TECHNISCHE UNIVERSITÄT WIEN

HEPHY INSTITUT FÜR HOCHENERGIEPHYSIK

DIPLOMARBEIT

Design of a Charge Sensitive Amplifier for Particle Detection in Silicon Carbide CMOS Technology

zur Erlangung des akademischen Grades

Diplom-Ingenieur

im Rahmen des Studiums

Technische Physik

eingereicht von

Sebastian Onder
Matrikelnummer 01621235

ausgeführt am Atominstitut
der Fakultät für Physik der Technischen Universität Wien
(in Zusammenarbeit mit dem Institut für Hochenergiephysik (HEPHY))

Betreuung
Betreuer: Ass.Prof. Dipl.-Ing. Dr. techn. Albert Hirtl (Atominstitut)
Mitwirkung: Univ.Lektor Dipl.-Ing. Dr. techn. Thomas Bergauer (HEPHY)

Wien, 13.09.2024

(Unterschrift Verfasser) (Unterschrift Betreuer)

Abstract

This thesis presents the design of a charge sensitive amplifier (CSA) circuit for the readout of particle sensors in high-energy physics using a 2 µm CMOS process in silicon carbide (SiC). This work aims to lay a first step towards testing the suitability of SiC CMOS electronics for detecting high-energetic particles. The knowledge gained by the design of the amplifier could aid the development of a Monolithic Active Pixel Sensor (MAPS) in SiC in the far future. MAPS offer excellent spatial and timing resolution as well as fast data acquisition paired with a compact and cost-efficient integration of the sensor and readout electronics on the same chip using CMOS technology. However, the silicon material conventionally used for MAPS shows considerable damage and defects after being exposed to high amounts of irradiation, which has a negative effect on the performance of the sensor. SiC, as a wide-bandgap material, potentially offers a better performance in terms of radiation hardness. The integration of MAPS in SiC using a SiC CMOS process could, therefore, combine the performance and compactness of MAPS with the potential endurance of SiC.

The design of the CSA was done using the Fraunhofer IISB 2 µm SiC CMOS process. A layout design of the CSA circuit was submitted as part of a multi-project wafer production run via Europractice and Fraunhofer IISB. The delivery time of the chip is scheduled after the completion of this thesis. Therefore, no design validation measurement could be conducted on the physical chip as part of this work.

Simulations of the circuit show that an equivalent noise charge of 107 e^- is achievable with an input capacitance of 1 pF. For the readout of a thin sensor using the epitaxial layer of a chip developed using the Fraunhofer IISB SiC CMOS process and an expected signal charge of 660 e^- for a minimum ionizing particle, a signal-to-noise ratio (SNR) of roughly 6 could be achieved. With a thicker sensor, the SNR could be improved significantly.

The technology is still in an early stage of development, resulting in relatively large feature sizes for electronic devices. This imposes a significant limitation on the circuit's bandwidth due to the introduction of high parasitic capacitances. With an achieved bandwidth of roughly 31 kHz at a gain of roughly 51 dB, there is still room for improvement when comparing the designed CSA to its silicon counterparts. But, with the maturing of SiC CMOS technology, a reduction in feature size and an improvement in the performance of future amplifier circuits can likely be expected.

Kurzfassung

Diese Arbeit befasst sich mit dem Design eines Ladungsverstärkers für das Auslesen von Teilchensensoren in der Hochenergiephysik unter Verwendung eines 2 µm CMOS-Prozesses in Siliziumkarbid (SiC). Ziel dieser Arbeit ist es, einen ersten Schritt zu setzen, um die Eignung von SiC-CMOS-Elektronik für den Nachweis hochenergetischer Teilchen zu testen. Die durch das präsentierte Design des Verstärkers gewonnenen Erkenntnisse könnten in Zukunft die Entwicklung eines monolithischen aktiven Pixelsensors (MAPS) in SiC unterstützen. MAPS bieten eine hervorragende räumliche und zeitliche Auflösung sowie eine schnelle Datenerverarbeitung, gepaart mit einer kompakten und kosteneffizienten Integration von Sensor und Ausleseelektronik auf demselben Chip mithilfe von CMOS-Technologie. Allerdings weist das normalerweise für MAPS verwendete Siliziummaterial nach hoher Bestrahlung erhebliche Schäden und Defekte auf, was sich negativ auf die Leistung des Sensors auswirkt. SiC, als ein Material mit einer größeren Bandlücke als Silizium, ist möglicherweise eine bessere Alternative in Bezug auf Resistenz gegen extreme Strahlenbelastung. Die Integration von MAPS in SiC unter Verwendung eines SiC-CMOS-Prozesses könnte daher die Leistung und Kompaktheit von MAPS mit der potenziellen Beständigkeit von SiC kombinieren.
Das Design des CSA wurde mit dem 2 µm SiC-CMOS-Prozess des Fraunhofer IISB durchgeführt. Ein Layoutentwurf der CSA-Schaltung wurde im Rahmen einer Multi-Project-Wafer-Produktion über Europractice und das Fraunhofer IISB eingereicht. Die Auslieferung des Chips ist nach der Fertigstellung dieser Arbeit vorgesehen. Daher konnten im Rahmen dieser Arbeit keine Design-Validierungsmessungen am physischen Chip durchgeführt werden.
Simulationen der Schaltung zeigen, dass mit einer Eingangskapazität von 1 pF eine äquivalente Rauschladung von 107 e^- erreicht werden kann. Für das Auslesen eines dünnen Sensors unter Verwendung der Epitaxieschicht eines im Fraunhofer IISB SiC CMOS-Prozess entwickelten Chips und einer erwarteten Signalladung von 660 e^- konnte ein Signal-zu-Rausch Verhältnis von etwa 6 erreicht werden. Mit einem dickeren Sensor könnte das Signal-zu-Rausch Verhältnis deutlich verbessert werden.
Die Technologie befindet sich noch in einem frühen Entwicklungsstadium, was aktuell relativ große elektronische Bauteile zur Folge hat. Dadurch wird die Bandbreite der Schaltung durch die Einführung hoher parasitärer Kapazitäten stark begrenzt. Mit einer erreichten Bandbreite von etwa 31 kHz bei einer Verstärkung von etwa 51 dB gibt es noch Raum für Verbesserungen, wenn man den entworfenen CSA mit seinen Silizium-Pendants vergleicht. Mit der Weiterentwicklung von SiC-CMOS-Technologie ist jedoch eine Verringerung der Bauteilgröße und eine Verbesserung der Leistung künftiger Verstärkerschaltungen zu erwarten.

Contents

1 Introduction

Semiconductor detectors have long been recognized as an indispensable part of particle detection systems in high-energy physics experiments, delivering exceptional particle tracking capabilities. To date, silicon has been the predominant material for semiconductor sensor fabrication, mostly due to the mature and advanced manufacturing processes available for it. Silicon-based sensors have continually shown that they can meet the strict requirements at large-scale collider experiments for many years. However, the landscape of particle colliders is evolving, with future colliders expected to operate under increasingly high particle rates and luminosities, which places greater demands on the performance of detection systems. The next generation of colliders will require sensors that offer not only superior spatial and temporal resolution as well as rapid data acquisition capabilities but also a higher radiation hardness to tolerate the increasing amount of particle intensities [1, 2, 3].
Silicon-based sensors, while generally capable of providing the necessary performance, face a significant degradation of performance due to damages to the silicon material when exposed to extreme radiation conditions. This has shifted interest towards alternative materials, particularly wide-bandgap semiconductors, which could potentially meet the requirements for radiation hardness in the face of high radiation exposure. As such, silicon carbide (SiC) is seen as a potential candidate for such a material [4, 5]. The demand for SiC in the industrial sector due to its convenient material properties has caused a better availability for high-quality SiC wafers and manufacturing processes [6, 7]. As a benefit for particle detection, SiC offers a lower leakage current and higher thermal conductivity than silicon, causing low noise also at room temperature and reducing the need for cooling. Even after irradiation, SiC can maintain leakage currents orders of magnitudes lower than irradiated silicon. A higher breakdown field and charge carrier saturation velocity could benefit the timing capabilities of detectors [5, 8, 9].
The push for enhanced lateral resolution and faster data acquisition has encouraged the development of pixel sensors. Currently, hybrid pixel sensors lead the way in terms of performance, as they decouple the fabrication of the sensing element from the readout electronics. This separation allows for optimized designs tailored to specific applications. However, the assembly of hybrid pixel sensors by combining the sensor and readout chip via bump-bonding is quite complex and expensive [1]. Monolithic sensor technologies, including monolithic active pixel sensors (MAPS), are emerging as alternatives to the currently dominant hybrid pixel architecture [10, 11]. These technologies integrate the sensor and readout electronics into a single semiconductor substrate with the use of CMOS integrated chip technologies, offering a more compact design by removing an additional layer of silicon and thus greater cost efficiency.
However, the silicon material used in CMOS processes for the development of MAPS is faced with the aforementioned challenges regarding radiation hardness. With the emergence of commercial SiC CMOS processes, a first step towards a SiC-MAPS with potentially better radiation hardness can be made. While these SiC-based processes do not yet offer the processing power of classic silicon CMOS processes, it is worth exploring the cost advantages of MAPS in combination with the promising properties of SiC. There already have been several studies on the development of integrated electronics [12] as well as sensors for the detection of temperature [13] and UV-light [14] in a SiC CMOS

process. For the application in the detection of particles in high-energy physics and the development of a MAPS, a first step has to be taken towards testing the capabilities of SiC-based CMOS electronics for the readout of particle detectors.
The work presented in this thesis is the design of a charge sensitive amplifier (CSA) for the detection of high-energetic particles with a commercial 2 µm SiC CMOS process. This includes the analysis of the used circuit topology, a definition of the circuit performance requirements for particle detection, and a documentation of the circuit and layout design process. The nominal performance of the circuit is discussed based on simulation results.

2 Semiconductor Particle Detectors

Semiconductor particle sensors are crucial tools in high-energy physics experiments, designed to detect and measure charged particles with high precision. The passage of a high-energy particle through the sensor material generates electron-hole pairs. By applying an electric field, the charges are collected and converted into an electrical signal, which can be processed through specialized readout electronics to determine the particle's properties, such as position, energy and momentum. As sensor material, silicon is by far the most widely used material. However, due to favorable material parameters, wide band-gap materials like SiC recently gained more attention. Since semiconductor particle detectors provide the electrical stimuli processed by the corresponding readout electronics, this section introduces their function principle and properties that need to be considered in the design process of a designated amplifier circuit.

2.1 Semiconductors and Diodes

Semiconductor materials can be distinguished from other solid materials by their band gap, i.e. the energy needed to promote atomic electrons from the valence band to the conduction band, which lies between the conductors and insulators. This is shown in Fig. 2.1. Insulators have a very large band gap, causing an effectively vanishing electric conductivity. On the other hand, conductors have an overlapping valence and conduction band, making them conductive also at low temperatures. Semiconductors have a band gap much smaller than the one of isolators with energies in the scale of a few eV. This allows electrons to occupy states in the conduction band by thermal excitation, which leads to a non-zero conductivity.

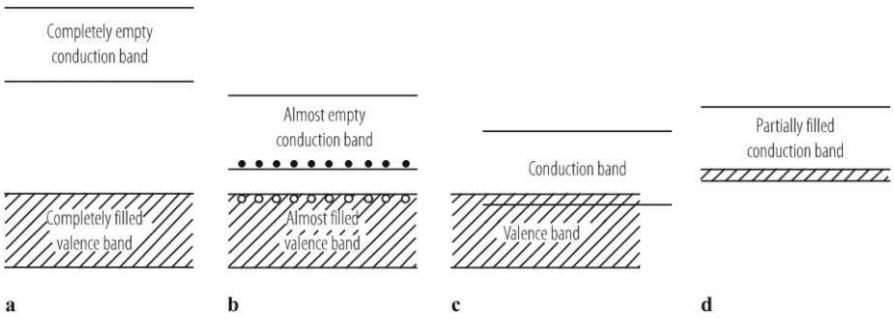


Figure 2.1: Visualization of the energy band structure of (a) insulators, (b) semiconductors, and (c), (d) conductors/metals. Image adapted from [15].

Introducing impurities into the semiconductor lattice structure can create energy states in the semiconductor band gap, allowing for easier promotion of electrons to the conduction band, leaving behind a so-called hole in the valence band. A conductive electron, along with an associated hole, is commonly referred to as an electron-hole pair. The intentional process of introducing substitutional atoms, taking the place of an atom of the original semiconductor, to alter the electric properties of the semiconductor is called doping. It distinguishes between two types of dopants: donors and acceptors.

Donor atoms introduce an excess of electrons relative to the semiconductor atoms. The excess electron cannot form covalent bonds with the electrons of the neighboring atoms in the lattice. It can, therefore, be easily removed from its atom, creating a higher number of free electrons in the conduction band. Semiconductors rich in donor atoms are called n-type semiconductors. Acceptor atoms introduce less electrons than the semiconductor atoms have, creating an excess of holes in the valence band, which free electrons can fill. Semiconductors rich in acceptor atoms are called p-type semiconductors.
Bringing an n-type semiconductor into contact with a p-type semiconductor forms a so-called pn-junction, which is the simplest form of semiconductor diode. The difference in concentration of electrons and holes across the pn-junction leads to a diffusion of free electrons from donor atoms across the junction. The free electrons leave behind positively charged donor ions. The electrons recombine with the holes in the p-type material, leaving behind negatively charged acceptor ions. The resulting oppositely charged ions on both sides of the pn-junction lead to a polarisation of the material and, consequently, to an electric field that counteracts the charge carrier diffusion. A thermal equilibrium is established when the electric field becomes strong enough to prevent further diffusion. This creates a region around the pn-junction called the depletion region, which is free of mobile charge carriers.
Fig. 2.2 depicts the formation of the depletion region without external voltage in the top picture. The width of the depletion region can be controlled by applying an external bias voltage to the opposite sides of the pn-junction, which is also shown in Fig. 2.2 in the center and bottom pictures. With forward bias, i.e. a positive voltage applied to the p-doped region and a negative voltage applied to the n-doped region, the external electric field compensates for the built-in field. Therefore, the depletion zone shrinks, and the charge is free to travel along the external electric field, making the diode conductive.

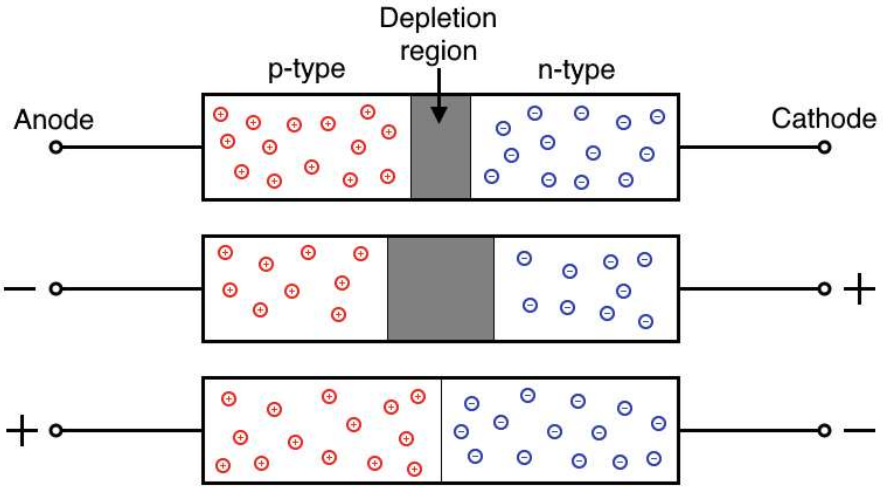


Figure 2.2: Manipulation of the depletion region of a diode without external bias (top), reverse bias (center) and forward bias (bottom).

With a reverse bias voltage, i.e. a negative voltage applied to the p-doped region and a positive voltage applied to the n-doped region, the external electric field forces more electrons to drift towards the p-doped material and more holes towards the n-doped

material. This depletes more free charge, the depletion region widens, and no current can flow across the pn-junction. However, free charge carriers nonetheless exist in the depletion region due to the thermal excitation of electrons. These electrons can drift along the externally applied electric field and produce a small so-called leakage current. As the depletion region of a pn-junction is filled with an electric field between two electrodes, it also holds an electric capacitance. The depletion region can be seen as a plate capacitor, with a capacitance C calculated by

$$C = \epsilon_{rel}\epsilon_0 \frac{A}{d}, \tag{2.1}$$

where ϵ_{rel} is the relative permittivity of the material, $\epsilon_0 = 8.854 \cdot 10 \cdot 10^{-12}\,\mathrm{AsV^{-1}m^{-1}}$ the dielectric constant, A the effective cross-sectional area of the depletion region and d the thickness of the depletion region.

2.2 Semiconductor Diodes as Particle Sensors

Semiconductor particle detectors that are used in high-energy physics are essentially diodes operated in reverse bias. A simple planar sensor pixel structure is shown in Fig. 2.3. On the top side a highly doped p^{++}-doped implant is brought into contact with a n-type bulk substrate with a doping concentration several orders of magnitudes lower than the implant. On the back-side, a n^{++}-doped implant limits the thickness of the n-bulk. The pn-junction is formed at the interface of the n-bulk and the p^{++}-implant. Due to the high difference in doping concentration, applying a reverse bias voltage causes the depletion region to mostly fill the n-bulk. The depletion region grows until the n^{++}-implant starts to get depleted, which dramatically slows down the growth of the depletion with increasing reverse bias voltage due to its high doping concentration. The n^{++}-implant is not fully depleted in practice.

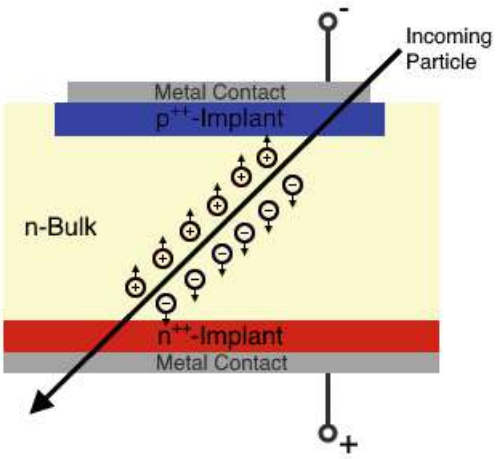


Figure 2.3: Structure of a simple semiconductor sensor pixel and the deposition of electron-hole pairs by an incoming high energetic particle.

A sensor with a depletion region filling the whole bulk substrate is referred to as fully

depleted. The bulk is free from charge carriers up to the small leakage current. However, charges can still be created by letting the sensor material interact with highly energetic particles. If such a particle traverses the sensor material, it scatters at atomic electrons or nuclei. The scattering at atomic electrons due to electromagnetic interactions can cause the particle to carry some of its kinetic energy over to the electron. If this energy is high enough to lift the electron from the valence band to the conduction band, it can be removed from the semiconductor atom. The atom becomes ionized and the electron leaves a hole behind. The now free negatively charged electron immediately experiences a drift along the externally applied electric field to the high potential electrode, i.e. towards the n^{++}-implant. The positively charged hole experiences a drift in the opposite direction as the electron towards the p^{++}-implant. This process is also sketched in Fig. 2.3.

According to the Shockley-Ramo theorem [16, 17], the drift of a charge carrier in an electric field between two or several electrodes induces a current i_S on the electrodes that can be measured as a signal:

$$i_S = -\frac{dQ}{dt} = -q\vec{E_w}\vec{v}, \tag{2.2}$$

where Q is the amount of charge induced on the electrode, q is the elementary charge of the drifting charge, and $\vec{E_w}$ is the weighting field describing the coupling of the charge to a specific electrode. $\vec{v}$ is the velocity of the charge carrier and can be determined by

$$v(\vec{x}) = \mu\vec{E}(x),$$

where μ is the mobility of the charge carrier in a specific material and $\vec{E}(x)$ is the electric field that the carrier experiences at position x. The current vanishes as soon as the charge carriers are collected at the electrodes, i.e. they stop to move through the electric field.

From equation 2.2, it can be seen that the time integral of the current induced by a charge recovers the amount of charge

$$Q = \int_{t_1}^{t_2} i_S(t)dt, \tag{2.3}$$

where $t_2 - t_1$ is the collection time of the charge carrier. The amount of current that can be measured is also proportional to the amount of charge carriers that drift simultaneously. Hence, integrating the signal current over the time until the last charge carrier is collected amounts to the entire charge deposited by a particle.

The number of electron-hole pairs created by a highly energetic particle can be determined by the energy loss of the particle along its trajectory in the sensor material. The average energy loss of a high energetic particle per unit length due to inelastic interactions with atomic electrons in a specific material can be modeled by the Bethe-Bloch equation:

$$-\left\langle \frac{dE}{dx} \right\rangle = 4\pi N_A r_e^2 m_e^2 c^2 z^2 \frac{Z}{A}\frac{1}{\beta^2}\left[\frac{1}{2}\ln\left(\frac{2m_e c^2\beta^2\gamma^2 T_{\max}}{I^2}\right) - \beta^2 - \frac{\delta(\beta\gamma)}{2}\right]$$

where,

$$\frac{dE}{dx} = \text{Energy loss per}$$
$$N_A = \text{Avogadro's constant}$$
$$r_e = \text{Electron radius}$$
$$m_e = \text{Electron mass}$$
$$z = \text{charge number of incoming particle}$$
$$Z = \text{Atomic number of material}$$
$$A = \text{Atomic mass of material}$$
$$\beta = \frac{v}{c}$$
$$\gamma = \frac{1}{\sqrt{1-\beta^2}} = \text{Lorentz factor}$$
$$T_{max} = \text{Maximum kinetic energy of a free electron after a collision}$$
$$I = \text{Mean excitation energy of material}$$
$$\delta = \text{Densitiy correction factor}$$

Fig. 2.4 shows the energy loss of different particles in several materials as a function of particle momentum as modeled by the Bethe-Bloch equation. The energy loss shows a minimum at approximately $3Mc$ for all particles and materials (M is the particle mass). Particles in this region are commonly referred to as minimum ionizing particles (MIP). Due to the lower energy deposition and, therefore, the smallest possible signal created in the sensor by a particle, they are often taken as the worst-case scenario, which a detector needs to be capable of detecting with reasonable confidence. While the energy loss of a particle in a material is still subject to statistical fluctuations, a material-specific number of electron-hole pairs per unit length created on average by a MIP can be used to estimate the amount of signal generated in a sensor.

2.3 Silicon Versus Silicon Carbide

Silicon has long been the material of choice for tracking sensors in high-energy physics due to its exceptional properties towards the detection of particles. However, its easy availability, mature fabrication processes, and the possibility of creating high-purity crystals have also contributed to its dominance in sensor technology, especially in large-scale accelerator experiments. The band gap of silicon of approximately 1.12 eV makes it a material quite sensitive material to ionizing radiation and allows for the generation of high numbers of electron-hole pairs even in the MIP regime of high energetic particles. Silicon also features a high mobility of charge carriers, which allows for a relatively fast signal response, which is attractive for timing applications.
While silicon sensors have proven highly effective, they face challenges in harsh radiation environments, such as those encountered in large-scale accelerator experiments or hadron radiation therapy, where the silicon is exposed to high amounts of radiation over long time periods. This kind of exposure leads to damage in the silicon lattice, causing defects that degrade the sensor's performance over time. This radiation damage reduces

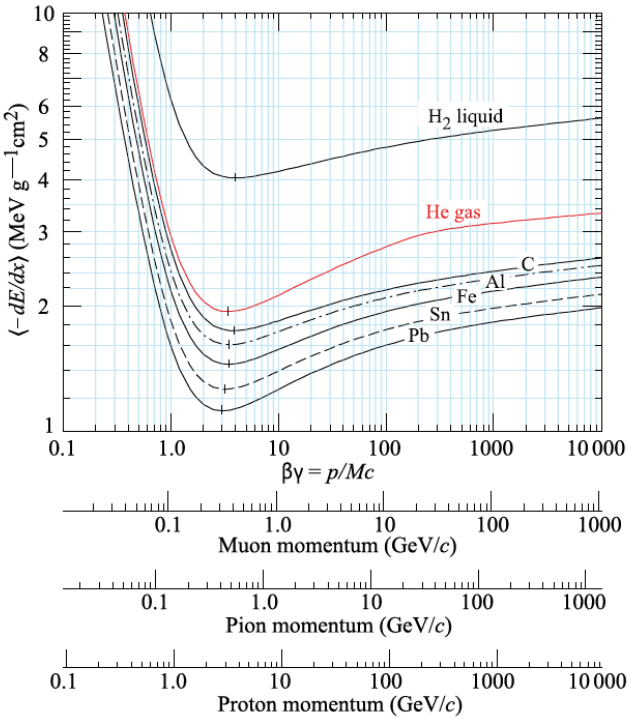


Figure 2.4: Mean energy loss of a muon, pion and proton in different materials as a function of the particle momentum modeled by the Bethe-Bloch equation. Image Adapted from [18].

charge collection efficiency, increases leakage currents, and reduces the overall lifetime of sensors [5, 8]. As experiments push toward higher luminosities and more intense particle fluxes, the limitations of radiation hardness become increasingly significant, shifting attention towards more radiation-resistant materials.

Silicon carbide (SiC) has emerged as a potential alternative to silicon for sensors in high-energy physics, particularly in environments with high exposure to radiation. Due to a few of its material properties that are beneficial for its application in harsh environments, it has seen a surge in popularity in the power electronics industry. Consequently, its accessibility via the semiconductor industry has also improved [6, 7].

SiC possesses a wider band gap of about 3.26 eV and features a higher displacement threshold of lattice atoms than silicon, which could potentially reduce its radiation tolerance over silicon. The wider band gap also results in significantly fewer thermally excited electron-hole pairs, reducing leakage current even after irradiation. Additionally, the higher thermal conductivity compared to silicon enables it to better manage heat dissipation, which potentially makes the performance of a SiC sensor independent of temperature over a high range of temperatures [5, 8]. Tab. 2.1 lists and compares selected material properties of silicon and 4H-SiC. 4H-SiC is one of several polytypes of SiC, which are distinguished by their different amorphous and crystalline structuring. Since the technology used in the work presented in this thesis is based on 4H-SiC, the term SiC is used synonymous with 4H-SiC throughout the remainder of the thesis.

