is executed and a script containing the evaluation is started. Octave is used to analyze and plot the input/output signals from the extracted data and is compatible with MATLAB. In other words, Octave is an alternative to MATLAB. The entire process is done automatically and is an essential part of the characterization procedure. If the USCT III system is used in a clinical study, several ASICs would have to be accurately characterized. A simplified block diagram of the characterization procedure is shown in figure 6.33. A waveform generator is used to generate an input signal for the 3-stage low-noise amplifiers. Different signals can be applied such as the chirp signal to evaluate bandwidth of the USCT9C ASIC. A chirp CSV file is provided by the USCT team which can be implemented in the Scala application. The application can choose the input signal type over the VISA driver functions. The measurements have shown that a Scala implementation

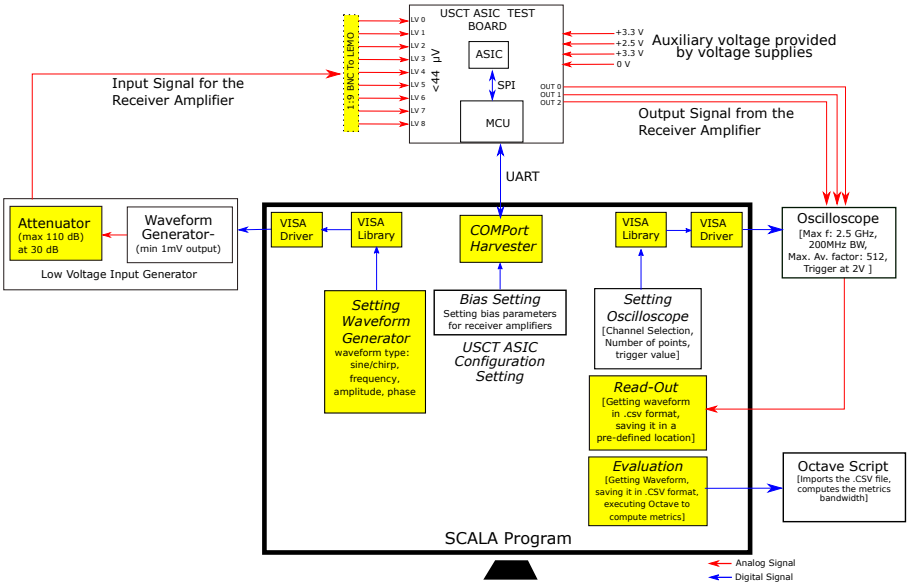

Figure 6.33: Simplified block schematic of the characterization procedure. Scala application handles the USCT9C bias setting and configuration the ASICs bias block with a shift register interface. Figure reproduced from ref. [82].

for the evaluation of several ASICs for USCT III is not advantageous. The bias setting configuration takes several tens of milliseconds until all registers in the USCT9C are configured. It has also been observed that the Scala application is not stable. To test a large number of USCT9C ASICs, it would be important to develop an application that is more stable and faster during the bias configuration sequence. A development was started to port the Scala application in a C/C++ application. The advantages of this new C/C++ application over the Scala application were immediately apparent.

New Characterization Application As it was shown before, a faster characterization is needed to speed up the USCT9C ASIC full characterization. There are 16 parameters in the USCT9C that can be set in the bias block. Doing an ASIC characterization of all parameters results in a huge amount of parameter sweeps. The sweep parameters are the following:

- VHVNAMP (high-voltage amplifier: bias for transconductor)
- VNSF1, VNSF2, VNSF3 (low-noise amplifier: bias for NMOS source follower stage 1, 2, 3)
- VPLOAD1, VPLOAD2, VPLOAD3 (low-noise amplifier: bias for PMOS load stage 1, 2, 3)
- VNFB1 (low-noise amplifier: bias for current mirror used for the resistor feedback)
- VN1, VN2, VN3 (low-noise amplifier: bias for current source for folded amplifier stage 1, 2, 3)

If this is done with the old setup a full characterization can take nearly a full day for one ASIC. For instance, the bias block configuration of the parameters is adjusted manually. Manually means that every parameter in the old Scala implementation is manually hard-coded inside the parametrization source. The Scala implementation requires a recompilation of the software for every parameter change. The aforementioned ASIC configuration with hex string makes the parametrization slow. A large amount of overhead is created when transmitting data. A redesign was implemented to reduce the overhead by caching the parameters on the micro-controller instead of sending every time. Measurements have shown that a data reduction of 50% was achieved [17]. In the new characterization application the manual steps were reduced to a minimum. The new device connection (old implementation done by gathering over COMport Harvester) and data acquisition can be controlled by the new C/C++ implementation. The new implementation also allows data to be read from devices, like oscilloscopes and waveform generators. As in the old implementation, the VISA library makes it possible to use C/C++ language to implement special functions for these devices. A new interface design was developed with Qt environment. This GUI offers a clear structure of the ASIC configuration but also the parametrization and data output generated files. The generated output files are in CSV format as this file format can be easily imported into other software such as MATLAB or Octave. For further information on the new characterization application, see ref. [17].

Measurements results of the 3-stage low-noise amplifier All measurements were done with the Scala application. The first measurement was done with stage 2 (set bypassing bit for stage 3) and the second refers to the stage 2 and 3. The gain was varied only in stage 2 from 1 to 9 and stage 3 set to a fixed gain factor of 2. The overall gain is between 40 and 720. The amplifier was excited with input chirp signal of 1 mV_{pp} generated by the arbitrary waveform generator with a length of 100 µs and a bandwidth of 5 MHz. Figure 6.34 shows an example transient chirp signal input and the signal in the frequency domain. The transient response of stage 2 and stage 3 to a chirp signal input of V_{pp} = 1 mV and a signal length of 100 µs is shown in figure 6.35. The measurement shows that the amplified output of stage 2 follows the input signal. The output amplitude increases over time for higher frequencies. After a certain time the amplitude drops. Distortions were observed in higher frequencies. The stage 3 also shows distortions at the output due to the distorted

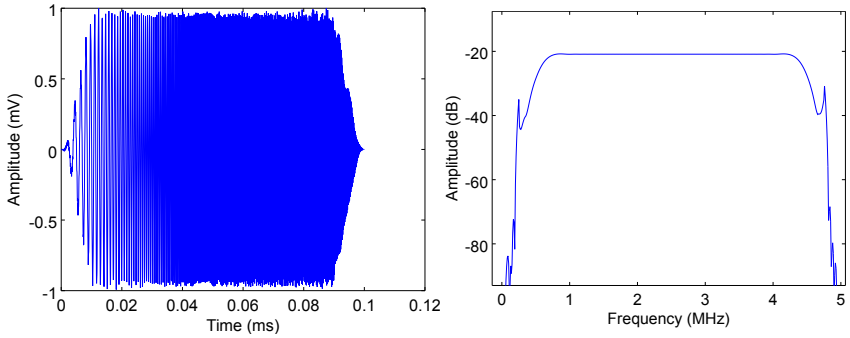

Figure 6.34: The transient chirp input signal with a length 100 µs and $V_{pp} = 2$ mV. Figure reproduced from ref. [46].

signal from stage 2. It is confirmed by the simulation that the stage 2 signal should follow the input chirp signal. A phase shift of 180 degree between stage 2 and stage 3 is expected. The amplitude of stage 2 follows the input with an amplified signal of 40 and the stage 3 is phase shifted by 180 degree with an amplifying factor of 80. Figure 6.36 shows a simulation of a chirp signal where the amplitude varies with time and is not constant compared to the chirp signal used in the measurement. All amplitudes in this simulation are normalized. Nevertheless the behavior of stage 2 (following the input signal) and stage 3 (phase shift of 180 degree) does not change.

It was also observed that the output is more distorted for a higher gain setting. The distortion is observed in the output for gain values above 400. Other simulations showed the similar amplitude clipping for signal input values in the regions 100 µV and 1 mV . The expected TAS signal in the range of 10-20 µV. Another point would be to vary the gain between stage 2 and stage 3.

For the bandwidth measurement the data were analyzed with Octave. The center frequency (f_C) determines the frequency at the maximum amplitude. The -3 dB cut-off frequency lower than f_C is defined as f_L and higher than f_C is defined as f_H. Generally the bandwidth can be defined $BW = f_H - f_L$. The frequency response of the input signal, stage 2 and stage 3 is plotted in figure 6.37. Equation $BW = f_H - f_L$ is used to calculate the bandwidth. The bandwidth of the input signal is 3.7 MHz where $f_L = 600$ kHz and $f_H = 4.3$ MHz.