Table 2.1: Selected mechanical and electronic properties of silicon and 4H-SiC at 300 K. Values were taken from [8, 19, 20, 21].

Property	Si	4H-SiC
Density (gcm^{-3})	2.33	3.22
Thermal Conductivity ($Wcm^{-1}K^{-1}$)	1.5	4.9
Band Gap Energy (eV)	1.12	3.23
Ionization Energy (eV)	3.6	5-8
Mean generated e-h pairs/µm (MIP)	80	55-57
Relative Permittivity	11.9	9.7
Breakdown Field ($MVcm^{-1}$)	~ 0.3	~ 3
Electron Mobility ($cm^2V^{-1}s^{-1}$)	1300-1450	800-1000
Hole Mobility ($cm^2V^{-1}s^{-1}$)	450-460	115-120
Saturated Electron Drift Velocity ($10\,cms^{-1}$)	0.8 - 1.0	2.0 - 2.2

While silicon carbide offers significant advantages regarding radiation hardness and thermal stability, it also presents a few challenges for particle detection. The wider band gap and higher electron-hole pair creation energy only allow for a lower average number of electron-hole pairs generated by a MIP per µm, potentially reducing the generated signal. Further, manufacturing processes for SiC are not yet able to fabricate very high purity and thick bulk SiC like it is possible for silicon. Instead, epitaxial growth needs to be used to create a substrate with a low concentration of impurities and defects that can be used as the active sensor volume. For this, a high-purity epitaxial layer is externally precipitated onto a thick, less pure, and more highly doped bulk substrate from, e.g. a highly pure gas phase [5, 8, 21]. Fig. 2.5 shows the cross-section of a SiC sensor diode with an n-type bulk substrate and epitaxial layer (epi-layer). The depletion region is created from a highly doped p^{++}-implant and extends into the epi-layer. The possible thickness of these epitaxial layers are commonly reported to values up to 50 µm for SiC [5, 8, 9]. This further contributes to a lower signal generation in SiC sensors, considering that silicon sensors can have depleted sensor volumes up to hundreds of µm.

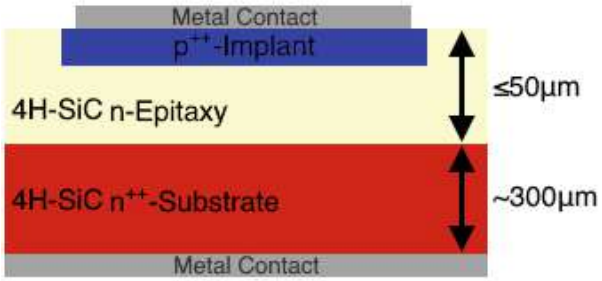


Figure 2.5: Cross-section of a p-in-n SiC particle sensor diode consisting of a n-doped epitaxial layer as the depleted sensor volume grown on a thick and highly doped n-type bulk substrate.

Tab. 2.1, on the other hand, shows quite promising properties for manufacturing elec-

tronics in SiC over electronics in silicon. The wider band gap reduces noise, while the thermal conductivity helps the performance to stay constant over a wide temperature range. The higher breakdown field allows for operation at higher voltages and, together with the higher saturated electron velocity, allows for fast drift velocities, increasing the speed of electronic devices. The downside is that the manufacturing of SiC electronics currently only features relatively large feature sizes for electronic devices, making it still slower than contemporary silicon technology.

2.4 Readout Electronics

Between the formation of the signal in a sensor and the observation of the signal by a human, a whole chain of readout electronics has to be employed to process the sensor signal into an observable form. A typical readout chain is divided into the front-end for processing the analog sensor signal and the back-end for digitization and processing the digitized signal. These are shown in Fig. 2.6.

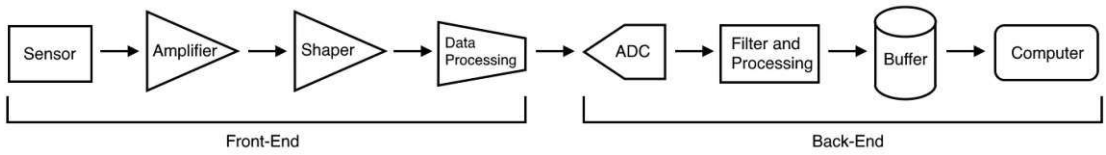


Figure 2.6: Schematic of a typical readout chain for particle detectors.

The front-end electronics start with an amplification stage, which turns the relatively small sensor current signal into an amplified voltage signal proportional to the collected charge. This is usually done by an integrating amplifier like a charge-sensitive amplifier (CSA). A shaper then processes the amplified signal into a pulse with a pre-defined shape, which also involves the filtering of noise to increase the signal-to-noise ratio (SNR). This is followed by some kind of data processing, like threshold discrimination, to remove signals with unwanted features and store only relevant signals for further processing.
The back-end first turns the still analog signal into a digital one using an analog-digital converter (ADC). The digital signal is then further processed by further digital filtering and data processing involving logic electronics. The signal can then be stored in a buffer and later on be read out by a computer or oscilloscope to make it accessible to the human eye. It should be noted that parts of the chain can be adapted or left out depending on the requirements of the specific detection application.
The implementation of the readout electronics or parts of the readout electronics in many experiments either happens on a printed circuit board (PCB), application-specific integrated circuit (ASIC) chip, or a combination of both. The ASIC chips are usually manufactured using CMOS technology and can carry one or several of the aforementioned blocks in the readout chain. The work presented in this thesis focuses on developing an amplifier, i.e. the first block of the readout chain, in the form of an ASIC.
The combination of the electronics with the sensor can generally be done in two ways. The first one is to have the sensor on a separate chip from the electronics, which can be located on a PCB or separate ASIC chip. The combination can then be accomplished using wire bonds on a PCB or bump bonds in the case of highly granular pixel sensors, where a large number of wire bonds is not feasible. Detectors using this approach are referred to as hybrid detectors. The other possibility is the integration of the sensor and electronics on the same chip using CMOS technology. This approach allows for the development of so-called monolithic sensors.

2.5 Monolithic Active Pixel Sensors

MAPS feature a highly granular pixel sensor along with the associated readout electronics on the same semiconductor chip. The pixel sensor is a segmented sensor, which

features neighboring sensor pixels with their own active depletion region separated by additional guarding structures implanted on the top side of the sensor. Each pixel possesses its own readout channel located on the same chip next to the pixel array. One readout channel can consist of the readout chain up to the data transmission to the computer processing. The segmentation and separate readout of the pixels allow for precise 2-dimensional position resolution of the particle trajectory through the sensor by sharing the deposited charge among several pixels.
The electronics can be implemented using contemporary CMOS manufacturing processes. However, they need to be shielded from the high voltage needed to deplete the sensor pixels. This can be achieved by the introduction of additional deep implants that establish a pn-junction with the depleted sensor area acting as a barrier for the high voltage [1]. A simplified structure of a MAPS with CMOS electronics is shown in Fig. 2.7b.

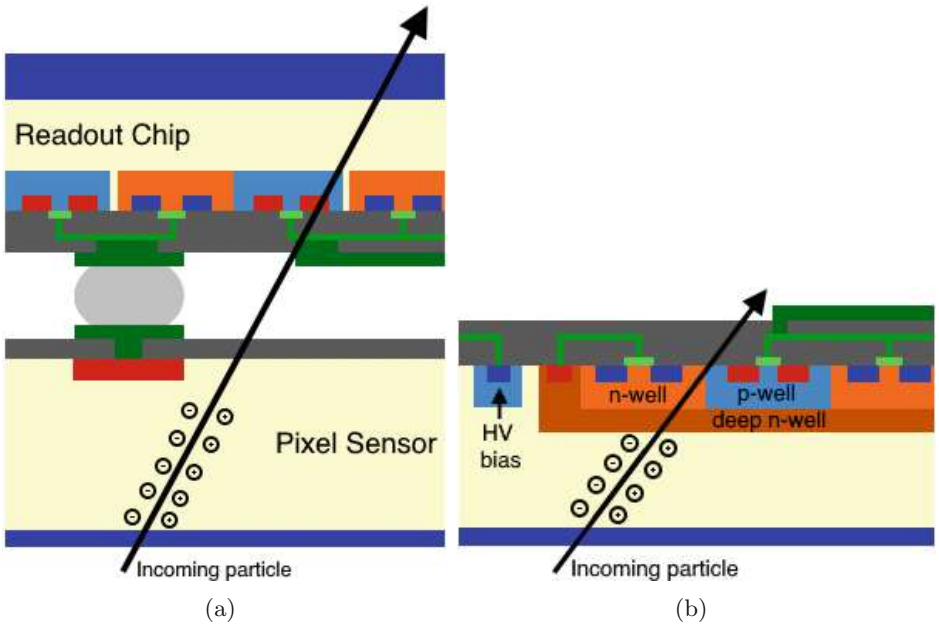


Figure 2.7: Schematic cross section of a (a) hybrid pixel detector consisting of separate chips for the sensor and readout electronics and (b) monolithic active pixel sensor integrating the sensor and readout on the same chip.

The MAPS approach stays in contrast to hybrid pixel sensors, which are conventionally used in high-energy physics at large-scale accelerator experiments like the LHC. In hybrid pixel sensors, the pixel sensor and the electronics are located on separate chips. The separate chips are combined using so-called bump bonding, where a solder bump is applied to the top of the sensor pixel, and the electronics chip is connected on the opposite side of the bump. This is shown in Fig. 2.7a.
While this approach offers the freedom of optimizing the separate chips using two different manufacturing processes that are beneficial for either task, the bump bonding process is quite expensive and complex [1]. MAPS feature a compacter design by removing an

additional chip layer, which makes the material budget more efficient and also removes additional and unnecessary particle scattering on a separate electronics chip.
Radiation hardness in MAPS is, however, still an actively addressed challenge [4]. While developments have been shown to fulfill the requirements for the current generation of colliders [1, 22, 23], this is not necessarily true for future generations [3]. The emergence of SiC CMOS processes offers a promising approach to address this challenge by combining the compactness of MAPS with the potential radiation hardness of SiC.

3 Metal-Oxide-Semiconductor Field Effect Transistors

Metal-oxide-semiconductor field effect transistors (MOSFETs) are the essential building blocks of CMOS (Complementary MOS) electronics. These electronic devices are widely used in electronics as voltage-controlled switches and current sources, which can be used to build amplifier circuits on ASIC chips. This section gives an introduction to the physics and function principle of MOSFETs as well as how to model them in small-signal circuit analysis.

3.1 Structure and Channel Formation

A MOSFET is a type of transistor that controls the flow of current using an electric field. It has four main structures to which voltages can be applied: the source, drain, gate, and bulk (sometimes also called substrate or body). They are referred to as the terminals of the MOSFET.
The MOSFET is built on a doped semiconductor substrate and has two separated implants with a doping opposite of the substrate doping. These implants form the source and drain. A thin insulating layer of silicon dioxide (SiO_2) separates the gate from the semiconductor. The gate is usually made of a layer of metal or polycrystalline silicon (poly-silicon). The bulk connects to the substrate of the MOSFET.
Depending on the doping of the substrate and the implants, there are two types of MOSFETs used in CMOS technology. An n-channel MOSFET, or NMOS for short, has n-doped source and drain implants and a p-doped substrate. A p-channel MOSFET, or PMOS, has p-doped implants and an n-doped substrate. The basic structure of an NMOS and a PMOS is shown in Fig. 3.1 along with their circuit symbols in analog electronics.

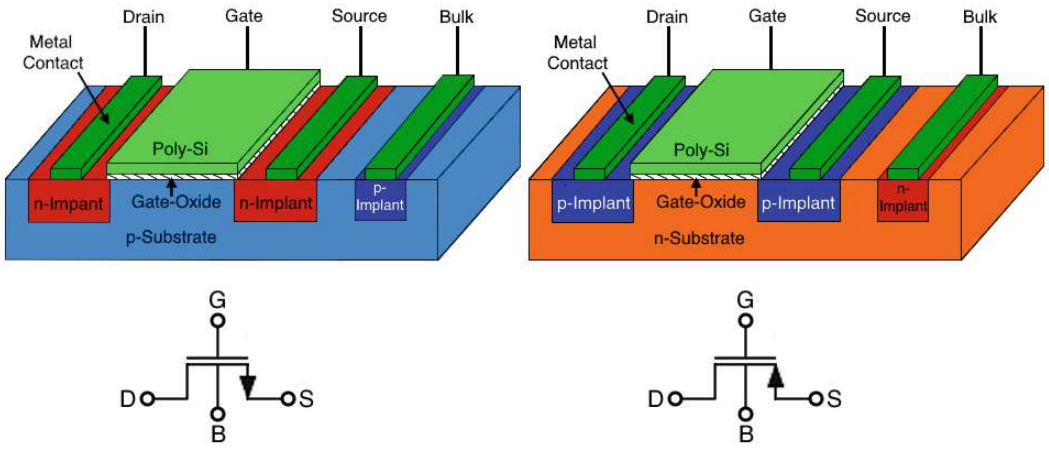


Figure 3.1: Cross section and circuit symbol of an n-channel MOSFET (left) and p-channel MOSFET (right)

The current flowing through a MOSFET is controlled by the formation of a conducting channel between the source and drain. This channel is established by an electric field that forms when a voltage is applied to the gate of the MOSFET. Assume that the source, drain and bulk are connected to ground. If no voltage or a negative voltage is

applied to the gate of an NMOS, its gate-source voltage $V_{GS} = 0\,\mathrm{V}$ or $V_{GS} < 0\,\mathrm{V}$. In this case, the p-doped substrate under the gate oxide is populated by holes as the majority carries as holes are attracted (or rather electrons are expelled) in the case of a negative voltage. The channel is not conductive.
If a positive voltage is applied to the gate, electrons are attracted to the space beneath the gate oxide and recombine with holes, which forms a depletion region. As the voltage gets higher than the so-called threshold voltage i.e. $V_{GS} = V_{TH}$, no holes are available for the electrons to recombine with. The conduction band near the gate drops close to the Fermi level, and an excess of easily excitable electrons is established. The p-doped substrate near the gate turns into an n-type.
This process is called inversion. The inversion layer extends from source to drain and forms a conducting channel between both n-doped implants. If now a positive voltage is applied to the drain, such that the drain-source voltage $V_{DS} > 0$, then a current can flow from source to drain via the conducting channel. This current is referred to as drain current I_D.
The channel formation of a PMOS works analogously, however the gate-source voltage needs to be negative (i.e. higher potential at the source) to attract holes and turn the space beneath the gate from an n-type semiconductor to a p-type. The drain-source voltage also needs to be negative to conduct a drain current. The gate-source and drain-source voltages of PMOS transistors will nonetheless be presented as positive values throughout the thesis while implicitly assuming that the potential at the source is higher than at the gate and drain.

3.2 I/V Characteristics

In practice, it can be observed that for $V_{GS} < V_{TH}$, still a small current can flow for $V_{DS} > 0$, which behaves exponentially with V_{GS} approaching V_{TH}. This conducting sub-threshold state is referred to as weak inversion. Well above the threshold voltage, where the inversion layer is properly formed, the conducting state is called strong inversion, and the current can be approximated by a square law. Close to the threshold voltage, a transition stage exists between weak and strong inversion, which is termed moderate inversion.
Apart from the inversion regions, three other regions of operation can be defined, depending on the relationship between V_{GS}, V_{DS} and V_{TH}. A visualization of these operation regions can be found in Fig. 3.2. At $V_{GS} < V_{TH}$, where no inversion layer is formed, the cut-off region can be found. For $V_{GS} > V_{TH}$ and $V_{DS} < V_{GS} - V_{TH}$ the drain current is dependent on V_{GS} and V_{DS}:

$$I_D = \mu_n C_{\mathrm{ox}} \frac{W}{L} \left[(V_{GS} - V_{TH}) V_{DS} - \frac{V_{DS}^2}{2} \right], \tag{3.1}$$

where μ_n is the electron mobility, C_{ox} the gate oxide capacitance per unit area, W the width of the channel and L the length of the channel. Note that the length of the channel is usually measured by the distance between the source and drain implant. W is measured by the length of the source and drain implants perpendicular to the channel length. $V_{GS} - V_{TH}$ is referred to as overdrive voltage.
The MOSFET behaves like a resistor with variable resistance governed by V_{GS}. For low

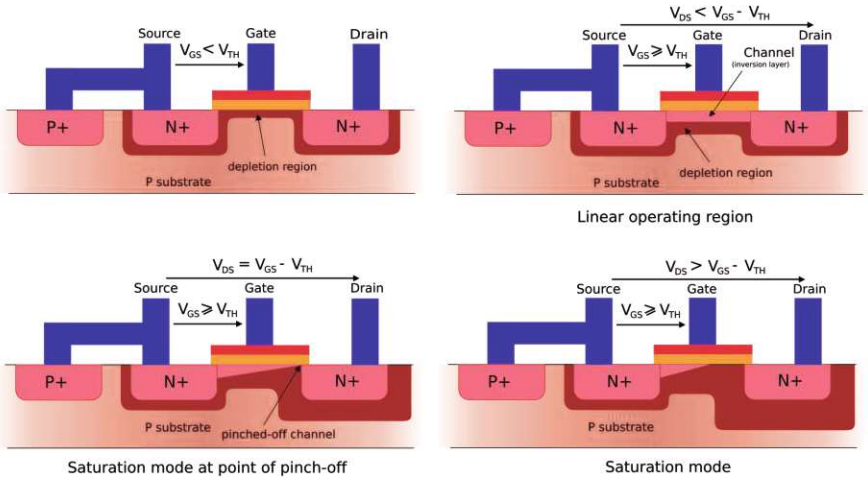


Figure 3.2: Visualization of the operation regions of a MOSFET: cut-off (top-left), linear region (top-right), saturation (bottom left), saturation after pinch-off (bottom right). Image adapted from [24].

values of V_{DS}, I_D behaves approximately linear, which is why this region is often termed linear region.
Equation 3.1 has a maximum at $V_{DS} = V_{GS} - V_{TH}$. However, in practice, the drain current does not fall again for $V_{DS} > V_{GS} - V_{TH}$. Instead, as V_{DS} approaches the overdrive voltage and the voltage difference between the gate and drain becomes smaller, the electrons in the inversion layer feel less attraction towards the drain. The inversion layer, therefore, begins to get narrower at the drain until pinch-off occurs (see Fig. 3.2). The pinch-off causes I_D to not increase significantly with increasing V_{DS}. However, a current can still flow as electrons can drift along the electric field in the depletion region near the drain.
This operation region, where $V_{GS} > V_{TH}$ and $V_{DS} > V_{GS} - V_{TH}$, is called saturation region. A visualization of the behavior of the drain current can be seen in Fig. 3.3 as the MOSFET transitions from linear to saturation region. The drain current behaves approximately constant with a change in V_{DS} but depends quadratically on the overdrive voltage:

$$I_D = \frac{1}{2}\mu_n C_{\text{ox}} \frac{W}{L}(V_{GS} - V_{TH})^2. \tag{3.2}$$

The benefit of driving a MOSFET in the saturation region is that its drain current does not significantly change with changes in its drain-source voltage. Hence, even in signal processing circuits, MOSFETs can be treated as constant current sources when they are driven deep enough in saturation.
For the design of amplifiers, where a time-dependent voltage signal is often applied to the gate of a MOSFET, the sensitivity of I_D on the change of V_{GS} is often of significant importance. This is called transconductance g_m, and for a specific value of V_{DS}, it can be expressed as

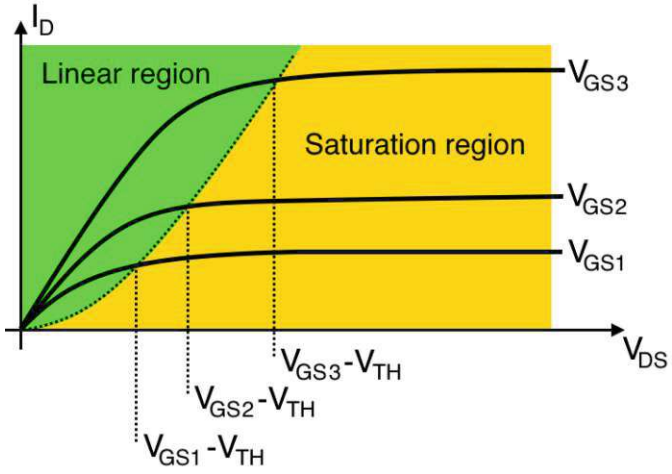


Figure 3.3: Visualization of the output characteristics of a MOSFET, i.e. the drain current I_D as a function of its drain-source voltage V_{DS} for different gate-source voltaged V_{GS}. I_D begins to saturate as V_{DS} approaches $V_{GS} - V_{TH}$.

$$g_m = \frac{\partial I_D}{\partial V_{GS}}\Big|_{V_{DS}=const.} = \frac{1}{2}\mu_n C_{\text{ox}} \frac{W}{L}(V_{GS} - V_{TH}). \tag{3.3}$$

3.3 Channel Length Modulation and Body Effect

As V_{DS} is increased in the saturation region and the inversion layer is shifted more and more towards the source due to pinch-off, the channel of the MOSFET begins to effectively appear shorter. This can be observed in a slight increase in drain current as V_{DS} increases in the saturation region. This increase can be explained by the fact that the drain current is inversely proportional to the channel length L (see equation 3.2). This process is called channel length modulation. Taking channel length modulation into account, the drain current in saturation given by equation 3.2 needs to be corrected:

$$I_D \approx \frac{1}{2}\mu_n C_{\text{ox}} \frac{W}{L}(V_{GS} - V_{TH})^2(1 + \lambda V_{DS}), \tag{3.4}$$

where *lambda* is the channel length modulation coefficient, which represents the rate of how much the channel length changes with increasing V_{DS}. Without channel-length modulation, the drain current would not change with voltage in the saturation region, making the MOSFET an ideal current source with an infinite output resistance. With channel-length modulation, however, this is not true anymore, which adds a finite output resistance across the drain-source path. This output resistance r_o can be expressed as

$$r_o = \frac{\partial V_{DS}}{\partial I_D}\Big|_{V_{GS}=const.} = \frac{1}{\frac{1}{2}\mu_n C_{\text{ox}} \frac{W}{L}(V_{GS} - V_{TH})^2 \cdot \lambda}. \tag{3.5}$$

Another effect that needs to be considered is the so-called body effect, which occurs when a voltage is applied to the bulk (also called body) of the MOSFET, other than the voltage applied to the source. This results in a negative (PMOS) or positive (NMOS)

source-bulk voltage V_{BS}. The voltage applied to the bulk counteracts the accumulation of minority charge carriers under the gate oxide originally exerted by the voltage applied to the gate. Consequently, a higher gate-source voltage needs to be applied to establish the inversion layer. This can be modeled by an increase in threshold voltage V_{TH}. The threshold voltage therefore becomes a function of V_{BS}
If hypothetically $V_{BS} > V_{GS}$, minority carriers would be attracted to the bulk instead of to the gate. Hence, the bulk terminal could be seen as a second gate controlled by V_{BS}. As a result, similar to the transconductance, the sensitivity of the drain-current to the change in bulk-source voltage can be determined:

$$g_{mb} = \frac{\partial I_D}{\partial V_{BS}} = \mu_n C_{\text{ox}} \frac{W}{L} (V_{GS} - V_{TH}) \left(-\frac{\partial V_{TH}}{\partial V_{BS}} \right), \tag{3.6}$$

where g_{mb} is the body-effect transconductance.
Note that I_D, g_m, r_o and g_{mb} all linearly depend on the ratio of the channel width and the channel length $\frac{W}{L}$. This can be used later on for the design of the MOSFET circuits.

3.4 Small-Signal Analysis

The formal analysis of MOSFETs is, in practice, not trivial, as MOSFETs do not behave linearly. However, for very small changes around the DC operating point of a MOSFET, e.g. due to small signals that do effectively not change the DC operation point, a linearized approximation can be applied. This becomes clear by e.g. looking at the Taylor-expansion of the drain-current I_D at a specific drain-source voltage V_{DS} and gate-source voltage V_{GS} with a very small perturbation in the gate-source voltage $\Delta V_{GS} = V_{GS} - V_{GS0}$, where $V_{GS0} = V_{GS} + \delta$ with a very small perturbation δ:

$$\begin{aligned} I_D &= I_{D0} + \frac{\partial I_D}{\partial V_{GS}}\Big|_{V_{GS0}} \Delta V_{GS} + \mathcal{O}(\Delta V_{GS}^2) \\ &\approx I_{D0} + g_m \Delta V_{GS}, \end{aligned}$$

where I_{D0} is the drain current at V_{GS0}. Terms of second order and above were neglected as higher powers of ΔV_{GS} effectively vanish. It was also used that the derivative in the first order term is equal to the transconductance gm of the MOSFET at V_{GS0}. The result is a linearized approximation of the drain current around the operating point with the g_m as the slope. Such an approximation is called small-signal approximation.
The drain current can, therefore, be modeled as a voltage-controlled current source $V_{GS} g_m$. If the same approximation is applied to the change of I_D by V_{DS}, the drain current can be modeled by a resistance r_o with a voltage drop of V_{DS}. Analogously, the change with V_{BS} can be linearized by using another voltage-controlled current source $V_{BS} g_{mb}$. Summarized in a circuit diagram, the equivalent small-signal model circuit of a MOSFET can be obtained, which is shown in Fig. 3.4.

3.5 Parasitic Impedances

By analyzing the structure of a MOSFET, it can be realized that between every pair of the four terminals a parasitic capacitance can be found. The capacitance between the

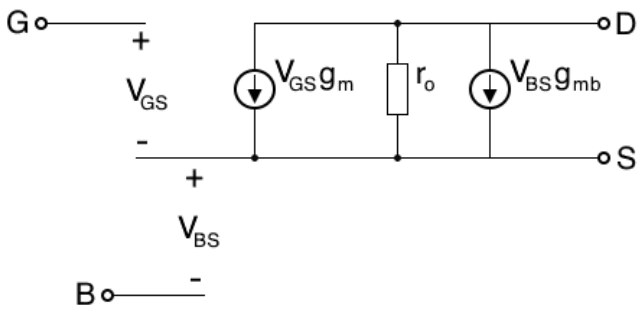


Figure 3.4: Small-signal equivalent circuit of a MOSFET.

gate and the bulk C_{GB} depends on the inversion region of the channel. If a depletion region has formed, this capacitance is a series combination of the capacitance of the gate oxide and the depletion region. In strong inversion, where the inversion channel has formed, the gate-bulk capacitance becomes negligible as the inversion layer shields the gate from the bulk.
The capacitance between the bulk and the source C_{SB} and drain C_{DB}, respectively, originate from the formation of the pn-junction at the implant-substrate interface. While the capacitance between the gate and the source C_{GS} and drain C_{GD}, respectively, originate from the overlap of the gate with the source and drain implants. C_{GS} and C_{GD} are, therefore, a linear function of the overlap area (plate capacitor approximation), which is directly proportional to the channel width W.
The capacitance between source and drain C_{DS} is usually negligible for devices with long channels, except for the special case where the bulk is shorted to the drain. In that case, the capacitance equals the pn-junction capacitance of the drain with the bulk $C_{DS} \approx C_{DB}$.
In the circuit design process presented below, only C_{GS} and C_{GD} were explicitly considered. As the parasitic capacitances become relevant at high signal frequencies, they need to be added to the small-signal equivalent circuit of a MOSFET. This is shown in Fig. 3.5.

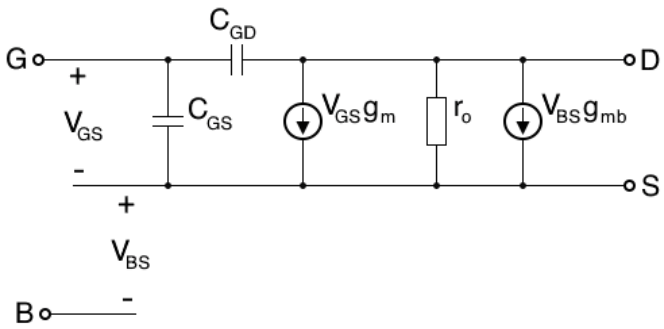


Figure 3.5: Small-signal equivalent circuit of a MOSFET at high signal frequencies.

In modern CMOS technology, MOSFETs are usually embedded in a doped epitaxial

layer. The bulk substrate is then formed by implanting a so-called well into this epitaxial layer, which surrounds the more highly doped implants of the source, drain and bulk terminals. Depending on the doping type of the epitaxial layer, a pn-junction is then formed with the well of either the NMOS or PMOS devices. E.g. an NMOS device is embedded in an n-doped epi-layer. Hence, the p-doped well forms a pn-junction with the epi-layer, which carries a quite high capacitance relative to the ones between the terminal implants and the well due to the bigger interface area. This capacitance usually then lies between the bulk and the potential applied to the epi-layer. If a source is connected to the bulk, this capacitance is felt at the source. Hence, if the epi-layer is connected to either ground or the supply voltage of the circuit and the source is not, it is advisable to not connect the source and bulk to remove the additional capacitance from the source.
Next to the parasitic impedances, the resistance of the gate poly-Si also needs to be considered. The wider the gate, the longer this layer of poly-Si becomes, which increases its resistance along the gate. In combination with the parasitic capacitances of the MOSFET, in particular the ones connected to the gate, the resistance can form an RC-circuit with a cut-off frequency of $f_{cut-off} = \frac{1}{2\pi RC}$, where R is the resistance of the poly-Si and C is any of the parasitic capacitance between the gate and the other terminals. For long channel widths, the resistance and capacitances can get significantly high, and either causes an unnecessary voltage drop along the gate or reduce the cut-off frequency to low values. The latter degrades the frequency response of an amplifier (see next section).
To combat this, MOSFETs are often implemented into a so-called finger structure. A MOSFET with a width of W is divided into n MOSFETs with a width of $\frac{W}{n}$, which then are put in parallel i.e. the source, drain, gate and bulk are respectively shorted among the n MOSFETs. The single gates of the MOSFETs are then termed fingers. This way, the gate resistance and the parasitic capacitances of every finger can be divided by a factor of n, potentially removing the above-mentioned issues.

4 Amplifiers

The main contribution of this thesis is the design of an amplifier circuit. To understand the design process of this amplifier, relevant concepts and parameters associated with the performance of the amplifier need to be explained. This is done in this section. First, an introduction to the DC characteristics of amplifiers is given, followed by an explanation of the AC behavior and feedback. The section closes with a description of a general charge sensitive amplifier.

4.1 DC Characteristics

The design of the CSA presented in this work consists of two types of amplifiers: a single-ended amplifier and a differential amplifier with single-ended output. These two are shown in Fig. 4.1. Assume that the voltage gain of both amplifiers A_v is the same. The single-ended amplifier simply amplifies the input voltage V_{in} applied to the input to produce an output voltage $V_{out} = A_v V_{in}$. The differential amplifier, on the other hand, has two inputs: a non-inverting input (+) and an inverting input (−). A differential amplifier amplifies the difference in voltages applied to the inverting V_- and non-inverting V_+ inputs to produce the output voltage: $V_{out} = A_v(V_+ - V_-)$. If ground is applied to the inverting input, then $V_{out} = A_v V_+$, so if $V_+ = V_{in}$, then the output voltage can be calculated similarly to the single-ended amplifier.

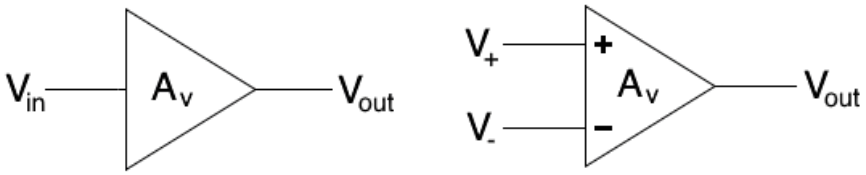


Figure 4.1: Single-ended amplifier (left) and differential amplifier with single-ended output (right) with a voltage gain of A_v.

It can easily be seen that $A_v = \frac{V_{out}}{V_{in}}$. This can be rewritten as $A_v = \frac{V_{out}}{I_{out}} \frac{I_{out}}{V_{in}}$, with I_{out} being the outgoing current. The first fraction has the unit of a resistance. It is referred to as the output resistance R_{out} of the amplifier, i.e., the resistance looking into the output of the amplifier. The second fraction has the unit of a conductance and is referred to as equivalent transconductance G_m, i.e. the sensitivity of the output current to an input voltage. Hence, the voltage gain can be expressed as the product of the output resistance R_{out} and equivalent transconductance G_m:

$$A_v = G_m \cdot R_{out}. \tag{4.1}$$

For an inverting amplifier, the gain is negative, i.e. $A_v < 0$. In that case, the signal is still amplified, but changes sign at the output of the amplifier. An inverting amplifier is usually identified in its circuit symbol by a small circle at the output of the amplifier.

4.2 Frequency Response

At high signal frequencies f, the impedance of capacitances $Z = \frac{1}{j\omega C}$ in an amplifier circuit begins to become smaller, where $\omega = 2\pi f$ is the angular frequency. Hence, for MOSFET amplifiers, parasitic capacitances become relevant and will eventually allow short connections between terminals. This degrades the functionality of an amplifier with increasing frequency. The gain of an amplifier, therefore, needs to be handled as a function of frequency, i.e. $A(\omega) = \frac{V_{out}(\omega)}{V_{in}(\omega)}$ and is often referred to as transfer function. The magnitude of the transfer function $|A(\omega)|$ gives the magnitude of the gain at a specific frequency of the amplifier. The frequency-dependent gain is usually given in units of dB, i.e.

$$|A(\omega)| = 20 \cdot \log_{10}\left(\left|\frac{V_{out}(\omega)}{V_{in}(\omega)}\right|\right).$$

A simple form of transfer function $A(\omega)$ can be found in a simple low-pass RC-circuit, which is shown in Fig. 4.2.

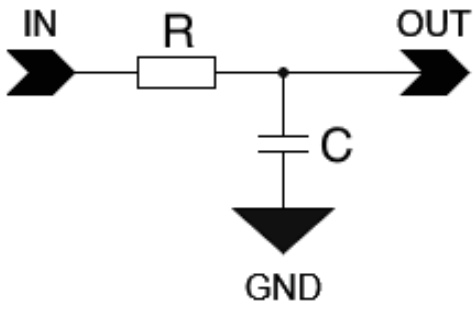


Figure 4.2: Circuit of a low-pass RC-circuit with a resistor R and a capacitance C.

The associated transfer function is

$$A(\omega) = \frac{1}{1 + i\omega RC}. \tag{4.2}$$

The magnitude of this transfer function can be determined as

$$|A(\omega)| = \frac{1}{\sqrt{1 + (\omega RC)^2}}. \tag{4.3}$$

The transfer function in 4.2 has a pole at $-i\omega = \frac{1}{RC}$. The magnitude of this is $\omega = \frac{1}{RC}$. Plugging this into equation 4.3, it can be seen that $|A(\omega)| = \frac{1}{\sqrt{2}}$. Hence, at a frequency of

$$f_c = \frac{1}{2\pi RC}, \tag{4.4}$$

where f_c is called the cut-off frequency, the magnitude of the transfer function drops to $\frac{1}{\sqrt{2}}$. So, for an input voltage signal with a frequency of f_c, the output voltage is therefore reduced by a factor of $\frac{1}{\sqrt{2}}$, which corresponds to a drop of $-3\,\mathrm{dB}$. Beyond the cut-off

frequency, the gain magnitude keeps falling with a slope of $-20\,\mathrm{dB}$ per decade.
The cut-off frequency f_c is often referred to as the bandwidth of a circuit, i.e. the frequency until which the amplifier can provide an appropriate gain to an input signal. In the context of an amplifier circuit with a low-frequency gain of A_v, i.e. the gain below the bandwidth, the transfer function can be expressed as

$$A(\omega) = \frac{A_0}{1 + \frac{\omega}{\omega_c}}.$$

The bandwidth can be determined by the pole $-i\omega = \omega_c$, and below the bandwidth, the gain is equal to A_v. Fig. 4.3 shows an example of the gain magnitude as a function of frequency with a logarithmic frequency axis. A pole in the plot can, therefore, be seen by the gain roll-off after the cut-off frequency associated with the pole. In the context of this thesis, ω_c can always be seen as $\frac{1}{RC}$.

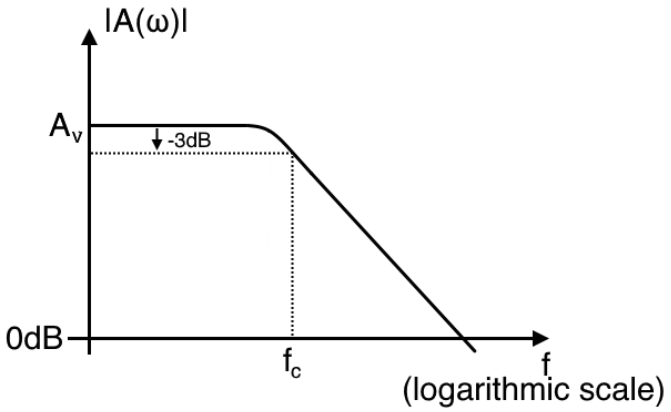


Figure 4.3: Visualization of the gain magnitude $|A(\omega)|$ as a function of frequency f with a logarithmic frequency scale. For $f << f_c$, the gain equals the DC voltage gain A_v.

The calculation until now, assumed a system with only one pole in its transfer function, however this is not always the case in reality. A transfer function of e.g. a two-pole system can be expressed as

$$A(\omega) = \frac{A_0}{(1 + \frac{\omega}{\omega_{c1}})(1 + \frac{\omega}{\omega_{c2}})}.$$

With the poles given by ω_{c1} and ω_{c2}. The pole with the lower resulting frequency is referred to as the dominant pole and determines the bandwidth of the amplifier. Every additional pole adds another gain roll-off of $-20\,\mathrm{dB}$ per decade.

4.3 Feedback and Stability

Adding feedback to an amplifier alters the amplification mechanism of the circuit depending on the feedback network used. A basic schematic of an amplifier with a feedback network is depicted in Fig. 4.4, where a portion of the output signal is fed back and added to the input signal.

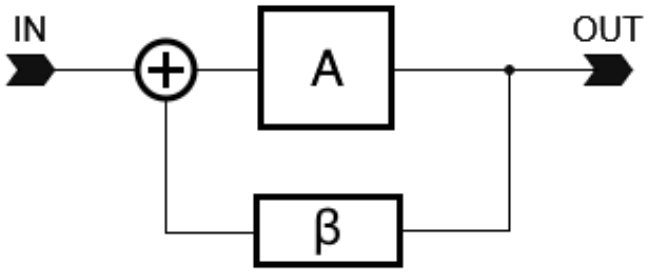


Figure 4.4: Generalized schematic of an amplifier A with a feedback network with the feedback factor β.

β is the so-called feedback factor and determines the ratio of the output signal that is fed back to the input. The gain of this circuit is given by

$$A_{CL} = \frac{A}{1 + \beta A}, \tag{4.5}$$

where A_{CL} is the closed-loop gain, A is the open-loop gain and β is referred to as loop gain. The open-loop gain is equal to the transfer function of the amplifier from the previous section. The loop gain corresponds to the frequency-dependent gain that is applied to a signal going around the feedback loop. It can be determined by breaking up the feedback loop of an amplifier at any point, injecting a signal into the feedback loop, and observing the amplification of the signal as a function of frequency coming back to the breaking point.
If $\beta > 0$, then $A_{CL} < A$, which means that a fraction of the output is subtracted from the input, which is called negative feedback. However, for $\beta < 0$, the closed-loop gain begins to increase above the open-loop gain. This means a fraction of the output is added to the input, which can destabilize the system and cause oscillations. This is called positive feedback. As β approaches $-\frac{1}{A}$, the closed-loop gain goes to infinity, and even in the absence of a signal, noise can get amplified to the point of oscillation.
To maintain stability and ensure that the feedback does not cause instability, the Barkhausen stability criterion can be applied to the amplifier. This theorem first states that to have oscillation $|\beta A| > 1$, where $|\beta A|$ is the magnitude of the loop gain. Hence, to maintain an oscillation, the signal must not be damped around the feedback loop. Secondly, the phase shift of the signal around the feedback loop needs to be a multiple of 2π, i.e. $\angle\beta A = 2\pi n$, where n is an integer. So, if the phase shift of the input signal at a specific frequency around the feedback loop is sufficient to bring the signal back in phase with the input signal, the signal can reinforce itself (positive feedback), which can lead to instability.
In the case of an inverting amplifier, where a 180° phase shift is automatically included by the amplifier, only an additional phase shift of 180° by the feedback loop is needed for oscillation. In that case, it is customary to require the phase angle to be removed from 180° by a specific so-called phase margin for all frequencies where $|\beta A| > 1$. The phase margin is usually measured at the frequency, where $|\beta A| > 1$ and is typically considered to be optimal around a value of 65°.

4.4 Noise

Noise refers to any unwanted electrical signals that interfere with the desired signal, degrading the performance of the system. Noise can originate from several sources, internal and external. Often observed types of noise are thermal noise, which is caused by random thermal excitation of charges in electronic devices, or flicker noise, which occurs from random fluctuations in current and voltage at low frequencies.
Noise can either be modeled by a random stochastic process or by the analysis of the voltage density of the noise over a frequency spectrum. Integrating the square root of the noise density frequency spectrum up to the bandwidth of an amplifier returns the total root-mean-square noise that is sufficiently amplified in the circuit so as not to be neglected.
In a circuit, all noise sources inside the circuit can be summarized by a single equivalent noise voltage source that is either applied at the input or output of the circuit. At the input, it models the noise created in the circuit via an equivalent input noise that would produce the same output noise as the actual noise sources within the circuit. Similarly, such an equivalent noise source can be referred to the noise observed at the output of the amplifier.
By comparing the output signal amplitude V_signal of an amplifier to the total root-mean-square noise voltage V_noise, the signal-to-noise ratio (SNR) can be obtained:

$$\mathrm{SNR} = \frac{V_{signal}}{V_{noise}}. \tag{4.6}$$

The higher the SNR is, the more clearly visible the signal can be measured.

4.5 Charge Sensitive Amplifier

The circuit designed during the course of this thesis is a so-called charge sensitive amplifier. As the name suggests, its purpose is to amplify an incoming charge signal and deliver an output that is still proportional to this signal charge. The CSA consists of an inverting amplifier with approximately infinite input resistance and a feedback capacitance C_f parallel to a feedback resistor R_f, as shown in Fig. 4.5.

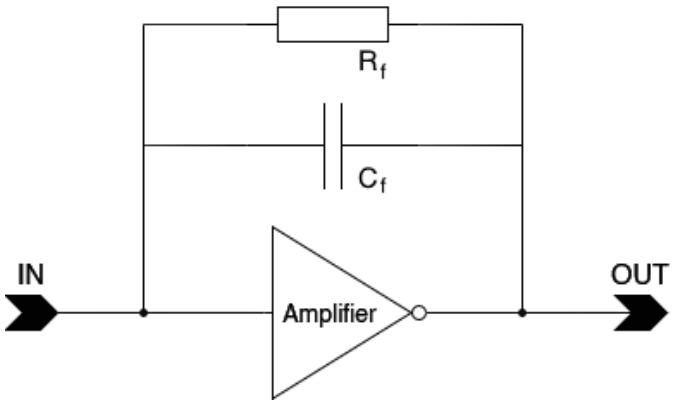


Figure 4.5: Schematic of a charge sensitive amplifier.

The output voltage of this constellation in the time domain can be shown to be

$$V_{out} = \frac{Q_{in}}{C_f} \mathrm{e}^{-\frac{t}{R_f C_f}},$$

where Q_{in} is the signal charge and t is the time. At $t = 0\,\mathrm{s}$, the amplitude of the output voltage is proportional to the signal charge. At positive times, the signal decays exponentially with a time constant of $\tau = R_f C_f$.
A CSA is an integrating amplifier. This becomes clear by realizing that the signal charge is equal to the integral of the current pulse over the pulse duration. The output voltage of the CSA is, therefore, proportional to the integral of the signal current.

5 Circuit Blocks

This section presents the architecture and function principle of the designed CSA circuit. First, an overview of the circuit is given, followed by a detailed discussion of the functionality of the single building blocks of the CSA. This also includes circuit blocks applied at the input and output of the circuit, as well as structures implemented in the circuit layout.

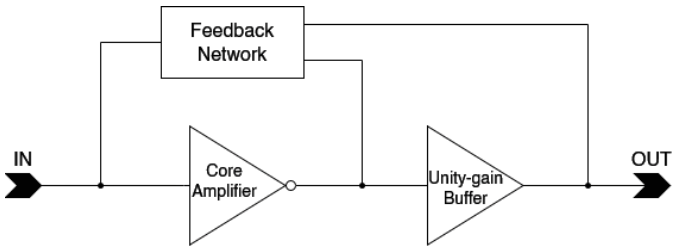


Figure 5.1: Simplified overview of the CSA, consisting of the core amplifier, the output driver stage (unity-gain buffer) and the feedback network.

Fig. 5.1 shows a simplified schematic of the CSA and its building blocks. It includes an inverting voltage amplifier, which is referred to as the core amplifier throughout this thesis. The core amplifier is followed by an output-driver stage implemented as a unity-gain buffer amplifier. The output-driver stage isolates the voltage amplifier from a high-impedance output load and is, hence, designed to have a low output resistance. A feedback network is implemented, which includes the feedback capacitance and resistance typical for a CSA. However, the CSA design in this work comprises two feedback paths: one from the voltage amplifier output and one from the buffer amplifier output, respectively, to the voltage amplifier input. The main purpose of the feedback path from the buffer output is to provide biasing for the feedback network. The additional feedback loop comes as a second-hand effect. It provides additional negative feedback, suppressing the buffer amplifier's noise by capacitively coupling the output back to the CSA input.
Fig. 5.2 shows the full schematic of the CSA, where its essential circuit blocks are highlighted by colored boxes (blue: core amplifier, red: output-driver/unity-gain buffer, green: feedback network). The core amplifier is a regulated folded cascode amplifier with a PMOS (M1) as the input device and NMOS as the output device (M2). M3 and M4 act as constant current sources, which fix the DC current of M1 and M2.
The output-driver is a buffered two-stage operational amplifier with a gain of 1. The first stage consists of an NMOS differential pair (M13 and M14) with a current mirror load (M11 and M12) and M15 as a constant current supply for proper biasing of this stage. The second stage is a common source amplifier with a PMOS as an input device (M16) and with M17 as a constant current supply for biasing. From MC2, the capacitance between its gate and shorted source, drain, and bulk is used as a capacitor. It is employed for frequency compensation, stabilizing the response of the buffer amplifier.
The feedback network employs the capacitance between the gate and shorted source and drain of MC1 as the feedback capacitor. MR is used as a high-valued resistor, and to achieve that, it is driven in the linear region. The gate-source voltage of MR directly

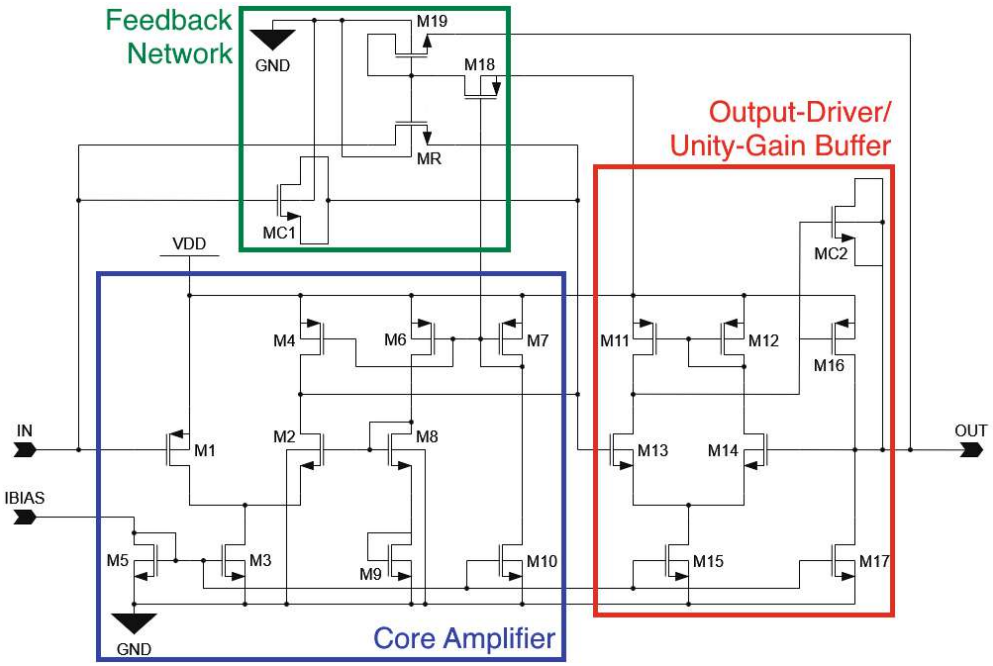


Figure 5.2: Full schematic of the CSA.

steers the effective resistance of MR. Its resistance is, therefore, sensible to changes in the DC voltage at the gate and source of MR. To regulate the gate-source voltage of MR under such changes, the biasing network consisting of M18 and M19 needs to be introduced. The second feedback path is established over the parasitic capacitances of M19 and MR.
The circuit blocks are biased using a MOSFET current mirror network, which is driven by an externally supplied biasing current. This network comprises three interconnected current mirrors with M5, M7 and M8 as the respective reference devices.

5.1 Sensor and Load

The sensor is considered the primary source of signal charge, typically generated through the interaction of a MIP with a semiconductor material like Si or SiC. This signal charge is the stimulus that the CSA circuit is designed to process and amplify. The load, on the other hand, represents the subsequent stages of electronics that process the output of the CSA. These could include analog-to-digital converters (ADCs), further buffer stages or the input for an oscilloscope. The load is essential for testing the performance of the CSA, as it simulates the real-world conditions under which the circuit will operate when combined with a full electronic readout chain.
The sensor connected to the input of the CSA is a single reverse-biased SiC diode with a depletion region carrying a capacitance determined by equation 2.1. This diode can be represented by a single SiC sensor pixel or strip. A charged particle, which will, for all purposes, be seen as a MIP, deposits charge in the substrate of the sensor. The amount of charge is subject to fluctuations and depends on the semiconductor material. For

SiC the average amount of generated electron-hole pairs per µm is $55\,e^{-}\mu m^{-1}$ [20]. The deposited charge induces a current on the readout electrode of the sensor, while being subjected to a drift enabled by the electric field in the depletion region. The current reduces as charges are collected at the electrode, which creates a characteristic current pulse that is fed into the amplifier circuit for amplification. As the signal current is equal to the collected charge, when integrated over the time of the current pulse, the amplified signal of the CSA is proportional to the input charge.
The equivalent circuit of the signal current pulse created by a MIP in a sensor is a pulsed current source. If the pulse width is sufficiently smaller than the time constant of the CSA, the amplifier sees the pulse as a Dirac pulse. If this is true, the form of the pulse is not important and can be approximated by a rectangular pulse with the signal charge Q_{in} given by the product of pulse height and width.
The depletion region in the substrate of the sensor is filled by an electric field and, therefore, carries a capacitance, which can be determined by the sensor material as well as the active area and thickness of the depletion region. Without knowing the active area and thickness exactly, the sensor capacitance is often approximated as a plate capacitor with the pixel pitch as the active area and the thickness of the epi-layer as the active thickness.
The sensor capacitance C_{in} loads the input of the amplifier and, therefore, needs to be included in the circuit analysis. The equivalent circuit for this capacitance is simply a capacitor carrying the sensor capacitance. It is customary to put the sensor capacitance in parallel to the current pulse source, as shown on the left side of Fig. 5.3. This will be used as the equivalent circuit for a sensor with a specific sensor capacitance and signal charge.