The gain settings were varied to reach the bandwidths of the amplifiers. With regard to the 3-stage low-noise amplifier measurements, the expected amplifications have been achieved. The output signal shows distortions. At higher amplification factors, the distortions can be observed even more in the output signals of both stages. Further simulations have shown, when amplifiers get to its limits, the output signal is clipped. It is difficult to estimate signal magnitude of the generated voltage of the real transducer. Probably the ultra sound pressure is producing on the transducer voltages is in the range of 10-20 µV. Further investigations of simulations using a more realistic load (output load of the piezoelectric sensor) are considered in the second generation of the USCT9C.

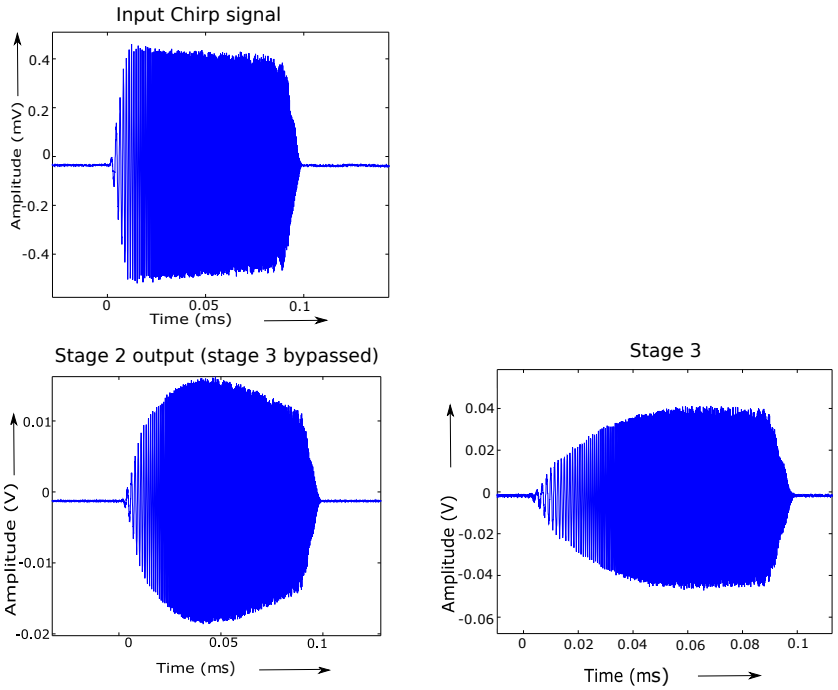

Figure 6.35: Transient response stage 2 and stage 3. Stage 2 has a gain factor of 40 and the amplified signal does not follow the input signal (signal distortions). The output signal of the stage 3 is also distorted due to the distorted signal in stage 2. The gain with stage 3 is 80. Figure reproduced from ref. [46].

The second generation USCT9C ASIC or the so-called USCT9CV2 was redesigned in 2018. Furthermore, simulation model was improved in order to simulate with a more realistic transducer load and to observe the behavior of the amplifiers. The measurement results of USCT9CV2 will be presented in a separate paper.

Measurement Results of the High-Voltage Amplifier As already explained in the chapter 6.2.1 about functionality of the high-voltage amplifier, the amplifier has an external gain of 20 which is set by two resistors. Before applying a chirp input signal, a sine wave and a square wave were provided by the waveform generator. A reference signal of 1.5 V is also connected to the InRef input. The reference signal can be varied by using the potentiometer placed on the interface test board. The high-voltage part must be supplied with TransGND, TransVDD, TransCasc (adjusted by a potentiometer), HV_Plus and HV_Minus. Another important point is that the input signal must have an offset that is roughly equal to InRef. To start an initial test of the high-voltage amplifier, a square input signal with V_{pp}=1 V and a frequency of 1 kHz was provided by the waveform generator. A frequency sweep between 1kHz and 5 MHz was done. The current consumption of the high-voltage path

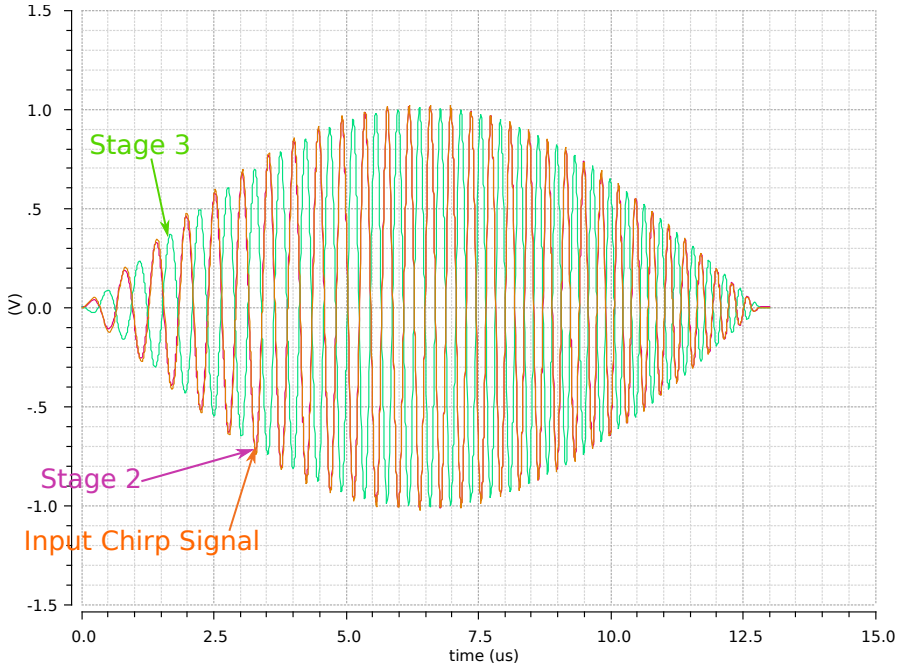

Figure 6.36: Simulation with a chirp signal input whose amplitude is not constant over time. Stage 2 follows the input and stage 3 is 180 degree phase shifted relative to stage 2. All amplitudes are normalized.

and the output amplitude were measured. If a high-voltage of $\pm$ 5 V is applied, an output swing of 10 V can be expected. The maximum output amplitude equals the input amplitude multiplied by gain factor 20. Because of the limitation in the high-voltage the output is 10 V. The bias parameter VHVNAMP plays an important role here. It influences the speed and the slope of the output signal. This bias parameter VHVNAMP can be set (like other parameters in the USCT9C ASIC) from 0 to 63 by the Scala application. All the measurements were done with the Scala application. As usual, the output is measured using the oscilloscope. The output signal is between HV_Minus and HV_Plus. After applying all voltages, the ASIC was configured with a VHVNAMP of 20. Figure 6.38 shows the currents of HV_Plus and HV_Minus as well as the V_{Outpp} as function of frequency. The currents are symmetric because the high-voltage amplifier has a p and n part (symmetric circuit topology) and the currents are also symmetric. The V_{Outpp} reaches a maximum output swing of 9.28 V with a frequency range of up to 2 MHz. After passing 2 MHz the voltage drops to 7.4 V.