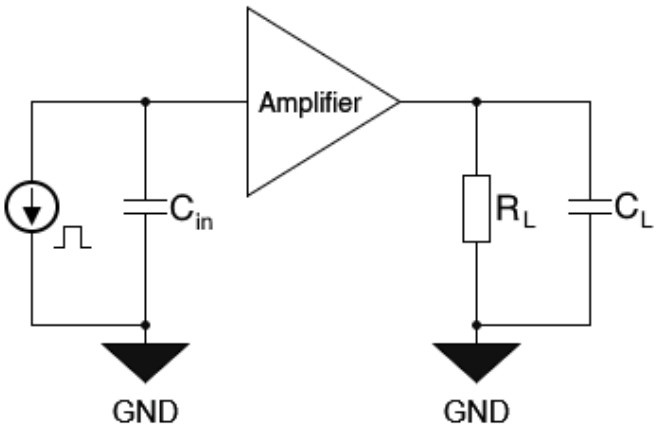


Figure 5.3: Schematic of the equivalent circuit of an amplifier with a particle sensor with sensor capacitance C_{in} at the input and load impedance R_L, C_L at the output.

The current pulse created by the sensor charges the sensor capacitance, which creates a voltage jump at the input of the charge sensitive amplifier given by

$$\Delta V = \frac{Q}{C},$$

where ΔV is the amplitude of the voltage jump, Q is the charge injected by the current

pulse, and C is the capacitance parallel to the current source. This voltage jump is hence proportional to the signal charge and is amplified by the amplifier with a specific voltage gain. The voltage jump decays over time while the capacitor discharges. Note that the amplitude of the voltage jump is also affected by the input capacitance of the amplifier, which adds to the sensor capacitance. The load impedance represents the resistances and capacitances applied to the output of the CSA due to external influence. These impedances mostly originate from the input impedances of further electronics that come after the CSA or due to parasitic impedances of e.g. wiring or pads. The load is simply modeled by a load resistance R_L parallel to a load capacitance C_L applied to the output of the amplifier, as shown in Fig. 5.3.

5.2 Core Amplifier

The core amplifier circuit is shown in Fig. 5.4. It corresponds to a regulated folded cascode amplifier [25, 26, 27]. This is essentially a folded cascode amplifier with a PMOS input device M1 and NMOS output device M2 as well as current source bias represented by M3 and current source load represented by M4. The input of the amplifier is located at the gate of M1 (very high input resistance). The output is located at the drain of M2 and M4. Between the input and output, the folded node is located at the drain of M3.

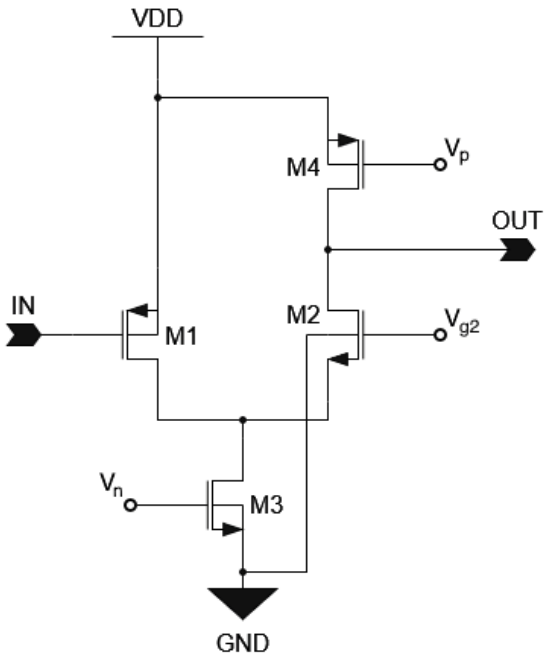


Figure 5.4: Schematic of the core amplifier.

The cascode amplifier is regulated in the sense that the DC currents of M1 and M2 are fixed by the biasing over M3 and M4. The drain current of M3 I_{D3} is the sum of the drain currents of M1 I_{D1} and M2 I_{D2}, hence the drain current of M1 is given by $I_{D1} = I_{D3} - I_{D2}$. However, $I_{D2} = I_{D4}$, provided that the output of the amplifier is not loaded. Hence, $I_{D1} = I_{D3} - I_{D4}$.

A positive incoming voltage signal reduces the gate-source voltage V_{GS1} of M1, as the potential at the source of M1 is fixed to V_{DD}. To maintain the same drain current I_{D1}, the drain-source voltage of V_{DS1} of M1 needs to increase, which causes the potential at the folded node to drop. This causes the gate-source voltage V_{GS2} of M2 to increase, as the potential at the gate of M2 V_{G2} is ideally fixed. The drain current of M2 ideally stays constant as well due to the current provided by M4 and $I_{D2} = I_{D4}$. Consequently, the drain-source voltage V_{DS2} of M2 has to decrease, which lowers the potential of the output of the amplifier. For a negative input signal, the process is analogous, but the voltages change in the opposite direction. This also assumes that the drain current of M3 and M4 is independent of the changes in the potentials of the folded node and output node, which is guaranteed if they are biased far enough into the saturation region.
To guarantee a voltage gain A_v, the change in voltage at the output of the amplifier has to be higher than the change introduced by the input signal. This is guaranteed if the

5.2.1 Determination of the DC Voltage Gain

An analytical expression for the voltage gain A_v of the core amplifier can be gained by applying small signal analysis and determining the small-signal model of the circuit. This work follows the methodology that is presented in [28]. The condition for the small-signal model of the circuit to approximate the voltage gain of the amplifier is that any signal is small enough not to alter the DC operating point of the circuit.

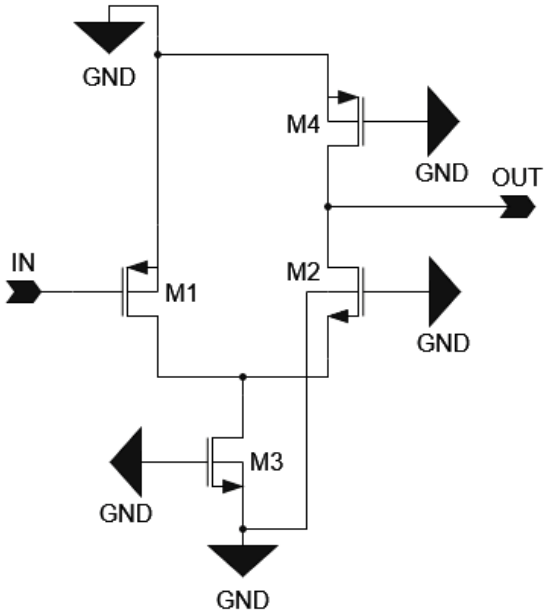


Figure 5.5: Setting constant potentials to ground for the determination of the small-signal model of the core amplifier.

To obtain the small-signal model of the core amplifier, first, all constant potentials need to be shorted to AC ground, i.e. the gates of M2, M3 and M4 as well as V_DD as shown in Fig. 5.5. The gate-source voltages of M3 and M4 and the bulk-source voltage M1, M3

and M4 drop to 0 V. Then, the MOSFETs need to be converted to their small-signal equivalent circuits, which were explained in section 3.4. Fig. 5.6 shows this conversion after the circuit was simplified by removing the voltage-controlled current sources with vanishing gate-source or bulk-source voltage.

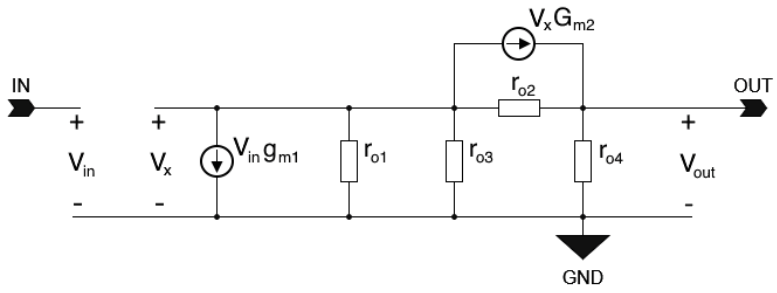


Figure 5.6: Small-signal model circuit of the core amplifier.

The circuit in Fig. 5.6 corresponds to the low-frequency small-signal model, i.e. excluding parasitic capacitances. g_m1 is the transconductance of M1, $G_{m2} = g_{m2} + g_{b2}$ is the sum of the transconductance and body-effect transconductance of M2 and r_o1, r_o2, r_o3, r_o4 are the output resistances of M1 to M4. V_x is the potential at the folded node.
The DC voltage gain $A_v = \frac{V_{out}}{V_{in}}$ of an amplifier is given by the product of the equivalent transconductance $G_m = \frac{I_{out}}{V_{in}}$ of the circuit and the output resistance $R_{out} = \frac{V_{out}}{I_{out}}$, i.e.

$$A_v = G_m \cdot R_{out}. \tag{5.1}$$

R_{out} can be determined by shorting the input node to ground as well and applying a voltage to the output of the circuit (see Fig. 5.7).

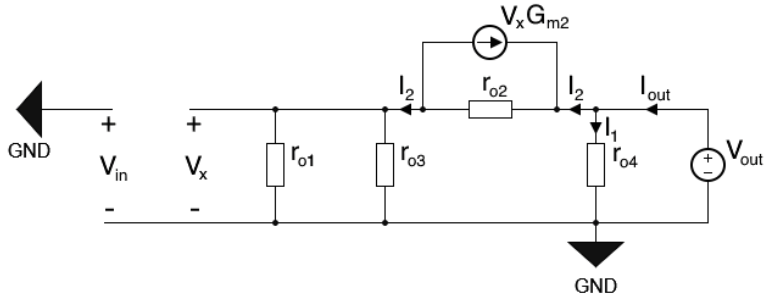


Figure 5.7: Grounding the input node and applying a voltage V_{out} to the output node in the small-signal model, to determine the output resistance of the core amplifier.

The current source $V_{in}g_{m1}$ vanishes as $V_{in} = 0\,\text{V}$. The current I_{out} flowing into the circuit can be calculated by $I_{out} = I_1 + I_2$. The current flowing through r_{o4} is

$$I_1 = \frac{V_{out}}{r_{o4}}. \tag{5.2}$$

I_2 can be calculated by the voltage drop across the left side of the circuit. The voltage drop across r_{o2} is $V_{out} - V_x = (I_2 + V_x G_{m2}) r_{o2}$ and across $r_{o1} \parallel r_{o3}$ it is $V_x = I_2(r_{o1} \parallel r_{o3})$.

Hence,

$$\begin{aligned} V_{out} &= (I_2 + V_x G_{m2}) r_{o2} + I_2 (r_{o1} \parallel r_{o3}) \\ &= [I_2 + I_2 (r_{o1} \parallel r_{o3}) G_{m2}]\, r_{o2} + I_2 (r_{o1} \parallel r_{o3}) \\ \Leftrightarrow I_2 &= \frac{V_{out}}{[1 + (r_{o1} \parallel r_{o3}) G_{m2}]\, r_{o2} + (r_{o1} \parallel r_{o3})} \end{aligned} \tag{5.3}$$

Using equations 5.2 and 5.3, I_{out} is determined as

$$I_{out} = \frac{V_{out}}{r_{o4}} + \frac{V_{out}}{[1 + (r_{o1} \parallel r_{o3}) G_{m2}]\, r_{o2} + (r_{o1} \parallel r_{o3})}. \tag{5.4}$$

R_{out} is determined by manipulating equation 5.4 and remembering that $R_{out} = \frac{V_{out}}{I_{out}}$:

$$\begin{aligned} \frac{I_{out}}{V_{out}} = \frac{1}{R_{out}} &= \frac{1}{r_{o4}} + \frac{1}{[1 + (r_{o1} \parallel r_{o3}) G_{m2}]\, r_{o2} + (r_{o1} \parallel r_{o3})} \\ \Leftrightarrow R_{out} &= r_{o4} \parallel \{[1 + (r_{o1} \parallel r_{o3}) G_{m2}]\, r_{o2} + (r_{o1} \parallel r_{o3})\}. \end{aligned}$$

Assuming that $(r_{o1} \parallel r_{o3}) G_{m2} \gg 1$, then

$$R_{out} \approx r_{o4} \parallel G_{m2} r_{o2} (r_{o1} \parallel r_{o3}). \tag{5.5}$$

The equivalent transconductance G_m of the small-signal circuit is found by grounding the output node of the small-signal circuit and assuming an output current I_{out} flowing into the circuit from the shorted output node as shown in Fig. 5.8. r_{o4} connects from ground to ground and therefore can be left out.

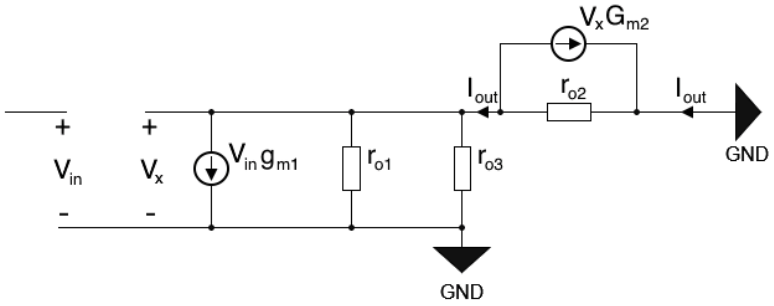


Figure 5.8: Grounding the output node with an output current I_{out} flowing into the circuit of the small-signal for the determination of the equivalent transconductance of the core amplifier.

The voltage drop across r_{o2} is $r_{o2}(I_{out} + V_x G_{m2})$ and across $r_{o1} \parallel r_{o3}$ it is $(r_{o1} \parallel r_{o3})(V_{in} g_{m1} - I_{out})$. Since the voltage drop across r_{o2} and $r_{o1} \parallel r_{o3}$ to ground is the same, it holds that

$$r_{o2}(I_{out} + V_x G_{m2}) = (r_{o1} \parallel r_{o3})(V_{in} g_{m1} - I_{out}). \tag{5.6}$$

After solving equation 5.6 for I_{out} and factoring out V_{out}, $G_m = \frac{I_{out}}{V_{in}}$ can be determined as:

$$G_m = -\frac{g_{m1}(r_{o1} \parallel r_{o3})}{(r_{o1} \parallel r_{o3}) + \left(\frac{1}{G_{m2}} \parallel r_{o2}\right)} \tag{5.7}$$

Plugging equations 5.7 and 5.5 into equation 5.1, the DC voltage gain of the core amplifier can be determined as

$$A_v = -\frac{g_{m1}(r_{o1} \parallel r_{o3})}{(r_{o1} \parallel r_{o3}) + \left(\frac{1}{G_{m2}} \parallel r_{o2}\right)}[r_{o4} \parallel G_{m2} r_{o2}(r_{o1} \parallel r_{o3})]. \tag{5.8}$$

The minus sign in front of the equation represents that the core amplifier is an inverting amplifier. For the remainder of the thesis, this minus sign will, however, be ignored when presenting numerical values for the gain.
Analyzing equation 5.8, it can be seen that the first term in the product, i.e. G_m, can be maximized if $(r_{o1} \parallel r_{o3}) \gg \left(\frac{1}{G_{m2}} \parallel r_{o2}\right)$, as then $G_m \approx g_{m1}$.

5.2.2 Determination of the Bandwidth

The bandwidth is determined by adding the parasitic capacitances of the MOSFETs into the small-signal model. Fig. 5.9 shows the core amplifier circuit with constant potentials shorted to ground and all the parasitic capacitances explicitly indicated.

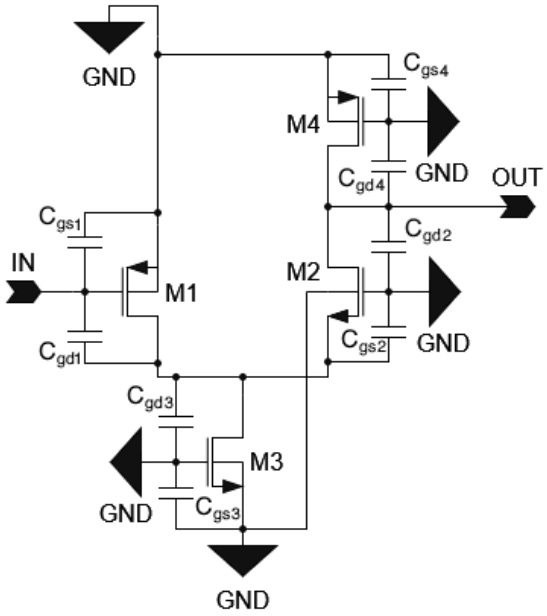


Figure 5.9: Setting constant potentials to ground and including parasitic capacitances for determining the high-frequency small-signal model of the core amplifier.

It can be seen that C_{GS3} and C_{GS4} can be neglected as they connect from ground to ground. By converting the MOSFETs to their small-signal equivalent circuits and

simplifying the circuit in Fig. 5.9, the small-signal model in Fig. 5.10 can be determined.

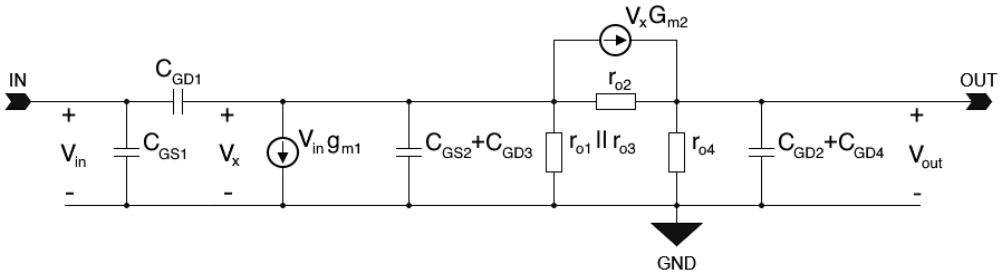


Figure 5.10: High frequency small-signal model of the core amplifier.

The bandwidth of the circuit can be found by calculating the poles of the frequency-dependent transfer function of the circuit. For multi-pole systems, this is, however, far from trivial. Hence, they were approximated by associating every node with a pole and calculating the corresponding time constant RC by determining the resistance R and capacitance C from a node to ground. From the time constant, the cut-off frequency of the RC circuit could be calculated by equation 4.4. The lowest pole frequency could then be used to estimate the bandwidth.
In the small-signal model of the core amplifier, three nodes were identified: the input node, the folded node and the output node. The input node could be neglected, as no resistance is applied to this node. The resistance of the output node is equal to the output resistance of the core amplifier (see equation 5.5). The capacitance equals the parallel combination of C_{GD2} and C_{GD4}. Hence, the pole frequency at the output node was determined using equation 4.4:

$$f_{out} = \frac{1}{2\pi R_{out}(C_{GD2} + C_{GD4})}. \tag{5.9}$$

The resistance looking into the folded node R_x was determined by using the circuit shown in Fig. 5.11. It uses the same approach as for the output resistance of the core amplifier, namely grounding the input as well as the output and applying a voltage V_x with a current I_x to the folded node. The current source $V_{in}g_{m1}$ again can be neglected and r_{o4} vanishes as it connects from ground to ground.
I_x can be determined by $I_x = I_1 + I_2$. I_1 and I_2 can be determined by $I_1 = \frac{V_x}{r_{o1} \| r_{o3}}$ and $I_2 = \frac{V_x}{r_{o2}} + V_x G_{m2}$. Hence,

$$\begin{aligned} I_x &= \frac{V_x}{r_{o1} \parallel r_{o3}} + \frac{V_x}{r_{o2}} + V_x G_{m2} \\ \Leftrightarrow \frac{I_x}{V_x} &= \frac{1}{R_x} = \frac{1}{r_{o1} \parallel r_{o3}} + \frac{1}{r_{o2}} + G_{m2}, \end{aligned}$$

and R_x is therefore

$$R_x = (r_{o1} \parallel r_{o3}) \parallel \left(\frac{1}{G_{m2}} \parallel r_{o2} \right). \tag{5.10}$$

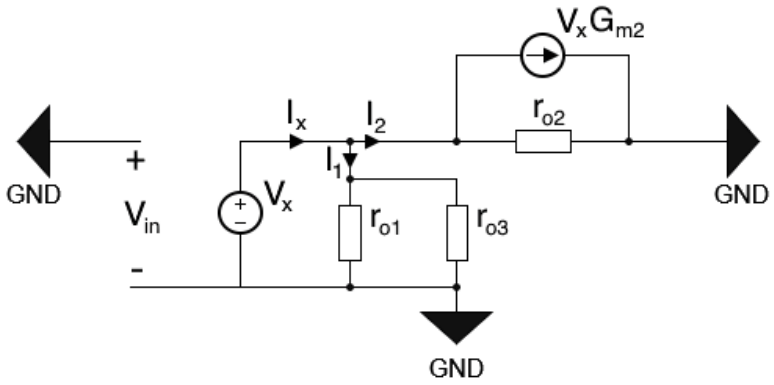


Figure 5.11: Grounding the input and output node and applying a voltage V_x to the folded node in the small-signal model to determine the resistance looking into the folded node of the core amplifier.

The associated capacitance is the parallel combination of C_{GS2} and C_{GD3}. The pole frequency at the folded node is, therefore, determined by

$$f_x = \frac{1}{2\pi R_x (C_{GS2} + C_{GD3})}, \tag{5.11}$$

where R_x corresponds to equation 5.10.
From the gain of the amplifier it could be concluded that $(r_{o1} \parallel r_{o3}) \gg \left(\frac{1}{G_{m2}} \parallel r_{o2}\right)$ would be needed to maximize the gain, which makes $R_x \approx \left(\frac{1}{G_{m2}} \parallel r_{o2}\right)$. Also, R_{out} should be chosen relatively high for a high gain. It can, therefore, be argued that $R_{out} \gg R_x$. Assuming that the capacitances are on the same order of magnitude, the pole at the output of the core amplifier with a cut-off frequency given by equation 5.9 could be identified as the dominant pole limiting the bandwidth.

5.3 Output Driver Stage

For the core amplifier to have a DC high gain, it needs a high output resistance, as can be concluded from equation 5.8. The output resistance, along with the parasitic capacitances at the output, add a pole, potentially limiting the bandwidth of the circuit. However, directly loading the output of the core amplifier with a high load impedance in the scale of the output impedance of the core amplifier could have an unwanted influence on the performance of the circuit. A high load resistance in the scale of the output resistance would draw significant amounts of current out of the amplifier, which causes an offset in the DC operating point of the amplifier circuit. On the other hand, any non-negligible load capacitances add to the parasitic capacitances at the output of the amplifier, decreasing the cut-off frequency introduced by the pole at the output. A high load capacitance would then cause a serious limitation of the bandwidth.
To prevent this from happening, a buffer amplifier with an almost infinite input resistance (gate of a MOSFET) and low output resistance is added to the output of the core amplifier as an output driver. The high input resistance does not load the output of the

core amplifier, causing no DC offset. The input capacitance of the buffer amplifier still loads the output of the core amplifier, reducing the bandwidth slightly. However, this reduction is far less than that caused by a much higher load capacitance.
This way, the core amplifier can be removed from the load impedance. The low output resistance of the buffer amplifier prevents an offset of the buffer amplifier's DC operating point and a limitation of the bandwidth if a high load impedance is applied to it. The latter assumes that the buffer amplifier does not introduce a dominant pole to the whole circuit.

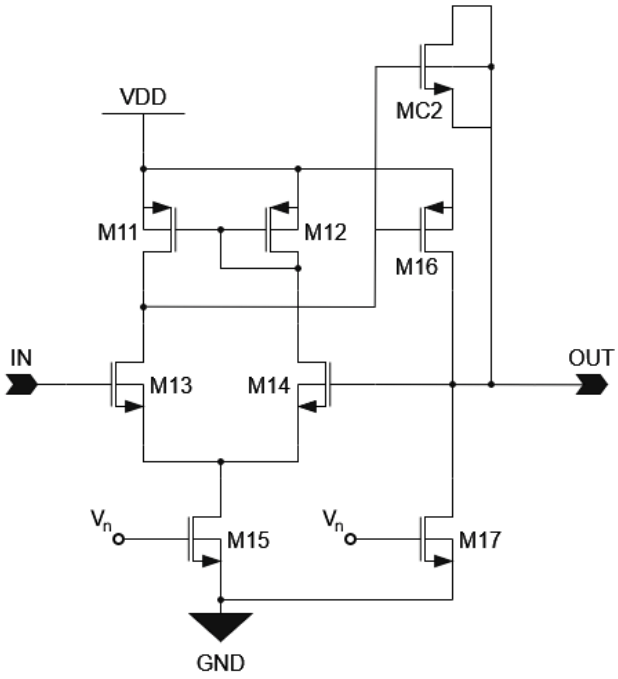


Figure 5.12: Circuit schematic of the unity-gain buffer amplifier.

The buffer amplifier employed as the output driver is shown in Fig. 5.12 [29, 28]. It is a two-stage amplifier with a simple differential amplifier with an NMOS differential pair (M13 and M14) and current mirror load (M11 and M12) as the first stage, which is biased by a current source transistor (M15). The second stage is a common source amplifier with a PMOS input device (M16) and a current source transistor bias (M17). The capacitance of the gate and shorted gate, as well as the source and bulk of MC2, was introduced as a capacitor for frequency compensation. Frequency compensation is the intentional introduction of a capacitance to stabilize a circuit by shaping its frequency response and preventing instability at higher frequencies. In this case, MC2 simply filters these higher frequencies.
The open-loop gain of the amplifier is given by A_1A_2, where A_1 is the gain of the first stage and A_2 is the gain of the second stage. The closed-loop gain is calculated by equation 4.5. The feedback loop connects the output of the second stage with the non-inverting input of the first stage without any impedance in between. The signal is, therefore, fed back without any alteration and $\beta = 1$. Hence,

$$A_{CL} = \frac{A_1 A_2}{1 + A_1 A_2} \approx 1, \tag{5.12}$$

assuming that $A_1 A_2 \gg 1$. The buffer amplifier is, therefore, a unity-gain buffer.
The output resistance of the amplifier, looking into the output of the second stage, is determined by assuming all constant potentials and the input shorted to ground. By replacing the MOSFETs in the second stage with their small-signal equivalent circuits and simplifying the circuit, the small-signal model in Fig. 5.13 can be obtained. The first stage is simplified as a differential amplifier with a gain of A_1 and output voltage V_1, while a voltage V_{out} is applied to the output of the second stage.