In the next measurement a chirp signal was generated by the waveform generator and applied to the input of the high-voltage amplifier. VHVNAMP was set to 40. The high-voltages are set to $\pm$ 20 V (40 V). In this measurement the output chirp amplitude decreases for higher frequencies. Time and a frequency domain representation of the chirp output is

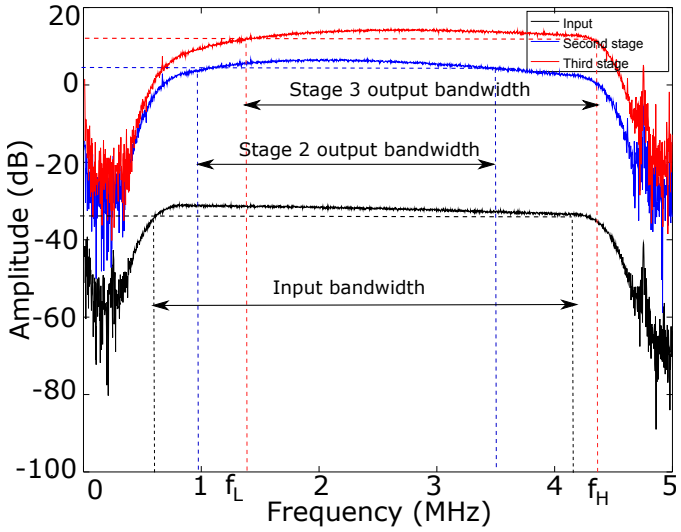

Figure 6.37: Bandwidth calculated with Octave. -3dB bandwidth of the chirp input signal (black) and the outputs of stage 2 (blue) and stage 3 (red). The bandwidth of stage 2 is $\approx$ 2.5 MHz and of stage 3 is $\approx$ 3 MHz. Figure is reproduced from ref. [46].

shown in figure 6.39. The high-voltage output response reaches maximum values up to $\pm$ 20 V (40 V output swing). This means 2 V input signals are amplified to an output swing of 40 V. This can be seen in figure 6.40. It can be seen that the output is not symmetric. The signal became more distorted as the values of VHVNAMP increased. The observation is not consistent with the expected behavior. When the VHVNAMP increases, the output signal should improve. A further observation showed that the current consumption for HV_plus and HV_minus increased strongly for larger values of VHVNAMP and that the chip was no longer functional. It was then simulated that in the transconductor amplifier (see figure 6.41) the cascode transistors sources (M4 or M5) can drift to the high-voltage potential. If a high-voltage potential is applied to HVPlus then the source of M5 can increase beyond 3.3 V, which is the maximum rating for safe operation. This may lead to transistor damage. We observed an unsymmetrical behavior of the current due to the rise in temperature inside the ASIC. In this version of the ASIC all nine channels are simultaneously switched on. In the next ASIC version only one of nine channels can be switched on in order to prevent the high power dissipation inside the ASIC. For instance, if a current of 100 µA flows through the high-voltage (applied high-voltage 100 V) channel, the following power dissipation can be calculated:

$$P = 100\,\mu A \cdot 100\,V \cdot 9 = 90\,mW \tag{6.24}$$

This power dissipation causes the destruction of the high-voltage channels. If we apply a high-voltage of $\pm$5 V the distortions do not occur probably due to the lower power dissipation. Another point would be that the current sources of the high-voltage amplifier

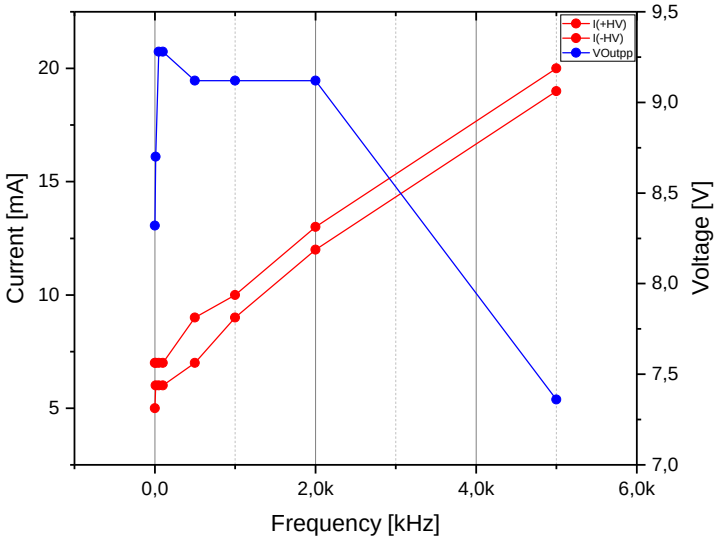

Figure 6.38: Measurement of the current consumption of high-voltage amplifier and output voltage swing as function of frequency. The applied input signal is a square wave with 1 mV_{pp}. A frequency sweep of the input signal was accomplished.

load the capacitances and in the case of a high-voltage applied, the amplifier is slew rate limited. This could explain why the output signal of the ASIC for high-voltages of $\pm$ 5 V looks much better than for $\pm$ 20 V. The capacitances are loaded more slowly for $\pm$ 5 V instead of high-voltages of $\pm$ 20 V. A high-voltage of $\pm$ 20 V leads to high current consumption.

A chirp signal with a constant amplitude over the frequency as excitation signal and low VHVNAMP values were used. The high-voltage amplifier produces the desired behavior. The applied (generated by the waveform generator) input chirp signal has an amplitude of 100 mV_{pp} and the length is 4000 samples using a sample rate of 40 MS/s which results in 100 µs length. The output signal is phase shifted by 180 degrees which corresponds to the behavior from the simulation shown in chapter 6.2.2. Figure 6.42 shows the input chirp signal (a.) and the output generated by the high-voltage amplifier (b.). The amplified signal has no distortions because VHVNAMP was not set to a too high value. A calculated gain factor of 17.5 was reached.

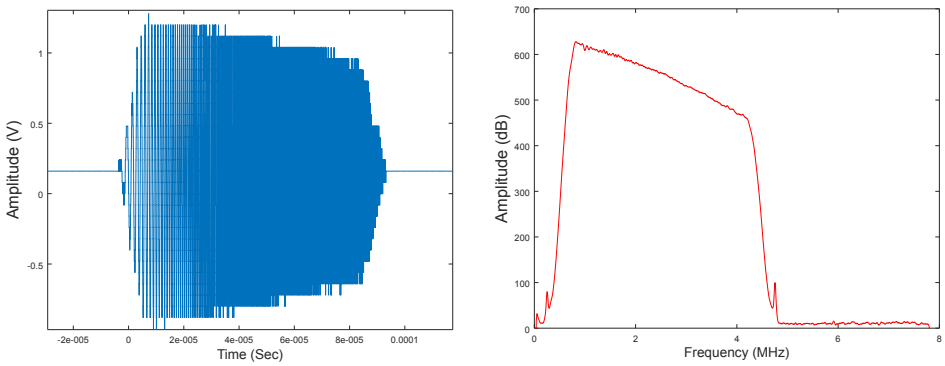

Figure 6.39: Generated input chirp signal from the waveform generator and the calculated signal in the frequency domain. The amplitude of the chirp signal drops for higher frequencies.

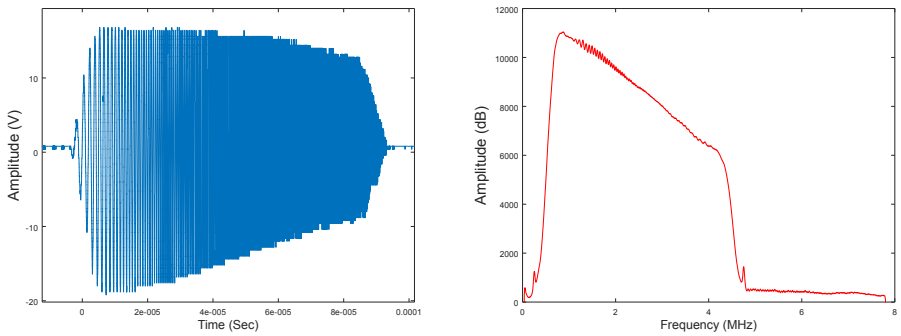

Figure 6.40: Produced high-voltage output of the high-voltage amplifier and calculated output in the frequency domain.

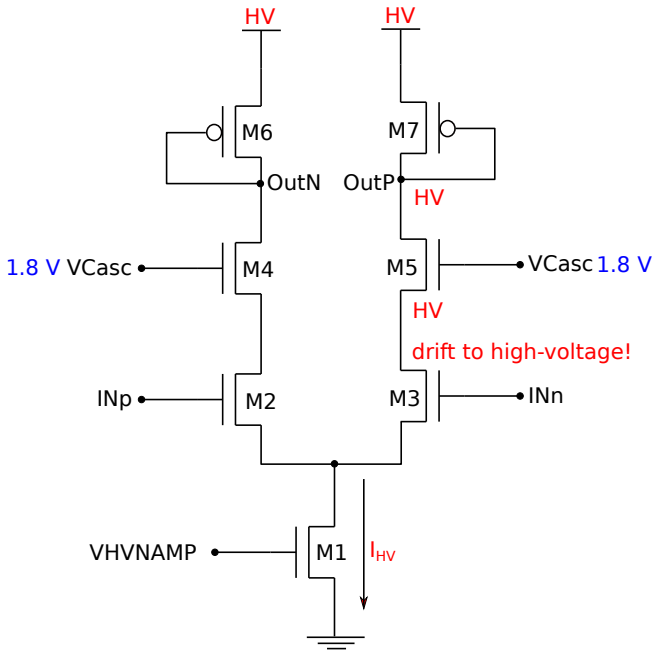

Figure 6.41: Transconductor amplifier of the high-voltage channel. The low-voltage transistors M5 and M3 drift to high-voltage.

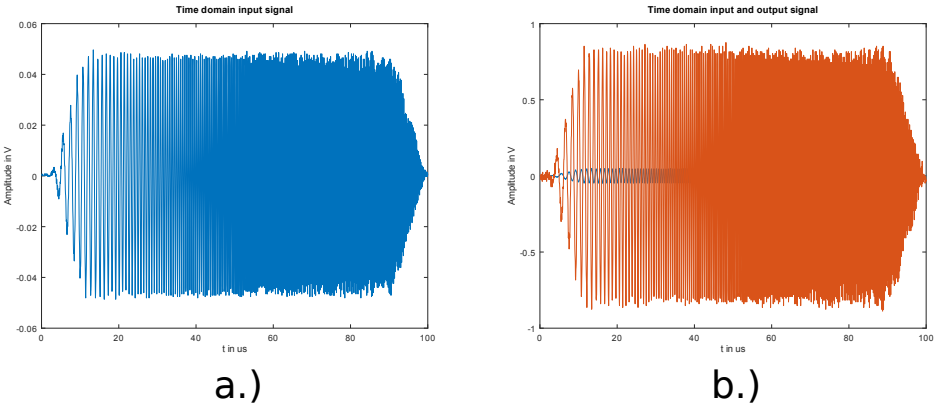

Figure 6.42: Input chirp signal a.) and comparison between input and amplified output b.). The input has a constant amplitude after ≈ 16 µs. The amplitude is 100 mV_{pp}. In b.) a calculated gain factor of 17.5 was reached.

7 Conclusion

The main focus of this work was the development of ASICs for two big experiments such as TRISTAN, Belle II. A new ASIC has been developed in this work for the TRISTAN experiment. The design of low-noise amplifier stages and a pipelined ADC was a very challenging task. Two ASICs (DCD and Switcher) have been developed for the Belle II experiment which are essential for this experiment and guarantee its success. A chip probe station has been developed for testing both ASICs. Furthermore, this thesis presents the development of a new monolithic single photon detector ASIC with multipurpose in-pixel electronics for various applications. A transceiver ASIC has been developed in this thesis for the KIT project "Ultrasound Computer Tomography".

- **ASICs for the TRISTAN experiment** have been designed in 350 nm high-voltage CMOS process. The development started in 2015. The ASIC includes low-noise amplifier stages and a 9-bit pipelined ADC at 25 MS/s. The ASIC is composed of the following analog blocks: bias block, amplifier stages, 9-bit pipelined ADC. The TRISTAN amplifier was designed by Professor Ivan Perić and Richard Leys. The pipelined ADC was designed by me. A firmware based on Verilog was implemented by me to control and handle the data between FPGA and ASIC. Scala programming language has been used. My colleague Richard Leys supported me with the software development in Scala. A software correction has been developed for the fully differential pipelined ADC to improve the linearity. The ASIC is fully functional. First measurements show promising results. The sensor used for the measurements is an SDD sensor developed by the company FBK in Italy. This SDD sensor detects the electrons generated by the Fe-55 source. A low-noise performance of 28 electrons has been reached. The noise was measured by taking 50k amplified analog outputs and measured with an Fe-55 source. Measurement results of the fully differential pipelined ADC are promising. A corrected ADC transfer curve was measured with an INL value between -1 to 0.8 LSB. All measurements used 12.5 MS/s. It is planned to improve the readout part to achieve 25 MS/s and reach a higher ADC resolution.
- **A new monolithic single photon avalanche detector ASIC (H35-SPADV1)** with multipurpose in-pixel electronics has been designed in 350 nm high-voltage CMOS process. The ASIC was developed by me with the support of Professor Ivan Perić. The ASIC is composed of a single-photon avalanche diode and readout electronics based on active quenching circuitry. The high-voltage CMOS single-photon avalanche diode (SPAD) is implemented as p+ diffusion in n-well diode. A p+ implant, acting as the anode, and a shallow n-well define the high field region in the active area. The SPAD signal produced is fed directly to a comparator. A monostable circuit can adjust the time window for quenching and disabling the diode (inhibition). Inhibition and quenching time can be set by external electronics. Furthermore, the pixel features a 1-bit hit flag for storing the hit information produced by the photons.

Single pixel, full matrix and self triggered readout is possible. The H35-SPADV1 ASIC is fully functional. For this project an interface board and an FPGA readout were also used. The readout was developed in C/C++ and the firmware was written with Verilog. The software was developed by me with help of Christian Krämer. The author was responsible for the entire development of the ASIC, readout and measurements. The measurements show promising results. The breakdown voltage values have been determined from the current-voltage characteristics, using an SMU unit. A breakdown voltage of 11.7 V was measured at 24 °C for a SPAD pixel size of 3496 μm^2. At 12.0 V a dark count rate of 610 kHz, corresponding to 174 Hz/μm^2, was measured. A special laser setup was developed to measure the detection efficiency of the SPAD sensor. These measurements have shown that the SPAD is not able to measure single photons (primary electrons). It was only possible to measure a few hundred primary electrons with a detection efficiency of 30 % with blue LED light ($\lambda \approx 432$ nm). During the measurements it was observed that the DCR is weakly dependent on temperature. The trap-assisted band-to-band tunneling is responsible for the generation of free charge carriers. The breakdown voltage shows the expected temperature sensitivity of ≈ 9 mV$^\circ C^{-1}$. Two more ASICs are planned to reduce the DCR and increase the detection efficiency. I started a new development (circuit as well as layout design) in LFoundry 150 nm high-voltage CMOS technology. A completely new ASIC with different SPAD sizes and readout electronics was developed by Professor Ivan Perić based on TSI 180 nm technology.

- **A DCD, Switcher ASIC and a Chip probe station** have been designed. The final version of DCD ASIC has been designed in UMC 180 nm technology using radiation hard design techniques. The final version of Switcher ASIC has been designed in 180 nm high-voltage CMOS. The DCD and Switcher ASICs were designed by Professor Ivan Perić. My contribution with the support of Richard Leys was the development of a chip probe station to test and characterize the DCD and Switcher ASICs for the Belle II experiment. Both ASICs are essential for the Belle II experiment. Each DCD is composed of 256 analog channels. Each channel contains one input amplifier. The analog signal is digitized using current-mode pipelined ADCs. The Switcher can generate fast voltage pulses up to 20 V amplitude to activate DEPFET rows and charges deposited in the DEPFET sensor. A Chip Probe Station was developed at KIT to evaluate and test the ASICs for the Belle II experiment. About 1000 ASICs are required for the construction of the whole PXD detector. Extensive tests must first be carried out. The following tests must be performed for the DCD: digital readout test, ADC curve measurement (characteristic curve), INL, DNL noise measurement. The following tests are also essential for the Switcher ASIC: digital configuration and testing of all high-voltage outputs. The developed chip probe station consists of a mechanical mounting and support, step motors, single and multiple vacuum chip carriers, wafer vacuum chuck, optics and probe cards. The chip probe station is fully configurable. The monitoring software control environment is based on a Java/Scala and web server application. The chip probe station tested up to 1000 DCD and Switchers. The chip probe station is capable of full DCD characterization (linearity, INL, DNL). The measurements show a high yield of the DCD ASICs. The Switcher also has a high yield. Rarely have more than three ASICs failed through all the tests. The tested ASICs have been mounted on the PXD module. The next step

is to improve the mechanical construction of the chip probe station. The use of an air-suspended mechanical mounting would absorb the vibrations. Furthermore, the probe cards would have to be improved. The needles for the connection between ASIC and probe card in particular should be improved.