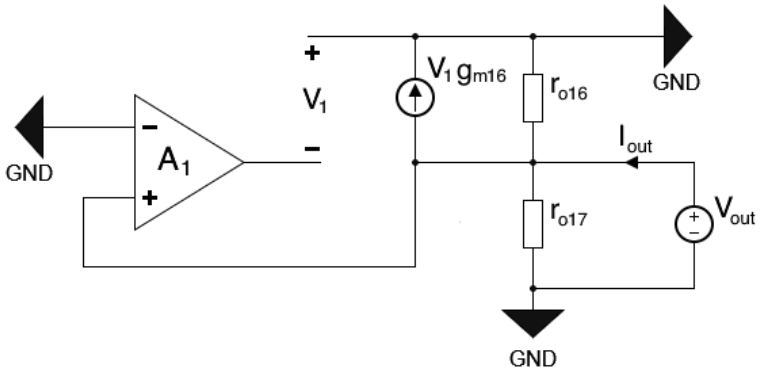


Figure 5.13: Small signal circuit for the calculation of the output resistance of the buffer amplifier.

The equivalent circuit of the second stage reminds of the circuit for the calculation of R_x (equation 5.10) seen in Fig. 5.11, where r_{o17} can be exchanged with $r_{o1} \parallel r_{o3}$, r_{o16} with r_{o2} and $V_x G_{m2}$ with $V_1 g_{m16}$. It can be seen that $V_1 = A_1 \cdot V_{out}$, due to the feedback path and the inverting input of the first stage being shorted to ground. Hence, $V_1 g_{m16} = A_1 \cdot V_{out} g_{m16}$.
The calculation of I_{out} is analogous to I_x and results in

$$R_{out} = \left(\frac{1}{A_1 g_{m16}} \parallel r_{o16} \right) \parallel r_{o17} \approx \frac{1}{A_1 g_{m16}}, \tag{5.13}$$

where it was assumed that $\frac{1}{A_1 g_{m16}} \ll r_{o16}, ro17$.
An exact calculation of A_1 can be found in [28]. Using the nomenclature introduced in this thesis, A_1 can be expressed as

$$A_1 = g_{m14}(r_{o12} \parallel r_{o14}) \frac{2 g_{m12} r_{o12} + 1}{2(g_{m12} r_{o12} + 1)} \approx g_{m14}(r_{o12} \parallel r_{o14}),$$

assuming that $g_{m12} r_{o12} \gg 1$. Plugging this into equation 5.13, R_{out} can be found to be

$$R_{\text{out}} = \frac{1}{g_{m14} g_{m16} (r_{o12} \parallel r_{o14})}. \tag{5.14}$$

5.4 Biasing Network

The biasing of the MOSFETs is accomplished using a network of MOSFET current mirrors [30]. The basic concept behind a current mirror is to copy the drain current I_{ref} flowing through a transistor, referred to as the reference transistor, into another transistor. The device copying the current is referred to as the mirroring device with a mirrored current $I_{mir} = I_{ref}$. The operation of a MOSFET current mirror relies on the fact that if two MOSFETs with the same channel length and width share the same gate-source voltage, they will conduct the same current. This, however, assumes that the MOSFETs are biased far enough in saturation. Fig. 5.14 shows the basic circuit of an NMOS current mirror.

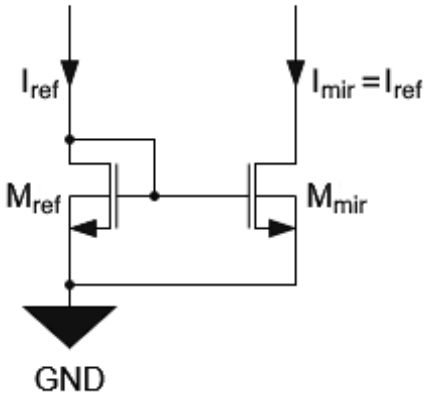


Figure 5.14: Basic schematic of a current mirror.

In a typical MOSFET current mirror, the reference device Mref has its drain connected to its gate and the gate of the mirroring device Mmir is connected to the gate of Mref. This configuration ensures that both transistors have the same gate-source voltage, i.e. $V_{GSref} = V_{GSmir}$. If the reference current I_{ref} is the drain current of Mref and the mirrored current I_{mir} is the drain current of Mmir, then – assuming both transistors in saturation and neglecting channel-length modulation –

$$\begin{aligned} I_{ref} &= \frac{1}{2}\mu_n C_{\text{ox}} \frac{W_{ref}}{L_{ref}} (V_{GSref} - V_{TH})^2 \\ I_{mir} &= \frac{1}{2}\mu_n C_{\text{ox}} \frac{W_{mir}}{L_{mir}} (V_{GSmir} - V_{TH})^2 \\ &\Rightarrow I_{mir} = \frac{W_{mir}/L_{mir}}{W_{ref}/L_{ref}}, \end{aligned} \tag{5.15}$$

where W_{ref}, L_{ref} are the channel length and width of the reference device, and W_{mir}, L_{mir} are the channel length and width of the mirroring device. $I_{ref} = I_{mir}$ holds if the channel dimensions of both devices are equal. From equation 5.15, it can also be concluded that the mirroring current can be scaled depending on the relation of the channel dimensions of the devices. I.e. if $L_{mir} = L_{ref}$, but $W_{mir} = x \cdot W_{ref}$, then

$I_{mir} = x \cdot I_{ref}$.
A current mirror can also be built with multiple mirroring devices. Fig. 5.15a shows a current mirror with two reference devices. Since the potential at the gate of Mref does not change due to the addition of an additional mirroring device, this can be exploited to build a current mirror network.

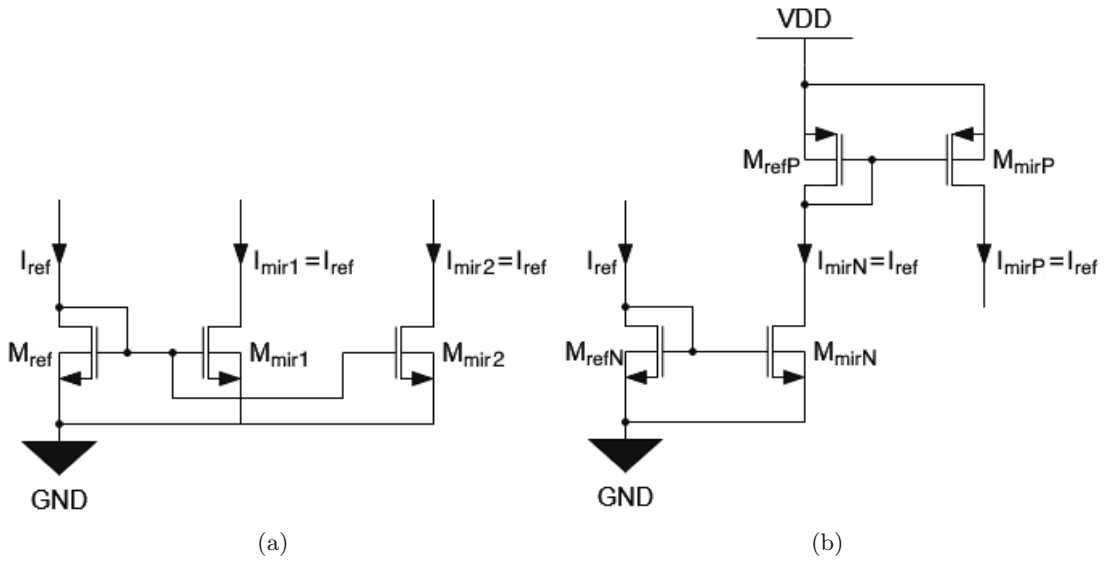


Figure 5.15: Schematic of (a) a current mirror with two mirroring devices and (b) an NMOS current mirror providing a reference current for a PMOS current mirror.

Note that the functionality of a PMOS current mirror is the same as that of an NMOS current mirror. However, typically, the reference current flows out of the drain of the reference device instead of into the drain. This can be used to further branch out the current mirror network by providing the reference current of a PMOS current mirror via an NMOS current mirror and vice versa. This is shown in the schematic in Fig. 5.15b. The biasing network of the CSA uses both the addition of multiple mirroring devices to one reference device as well as providing the reference current of a current mirror via a current mirror built from the opposite type of MOSFET.

5.5 Feedback Network

A CSA has a feedback capacitor parallel to a feedback resistance. However, as later on mentioned in section 6.1, the used CMOS process does not feature resistors or capacitors in its process development kit. In that case, a workaround using MOSFETs needs to be found. While the parasitic capacitances of a MOSFET can be used for the capacitor, the design of a feedback resistor is not as straightforward. Fig. 5.16 shows a rearranged version of the CSA circuit, showing the MOSFETs of the feedback network employed for the CSA, while the core and buffer amplifier are summarized using amplifier symbols. MC1 is used as the feedback capacitor. Its effective capacitance constitutes the parasitic capacitance between the gate and a shorted drain and source. MR acts as the feedback resistor [26, 31] and, as such, is driven in the linear region. The gate-source voltage of

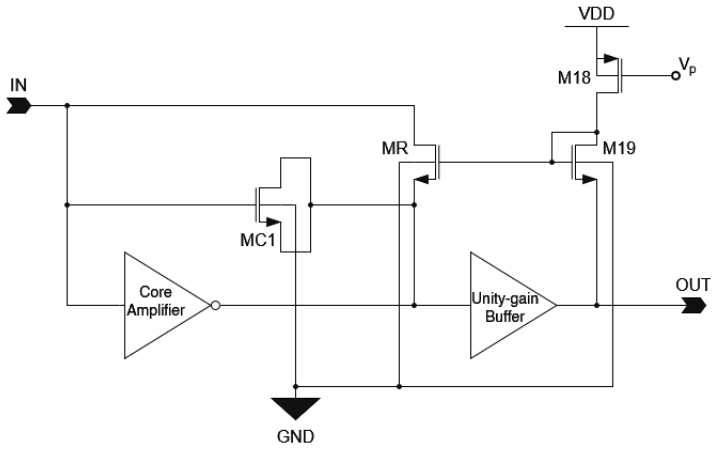


Figure 5.16: Schematic of the feedback network.

MR regulates its effective resistance, and for the sake of a high feedback resistance, MR has to be driven barely above the threshold. The current through MR can be assumed to be small, such that the drain-source voltage of MR is ideally a few mV to tens of mV at most.
Unwanted fluctuations of the DC operating caused by AC signals or by systematic changes like fluctuations in the power supply or temperature can cause changes in the potential at the source of MR and can hence alter the gate-source voltage and, consequently, the resistance. As a countermeasure, the biasing network consisting of MOSFETs M18 and M19 is introduced. Its purpose is to compensate for the change in the potential at the source of MR by introducing a proportional shift of the potential at its gate, such that the gate-source voltage ideally stays constant.
The circuit topology of M18, M19, and MR resembles a current mirror with M19 as the reference device and MR as the mirroring device. M18 acts as a constant current source supplying the reference current. However, since the drain-source voltage of MR is effectively 0, it is not capable of mirroring the current flowing through M19. Additionally, the source of MR and M19 are not connected to ground. MR connects to the output of the core amplifier. M19, on the other hand, connects to the output of the buffer amplifier, such that the drain current of M19 does not load the high resistive output node of the core amplifier. Since the gain of the buffer amplifier is equal to 1, any voltage fluctuations permeating through the buffer amplifier should be the same at its input and output.
Hence, any changes in the DC voltage at the source of MR should be sensed by M19 at its source. These changes cause a change in the drain-source and gate-source voltage of M19. To keep its drain current constant, which is supplied by M18, the drain-source and gate-source voltage of M19 need to change accordingly. Hence, the potential at the drain and gate of M19 and at the gate of MR experience a shift proportional to the sensed fluctuation. Consequently, the potential shift at the gate of MR compensates for the fluctuation appearing at its source.The quality of this feedback network could be negatively affected by either a buffer amplifier gain of not entirely 1 or by M18 not being far enough into saturation. The former causes M19 to sense a slightly altered

fluctuation, which potentially degrades the effect of the compensation effect. The latter could cause the drain current of M18 to be dependent on the fluctuation, which could negatively alter the potential shift at the drain and gate of M19.
M19, MC1 and MR have their bulks connected to ground instead of to their bulk connections. This way, the source implants have a pn-junction capacitance to the bulk of the MOSFETs. If the source were shorted to the bulk, the source would experience the junction capacitance between the p-Well and n-substrate of the epi-layer, which is connected to the supply voltage (bulks of all PMOS are connected to the supply voltage). The latter capacitance would be much higher than the former due to the larger pn-junction. In the case of the feedback network, the reduction of this capacitance potentially helps with feedback stability. The downside is that the shift of threshold voltage due to the body effect causes MR to need a high gate-source voltage, which causes the potential at the gate of MR to increase. The drain-source voltage of M18, therefore, could be reduced, making it impossible to properly get it into saturation.

5.6 Pads and ESD-Protection

The physical chip of the CSA, along with the sensor chip, will be placed on a PCB board, which includes interfaces from the physical power supplies to the electronics on the CSA chip. The powering of the CSA, as well as the signal transmission to and from the circuit, will be done via wire bond connections from the PCB to the chip. The wire bonds will be connected to the chip using bonding pads, which are placed next to the electronics.
Fig. 5.17 shows a layout of the pads that were used [32]. One pad is built from overlapping Metal1 and Metal2 sheets with dimensions of 90 µm × 80 µm connected by 8 10 µm × 10 µm ViaMetal1. The Metal2 layer is exposed through a 70 µm × 60 µm passivation opening for applying the wire bonds.

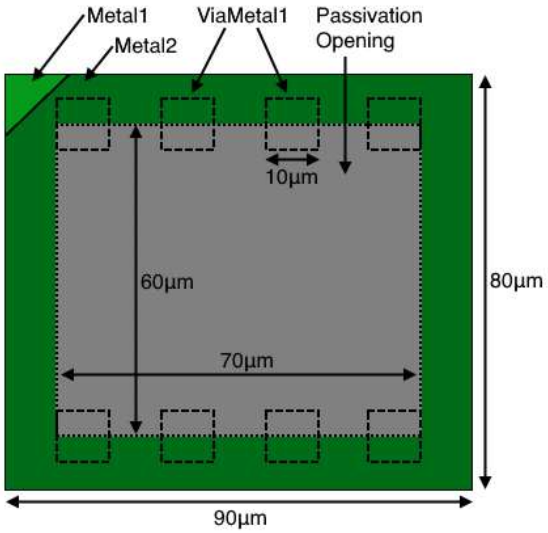


Figure 5.17: Layout of a pad used to bias the CSA and bring signals to and from the CSA.

For pads other than the ones that supply the supply voltage or the ground potential, electrostatic discharge (ESD) diodes were introduced to protect the circuit from ESD events. ESD events typically occur when charged objects are connected to the pads or simply touch them, causing a sudden pulse of high-valued voltage or current to interact with the circuitry on the chip, potentially destroying it. To prevent that, two ESD protection diodes can be introduced at a pad. These connect between a pad and, respectively, the supply voltage and ground rails or pads in reverse direction. In the case that a sudden ESD pulse above the breakdown voltage of the diode hits a pad, one diode becomes conductive, depending on the polarity of the voltage, and shunts the flowing current to either VDD or GND bypassing the circuitry.
The ESD diodes used for the layout were also taken from [32] and are shown in Fig. 5.18 along with how they are applied to opposite edges of a pad. One diode connects the pad to a p-implant, which forms the pn-junction with an n-well connected to the supply voltage. The other diode connects the pad to an n-implant, which forms the pn-junction with a p-well connected to ground.

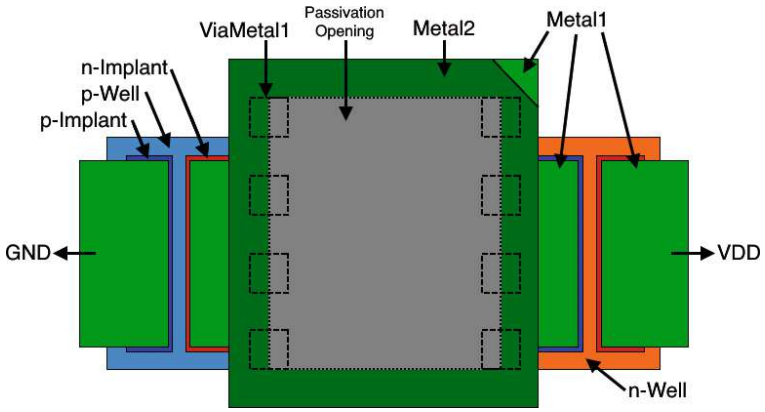


Figure 5.18: Layout of a pad with ESD diodes that can be connected to VDD and GND.

The pads and ESD diodes introduced parasitic capacitances that needed to be considered in the design of the CSA. The total parasitic capacitance constitutes the one between the Metal1 layer and the substrate below, as well as the depletion region of both diodes since they are reverse-biased under normal operation. The total estimated capacitance per pad amounts to roughly 1 pF.

6 SiC CMOS Technology

This section presents the CMOS technology used for the design of the CSA. First, an overview of the featured geometry and available layers, as well as the contents featured in the process development kit (PDK), are given. Then, an analysis of the featured SPICE models for the NMOS and PMOS transistors is presented, followed by an extraction of the threshold voltage from the models and a discussion of a methodology for the extraction of small-signal parameters as well as parasitic capacitances.
Note that this work includes and presents process-specific information regarding

- the thickness of the epitaxial layer,
- the layer structure,
- the sheet resistance of metal and poly-silicon layers,
- inter-layer capacitances,
- the accuracy of the Verilog-A models of the MOSFETs,
- and the temperatures and MOSFET dimensions, for which these models were experimentally validated.

This information was not publicly available at the time this thesis was published. However, officials of Fraunhofer IISB have confirmed with the author that the presented information can be disclosed and published in this thesis.

6.1 Technology Overview

The technology used to design the CSA is the Fraunhofer IISB 2 µm 4H-SiC CMOS process. It is developed on a SiC wafer with a 12 µm (reported by the manufacturer) thick n-doped epitaxial layer on top of a highly n-doped bulk substrate. Fig. 6.1 shows a visualization of the available implants and layers available in the process [12, 33] and for a possible implementation of NMOS and PMOS transistors.

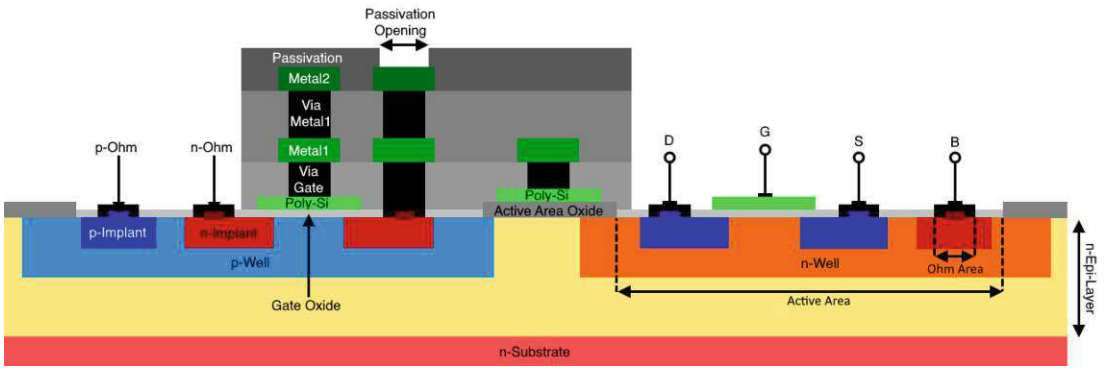


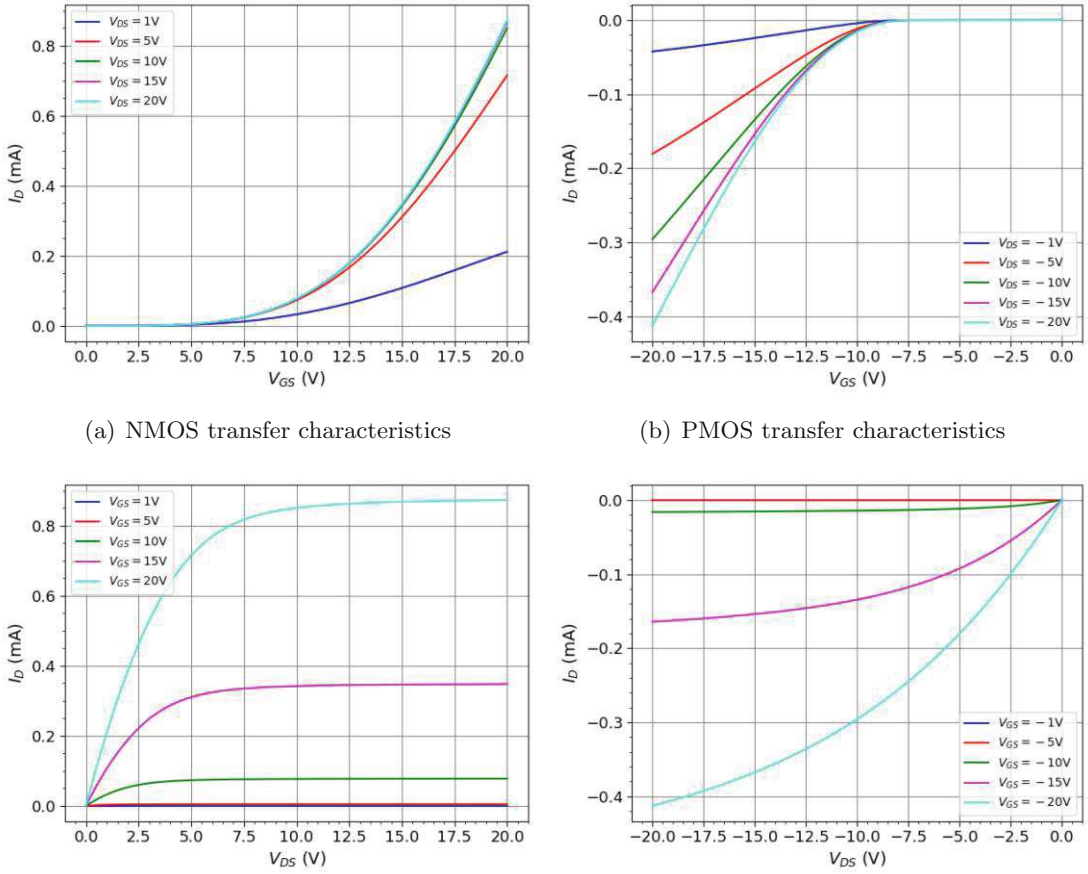
Figure 6.1: Visualisation of the used technology, including the available implants and layers and a possible implementation of NMOS and PMOS transistors.

The technology features n-doped implants and wells using nitrogen as a dopant, while the p-doped implants and wells use aluminum as a dopant. An active area layer forms an opening in the deepest SiO_2 passivation layer right above the epitaxial layer. Throughout the active layer, a gate oxide layer is placed, which can be contacted using a polycrystalline silicon (Poly-Si) layer, which is isolated from the SiC-substrate by the deepest SiO_2 passivation layer outside the active area. In the active area, the gate oxide layer can be opened up by specifying an ohmic area such that the substrate can be contacted from the first metal layer (Metal1). Additional metal implants in the ohm area between the Metal1 layer and the substrate below ensure ohmic contacts at the metal-substrate interface. The Metal1 layer lies above the Poly-Si layer and is isolated by another SiO_2 passivation layer. Above the Metal1 layer lies a second metal layer (Metal2). Metal1 and Metal1 are isolated by another SiO_2 intermetal isolation layer. Vertical interconnections between the Poly-Si and Metal1 layers can be established through metal-filled vias (ViaGate) that open up the isolating layer in between. The Metal1 and Metal2 layers can also be connected by vias (ViaMetal1). The Metal2 layer is isolated from the outside world by another SiO_2-passivation, which can be opened for external metal contacting by specifying a passivation opening area.
The essential building blocks of the technology are NMOS and PMOS transistors with peak voltages of up to 20 V. Even though the process is generally advertised as a 2 µm process, the minimal possible PMOS channel length is 6 µm. The process development kit (PDK) version 1.2 comes with compact SPICE models and ready-to-use schematic and layout cells of the transistors. The PDK features files for setting up layout versus schematic (LVS) checks in Cadence Virtuoso [34] and layout design rule checks (DRC) in KLayout.
The PDK version 1.2 does not include models or cells for passive components like resistors or capacitors. There are also no simulation possibilities for process variations, process corners, or device mismatch included. On top of that, an automated extraction of parasitic impedances from the layout is not included as well. Post-layout simulations would, therefore, only be possible if the parasitic impedances are extracted manually.

6.2 SiC MOSFET Models

The MOSFET models are based on the BSIM4SiC VerilogA model [35] with model parameters specifically tuned for the technology [13]. The models are experimentally verified (as reported by the manufacturer) for minimal channel widths of 3 µm for NMOS and 10 µm for PMOS as well as maximal widths of up to 100 µm for both NMOS and PMOS. The channel lengths are verified for values between and 2 µm and 10 µm for NMOS and between 6 µm and 10 µm for PMOS. Temperatures are validated in the range of 0 °C up to 300 °C. However, the manufacturer reports inaccuracies in the models when compared to the data measured on physical devices, especially for the shortest channel length of an NMOS of 2 µm and at high temperatures.
Fig. 6.2 show simulated transfer and output characteristics for an NMOS and a PMOS with a channel width of 100 µm and length of 6 µm [36]. The NMOS performance is significantly better than the PMOS by driving double the drain current at the same voltage biasing. Additionally, it seems that PMOS transistors reach strong inversion at

higher V_{GS} as well as saturate at higher V_{DS}. This indicates a higher threshold voltage than for an NMOS. The dependence of the drain current on V_{DS} in saturation also shows a higher slope for PMOS transistors.