- **A new ASIC (USCT9C) for the 3D USCT project** has been designed in 350 nm high-voltage CMOS technology. This ASIC is used for medical applications. The development started in 2015. My contribution was the design of the high-voltage amplifier. The 3-stage low-noise amplifier was designed by Professor Ivan Perić. The digital interface was mainly designed by Richard Leys. Garima Singhal and Bhavana Imanani developed the test system (software) and carried out the measurements with the support of Richard Leys and me. The first generation ASIC has only one channel consisting of a high-voltage and low-noise amplifier stage. The USCT9C ASIC has nine channels consisting of high-voltage and a 3-stage low-noise amplifier. Simulations, design, layout design and measurements are presented in this thesis. The ASIC acts as a transceiver. This means, it serves as emitting and receiver stage. The piezoelectric transducer produces ultrasound waves in water during the emitting phase. A high-voltage is applied on the piezoelectrical transducer in form of a chirp signal. After the emitting phase, the ASIC is switched for receiving. During receiving the reflected ultrasound waves produce signals of 10-20 μV. This small amplitude needs to be amplified and transmitted to the DAQ system. The idea is to integrate all the discrete transducer array system (TAS) electronics into an ASIC which consumes low power, and has a high bandwidth and S/N ratio. The use of an ASIC results in a significant reduction in the space occupied by the electronics. A readout environment was also developed. Scala programming language has been used. Many measurements were done with the Scala application. The second version of the readout software was implemented in C++ to optimize the speed of configuration. The measurements of the 3-stage low-noise amplifier show promising results. The amplifier was excited with an input chirp signal of 1 mV_{pp} and a length of 100 μs as well as a bandwidth of 5 MHz. The expected behavior was measured. It was also observed that the output is distorted for higher gain values. In the second generation of USCT9C ASIC, or the so-called USCT9CV2, several changes have been carried out to ensure more stability and more gain at the output stages. The high-voltage amplifier shows that it can reach 40 V of amplitude swing. As the VHVNAMP increases the output signal tends to be distorted, as already mentioned for the 3-stage low-noise amplifier. Unfortunately with USCT9C ASIC an output swing of 120 V was not reached. No distortions were observed for small VHVNAMP values. USCT9CV2 will be measured during December 2018 and January 2019. Measurement results with the real piezoelectric load are also planned. In summary, the USCT9C is functional for lower gain values.
- **Photon ASIC for photon counting** has also been developed. The photon ASIC was designed by Professor Ivan Perić. The test system and the measurements were carried out by me. The Photon ASIC is not described in this thesis.
- **The development of HV-CMOS sensors for ATLAS experiment** is the last project that was done by me. A monolithic pixel detector has been designed in LFoundry LFA15 150 nm process. My contribution was the circuit and layout design of the

digital triggered readout cell. The untriggered and pixel circuit as well as layout design were developed by Professor Ivan Perić and Hui Zhang. The ASIC is not described in this thesis.

List of Figures

List of Tables

Abbreviations

ACMC	Analog Common-Mode Correction
ADC	Analog-to-Digital Converter
ADL	ASIC and Detector Laboratory
APD	Avalanche Photodiodes
AQC	Active Quenching Circuit
BARC	Bottom Anti Reflective Coating
CDC	Central Drift Chamber
CMC	Current Memory Cell
CMOS	Complementary Metal Oxide Semiconductor
CSA	Charge Sensitive Amplifier
CTS	Clock Tree Synthesis
DAC	Digital-to-Analog Converter
DAQ	Data Aquisition
DCD	Drain Current Digitizer
DCR	Dark Count Rate
DEPFET	Depleted P-Channel Field-Effect Transistor
DHP	Data Handling Processor
DNL	Differential Nonlinearity
DSSD	Double-Sided Silicon Dector
ECL	Electromagnetic Calorimeter
ENC	Equivalent Noise Charge
FBK	Fondazione Bruno Kessler
FIFO	First In First Out
FSM	Finite State Machine
FTDI	Future Technology Devices International
FWHM	Full Width at Half Maximum
GDS	Graphic Database System
GUI	Graphical User Interface
HLL	Halbleiter Labor München
HV-CMOS	High-Voltage Complementary Metal Oxide Semiconductor
INL	Integral Nonlinearity
JFET	Junction Gate Field-effect Transistor
JTAG	Joint Test Action Group
KATRIN	Karlsruhe Tritium Neutrino Experiment
LASER	Light Amplification by Stimulated Emission of Radiation
LED	Light Emitting Diode
LHC	Large Hadron Collider
LOR	Line of Response

LVS	Layout versus Schematic
MAC-E	Magnetic Adiabatic Collimation combined with an Electrostatic Filter
MISO	Master In Serial Out
MOSI	Master Out Serial In
MRI	Magnetic Resonance Imaging
PDE	Photon Detection Efficiency
PET	Positron Emission Tomography
PMT	Photomultiplier Tube
PQC	Passive Quenching Circuit
PXD	Pixel Detector
RFG	Register File Generator
ROI	Region of Interest
RTL	Register Transfer Level
SDD	Silicon Drift Detector
SMU	Source Measurement Unit
SPAD	Single Photon Avalanche Diode
SPI	Serial Peripheral Interface
SVD	Silicon Vertex Detector
TAS	Transducer Array System
TCAD	Technology Computer Aided Design
TCL	Tool Command Language
TOF	Time of Flight
TRISTAN	Tritium Beta Decay to Search for Sterile Neutrinos
UART	Universal Asynchronous Receiver Transmitter
UCF	User Constraint File
USB	Universal Serial Bus
USCT	Ultra Sound Computer Tomography

Bibliography

[1] *MSP432P401R SimpleLink Microcontroller LaunchPad Development Kit*. http://www.ti.com/lit/ug/slau597f/slau597f.pdf, year = 2018.

[2] *CMOS Technology for SPAD / SiPM*, 2014. https://juser.fz-juelich.de/record/153873/files/FZJ-2014-03352.pdf.

[3] *Medical Imaging with 3D Ultrasound Computer Tomography for early breast cancer diagnosis*, 2015. https://publikationen.bibliothek.kit.edu/230100980/3816832.

[4] *Absorption in silicon*, 2017. https://www.pveducation.org/pvcdrom/materials/optical-properties-of-silicon.

[5] *Belle II*, 2017. https://www.belle2.org/.

[6] *Moore's Law - The number of transistors on integrated circuit chips*, 2017. https://ourworldindata.org/technological-progress.

[7] *LHC CERN*, 2018. https://home.cern/topics/large-hadron-collider.

[8] ABE, T. et al.: *Belle II Technical Design Report*. 2010.

[9] ACKERMANN, K.: *Schematic Drawing*, 2016. .

[10] ALAN MIGDALL, SERGEY POLYAKOV, J. F. J. B.: *Single-Photon Generation and Detection*. Academic Press, 1st ed., 2013.

[11] ALTIUM: *Altium Designer*, 2018. https://www.altium.com/altium-designer/.

[12] ANDRICEK, L., P. FISCHER, K. HEINZINGER, S. HERRMANN, D. HERZ, M. KARAGOUNIS, R. KOHRS, H. KRUEGER, P. LECHNER, G. LUTZ, H.-G. MOSER, I. PERIC, L. REUEN, R. RICHTER, C. SANDOW, M. SCHNECKE, F. SCHOPPER, L. STRUEDER, E. TOERNE, J. TREIS, M. TRIMPL, J. VELTHUIS, N. WERMES and S. WOELFEL: *The MOS-type DEPFET pixel sensor for the ILC environment*. Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, 565(1):165 – 171, 2006. Proceedings of the International Workshop on Semiconductor Pixel Detectors for Particles and Imaging.

[13] A.ROCHAS, A. PAUCHARD, L. A. P. R. R.: *Ultra compact CMOS single photon detector*. Vol. 6372, pp. 6372 – 6372 – 8, 2006.

[14] AUSTIN, L. and H. STARKE: *Ueber die Reflexion der Kathodenstrahlen und eine damit verbundene neue Erscheinung secundaerer Emission*. Annalen der Physik, 314(10):271–292.

[15] BAHAA E. A. SALEH, M. C. T.: *Fundamentals of Photonics*. John Wiley and Sons, Inc, 2nd Edition ed., 2007.

[16] BAKER, J.: *CMOS - Circuit Design, Layout and Simualtion*. IEEE Press, 3rd ed., 2010.

[17] BECKER, L.: *Automated Characterization of an ASIC for Ultrasound Transceivers*, 2018.

[18] BLANCO, R., C. KRAEMER and I. PERIC: *Silicon photomultiplier detector with multipurpose in-pixel electronics in standard CMOS technology*. Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, 2018.

[19] BLANCO, R., R. LEYS and I. PERIC: *Integrated readout electronics for Belle II pixel detector*. Journal of Instrumentation, 13(03):C03001, 2018.

[20] BOMBELLI, L., C. FIORINI, T. FRIZZI, R. ALBERTI and A. LONGONI: *CUBE, A low-noise CMOS preamplifier as alternative to JFET front-end for high-count rate spectroscopy*. In *2011 IEEE Nuclear Science Symposium Conference Record*, pp. 1972–1975, Oct 2011.