(a) NMOS transfer characteristics

(b) PMOS transfer characteristics

(c) NMOS output characteristics

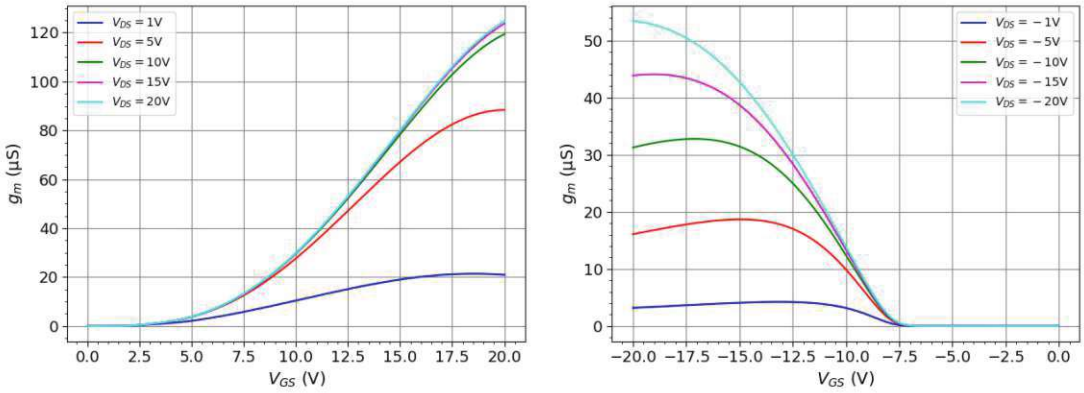
(d) PMOS output characteristics

Figure 6.2: Simulated (a),(b) transfer characteristics and (c),(d) output characteristics of SiC NMOS and PMOS transistors with a channel width of 100 µm and length of 6 µm at a temperature of 25 °C and bulk-source voltage $V_{bs} = 0$ V [36].

Modern sub-micrometer technologies in silicon can achieve the same currents with smaller MOSFET dimensions while also needing less voltage bias. As a result, circuits developed in the Fraunhofer IISB SiC process will generally be quite large in comparison. This introduces increased parasitic capacitances, which significantly limit the bandwidth performance.

Fig. 6.3 shows the simulated transconductance versus V_{GS} for different V_{DS} of an NMOS and PMOS [36]. The same dimensions as for the transfer- and output characteristics in Fig. 6.2 were used. The values move in the tens of µS for PMOS, while for NMOS, the transconductance can get > 100 µS. Modern silicon technologies can reach similar transconductances with smaller devices and lower operation voltages [28, 37]. This does not necessarily impose any significant constraints on the gain of an amplifier if the circuit

is designed properly. However, looking at equation 4.1 ($A_v = G_m \cdot R_{out}$), to reach high gain with a G_m in the scale of tens of µS, an amplifier needs quite high output resistance of at least a few MΩ, which makes driving high load impedances challenging. Even though a buffer amplifier allows the reduction of the output resistance seen by the load, the transconductance limitations also transfer to their output resistance. So to properly reduce the output resistance of a buffer amplifier a good choice of circuit is required.



(a) NMOS (b) PMOS

Figure 6.3: Simulated transconductance for different drain-source voltages V_{DS} of SiC NMOS and PMOS with a channel width of 100 µm and length of 6 µm at a temperature of 25 °C and bulk-source voltage $V_{bs} = 0$ V [36].

6.3 Extraction of Technology Parameters

The BSIM4SIC MOSFET models featured in the used PDK are empirical in nature and need a high number of model parameters as input to model the characteristics of a MOSFET properly. This also makes it non-trivial to extract typical MOSFET parameters from the models, like the threshold voltage, if these parameters are not directly accessible through the models themselves. When it comes to the SiC MOSFET models, the PDK features a full model card showing basic model parameters used as input. However, taking some of these parameters as they are specified in the PDK does not seem appropriate as they are dependent on the operating point. Hence, the are processed and overwritten during the simulation process due to e.g. the biasing of the MOSFET or the body effect. The effective parameter values resulting from the simulation are then not the same as the pre-specified input values. To the best knowledge of the author, the effective values are not directly accessible from the model via the simulation tools used for the design of the CSA. Hence, they need to be extracted manually via simulations. Apart from inherent model parameters like the threshold voltage, there are also MOSFET parameters, like the small-signal parameters, that are computed during a simulation and are therefore based on the input model parameters as well as the DC operating point. These are also not directly accessible via the used tools. Since the design approach used for designing the CSA circuit block requires several of these parameters, extracting them manually using simulations was necessary. Below is a listing of these parameters:

- Threshold voltage V_{TH}
- Small-signal parameters: transconductance g_m, body-effect transconductance g_{mb} and output resistance r_o
- Parasitic capacitance between gate and drain C_{GD}

Additionally, as already mentioned, the extraction of parasitic impedances introduced by the layout geometry also needed to be done manually. Therefore a method for doing so, needs to be discussed as well.
The remainder of this section presents the methods and partly the results of the extraction of the threshold voltage, small-signal parameters, and the parasitic impedances in the schematic (C_{GD} of a MOSFET) and layout.

6.3.1 Threshold Voltage

To extract the threshold voltage V_{TH}, the so-called "transconductance-ratio method" [38] was employed. This method uses the minimum of the derivative of the transconductance efficiency with respect to the gate-source voltage $\frac{\partial(g_m/I_D)}{\partial V_{GS}}$ as the indicator, at which V_{GS} the threshold voltage is located. Fig. 6.4 shows DC simulation results of the derivative of the transconductance-efficiency $\frac{\partial(g_m/I_D)}{\partial V_{GS}}$ as a function of the gate-source voltage of an NMOS and PMOS transistor with the bulk connected to the source, and an NMOS transistor with a bulk-source voltage of $V_{BS} = 6\,\mathrm{V}$. The threshold voltage of the latter is necessary, as M2 in the core amplifier is operated with a $V_{BS} = 6\,\mathrm{V}$ (see section 8.2). The widths of the transistors are 100 µm while their channel lengths are set to 5 µm for the NMOS devices and 6 µm for the PMOS device. For the NMOS and PMOS transistors with shorted source and bulk, the minimum of the respective $\frac{\partial(g_m/I_D)}{\partial V_{GS}}$-curves and therefore threshold voltages V_{TH} amount to

$$V_{\mathrm{TH,NMOS}} = 2.23\,\mathrm{V}$$
$$V_{\mathrm{TH,PMOS}} = 6.82\,\mathrm{V}.$$

For the NMOS with $V_{BS} = 6\,\mathrm{V}$ the body effect shifts the threshold voltage to

$$V_{\mathrm{TH,NMOS}} = 4.09\,\mathrm{V}.$$

6.3.2 Small-Signal Parameters

The extraction of the small-signal parameters was performed from the transfer- or output characteristics of a MOSFET. For the transconductance g_m, the drain-source voltage V_{DS} and source-body voltage V_{BS} were fixed, and a simulation of the DC operating point of the drain current I_D over a sweep of the gate-source voltage V_{GS} was done. As $g_m = \frac{\partial I_D}{\partial V_{GS}}\Big|_{V_{DS}=const.}$, the resulting transfer characteristics needed to be differentiated by V_{GS} to get a transconductance curve.
r_o was determined by keeping V_{GS} and V_{BS} constant and do another DC operating point simulation of I_D over a sweep of V_{DS}. By applying the derivative of the resulting output

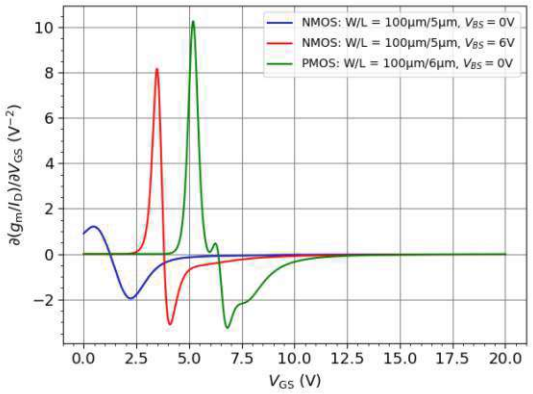
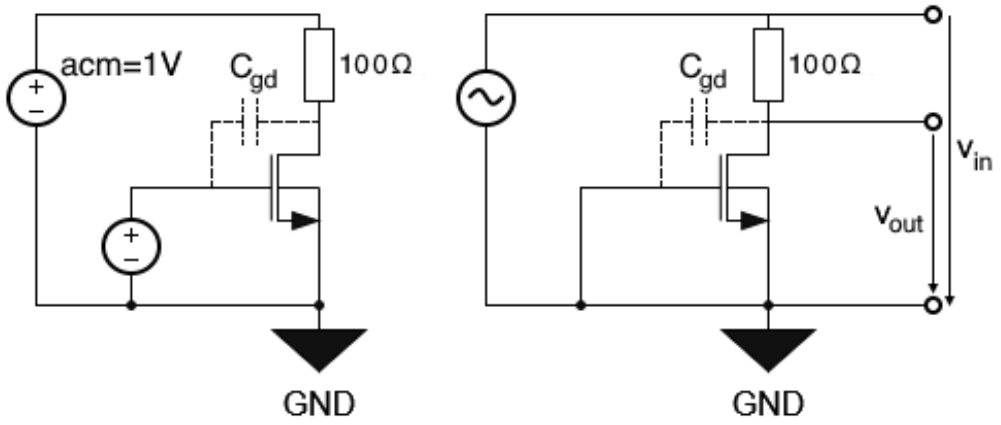


Figure 6.4: Derivative of the transconductance-efficiency $\frac{\partial(g_m/I_D)}{\partial V_{GS}}$ as a function of the gate-source voltage V_{GS} of an NMOS transistor, a PMOS transistor and an NMOS transistor with bulk-source voltage $V_{BS} = 6\,\mathrm{V}$ - the widths of all the transistors is $100\,\mu\mathrm{m}$, while the length is $5\,\mu\mathrm{m}$ for the NMOS transistors and $6\,\mu\mathrm{m}$ for the PMOS transistor, respectively.

characteristics, a curve for r_o was obtained by calculating the inverse of the derivative of V_{DS}, as $r_o = \frac{\partial V_{DS}}{\partial I_D}\Big|_{V_{GS}=const.}$.
The body-transconductance g_{mb} was obtained by doing yet another DC operating point simulation of I_D, but now V_{GS} and V_{DS} stay at a fixed value, while the V_{BS} was swept. By calculating the derivative of the resulting I_D versus V_{bs} curve, according to $g_{mb} = \frac{\partial I_D}{\partial V_{BS}}$, a curve for g_{mb} could be determined.
Note that for every curve of the three small-signal parameters, the actual value of the device had to be extracted by looking up the value that is located at the applied V_{GS}, V_{DS} or V_{BS} for g_m, r_o and g_{mb} respectively.

6.3.3 Parasitic Capacitances Between MOSFET Terminals

The strategy for extracting the parasitic capacitances between the gate and drain C_{gd} consisted of determining the cut-off frequency of a low-pass configuration with the parasitic capacitances as the capacitors. The idea was to fix the DC operating point of a MOSFET in simulation using ideal voltage sources and apply a small ideal resistor with $R = 100\,\Omega$ to the drain terminal. An equivalent circuit schematic for determining C_{gd} is shown in Fig. 6.5. The resistor was chosen to be rather small so as not to significantly influence the MOSFET's DC operating point, which could also influence the parasitic capacitance.
Using an AC simulation with a sweep over frequencies up to 100 GHz, the bode-diagram for the gain of the MOSFET could be obtained at the point between the resistor and parasitic capacitance (see v_{out} in Fig. 6.5b). Note that the stimulus for the AC simulation has to be created by the voltage source at the drain. The cut-off frequency $f_{\mathrm{cut-off}}$ of this low-pass configuration is found at the frequency $-3\,\mathrm{dB}$ of the maximum value in the Bode-diagram. Then, using equation 4.4, solving for C,



(a) (b)

Figure 6.5: Schematic used for the extraction of the parasitic gate-drain capacitance C_{gd} of an NMOS by inserting a small resistance of $10\,\Omega$ to the drain of the MOSFET forming a low-pass filter with the parasitic capacitances and applying an AC signal - (a) simulation schematic with a voltage source with an AC magnitude (`acm`) of $1\,\mathrm{V}$ (b) equivalent circuit in terms of AC signal i.e. static potentials are shorted to GND.

$$C = \frac{1}{2\pi R f_{\mathrm{cut-off}}},$$

and plugging $R = 100\,\Omega$ and the just determined $f_{\mathrm{cut-off}}$ into the resulting equation returned the value for the parasitic capacitance.

6.3.4 Parasitic Impedances from Layout

Before designing a corresponding physical layout, the SPICE simulations of a circuit usually only include parasitic impedances that are included in the models of the single devices in the circuit. However, the parasitic impedances introduced by wires and the relative position of devices to one another are not considered. As soon as a layout is developed, these parasitics can be determined from the layout geometry. The equivalent capacitances and resistances between nodes can then be included to simulate their influence on the circuit's performance. Modern silicon processes usually feature algorithms in their PDKs for automatic extraction and inclusion of parasitics in the circuit schematic to conveniently set up such post-layout simulations.
Unfortunately, the PDK used in this work did not feature automatic parasitic extraction. Hence, these needed to be extracted manually. The most notable parasitic contributions that needed to be determined were:

- Resistance of single metal wires
- Resistance of single Poly-Si wires
- Capacitance between metal or Poly-Si and the substrate

- Capacitance between Metal1 and Metal2
- Capacitance between Metal1/Metal2 and Poly-Si wires

The parasitic resistance R of a wire can be determined by knowing its material-dependent sheet resistance R_{sq}. If the wire has a length L and width W, the resistance is determined by

$$R = R_{sq}\frac{L}{W}. \tag{6.1}$$

R_{sq} is the resistance of a square-shaped sheet of material, i.e. where $W = L$.
The sheet resistance of the metal layers in the Fraunhofer IISB SiC process, as reported by the manufacturer, is given by $R_{\mathrm{sq,Metal}} = 0.8\,\frac{\Omega}{\mathrm{sq}}$. According to the design rules, the smallest wire width on Metal1 and Metal2 is limited to a few µm. Assuming, e.g., a width of 10 µm and length of 1 mm and using equation 6.1, the wire has a resistance of 80 Ω. Assuming a current of 100 µA, the voltage drop along a metal wire is 8 mV. This voltage drop can be seen as negligible, considering that the voltages of a single MOSFET in the Fraunhofer IISB SiC process are in the scale of a few V. Hence, it was ensured that only wires with a length $\leq$ 1 mm and a maximum current in the vicinity of 100 µA flowing through it were used. Under these circumstances, any wiring using the Metal1 and Metal2 layers could be done with the smallest possible width. This also helped minimize parasitic capacitances between the metal layers and other layers.
The sheet resistance of the Poly-Si layer is given by $R_{\mathrm{s,poly}} = 17\,\frac{\Omega}{\mathrm{sq}}$, as reported by the manufacturer. With a width of 10 µm and current of 100 µA a voltage drop of 8 mV (as in the example with the metal wires), a Poly-Si wire length of only approximately 47 µm is needed. So, for short wires with a length of a few µm, the voltage drop could also be seen as negligible. However, since the Poly-Si layer has the lowest distance to the substrate compared to the metal layers, it also creates the largest parasitic capacitance to the substrate, which can be critical at nodes important to the stability of the circuit. While parasitic resistances could be neglected for short wires, routing with the Poly-Si was still prevented as much as possible to preserve circuit stability.
The parasitic capacitances between the metal wires, Poly-Si wires and the substrate could be approximated as plate capacitors. Hence, they could be calculated using eq. 2.1 and by knowing the material and thickness of the dielectric between the respective layers. Tab. 6.1 shows the parasitic capacitances between the SiC-substrate, Poly-Si, Metal1 and Metal2 layers per µm^2. It shows that the parasitic capacitances with overlapping areas in the scale of µm^2 move in the range of a few fF. For higher overlaps, which mostly appear relative to the substrate, these capacitances could be scaled for the corresponding overlapping area and move up to capacitances of a few hundred fF. Combined with the resistance of $\leq$ 1 mm long metal wires and short Poly-Si wires, the resulting cut-off frequencies were in a scale $>$ 10 GHz. At this frequency scale, filtering effects by parasitic capacitances could be neglected for the most part.
Parasitic capacitances, however, that needed to be analyzed nonetheless were the ones that connect to high ohmic nodes, such as the output nodes of an amplifier. These parasitic capacitances add to the parasitic capacitances included by the MOSFET models, which causes additional phase shifts and reduction of the pole frequencies, which can alter the bandwidth and stability of the circuit. Any parasitic capacitances parallel to

Table 6.1: Parasitic capacitances between the SiC-substrate, Poly-Si, Metal1 and Metal2 layers per μm^2 in a layout.

Overlapping Layers	Capacitance ($fF\mu m^{-2}$)
Substrate/Poly-Si	0.08
Substrate/Metal1	0.04
Substrate/Metal2	0.03
Poly-Si/Metal1	0.09
Poly-Si/Metal2	0.04
Metal1/Metal2	0.08

the input and output capacitances needed to be considered if their magnitude was not negligible in comparison.
The routing between layers, using vias, also only introduces negligible parasitics. Since vias have a minimum cross-sectional area of a few µm and use the same material as the metal layers as conductor, they effectively have no resistance assuming currents in the scale of 100 µA. Since the minimal space between vias is also a few µm and their height in the range of nm, any horizontal capacitances between vias could also be neglected.
While applying long Poly-Si wires can generally be prevented quite easily, the long gate Poly-Si of wide MOSFETs cannot. Such gate Poly-Si can have lengths of hundreds of µm, if the MOSFET is not implemented using a finger structure. As an extreme example, assume an NMOS with a high channel width of 1 mm and a short channel length of 2 µm. The resistance of the gate Poly-Si is equal to 8.5 kΩ. This resistance can form a low pass with the non-negligible capacitances between the gate and the source or, respectively, the drain. Considering that these capacitances can be in the scale of hundreds of fF up to pF (for very wide MOSFETs and high V_{GS}, V_{GD}), the cut-off frequency in the scale of > 1 GHz. Bandwidth limitations via a long gate poly-Si could, therefore, be neglected as well. MOSFETs were still implemented using finger structures with a maximum width of 100 µm per finger. This limitation came from the maximum MOSFET widths, for which the SPICE models were experimentally validated.

7 Circuit Design Considerations and Methodologies

This section presents the basic requirements, considerations and methods that are used to design the CSA circuit using the Fraunhofer IISB CMOS process. First, the circuit specifications and performance requirements are discussed based on potential lab test scenarios and the limits imposed by the used CMOS technology. The view is then shifted from the whole circuit to single MOSFET devices and the requirements regarding their operation as well as sizing needed to reach the wanted circuit performance. The section closes with an explanation of the method used to determine the sizing of the MOSFETs, as well as the software tools that were leveraged to find the sizing and simulate the subsequent circuit performance.

7.1 Circuit Requirements

The main goal in the design of the CSA is to study the capabilities of the underlying SiC CMOS process for developing particle sensor readout electronics. A possible combination of the electronics with the sensor on-chip to develop a SiC MAPS is a future prospect to keep in mind, as mentioned already. At this stage, however, the separate design of the CSA presented in this work is seen as a proof of concept. Therefore, the focus lies mainly on an overall functional circuit, which can be used to confidently detect the presence or absence of particles when paired with an appropriate sensor. Under this condition, an SNR of $geq10$ for a MIP is typically required. For the stability of a circuit, a phase margin of around 65° is typically required. As parasitic impedances can degrade circuit stability and an extraction of parasitics from the layout is absent in the used PDK, a safety margin of 65° to 90° was applied as the stability requirement for the CSA. Additionally, no restrictions concerning the circuit size, other than the available space on the chip, the power consumption, or timing capabilities were defined.
Restrictions on the CSA from the planned testing environments could also be defined. One testing case uses it as a pre-amplifier for particle detection with further readout electronics following the CSA. The other one is the electronic characterization of the circuit itself using a signal generator and oscilloscope. Considering these applications along with the requirements for SNR and stability, the compatibility with different sensors and load impedances needed to be simulated. Further requirements for the voltage supply or performance capabilities could be defined from the used CMOS process.

7.1.1 Sensor Configurations

Two critical sensor configurations were considered for testing the CSA. First off, a sensor pixel developed along with the CSA in the Fraunhofer IISB SiC CMOS process with a pitch of $50\,\mu m \times 50\,\mu m$. The thickness of the epi-layer of $12\,\mu m$ was assumed as the active thickness of the sensor in full depletion. Considering that a MIP that interacts with a SiC sensor deposits a signal charge of $55\,e^{-}\mu m^{-1}$ on average [20], a minimal average signal charge of $Q_{in} = 660\,e^{-}$ was expected for such a pixel sensor. An SNR of at least 10, therefore, requires a minimum ENC of $66\,e^{-}$ from the CSA.
Treating the pixel pitch as the active area of the pixel, a sensor capacitance of approximately 180 fF could be estimated. However, since the sensor and CSA circuit will be

located on separate chips, they need to be connected using wire bonds via the bond pads explained in section 10.1.3. The bond pads add an additional capacitance, bringing the total input capacitance to roughly $C_{in} = 1\,\text{pF}$, considering two bond pads (one on the sensor chip and one on the CSA chip).
The second critical configuration was considered as an auxiliary sensor to the thinner pixel described above. The idea was to use another SiC-based sensor that is, however, thicker and already proven to work, in case the signal charge created in the thin pixel is not enough to satisfy the design requirements. Hence, an active thickness of 50 µm was expected, as this is the typical thickness of sensors grown in an epitaxial process. A signal charge of approximately $Q_{in} = 2754\,\text{e}^-$ could, therefore, be expected for a MIP, which is roughly four times the charge as for the thin pixel sensor. Therefore, the minimum noise requirement was set to an ENC of $275\,\text{e}^-$.
A high value was considered for sensor capacitance to cover a larger range of possibly compatible input capacitances. The capacitance of a strip sensor, which can typically go up to 4 pF [39], was chosen for this sensor configuration. However, the capacitance of the pad on the CSA chip needed to be added to make an input capacitance of roughly $C_{in} = 4.5\,\text{pF}$.
Due to the difference in active thickness of the two sensor configurations, the SiC pixel with a signal charge of $Q_{in} = 660\,\text{e}^-$ and input capacitance of $C_{in} = 1\,\text{pF}$ will be considered as the "thin" sensor configuration throughout the remainder of this thesis. The thicker sensor with a signal charge of $Q_{in} = 2754\,\text{e}^-$ and input capacitance of $C_{in} = 4.5\,\text{pF}$ will therefore be dubbed as the "thick" sensor configuration.

7.1.2 Load Impedance Requirements

The output load impedance requirements were determined by the possibility of using the CSA in a detector readout chain or by its electrical characterization via an oscilloscope. For the lower margin of load capacitances, ADCs and buffer amplifiers with low input capacitances of 4 pF were considered. For the upper margin of load capacitances a 10x oscilloscope probe was assumed, which typically have load capacitances up to 20 pF. The load capacitance requirements were therefore set to $C_L = 4\,\text{pF}$ up to 20 pF.
Since the output-driver stage of the CSA uses a low output resistance buffer amplifier, subsequent electronics with low input resistance loading the CSA were considered to be critical to the functionality of the CSA. Load impedances in the scale of the output resistance draw a significant amount of current out of the circuit, causing a drift of the DC operating point at the output and ultimately altering the circuit performance. To prevent this, a load resistance of $R_L = 1\,\text{M}\Omega$ was considered as the minimum. The resistance of a 10x oscilloscope probe can typically be found at $1\,\text{M}\Omega$ parallel to its capacitance. For electronics introducing a lower load resistance, a $1\,\text{M}\Omega$ resistor would therefore be needed to be applied off-chip.

7.1.3 Requirements from the CMOS Process

Further requirements for the CSA circuit arise from the used CMOS process. Compared to common CMOS processes using silicon in the sub-micrometer range, the Fraunhofer IISB process is quite large with a minimum channel length of 2 µm. The consequence

of feature sizes in this scale is slow electronics. The reason are parasitic capacitances that form between the MOSFET terminals, which can be found mostly in the scale of hundreds of fF as well as inverse MOSFET transconductances and MOSFET output resistances that move in the scale of a few up to tens of MΩ for MOSFETs with a width above 100 µm. Plugging a resistance of 1 MΩ and a capacitance of 100 fF into equation 4.4 an optimistic estimate for bandwidth limitations can be found at approximately 1.5 MHz. In practice, it turns out that especially parasitic terminal capacitances of MOSFETs with widths of hundreds of µm end up in hundreds of fF, such that realistically bandwidths in scales of tens to hundreds of kHz can be expected using the Fraunhofer IISB SiC process. A previous design study using this process reached an amplifier bandwidth of tens of kHz [33]. Hence, a design goal of at least 10 kHz for the bandwidth was applied. The corresponding gain was chosen as high as possible to make an SNR of at least 10 achievable.
Since the load capacitance of the CSA could go up to 20 pF, it had to be ensured that the output of the buffer amplifier does not introduce a new dominant pole. In order to do so, a requirement for the output resistance of the buffer amplifier was set to a value below 1000 Ω. This way, with a load capacitance of 20 pF, the pole frequency at the output of the buffer amplifier could be removed by 3 decades from the just estimated 10 kHz-range.
The peak voltage rating of the MOSFETs in the process was considered for requirements on the voltage headroom. It is specified as maximal gate-source and drain-source voltages of 20 V. Since no requirements on power consumption were defined, the supply voltage of the circuit is set to 20 V. In that way, the voltage budget is enough to run three MOSFETs with widths of hundreds of micrometers in series from VDD to GND and simultaneously drive them deep enough in saturation without the risk of overstepping the voltage limit on either MOSFET.
When it comes to requirements for currents, the discussions from section 6.3.4 about parasitic impedances in the layout were considered. It was defined that long wires should not exceed currents in the vicinity of 100 µA to make the voltage drop along the wire negligible. During the circuit design stage, however, it is not known a priori whether a wire in the schematic has to be represented by a long wire in the layout. Hence, the maximal currents flowing through a single wire were restricted to around 100 µA during the circuit design phase.
This also keeps the power consumption of the circuit manageable. Considering that the core amplifier consists of one stage and the buffer amplifier out of two, the maximum power consumption per stage would at a supply voltage of $V_{DD} = 20\,\mathrm{V}$ would be around 2 mW. Hence, in total a maximum power consumption in the vicinity of 6 mW could be expected.

7.2 Considerations for Single MOSFETs

The sizing of a MOSFET, combined with its operating point, determines its small-signal parameters and, therefore, directly influences the performance of a circuit, as it can be expressed as a function of the small-signal parameters. Requirements on the performance of the circuit can, therefore, limit the parameter space for the small-signal parameters

and, retroactively, also the sizing and biasing of the MOSFETs. Since the CSA consists of several MOSFETs and the requirements for the circuit are relatively loosely defined, the parameter space for the design is still quite large. In this section, in addition to the performance requirements, specific requirements for the sizing and DC operating point of individual MOSFETs are predefined to restrict the available parameter space further. With regard to the DC operating point, considerations regarding the appropriate operating region of the MOSFETs are discussed first. Subsequently, requirements for the channel lengths of the MOSFETs are defined based on considerations of the systematic circumstances and the simulation models of the MOSFETs. The channel widths can then be chosen based on these previous considerations and follows a method presented in the next section.

7.2.1 Operation Requirements

Whether a MOSFET is driven in weak-, moderate- or strong inversion brings different advantages and disadvantages. MOSFETs driven with low gate-source voltage, i.e. operated in weak inversion, have an overall lower drain current output and are therefore advantageous for applications where low power consumption is needed. However, to reach high drain currents that might be needed to meet the desired circuit performance, devices tend to have larger device dimensions, which results in lower bandwidth due to high parasitic impedances between MOSFET terminals. Large devices with lower currents, on the other hand, are not as prone to device mismatch or random process variations of the MOSFET channel dimensions.
In the context of the design of the CSA, very low power consumption is not of the highest priority. Additionally, the used process with a minimum channel length of 2 µm inherently introduces rather large devices at already low drain currents, which can be critical due to space constraints on the chip die as well as significant bandwidth limitations. Also, considering that the MOSFET models tend to be least precise in weak inversion when compared to experimental data, weak inversion will not be used.
Devices driven with high gate-source voltages, i.e. operated in strong inversion, allow for high drain currents using comparably smaller MOSFET dimensions. This results in higher bandwidth while sacrificing power consumption and stronger influences of process variations. Since the extent of the influence of process variations in the used CMOS process on the circuit performance is not known at this stage, fully driving in strong inversion does not seem ideal either.
The moderate inversion region, which lies between weak and strong inversion, is often regarded as a sweet spot that draws from the advantages (and disadvantages) of its neighboring regions. To keep the circuit size as well as influences from the process variations manageable, it seems appropriate to favorably choose an operation in moderate inversion. To additionally combat the impreciseness of the MOSFET models and bandwidth limitations in the weak inversion region, it further seems legitimate to choose the gate-source voltage such that MOSFETs still operate in moderate inversion but closer to strong inversion than to weak inversion.
Unfortunately, an exact voltage range for moderate inversion is not known in this process. The only known reference points are the threshold voltages determined in section 6.3.1. The threshold voltage is located in the center of the moderate inversion region

i.e. $V_{GS} > V_{TH}$ moves the operation region more towards strong inversion. Under the assumption that the supply voltage might vary by 10 %, i.e. a drift of the DC operation point of 2 V, a requirement for the gate-source voltage could be defined as $V_{GS} \geq V_{TH} + 2\,\mathrm{V}$ to keep the MOSFETs from operating closer to weak inversion and ensure that they are closer to strong inversion for the nominal supply voltage. However, since the lower boundary of strong inversion is not known, the gate-source voltages might end up in strong inversion. But from inspection of the transfer characteristics (see e.g. 6.2) of the MOSFETs, ensuring that $V_{GS} \leq V_{TH} + 4\,\mathrm{V}$, the drain currents are not placed in a significant increase of the curves, so they are probably at most at the boundary of moderate inversion and strong inversion, which is deemed as acceptable. Hence, the design margin for the gate-source voltage of a MOSFET was set to $V_{TH} + 2\,\mathrm{V} \leq V_{GS} \leq V_{TH} + 4\,\mathrm{V}$. For an NMOS with $V_{BS} = 0\,\mathrm{V}$ this means $4.23\,\mathrm{V} \leq V_{GS} \leq 6.23\,\mathrm{V}$, for an NMOS with $V_{BS} = 6\,\mathrm{V}$ it means $6.09\,\mathrm{V} \leq V_{GS} \leq 8.09\,\mathrm{V}$ and for a PMOS $8.82\,\mathrm{V} \leq V_{GS} \leq 10.23\,\mathrm{V}$.
Regarding output characteristics, it is generally advantageous to drive a MOSFET in saturation, where the drain current is less dependent on the drain-source voltage. The circuit performance stays more stable over variations of the DC operating point due to, e.g., variations of the voltage supply because the MOSFET can still provide the same current over a wider range of gate-drain voltages. This is especially important for MOSFETs that are required to provide constant currents. To ensure that a MOSFET is deep enough in saturation, i.e. still in saturation also with a DC operating point offset of up to 2 V, it is generally preferred to choose the drain-source voltage greater equal than the gate-source voltage i.e. $V_{DS} \geq V_{GS}$, which is hence defined as a design requirement.
The above-defined requirements were not seen as definite rules for the design of the corresponding values but rather as a guideline and were, therefore, applied as much as possible. However, a few special cases where it was impossible to satisfy them, or they were intentionally violated, are explicitly pointed out throughout section 8.

7.2.2 Choice of Channel Length

The smallest possible channel lengths of the MOSFET defined by the process are 2 µm for NMOS and 6 µm for PMOS devices. These lengths should be chosen consistently to keep the circuit size on the layout as small as possible. In practice, the choice was not quite so trivial.
Without process corners or device mismatch simulations, it is not possible to know how uncertainties in the fabrication of the devices might influence the deviation between the simulated performance and the physical performance of the circuit. Deviations in the MOSFET channel dimensions can cause the DC operating point as well as the drain current to drift from the wanted values. However, the drain current is linearly proportional to $\frac{W}{L}$. Therefore, choosing a larger channel length for the same drain current reduces the slope with which the drain current changes when the channel width deviates from the nominal value.
Additionally, the MOSFET models show a worse fit to the experimental data for smaller channel lengths. Choosing the minimal lengths, therefore, can result in a less realistic simulation of the physical circuit. For the minimal length of PMOS devices, this is,

however, not as critical as for NMOS devices.
Considering both considerations, for NMOS devices, the channel length was purposely chosen higher with 5 µm as the general guideline value. This value was generally chosen for MOSFETs that act as constant current sources. However, in rare cases, either lower values to reduce parasitic capacitances or higher values for MOSFETs that act as resistors were taken.
For PMOS devices, a slightly higher channel length of 7 µm was chosen if a device carried a low drain current and the width stayed reasonably small. For higher drain currents, it has to be taken into account that PMOS devices have a lower current output than NMOS devices at the same DC operating point and sizing. Therefore, if a high drain current unreasonably inflated a PMOS device's width, the minimal channel length of 6 µm was chosen.

7.3 MOSFET Sizing Methodology

The design of the core amplifier and the buffer stage employs a systematic approach that is inspired by the methodologies presented in [37]. The approach in this thesis employs considerations based on the operation regions (weak-, moderate- and strong inversion) to find appropriate biasing of MOSFETs in combination with the generation of look-up tables of MOSFET parameters, which are used to calculate e.g. the gain and bandwidth of a circuit as derived via the small-signal model presented in section 5.2. These MOSFET parameters, like the small-signal parameters, are tied to the MOSFET channel dimensions at a specific DC operating point. Hence, instead of looking for MOSFET width and lengths as well as proper biasing that result in the desired circuit performance with a large multi-dimensional sweep, the opposite can be done. Using parameters like the small-signal parameters, first, the performance characteristics (e.g. gain and bandwidth) of a circuit are calculated, and from a set of small-signal parameters that result in the desired performance, the corresponding MOSFET sizing can be determined. This exact procedure is explained in the remainder of this section.
A look-up table (LUT) is a multi-dimensional array that captures a MOSFET parameter such as the drain current or the small-signal parameters at various different DC operating points and MOSFET sizings. While one dimension of the array corresponds to a listing of the MOSFET parameter of interest, the remaining axes respectively contain a sweep over several values of every degree of freedom with which the parameter can be tuned. These degrees of freedom are usually the terminal voltages V_{GS} and V_{gd} and the MOSFET channel width W and the channel length L. So an LUT contains the corresponding MOSFET parameter for every permutation of the swept degrees of freedom.
The parameters of interest featured in an LUT can vary depending on the specific application. However, since the gain and bandwidth of the core amplifier from equations 5.8 and 5.9, as well as the output resistance from equation 5.14 of the buffer amplifier will be the most significant figures of merit for the design of the two circuit blocks, the following parameters of a single MOSFET are of interest:

- Drain current I_D
- Small-signal parameters: g_m, r_o and g_{mb}

- Gate-drain capacitance: C_{gd}

Here, the values for the length L as well as the voltages V_{GS} and V_{DS} of single MOSFETs will be defined according to process-specific considerations already before the generation of the LUTs, which will be explained in more detail later on. These values will, therefore, stay constant during the generation of the LUTs. The only degree of freedom then left, which can be used to tune the above-listed parameters, is the channel width W, which simplifies the LUTs to a 2-dimensional array.
The generation of the parameters for several W can e.g. be achieved by doing a simulation involving a sweep of values for W and recording the parameters of interest for every simulation. While this method produces values true to the MOSFET models and is, therefore, the most accurate way, it can prove to be tedious as it includes a high number of single simulations. This is especially true if the circuit features several MOSFET devices, as every MOSFET needs its own LUTs for every parameter of interest.
Alternatively, since all of the listed parameters have a dependence on W, they can be scaled according to their proportionality, as long as the parameters of a reference device are known. As explained in section 3 it can be seen that I_D, g_m, g_{mb} and C_{gd} scale linearly with W, while r_o scales with the inverse of W. Therefore, by knowing the value x_{ref} of a parameter for a reference device, the remaining values x can be calculated by simply normalizing x_{ref} to the reference width W_{ref} and scaling the normalized reference parameters by W. For direct proportionality, this means that a parameter x at a specific width W is

$$x = \frac{x_{\text{ref}}}{W_{\text{ref}}} W, \tag{7.1}$$

where x could be I_D, g_m, g_{mb} or C_{gd}.
For inverse proportionality, the scaling looks like the following:

$$x = \frac{x_{\text{ref}} W_{\text{ref}}}{W}, \tag{7.2}$$

where x could be r_o. Using these expressions instead of simulations for every parameter can speed up the generation of the LUTs if the calculations are automated using, e.g. a corresponding Python script.
If a lookup table is calculated for every MOSFET in a circuit, the performance characteristics for an arbitrary combination of MOSFET widths can be calculated. However, it should be noted that not every arbitrary set of MOSFET widths is usable, as the MOSFETs are analyzed separately and they can have arbitrary combinations of drain currents. Hence, in the context of a specific circuit topology, the drain currents of the single MOSFETs from an arbitrary set of widths do not necessarily obey Kirchhoff's current law i.e. the sum of ingoing and outgoing currents at every node vanishes. Therefore, relations between the widths of the MOSFETs based on this law need to be found to ensure that the ingoing and outgoing currents at every node are conserved. These relations can be very convenient as the parameters of a single MOSFET at a specific W might be used to derive other devices' corresponding widths and parameters. This also helps to restrict the possible space of parameters, which speeds up the calculation of the performance characteristics while at the same time ensuring that they obey Kirchhoff's current law. The relations between the MOSFET widths are specific for every circuit

topology and will, therefore, be derived for the core and buffer amplifier in the following sections.
One downside to this approach, however, is that the assumed linear or inverse dependencies of the MOSFET parameters on W are only approximately true for empirical models like the BSIM models. Plugging a set of widths derived from these ideal dependencies into SPICE simulations that use empirical models, the results will not necessarily reproduce the same performance characteristics, albeit similar ones. Therefore, the results from this approach can be used as a convenient way of finding a basis for MOSFET dimensions that result in similar performance characteristics as desired so that only minor adjustments are needed on the MOSFET widths to find the wanted circuit performance.

7.4 Software Tools

The overall design of the CSA circuit and layout is done using the Cadence Virtuoso IC design environment [34]. It is an industry-standard software for the design of CMOS integrated circuits and chip layouts. For the purpose of designing analog circuits, the Cadence Virtuoso analog design suite offers several tools for simulating analog circuits using a SPICE simulator and designing a corresponding physical layout.
The Virtuoso Analog Design Environment (ADE) contains a circuit schematic editor, which allows the creation of circuit schematics consisting of several electronic devices from different device libraries. The main library, which is used for electronic devices, is the one provided in the Fraunhofer IISB SiC process PDK consisting of cells for SiC NMOS and PMOS transistors, respectively, which have intrinsic parameters to define the channel length and width of a single device. Ideal resistors and capacitors from the default analog devices library are mainly used to represent the output load or input impedances and parasitic and extract MOSFET parameters. Ideal current or voltage sources from the default analog device library can be used to provide stimuli (e.g. input signal or AC signal) for the MOSFET circuits or biasing a MOSFET to extract parameters.
Using the Spectre SPICE simulator, different kinds of simulated circuit analyses can be carried out to extract different characteristics of the circuit. A single simulation run can be configured to carry out several such analyses. Among numerous options, the following are used in the design of the CSA: DC analysis, AC analysis, stability analysis, transient analysis and noise analysis.
The DC analysis is used to calculate the DC operating point of a circuit i.e. the static voltages at every node in the circuit and static currents flowing through the electronic devices, which helps in checking that the devices are properly biased. They also help in extracting the small-signal parameters of the MOSFETs, as explained in section 6.3.2.
AC analysis can be used to calculate the frequency of the circuit's response. This way, the gain, bandwidth and phase shift of amplifiers, as depicted in a Bode diagram, can be simulated. The AC simulation is a small-signal simulation that linearises the circuit around its DC operating point and calculates the gain magnitude and phase shift over a frequency sweep relative to either an AC signal emitted by a source or between two nodes. For an AC analysis, including at least one voltage or current source with an AC magnitude unequal to 0 V is important. Throughout the design process of the CSA, AC analyses only use an AC magnitude of 1 V.

A stability analysis calculates the loop gain and corresponding phase shift of a circuit with a feedback loop. Using a probe device (`IProbe`) from the analog device library, the feedback loop is broken, and the loop gain, the phase shift, is determined when an AC signal with swept frequency is sent around the broken loop. As a result, a Bode diagram of the loop gain and phase shift can be plotted, while also the stability of the feedback loop via the phase margin can be tested.
Using a transient analysis, the time evolution of the voltage at every node and the current through a device around the DC operating point can be simulated. Using transient simulations, the actual circuit response to a voltage or current input signal can be determined and plotted. The simulation is configured by defining a time frame in which the transient response should be simulated and a maximum time step width. The actual time steps are chosen by the simulation, but the granularity can be limited by specifying the step width.
Lastly, a noise analysis is used, which calculates the noise-frequency spectrum of a circuit. The noise analysis linearises the circuit around the DC operating point and determines the noise spectrum referred to the input by specifying an input voltage or current source or output by specifying the node of interest. The spectrum can be plotted as well as integrated over a range of frequencies to obtain the total root-mean-square noise of the circuit in that frequency range (e.g. up to the bandwidth frequency).
For the purpose of simulating random process variations, Spectre also has the possibility to do simulations with randomized parameters using the Monte-Carlo method. For that, a variable with an underlying probability distribution can be defined and used as an input for a device parameter (e.g. length or width of MOSFET). By then specifying a number of points and a random seed, a sample of random values from the underlying distribution is calculated for the variable and for every value in the sample a simulation with the specified analyses is carried out. Additionally, a sample for several random variables can also be created in parallel.
For the design of a layout, the Virtuoso environment also offers a layout editor. Similar to the schematic editor, the layout editor allows one to pick from ready-to-use cells contained in different libraries in order to arrange the layout geometry. The Fraunhofer IISB SiC process PDK comes with layout cells for NMOS and PMOS transistors, which again have parameters to change the length and width. These cells are built from instances in the several layers of the CMOS technology. Additionally to the MOSFET cells, the editor allows to draw arbitrarily shaped instances of all the layers in the technology e.g. metal wires for interconnecting the electronic devices.

8 Circuit Design Process

This section presents the design process of the CSA circuit step-by-step, applying the tools and considerations presented in the previous section. A short overview of the design workflow is given. Following that, the single design steps are explained in detail.

8.1 Circuit Design Workflow

The design process for the CSA circuit schematic was divided into 5 steps, each focussing on a specific circuit block. The first step was concerned with the core amplifier and its biasing network. First, the voltage biasing and MOSFET sizing were determined using the requirements and methods presented in the previous section. This design step was driven by the requirement of a gain as high as possible for the circuit. The bandwidth had to be chosen reasonably high as a safety precaution to not fall below the minimum required bandwidth of 10 kHz, due to adding the buffer amplifier later on. Next, the biasing network for the core amplifier was added.
The buffer amplifier followed the same steps as the core amplifier, only that the design was driven by its output resistance. The addition of the biasing network of the buffer amplifier meant the inclusion of the buffer to the biasing network of the core amplifier already designed in the previous step, as well as the addition of the core amplifier to the input of the buffer amplifier.
The feedback network was added to complete the closed-loop structure of the CSA. First, the channel lengths of the MOSFETs were determined using the considerations from the previous section. However, the DC operating point and the channel widths were determined over a simulation sweep of the device widths. The current, with which the feedback network loads the CSA output, the SNR and feedback stability were used as a figure of merit to find an appropriate setup.
Next, the closed-loop circuit configuration needed to be changed to include every MOSFET divided onto several fingers with a width of $\leq 100\,\mu m$ (see end of section 6.3.4). Every finger of a single MOSFET was then included in the schematic as a separate MOSFET, and the respective terminals were shorted to operate the fingers with the effective width of the undivided MOSFET.

8.2 Design of the Core Amplifier

The core amplifier is implemented as a regulated folded cascode amplifier, which is presented in section 5.2.
The design requirements for the core amplifier by itself were set to a high as possible gain at a bandwidth of at least 10 kHz. A design approach consisting of the following steps is adopted:

- Voltage biasing
- Sizing of the MOSFETs
- Simulate and adjust full schematic

- Add the biasing network and make final adjustments

The exact methodology and considerations of every step will be presented in this section.

8.2.1 Voltage Biasing

The first step in the design of the core amplifier is finding the voltage biasing of the MOSFETs based on the available voltage budget, the threshold voltage of the MOSFETs and the desired operation regions of the MOSFETs. As explained in section 7.2.1, the MOSFETs are preferably driven in moderate inversion closer towards strong inversion and in saturation. To accomplish this, first V_{GS} is chosen in the range $V_{TH} + 2\,\mathrm{V} \leq V_{GS} \leq V_{TH} + 4\,\mathrm{V}$. Then V_{DS} is chosen such that the MOSFETs operate in saturation, considering that the MOSFETs begin to saturate at $V_{DS} = V_{GS} - V_{TH}$. If possible, the MOSFETs have $V_{DS} \geq V_{GS}$ to ensure that they are deep enough in saturation. However, the voltage budget of $V_{DD} = 20\,\mathrm{V}$ needs to be taken into account as a limiting factor for the sum of drain-source voltages of the MOSFETs between V_{DD} and GND.
As explained in section 5.2, the gain of the core amplifier is directly proportional to g_{m1}, M1 is expected to have a high width and therefore high current to boost the gain of the circuit. The width of M1 needs to be high to achieve that, which increases its gate-source capacitance. These two capacitances add to the input capacitance of the circuit, as they both connect from the input node to GND in the small-signal approximation.
To naturally decrease them in the subsequent design, the gate-source voltage of M1 was chosen more towards strong inversion with $V_{GS1} = 10\,\mathrm{V}$, such that the width of M1 can be chosen lower as when placed closer to the threshold voltage. The drain-source voltage V_{DS1} of M1 was chosen later on as a result of the voltage biasing of the other three transistors in the core amplifier. However, since V_{DS1} is equal to the summed drain-source voltages of M2 and M4, it was assumed that V_{DS1} would, in any case, be high enough to put M1 into saturation.
The choice of V_{GS1} automatically sets the potential at the input node and the output node of the amplifier to 10 V considering that $V_{DD} = 20\,\mathrm{V}$ and both nodes are connected via a single feedback resistor. The drain-source voltage of M4 could then already be set to $V_{DS4} = 10\,\mathrm{V}$. The gate-source voltage of M4, on the other hand, is chosen lower than V_{DS4} with $V_{GS} = 9\,\mathrm{V}$, which places it more towards moderate inversion and further in saturation. This makes it less prone to process variations of the channel dimensions and DC operating point and, therefore, improves its performance as a constant current supply for the biasing of M2. An additional advantage is that values for r_{o4} naturally increase, which is beneficial for the gain of the circuit. However, this choice comes at the price of bandwidth, as the width of M4 needs to be higher for the same drain current as when set closer to strong inversion, increasing its gate-drain capacitance.
The gate-source voltage of M2 was therefore placed closer to strong inversion with $V_{GS2} = 7\,\mathrm{V}$ to compensate for this i.e. the width and corresponding gate-drain capacitance of M2 can be chosen to be smaller. But then setting its gate-drain voltage to $V_{DS2} = V_{GS2}$ in turn creates the problem that M3 cannot properly be placed in saturation, as only 3 V would be left for its drain-source voltage. This would pose a serious disadvantage as M3 is supposed to supply a constant current for the biasing of the core amplifier. Hence, it was chosen that $V_{DS2} = 4\,\mathrm{V}$ instead, which set the potential at the folded node and notably also the bulk-source voltage of M2 V_{BS2} to 6 V. The threshold voltage of M2 is hence

4.09 V, which brings its overdrive voltage to $V_{GS2} - V_{TH2} = 7\,\text{V} - 4.09\,\text{V} = 2.91\,\text{V}$. Albeit closer to the lower edge, this still allows to place M2 in saturation as $V_{DS2} > V_{GS2} - V_{TH2}$, while also leaving some voltage budget for M3.
As M3, similar to M4, is supposed to provide a constant current for biasing, it should be placed closer to moderate inversion. However, since its drain current is the sum of the drain currents of M1 and M2, it was assumed that its width would need to be quite high. To keep its width lower to save some space on the layout, it was placed closer to strong inversion with a gate-source voltage of $V_{GS3} = 6\,\text{V}$. With the potential at the folded node set to $V_{DS3} = 6\,\text{V}$ by the earlier choice of V_{DS2}, M3 can also be placed in saturation. Also, $V_{DS1} = 14\,\text{V}$ as a result.

8.2.2 MOSFET Sizing

Determine MOSFET Channel Lengths

The length of M1 was chosen to be as small as possible with $L_1 = 6\,\mu\text{m}$. This is due to its high expected drain current and, hence, to decrease its size as well as the parasitic capacitances it introduces to the input node. For M4, the length was chosen to be $L_4 = 7\,\mu\text{m}$, which increases its overall size at the same current and, therefore, gate-drain capacitance slightly, reducing the bandwidth but, at the same time, potentially reducing the effect of process variations, which is critical for M4 as mentioned above. The loss in bandwidth was compensated by keeping the length of M2 lower with $L_2 = 3\,\mu\text{m}$. While this reduces comparability to the physically measured devices, this risk was purposely taken to reduce the parasitic capacitance at the output of the core amplifier, where the dominant pole can be found. M3, on the other hand, is a constant current source. Its length was chosen to be $L_3 = 5\,\mu\text{m}$ to potentially reduce the effect of process variations of the MOSFET sizing. Since M3 expectedly carries a high drain current, this could potentially blow up the overall size of M3 taking up a lot of space on the layout, it was important to choose an acceptable value for its width in the next step.

Determine MOSFET Channel Widths Using LUTs

After the voltage biasing and the channel lengths of all MOSFETs were fixed, the LUTs were generated by finding a reference value of I_D and r_o for all four MOSFETs, g_m and C_{gd} for M1 and M2 and g_{mb} for M2. The reference values were found for devices, which all have a width of 100 µm. The small-signal parameters and parasitic capacitances are extracted as explained in sections 6.3.2 and 6.3.3. Using equations 7.1 and 7.2, the according LUTs for all the parameters of the four MOSFETs are generated.
From equation 5.8, it can be seen that the fraction, which is the equivalent transconductance G_m of the core amplifier circuit, is approximately equal to g_{m1} (transconductance of M1) if $(r_{o1} \parallel r_{o3}) \gg \left(\frac{1}{G_{m2}} \parallel r_{o2}\right)$. Therefore, by choosing G_m as high as possible, the output resistance R_{out} can be reduced relatively for the same gain ($A_v = G_m \cdot R_{out}$). This benefits the bandwidth, which is inversely dependent on R_{out}, while keeping the gain at the same value. Hence, another design requirement was set to $G_m \geq 0.9 g_{m1}$.
Before the generation of the LUTs, the circuit first needed to be analyzed concerning Kirchhoff's current law to generate sets of widths where the corresponding currents obey

said law. Since the sum of ingoing and outgoing currents at every node needs to be zero, it holds that

$$-I_{D2} + I_{D4} = 0 \Leftrightarrow I_{D2} = I_{D4} \tag{8.1}$$

at the output node and

$$I_{D1} + I_{D2} - I_{D3} = 0 \Leftrightarrow I_{D2} = I_{D3} + I_{D2}, \tag{8.2}$$

for the folded node, where I_{D1}, I_{D2}, I_{D3}, I_{D4} respectively are the drain currents of MOSFETs M1, M2, M3 and M4. Since the drain current is approximately linearly dependent on the channel width W of a MOSFET, one can define the drain current normalized to the channel width: $J_D = \frac{I_D}{W}$. These normalized currents can be determined for every MOSFET by dividing the already simulated drain currents of the reference devices and dividing them by the reference width 100 µm.
Plugging the normalized currents of every MOSFET into equations 8.1 and 8.2 the width W_4 of M4 can be calculated in terms of W_2 and the width W_3 of M3 in terms of W_1 and W_3:

$$J_{D2}W_2 = J_{D4}W_4 \Leftrightarrow W_4 = W_2\frac{J_{D2}}{J_{D4}} \tag{8.3}$$

$$J_{D2}W_2 = J_{D3}W_3 - J_{D1}W_1 \Leftrightarrow W_2 = \frac{J_{D3}W_3 + J_{D1}W_1}{J_{D2}}. \tag{8.4}$$

The drain currents of M3 were then swept over different values of its width W_3. For every W_3 a sweep of the width W_1 of M1 was done over values where $I_{D1} < I_{D3}$. For every combination of W_1 and W_3, the corresponding currents and widths for M2 and M4 are calculated using equation 8.2 and 8.1. For every subsequent set of widths, the other parameters needed for the calculation of the gain and bandwidth are scaled to the corresponding widths of the set. Subsequently, the gain using equation 5.8, the bandwidth using equation 5.9 and $\frac{G_m}{g_m}$ using equation 5.7 were calculated and compared to the design requirements. Tab. 8.1 shows the reference values for the width scaling of the necessary parameters of M1, M2, M3 and M4 determined at a width of 100 µm.