[21] CARUSONE, T. C., D. JOHNS and K. MARTIN: *Analog Integrated Circuit Design*. John Wiley and Sons, New York, 2011.

[22] CASANOVA, R. and S. GRINSTEIN: *A Verilog-A model of a charge sensitive amplifier for a HV-CMOS pixel sensor*. In *2016 13th International Conference on Synthesis, Modeling, Analysis and Simulation Methods and Applications to Circuit Design (SMACD)*, pp. 1–4, June 2016.

[23] CHAIWONGKHOT, K., T. KOHRIKI, N. TANIGUCHI, S. UNO, E. NAKANO, U. TIPPAWAN, S. MINEMURA, M. H. A. NOUXMAN and K. A. B. AZMI: *Inner chamber of the Belle II Central Drift Chamber*. In *2013 IEEE Nuclear Science Symposium and Medical Imaging Conference (2013 NSS/MIC)*, pp. 1–3, Oct 2013.

[24] CHRISTENSON, J. H., J. W. CRONIN, V. L. FITCH and R. TURLAY: *Evidence for the* 2π *Decay of the* K_2^0 *Meson*. Phys. Rev. Lett., 13:138–140, Jul 1964.

[25] COLLBORATION, B. I.: *Belle II PXD Mass Testing*, 2018. `https://elog.belle2.org/elog/PXD-Mass-Testing`.

[26] COVA, S., M. GHIONI, A. LACAITA, C. SAMORI and F. ZAPPA: *Avalanche photodiodes and quenching circuits for single-photon detection*. Appl. Opt., 35(12):1956–1976, Apr 1996.

[27] COVA, S., A. LACAITA and G. RIPAMONTI: *Trapping phenomena in avalanche photodiodes on nanosecond scale*. IEEE Electron Device Letters, 12(12):685–687, Dec 1991.

[28] D ASCENZO, N., X. ZHANG and Q. XIE: *Application of CMOS Technology to Silicon Photomultiplier Sensors*. Sensors, 17(10), 2017.

[29] DAPP, R.: *Abbildungsmethoden fuer die Brust mit einem 3D-Ultraschall-Computertomographen*. PhD thesis, 2013.

[30] DIGILENT: *Atlys Board Reference Manual*, 2013. `https://www.xilinx.com/support/documentation/university/XUP%20Boards/XUPAtlys/documentation/Atlys_rm.pdf`.

[31] DIGILENT: *Digilent Atlys Development Baord*, 2018. `https://store.digilentinc.com/atlys-spartan-6-fpga-trainer-board-limited-time-see-nexys-video/`.

[32] DIGILENT: *Digilent Nexys Development Baord*, 2018. `https://store.digilentinc.com/nexys-video-artix-7-fpga-trainer-board-for-multimedia-applications/`.

[33] D.RENKER and E.LORENZ: *Advances in solid state photon detectors*. Journal of Instrumentation, 4(04):P04004, 2009.

[34] EINSTEIN, A.: *Ueber einen die Erzeugung und Verwandlung des Lichtes betreffenden heuristischen Gesichtspunkt*. Annalen der Physik, 322(6):132–148.

[35] FELIX B. MUELLER, H. Y.: *Commissioning at KEK*, 2015. `https://indico.desy.de/indico/event/21497/session/6/contribution/8/material/slides/0.pdf`.

[36] FEYNMAN, RICHARD, L. R. S. M.: *The Feynman Lectures on Physics*. Basic Books, 250 West 57th Street, 15th Floor, New York, NY, 10107, Volume III ed., 2010.

[37] FIGUERAS, E. V.: *Feasibility of Geiger-mode avalanche photodiodes in CMOS standard technologies for tracker detectors*. PhD, p. 205, 2013.

[38] FISHBURN, M.: *Fundamentals of CMOS Single-Photon Avalanche Diodes*. Dissertation, Technische Universitaet Delft, Delft, 2012.

[39] FTDI: *FT2232H - Hi-Speed Dual USb UART/FIFO IC*, 2018. `http://www.ftdichip.com/Products/ICs/FT2232H.htm`.

[40] FTDI: *Future Technology Devices Internation Ltd.*, 2018. `https://www.ftdichip.com/`.

[41] GEMMEKE, H., L. BERGER, T. HOPP, M. ZAPF, W. TAN, R. BLANCO, R. LEYS, I. PERIC, and N. V. RUITER: *The new generation of the kit 3d usct*. In *Proceedings of the International Workshop on Medical Ultrasound Tomography: 1.- 3. Nov. 2017, Speyer, Germany. Hrsg.: T. Hopp*, pp. 271–282. KIT Scientific Publishing, Karlsruhe, 2018. 54.02.02; LK 01.

[42] GEMMEKE, H. and N. V. RUITER: *3 d ultrasound computer tomography for medical imaging*. 2007.

[43] HAIDL, J.: *Readout Optimization for the Belle II Pixel Detector - Characterization by ADC Transfer Curves*, 2016. `http://docs.belle2.org/record/352/files/`.

[44] HAMAMATSU: *MPPC and MPPC module for precision measurement*, 2016. `https://www.hamamatsu.com/resources/pdf/ssd/mppc_kapd0004e.pdf`.

[45] HERTZ, H.: *Ueber einen einfluss des ultravioletten lichtes auf die electrische entladung*. Annalen der Physik, 267(8):983–1000.

[46] IMANANI, B.: *Characterization of Multichannel ASIC for Ultrasound Computer Tomography*, 2016.

[47] KEMMER, J. and G. LUTZ: *New detector concepts*. Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, 253(3):365 – 377, 1987.

[48] KOCH, M.: *Development of a Test Environment for the Characterization of the Current Digitizer Chip DCD2 and the DEPFET Pixel System for the Belle II Experiment at SuperKEKB*. PhD thesis, Bonn U., 2011.

[49] KOLANOSKI, H. and N. WERMES: *Teilchendetektoren - Grundlagen und Anwendungen*. Springer-Verlag, Berlin Heidelberg New York, 2016.

[50] KORZECZEK, M.: *ev- and kev-sterile neutrino studies with katrin*. Master Thesis, 2016.

[51] KUBETSKY, L.: *Author's certificate no. 24040, priority date 4 august 1930*. 314(24040).

[52] LACAITA, A., S. COVA, A. SPINELLI, and F. ZAPPA: *Photon-assisted avalanche spreading in reach through photodiodes*. Applied Physics Letters, 62(6):606–608, 1993.

[53] LACAITA, A. and M. MASTRAPASQUA: *Strong dependence of time resolution on detector diameter in single photon avalanche diodes*. Electronics Letters, 26(24):2053–2054, Nov 1990.

[54] LACAITA, A., M. MASTRAPASQUA, M. GHIONI, and S. VANOLI: *Observation of avalanche propagation by multiplication assisted diffusion in pn junctions*. Applied Physics Letters, 57(5):489–491, 1990.

[55] LACAITA, A. L., F. ZAPPA, S. BIGLIARDI, and M. MANFREDI: *On the bremsstrahlung origin of hot-carrier-induced photons in silicon devices*. IEEE Transactions on Electron Devices, 40(3):577–582, March 1993.

[56] LEMARENKO, M., T. HEMPEREK, H. KRUEGER, M. KOCH, F. LUETTICKE, C. MARINAS, and N. WERMES: *Test results of the data handling processor for the depfet pixel vertex detector*. Journal of Instrumentation, 8(01):C01032, 2013.

[57] LEYS, R.: *Raising the abstraction level of hardware software co-designs*. PhD, p. 196, 2015.

[58] LILIENFELD, J.: *Method and apparatus for controlling electric currents*, 1930. `https://worldwide.espacenet.com/publicationDetails/originalDocument?FT=D&date=19300128&DB=&locale=de_EP&CC=US&NR=1745175A&KC=A&ND=1`.

[59] LUTZ, G.: *Semiconductor Radiation Detectors - Device Physics*. Springer Science and Business Media, Berlin Heidelberg, 2001.

[60] MAKIDA, Y., H. YAMAOKA, Y. DOI, J. HABA, F. TAKASAKI, and A. YAMAMOTO: *Development of a Superconducting Solenoid Magnet System for the B-Factory Detector (BELLE)*, pp. 221–228. Springer US, Boston, MA, 1998.

[61] MARTIN, P. A., B. G. STREETMAN, and K. HESS: *Electric field enhanced emission from non coulombic traps in semiconductors*. Journal of Applied Physics, 52(12):7409–7415, 1981.

[62] MATTIA ASSANELLI, ANTONINO INGARGIOLA, I. R. A. G. M. G.: *Photon-timing jitter dependence on the injection position in single-photon avalanche diodes*. Vol. 7681, pp. 7681 – 7681 – 11, 2010.

[63] MENSHIKOV, A.: *USCT Project Transduceer Array Head, Karlsruhe Institue of Technology, Insitute for Data Processing and Electronics*, 2008. .

[64] MERTENS, S., T. LASSERRE, S. GROH, G. DREXLIN, F. GLUECK, A. HUBER, A. POON, M. STEIDL, N. STEINBRINK, and C. WEINHEIMER: *Sensitivity of next-generation tritium beta-decay experiments for kev-scale sterile neutrinos*. Journal of Cosmology and Astroparticle Physics, 2015(02):020, 2015.

[65] MOORE, G.: *Cramming more components onto integrated circuits*. Electronics, 38, 1965.

[66] MOSER, H.-G.: *The belle ii depfet pixel detector*. Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, 831:85 – 87, 2016. Proceedings of the 10th International Hiroshima Symposium on the Development and Application of Semiconductor Tracking Detectors.

[67] MÜLLER, F.: *Characterization and optimization of the prototype depfet modules for the belle ii pixel vertex detector*. Juli 2017.

[68] MULLER, R. S., T. I. KAMINS, and M. CHAN: *Device Electronics for Integrated Circuits* -.

Wiley, New York, 2003.

[69] PELLION, D., K. JRADI, N. BROCHARD, D. PRÃªLE, and D. GINHAC: *Single-photon avalanche diodes (spad) in cmos 0.35um technology*. Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, 787:380 – 385, 2015. New Developments in Photodetection NDIP14.

[70] PERIĆ, I.: *Design and realisation of integrated circuits for the readout of pixel sensors in high-energy physics and biomedical imaging*. PhD, p. 191, 2004.

[71] PERIC, I., T. ARMBRUSTER, M. KOCH, C. KREIDL, and P. FISCHER: *Dcd the multi channel current mode adc chip for the readout of depfet pixel detectors*. IEEE Transactions on Nuclear Science, 57(2):743–753, April 2010.

[72] PERIC, I., P. FISCHER, J. KNOPF, and T. H. H. NGUYEN: *Dcdb and switcherb, the readout asics for belle ii depfet pixel detector*. 2011 IEEE Nuclear Science Symposium Conference Record, pp. 1536–1539, 2011.

[73] PIERRE CURIE, J.: *Development, via compression, of electric polar- ization in hemihedral crystals with inclined faces*. Bulletin de la Societe minerologique de France, pp. 90–93, 1880.

[74] QIN, B., M. ZAPF, E. KRETZEK, X. MA, and N. V. RUITER: *Fast detection of breast position for 3d usct*. In *2014 IEEE International Ultrasonics Symposium*, pp. 1896–1899, Sept 2014.

[75] RAZAVI, B.: *Fundamentals of Microelectronics, 2nd Edition* -. Wiley Global Education, 2013.

[76] RAZAVI, B.: *Design of Analog CMOS Integrated Circuits*. McGraw-Hill Professional, 2 Penn Plaza. New York, NY 10121, 2nd uk ed. ed., 2016.

[77] RECH, I., A. INGARGIOLA, R. SPINELLI, I. LABANCA, S. MARANGONI, M. GHIONI, and S. COVA: *Optical crosstalk in single photon avalanche diode arrays: a new complete model*. Opt. Express, 16(12):8381–8394, Jun 2008.

[78] RIVETTI, A.: *CMOS Front-End Electronics for Radiation Sensors*. CRC Press, 2015.

[79] RUMMEL, S.: *Investigation of DEPFET as Vertex Detector at ILC Intrinsic properties, radiation hardness and alternative readout schemes*. Dissertation, Technische Universitaet Muenchen, Muenchen, 2009.

[80] SCHWARZENBERG, G. F.: *Untersuchung der Abbildungseigenschaften eines 3D-Ultraschall-Computertomographen zur Berechnung der 3D-Abbildungsfunktion und Herleitung einer optimierten Sensorgeometrie*. PhD thesis, 2009.

[81] SHOCKLEY, W.: *A unipolar "field-effect" transistor*. Proceedings of the IRE, 40(11):1365–1376, Nov 1952.

[82] SINGHAL, G.: *Extending a Characterization Set-Up for USCT ASIC on both Hardware and Software*, 2017.

[83] SZE, S. M. and K. K. NG: *Physics of Semiconductor Devices* -. John Wiley Sons, New York, 2006.

[84] TAKEUCHI, K. and M. FUKUMA: *Effects of the velocity saturated region on mosfet charac-*

teristics. IEEE Transactions on Electron Devices, 41(9):1623–1627, Sep 1994.

[85] TISA, S., F. GUERRIERI, and F. ZAPPA: *Variable-load quenching circuit for single-photon avalanche diodes*. Opt. Express, 16(3):2232–2244, Feb 2008.

[86] TSIVIDIS, Y.: *Operation and Modeling of the MOS Transistor*. WCB/McGraw-Hill, Boston, 1999.

[87] VINCENT, G., A. CHANTRE, and D. BOIS: *Electric field effect on the thermal emission of traps in semiconductor junctions*. Journal of Applied Physics, 50(8):5484–5487, 1979.

[88] WESTE, N. H. E. and D. M. HARRIS: *CMOS VLSI Design - A Circuits and Systems Perspective*. Addison Wesley, Amsterdam, 2011.

[89] XU, H., L. PANCHERI, L. H. BRAGA, G.-F. D. BETTA, and D. STOPPA: *Crosstalk characterization of single-photon avalanche diode (spad) arrays in cmos 150nm technology*. Procedia Engineering, 87:1270 – 1273, 2014. EUROSENSORS 2014, the 28th European Conference on Solid-State Transducers.

[90] YAACOB, M., M. R. ARSHAD, and A. ABD MANAF: *Modeling of circular piezoelectric micro ultrasonic transducer using cual10ni5fe4 on zno film for sonar applications*. 57:151–158, 03 2011.

[91] ZAPPA, F., A. LOTITO, A. C. GIUDICE, S. COVA, and M. GHIONI: *Monolithic active-quenching and active-reset circuit for single-photon avalanche detectors*. IEEE Journal of Solid-State Circuits, 38(7):1298–1301, July 2003.

Publications

[AAA+18] F. Abudinen, K. Ackermann, P. Ahlburg, M. Albalawi, O. Alonso, L. Andricek, R. Ayad, V. Babu, Y. Bai, T. Bilka, R. Blanco, M. Boronat, A. Bozek, C. Camien, A. Caldwell, V. Chekelian, B. Deschamps, A. Dieguez, J. Dingfelder, Z. Dolezal, D. Esperante, M. Fras, A. Frey, J. Fuster, M. Gabriel, K. Gadow, U. Gebauer, L. Germic, T. Gessler, D. Getzkow, L. Gioi, P. Gomis, M. Heck, T. Hemperek, M. Hensel, M. Hoek, J. Kandra, P. Kapusta, C. Kiesling, B. Kisielewski, D. Kittlinger, D. Klose, P. Kodys, C. Koffmane, I. Konorov, S. Krivokuca, H. Krueger, T. Kuhr, W. Kuehn, P. Kvasnicka, C. Lacasta, J.S. Lange, K. Lautenbach, U. Leis, P. Leitl, D. Levit, G. Liemann, Z. Liu, F. Luetticke, L. Macharski, C. Marinas, S. Mccarney, H.G. Moser, D. Moya, F.J. Mueller, F. Mueller, D. Muenchow, C. Niebuhr, J. Ninkovic, U. Packheiser, B. Paschen, S. Paul, I. Peric, F. Poblotzki, A. Rabusov, S.P. Reiter, R. Richter, M. Ritter, M. Ritzert, S. Rummel, J.G. Sanchez, B. Scavino, G. Schaller, M. Schnecke, F. Schopper, H. Schreeck, B. Schwenker, R. Sedlmeyer, C. Sfienti, F. Simon, S. Skambraks, Y. Soloviev, B. Spruck, R. Stever, U. Stolzenberg, M. Takahashi, E. Tafelmayer, I. Vila, A.L. Virto, S. Vogt, M. Vos, C. Wang, N. Wermes, C. Wessel, P. Wieduwilt, H. Windel, H. Ye, and J. Zhao. Depfet pixel detector in the belle ii experiment. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 2018.

[BKP18] R. Blanco, C. Kraemer, and I. Peric. Silicon photomultiplier detector with multipurpose in-pixel electronics in standard cmos technology. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 2018.

[BLP18] R. Blanco, R. Leys, and I. Peric. Integrated readout electronics for belle ii pixel detector. *Journal of Instrumentation*, 13(03):C03001, 2018.

[BZK+17] R. Blanco, H. Zhang, C. Kraemer, F. Ehrler, R. Schimassek, R. Casanova Mohr, E. Vilella Figueras, F. Guezzi Messaoud, R. Leys, M. Prathapan, A. Weber, and I. Peric. Hvcmos monolithic sensors for the high luminosity upgrade of atlas experiment. *Journal of Instrumentation*, 12(04):C04001, 2017.

[EBLP16] F. Ehrler, R. Blanco, R. Leys, and I. Peric. High-voltage cmos detectors. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 824:400 – 401, 2016. Frontier Detectors for Frontier Physics: Proceedings of the 13th Pisa Meeting on Advanced Detectors.

[FBS+15] Peter Fischer, Roberto Blanco, Ilaria Sacco, Michael Ritzert, and Sascha Weyers. Spad array chips with full frame readout for crystal characterization. *EJNMMI Physics*, 2(1):A3, May 2015.

[GRH+18] HH. Gemmeke, N. V. Ruiter, T. Hopp, M. Zapf, I. Peric, L. Berger, W. Y. Tan, R. Blanco, and R. Leys, editors. *Proceedings of the International Workshop on Medical Ultrasound Tomography: 1.- 3. Nov. 2017, Speyer, Germany*. KIT Scientific Publishing, Karlsruhe, 2018. 54.02.02; LK 02.

[KAA+18] P. Kodys, F. Abudinen, K. Ackermann, P. Ahlburg, H. Aihara, M. Albalawi, O. Alonso, L. Andricek, R. Ayad, T. Aziz, V. Babu, S. Bacher, S. Bahinipati, Y. Bai, E. Barberio, Ti. Baroncelli, To. Baroncelli, A.K. Basith, G. Batignani, A. Bauer, P.K. Behera, V. Bertacchi, S. Bettarini, B. Bhuyan, T. Bilka, R. Blanco, F. Bosi, M. Boronat, L. Bosisio, A. Bozek, F. Buchsteiner, C. Camien, A. Caldwell, G. Caria, G. Casarosa, M. Ceccanti, D. Cervenkov, V. Chekelian, T. Czank, N. Dash, M. De Nuccio, B. Deschamps, A. Dieguez, J. Dingfelder, Z. DoleÅ¾al, D. Esperante, P. Fischer, F. Forti, M. Fras, A. Frey, M. Friedl, J. Fuster, M. Gabriel, K. Gadow, U. Gebauer, L. Germic, T. Gessler, D. Getzkow, L. Gioi, B. Gobbo, P. Gomis, J.A.M. Grimaldo, K. Hara, M. Heck, T. Hemperek, M. Hensel, T. Higuchi, M. Hoek, C. Irmler, A. Ishikawa, H.B. Jeon, C. Joo, M. Kaleta, J. Kandra, N. Kambara, K.H. Kang, P. Kapusta, C. Kiesling, B. Kisielewski, D. Kittlinger, D. Klose, C. Koffmane, T. Kohriki, S. Koike, I. Komarov, I. Konorov, S. Krivokuca, H. Krueger, T. Kuhr, W. Kuehn, M. Kumar, R. Kumar, W. Kun, P. Kvasnicka, C. La Licata, C. Lacasta, K. Lalwani, L. Lanceri, J.S. Lange, K. Lautenbach, J.Y. Lee, S.C. Lee, U. Leis, P. Leitl, D. Levit, J. Libby, G. Liemann, Z. Liu, T. Lueck, F. Luetticke, L. Macharski, P. Mammini, C. Marinas, A. Martini, S.N. Mayekar, S. Mccarney, G.B. Mohanty, T. Morii, H.G. Moser, D. Moya, F.J. Mueller, F. Mueller, D. Muenchow, K.R. Nakamura, Z. Natkaniec, C. Niebuhr, J. Ninkovic, Y. Onuki, W. Ostrowicz, U. Packheiser, A. Paladino, E. Paoloni, H. Park, B. Paschen, S. Paul, I. Peric, F. Poblotzki, K. Prasanth, A. Profeti, A. Rabusov, I. Rashevskaya, K.K. Rao, S.P. Reiter, Resmi P.K., R. Richter, M. Ritter, M. Ritzert, G. Rizzo, M. Rozanska, S. Rummel, D. Sahoo, J.G. Sanchez, J. Sasaki, N. Sato, B. Scavino, G. Schaller, M. Schnecke, F. Schopper, H. Schreeck, S. Schultschik, C. Schwanda, B. Schwenker, R. Sedlmeyer, C. Sfienti, F. Simon, S. Skambraks, Y. Soloviev, B. Spruck, R. Stever, U. Stolzenberg, J. Stypula, J. Suzuki, E. Tafelmayer, M. Takahashi, S. Tanaka, H. Tanigawa, G.N. Taylor, R. Thalmeier, T. Tsuboyama, P. Urquijo, I. Vila, A.L. Virto, L. Vitale, S. Vogt, M. Vos, C. Wang, S. Watanuki, M. Watanabe, I.J. Watson, J. Webb, N. Wermes, C. Wessel, J. Wiechczynski, P. Wieduwilt, S. Williams, H. Windel, H. Ye, H. Yin, L. Zani, and J. Zhao. The belle ii vertex detector integration. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 2018.

[PBM+17] I. Peric, R. Blanco, R. Casanova Mohr, F. Ehrler, F. Guezzi Messaoud, C. Kraemer, R. Leys, M. Prathapan, R. Schimassek, A. Schoening, E. Vilella Figueras, A. Weber, and H. Zhang. Status of hvcmos developments for atlas. *Journal of Instrumentation*, 12(02):C02030, 2017.

[PELB15] I. Peric, F. Ehrler, R. Leys, and R. Blanco. Hvcmos pixel sensors. In *2015 IEEE SENSORS*, pages 1–4, Nov 2015.

[SBC+18] R. Schimassek, R. Blanco, R. Casanova, F. Ehrler, I. Peric, E. Vilella, and H. Zhang. Monolithic sensors in lfoundry technology: Concepts and measure-

ments. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 2018.

Supervised Student Research

[Bec18] BECKER, LUKAS: *Automated Characterization of an ASIC for Ultrasound Transceivers*, 2018. Bacherlor Thesis, Karlsruhe Institute of Technology and University of Applied Sciences Karlsruhe.

[Hau18] HAUER, SVEN: *Development of a High-Bandwidth Current Sensor for High-Frequency Power Applications*, 2018. Master Thesis, Karlsruhe Institute of Technology.

[Kra16] KRAEMER, CHRISTIAN: *Development of Capacitive Coupled CMOS Sensors*, 2016. Bachelor Thesis, Karlsruhe Institute of Technology.

[Sin17] SINGHAL, GARIMA: *Extending a Characterization Set-Up for USCT ASIC on both Hardware and Software*, 2017. Internship Report, Karlsruhe Institute of Technology.

Teaching

- WS 2016/17: Exercises for Analog Circuit Design
- WS 2017/18: Exercises for Analog Circuit Design
- WS 2018/19: Exercises for Analog Circuit Design

Content:

- Integrated components (MOSFET)
- Amplifier design and feedback theory
- DC characteristics of NMOS and PMOS transistors
- Design of input, output stages and current mirrors
- Frequency response and stability criterion
- Noise in integrated circuits

KSETA Attended Courses

KSETA is the graduate school associated with the KIT Center Elementary Astroparticle Physics (KCETA). It combines experimental and theoretical research and education at the interface between astronomy, astrophysics, elementary particle physics and cosmolology. The following courses were attended during the time at the graduate school KSETA:

- **Introduction to String theory**
- **Introduction to Cosmology**
- **Particle Propagation in Astroparticle Physics**
- **Memorizing, Reading and Working Strategies**
- **Particle Physics for Engineers**
- **Data Visualization and Presenting**