Table 8.1: Drain current I_D, transconductance g_m, MOSFET output resistance r_o and gate-drain capacitance C_{GD} of MOSFETs M1, M2, M3 and M4 used as reference values for the generation of LUTs at a reference width of 100 µm.

	I_D (µA)	g_m (µS)	r_o (MΩ)	C_{GD} (fF)
M1	15.15	13.06	5.54	-
M2	2.478	4.036	15.97	51.79
M3	12.05	-	9.57	-
M4	3.117	-	15.37	51.68

First, the set of widths was filtered by the highest possible gains and bandwidths. The highest possible gain amounted to roughly 49 dB, while the highest bandwidths were roughly 100 kHz. Using these values, the sets with values in the vicinity of both values

were found. By roughly fixing the bandwidth across all sets, it could be observed that sets with generally low widths for M2 and M4 i.e. low parasitic capacitance at the output of the amplifier, also have an output resistance to maintain the same bandwidth. While on the other hand, sets with generally high widths for M2 and M4 have high parasitic output capacitances but low output resistance. In the end, a set with intermediate values for the widths was chosen to compromise between output capacitance and output resistance. The chosen widths were

$$W_1 = 551\,\mu\text{m} \quad W_2 = 395\,\mu\text{m}$$
$$W_3 = 774\,\mu\text{m} \quad W_4 = 314\,\mu\text{m}.$$

A DC analysis using these widths of every MOSFET when still isolated, with ideal voltage biasing and using the MOSFET models yields the following drain currents $I_{D1} = 94.35\,\mu\text{A}$, $I_{D2} = 9.73\,\mu\text{A}$, $I_{D3} = 92.63\,\mu\text{A}$ and $I_{D4} = 11.69\,\mu\text{A}$.
It seems that using widths calculated from the LUTs and using them in a simulation with the MOSFET models, the drain currents do not satisfy Kirchoff's current law. This was expected since the models do not have the width scaling behavior of the drain currents that was assumed for the generation of the LUT. However, these widths could be used as a starting point to find an appropriate MOSFET sizing that works with the SPICE models by first manually adjusting them using DC analyses such that they satisfy Kirchhoff's current law. For that, the widths of the transistors were tweaked for the drain currents to be $I_{D1} = 94\,\mu\text{A}$, $I_{D2} = I_{D4} = 10\,\mu\text{A}$ and $I_{D3} = 104\,\mu\text{A}$. The resulting widths are

$$W_1 = 549\,\mu\text{m} \quad W_2 = 407\,\mu\text{m}$$
$$W_3 = 869\,\mu\text{m} \quad W_4 = 273\,\mu\text{m}.$$

8.2.3 SPICE Simulation of Core Amplifier Schematic

Until this step, the MOSFETs have been analyzed separately in simulation, assuming that the DC operating point for every MOSFET stays fixed at the chosen voltages. This is, of course, not the case if the MOSFETs are plugged into the actual schematic of the core amplifier. Therefore, they need to be inserted into a SPICE simulation using the full circuit schematic. Fig. 8.1 shows the schematic used to simulate the CSA by itself. At this stage, the biasing is still provided using only ideal voltage sources at the gates of M2, M3, and M4. To ensure that the potential at the input node (gate of M1) and output node is the same, an ideal resistor is placed between them in the simulation. The resistance is chosen high with $100\,\text{G}\Omega$, such that the output is not loaded and no current can flow between input and output.
DC and AC analyses then determine whether the MOSFET models and SPICE simulation reproduce the pre-calculated circuit performance or if adjustments still need to be applied to the DC operating point and MOSFET channel dimensions. To inject an AC signal for the AC analysis, a voltage source with AC magnitude of 1 V is AC-coupled to the input of the core amplifier. This way, the DC operating point is determined only by the supply voltage and the voltage sources at the gates of M2, M3 and M4. An AC

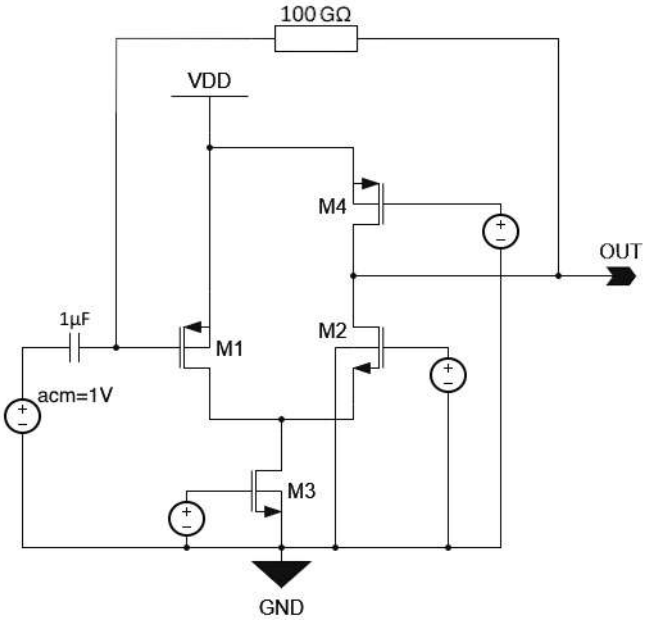


Figure 8.1: Schematic used for the simulation of the isolated core amplifier. Biasing is provided by voltage sources at the gates of M2, M3 and M4. M1 is biased by a high ohmic connection of its gate to the output node. For AC analysis, a voltage source with AC magnitude (`acm`) of 1 V is AC coupled to the input of the amplifier.

signal is then superimposed on this DC operating point through a high-valued capacitor of 1 µF for the AC analysis. This configuration resembles a differentiator, however it is only used to determine the open loop-gain of the core amplifier by comparing the voltage signal amplitude at the input and output of the amplifier in the AC simulation.
The channel width and lengths of M1 to M4 determined in the previous steps and the gate potentials of M2, M3 and M4 were plugged into the circuit simulation without biasing network, using the schematic shown in Fig. 8.1. The DC analysis yields a DC operating point very similar to the one pre-determined above with the isolated MOSFETs and with only very small deviations lower than a percent. The gain and bandwidth result in 49.77 dB and 98.42 kHz respectively, which is quite similar to the values obtained from the LUTs and the small-signal model calculations.
Next, the widths of M2 and M4 were readjusted to increase the gain and bandwidth. First the drain current through them was reduced to $I_{D2} = I_{D4} = 7\,\mu\text{A}$. This allowed to reduce W_2 and W_4, which decreased the parasitic capacitances at the output node and increased most notably r_{o4}. This, in turn, increased the output resistance, positively affecting the gain but counteracting the decrease in output capacitance such that the bandwidth did not change much. So, to further increase the bandwidth, the width of M2 was again individually reduced by increasing its gate-source voltage by 0.3 V to $V_{GS2} = 7.3\,\text{V}$. The new widths for M2 and M4 resulted in

$$W_2 = 203\,\mu\text{m} \quad W_4 = 197\,\mu\text{m}.$$

The reduction of the current through M2 and M4 also reduced the current through M3

to $I_{D3} = 101\,\mu A$ and therefore also reduced the width of M3 to

$$W_3 = 845\,\mu m.$$

Plugged into the simulation, the new gain and bandwidth result in 51.97 dB and 121.20 kHz respectively.

8.2.4 Adding the Biasing Network and Final Adjustments

Until this step, the biasing of MOSFETs is done via ideal voltage sources. However, as already explained in section 5.4, in practice, the biasing is done via a biasing network consisting of current mirrors using one externally supplied biasing current as a reference. Fig. 8.2 shows the core amplifier schematic with the biasing network.
The idea is to have an externally provided biasing current I_b flow through M5, making it the reference device of the first NMOS current mirror network with M3 and M10 as the mirroring devices with equal channel lengths. Since the current and width for M3 were already chosen in the first few design steps, the channel dimensions and current for M5 could be chosen to be the same as M3. Hence, a value for the biasing current could already be fixed:

$$I_b = 101\,\mu A.$$

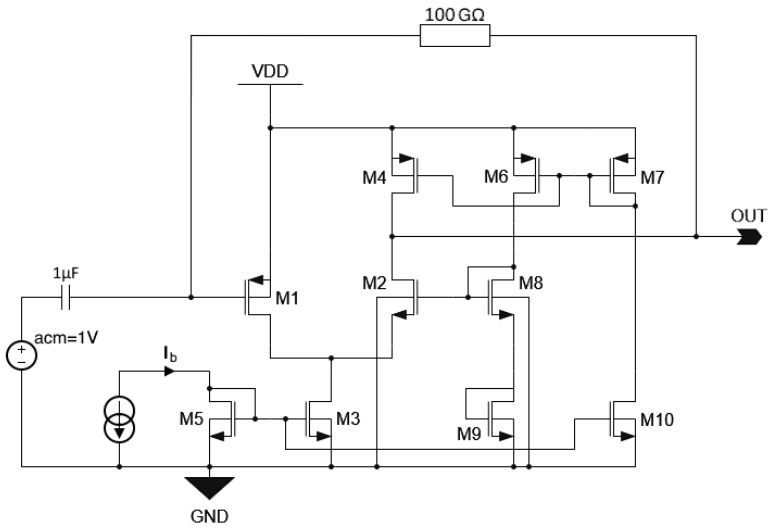


Figure 8.2: Schematic used for the simulation of the isolated core amplifier with biasing network. Biasing is provided by a biasing network supplied by externally provided biasing current I_b, which is represented as a DC current source in the schematic.

The purpose of M10 is to provide a reference current for the PMOS current mirror network, which has M7 as the reference device. The widths of M7 and M10 have to be scaled according to the currents needed by M4 and M6, which are the mirroring devices of M7 with equal channel lengths. Since the width and current of M4 are already known from previous design steps, the width of M7 needs to be scaled as a multiple of the width

of M4, which also carries the current provided by M10. The width and current of M10 can be chosen rather arbitrarily, with the requirement that the current provided to M7 does not inflate the width of M7 too much, considering that PMOS devices have worse current output than NMOS devices.
M6 is used to provide a reference current to M8 and M9, which set the operating point of the cascode transistor M2. This current mirror is used to apply the chosen gate-source voltage to M2. In order to do so, the dimensions of M6 are set to the same as M4, providing the same current to M8 as flows through M4 and M2. The dimensions of M8 are chosen to be the same as M2, such that they have approximately the same operating point. The diode-connected MOSFET M9 shifts the source voltage of M8 up, resulting in a higher gate voltage on M2. Since M8 and M9 have the same constant current flowing through them (as provided by M6), the width of M9 can directly scale the potential at the source of M8. A wider M9 means that it needs less gate-source voltage to conduct the same current and vice-versa. If the potential at the source of M8 changes, its gate-source and drain-source voltages have to change at the same rate, such that its drain current stays the same.
Tab. 8.2 summarizes the MOSFET sizing and DC operating point of the MOSFETs M1 to M10 in the core amplifier, as determined by a DC analysis.

Table 8.2: Width W, length L, drain-source voltage V_{DS}, gate-source voltage V_{GS} and drain current I_D of M1 to M10 in the isolated core amplifier with biasing network.

	W (µm)	L (µm)	V_{DS} (V)	V_{GS} (V)	I_D (µA)
M1	549	6	14.00	10.00	94.04
M2	203	3	4.00	7.29	6.96
M3	845	5	6.00	6.00	101.00
M4	197	7	10.00	9.00	6.96
M5	845	5	6.00	6.00	101.00
M6	197	7	6.71	9.00	6.35
M7	697	7	9.00	9.00	26.60
M8	203	3	7.22	7.22	6.35
M9	108	10	6.07	6.07	6.35
M10	218	5	11.00	6.00	26.20

An AC analysis yields the gain as a function of the signal frequency shown in Fig. 8.3. The final gain A_v and bandwidth f_{BW} of the core amplifier with its biasing network then result to:

$$A_v = 52.04\,\mathrm{dB}$$
$$f_{BW} = 90.60\,\mathrm{kHz}.$$

The addition of the biasing network reduced the bandwidth by roughly 30 kHz.

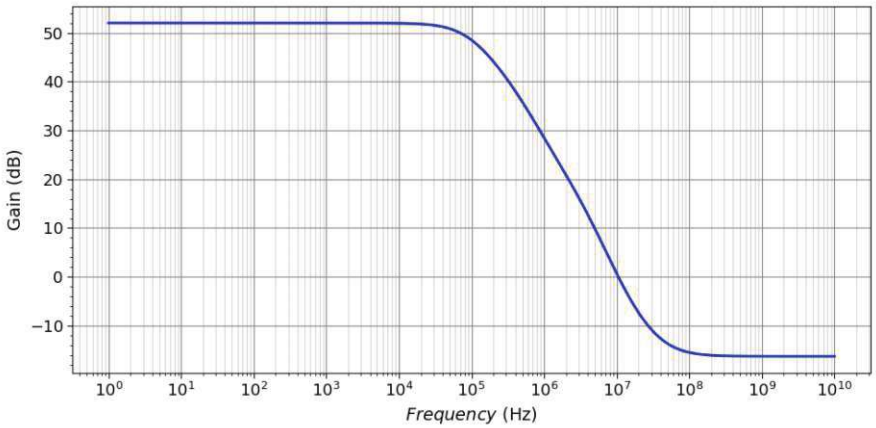


Figure 8.3: Gain as a function of frequency for the core amplifier.

8.3 Design of the Output-Driver Stage

The output driver stage is implemented as a unity-gain buffer amplifier with low output resistance, discussed in section 5.3. The design requirement for the buffer amplifier is a maximum output resistance of 1 kΩ. A requirement for the bandwidth is omitted as it is assumed that the buffer amplifier does not introduce a pole frequency lower than that of the dominant pole of the core amplifier. This is provided as the parasitic capacitances at every node will be in the same scale of 100 fF, since no more than three MOSFETs are connected to any node, and by choosing the drain current of any of the MOSFETs higher than the current of M4 and M2, which ensures lower $\frac{1}{g_m}$ and r_o.
The design steps for the buffer amplifier are essentially the same as for the core amplifier. However, since their execution differs from the core amplifier because of different design requirements, the design steps for the buffer amplifier are discussed again in this section.

8.3.1 Voltage Biasing

As with the core amplifier, the first step in the design of the buffer amplifier is the determination of the voltage biasing. Since the input of the buffer amplifier is connected with the output node of the core amplifier, the potential at the gate of M13 is restricted to 10 V. The potential at the buffer output is then also restricted to the same potential, as the buffer amplifier must have a symmetric DC operating point at the input and output nodes to function properly (see section 5.3).
M15 and M17 act as constant current supplies to bias the first and second amplifier stages, respectively, similar to M3 in the core amplifier. Hence, they can later be added to the NMOS current mirror network with M5 as the reference device. Their gate-source voltages therefore are $V_{GS15} = V_{GS17} = 6\,\text{V}$. The drain-source voltage of M15 is chosen equal to that of M3 with $V_{DS15} = 6\,\text{V}$, which also sets the potentials at the folded node to 6 V. The drain-source voltage of M17, on the other hand, is automatically given by the potential of the buffer output node, i.e. $V_{DS17} = 10\,\text{V}$.
The choice of the potential at the folded node with 6 V set the gate-source voltage of M13 and M14 to $V_{GS13} = V_{GS14} = 4\,\text{V}$. This is slightly below the requirement for the

gate-source voltage of an NMOS with $V_{BS} = 0\,\text{V}$ that $V_{GS} \leq 4.23\,\text{V}$. This sets both M13 and M14 deeper into moderate inversion, implying that both transistors get quite large, increasing their inter-terminal parasitic capacitances. Since they add to the capacitance at the output node of the core amplifier, where the dominant pole is located, a significant reduction in the bandwidth of the circuit could be expected. This could be solved by reducing V_{DS15}, but instead, it was kept the same to have M15, as a constant current supply, in saturation. This means prioritizing an overall functioning circuit over the bandwidth, as doing so should make the circuit less prone to variations of the DC operating point.
To have M13 and M14 in saturation, their drain-source voltage was chosen to be equal to their gate-source voltage, i.e., $V_{DS13} = V_{DS14} = 4\,\text{V}$. This sets the potential at the drains and gates of M11 and M12 and the gate of M16 automatically to 10 V. Therefore their gate-source and drain-source voltages, considering again that $V_{DD} = 20\,\text{V}$, are given by $V_{GS11} = V_{DS11} = V_{GS12} = V_{DS12} = V_{GS16} = V_{DS16} = 10\,\text{V}$.

8.3.2 MOSFET Sizing

Determine MOSFET Channel Lengths

For the NMOS transistors, the lengths were chosen to be $L_{13} = L_{14} = L_{15} = L_{17} = 5\,\mu\text{m}$, to make the subsequent simulations using the MOSFET models better fitting to the measurement data of the physical devices. For the lengths of M11 and M12, a similar reasoning was chosen as for M4, i.e. to make them less prone to process variations of the channel width and the DC operation point, their lengths were chosen slightly higher than the minimal possible length: $L_{11} = L_{12} = 7\,\mu\text{m}$. The length of M16 was chosen to be $L_{16} = 6\,\mu\text{m}$.
To reduce the output resistance R_{out} of the buffer amplifier, a high drain current is needed since $R_{\text{out}} \propto \frac{1}{g_{m16}}$. A high drain current, in turn, means a high width, and to prevent the size of M16 from inflating too much and save some space on the layout, the length was chosen as small as possible. M17 is the constant current source that supplies this high drain current through M16. Since it is an NMOS current source, it will later be added to the current mirror network with M5 as the reference device. Analogously, the length is determined to be $L_{17} = 5\,\mu\text{m}$.
MC2 acts as a capacitor using the capacitance between the gate and a short source, drain and bulk. The length was chosen to be as high as possible with $L_{C2} = 10\,\mu\text{m}$ to increase the active area of the capacitance between the gate and the bulk. This way, the capacitance could be maximized with the choice of width.

Determine MOSFET Channel Widths Using LUTs

Reference values will again be taken for MOSFETs with a channel width of 100 µm. The basic design requirement is a maximal output resistance of 1 kΩ, which can be calculated using equation 5.14. Hence, the following reference parameters are extracted: g_{m14}, g_{m16}, r_{o12} and r_{o14}.
The channel widths of M11 and M12 and, respectively, M13 and M14 are the same to

ensure the symmetric DC operating point. Because of Kirchhoff's current law, they, therefore, all need to have the same drain current ideally:

$$I_{D11} = I_{D12} = I_{D13} = I_{D14}. \tag{8.5}$$

Using Kirchhoff's current law, the drain current of M15 can be written as the sum of the drain currents of M12 and M13:

$$I_{D15} = I_{D13} + I_{D14}. \tag{8.6}$$

By plugging the formula for normalized currents $J_D = WI_D$ into 8.6, the width of M15 is then given by:

$$W_{15} = \frac{J_{D13}W_{13} + J_{D14}W_{14}}{J_{D15}}. \tag{8.7}$$

Assuming no load at the output, M16 and M17 should have the same drain current, which is why the width of M16 can be expressed over the channel width of M17 and the normalized currents:

$$J_{D16}W_{16} = J_{D17}W_{17} \Leftrightarrow W_{16} = W_{17}\frac{J_{D17}}{J_{D16}}. \tag{8.8}$$

By doing a sweep of the channel widths of M12, M13 and M17, the widths of M11 and M14 are automatically known since $W_{11} = W_{12}$ and $W_{13} = W_{14}$, while W_{15} and W_{16} can be calculated using equations 8.7 and 8.8. This way, the drain currents of the MOSFETs should automatically satisfy Kirchoff's current law. For every possible set of channel widths, the parameters needed to calculate the output resistance are determined by scaling the reference values with the corresponding widths. The resulting output resistances can be compared to the requirement of$R_{\text{out}} < 1\,\text{k}\Omega$ to find appropriate sets of channel widths that can be used for the MOSFETs in the buffer amplifier. Tab. 8.3 shows the reference values for the width scaling of the necessary parameters at the reference width of 100 µm.

Table 8.3: Drain current I_D, transconductance g_m and MOSFET output resistance r_o of MOSFETs M12, M14, M15, M16 and M17 used as reference values for the generation of LUTs at a reference width of 100 µm for the buffer amplifier.

	I_D (µA)	g_m (µS)	r_o (MΩ)
M12	10.46	-	4.719
M14	2.248	2.317	21.55
M15	12.05	-	-
M16	14.25	12.42	-
M17	12.249	-	-

The generated sets of widths were filtered by the requirement of $R_{\text{out}} < 1\,\text{k}\Omega$. However, a second filter condition was introduced, namely that the drain currents of M15 and M17 cannot be higher than the biasing current $I_b = 101\,\mu\text{m}$. This requirement was introduced to limit the size of the circuit. Also, since the gate-source voltage of M13 is quite low

with 4 V, a very high current through the first stage of the buffer amplifier (measured by the drain current of M15) would inflate the width and parasitic capacitance of M13. This would cause a significant drop in bandwidth.
Using these restrictions, the lowest possible output resistance that was found was $R_{\text{out}} = 451.29\,\Omega$. For all these sets, M17 carries a drain current of $I_{D17} = 101\,\mu\text{m}$, i.e., the maximal value allowed by the just mentioned restriction on the drain current. This was expected as $R_{\text{out}} \propto \frac{1}{g_{m16}}$.
A set was chosen with calculated drain currents of $I_{D15} = 20\,\mu\text{m}$ as a compromise between the minimization of the size of the first stage of the buffer amplifier and the current through M11 and M13 being slightly higher than that through M2 and M4 such that the output resistance of the first buffer stage does not introduce a new dominating pole to the circuit. The drain current of M17 was chosen to be $I_{D17} = 101\,\mu\text{m}$ to minimize the output resistance of the buffer amplifier. The corresponding widths are

$$\begin{aligned} W_{11} = W_{12} = 94\,\mu\text{m} \quad & W_{13} = W_{14} = 437\,\mu\text{m} \\ W_{15} = 163\,\mu\text{m} \quad & W_{16} = 708\,\mu\text{m} \\ W_{17} = & 824\,\mu\text{m}. \end{aligned}$$

The channel lengths and widths chosen above were entered into a simulation using the MOSFET models. A DC analysis of every MOSFET, when isolated and biased around the above-chosen voltage biasing, was carried out to compare the calculated drain currents of the MOSFETs with the ones produced via the MOSFET models. The DC analysis yielded the following currents: $I_{D11} = I_{D12} = 9.74\,\mu\text{A}$, $I_{D13} = I_{D14} = 9.78\,\mu\text{A}$, $I_{D15} = 19.58\,\mu\text{A}$, $I_{D16} = 114.71\,\mu\text{A}$ and $I_{D17} = 98.60\,\mu\text{A}$. The pre-calculated widths for the first stage of the buffer amplifier seemed to compare quite well to the MOSFET models. They only needed slight tweaking using DC analysis to bring the currents closer to $I_{D11} = I_{D12} = I_{D13} = I_{D14} = 10\,\mu\text{A}$ and $I_{D15} = 20\,\mu\text{A}$. For the second stage, the pre-calculated and simulated drain currents do not compare as well, showing that scaling the currents linearly with the channel width deviates more for higher currents. The widths for M16 and M17 were both adjusted such that their currents approximately amounted $I_{D16} = I_{D17} = 101\,\mu\text{A}$. The new widths obtained from simulation and satisfying Kirchoff's current law amounted to:

$$\begin{aligned} W_{11} = W_{12} = 97\,\mu\text{m} \quad & W_{13} = W_{14} = 447\,\mu\text{m} \\ W_{15} = 167\,\mu\text{m} \quad & W_{16} = 626\,\mu\text{m} \\ W_{17} = & 844\,\mu\text{m}. \end{aligned}$$

Note that, while M3 and M17 are supposed to carry the same drain current, W_{17} is slightly higher because $V_{DS17} > V_{DS3}$.
The width of MC2 was chosen independently from the LUT method. Instead, it was manually chosen according to the necessary requirements in the steps later on to stabilize the output signal response of the CSA. The width was, however, limited by the available space left on the layout of the CSA circuit. The value chosen in the end was 1500 µm. Although this value was originally determined later on in the design process, for the sake of simplicity, it will already be included in the explanation of the design buffer amplifier.

8.3.3 SPICE Simulation of Buffer Amplifier Schematic

Similar to the core amplifier, the widths needed to be plugged into a SPICE simulation of the circuit to check whether the design requirements were satisfied also using the MOSFET models and the full circuit schematic or adjustments needed to be applied. Fig. 8.4 shows the schematic used to simulate the buffer amplifier by itself. A load resistance of $R_L = 1\,\mathrm{M\Omega}$ was used. For the load capacitance, values of $C_L = 4\,\mathrm{pF}$ and $C_L = 20\,\mathrm{pF}$, which correspond to the lower and upper design requirements were tested. In this step, the biasing is still done using ideal voltage sources at the gates of M13 (10 V), M15, and M17 (6 V).

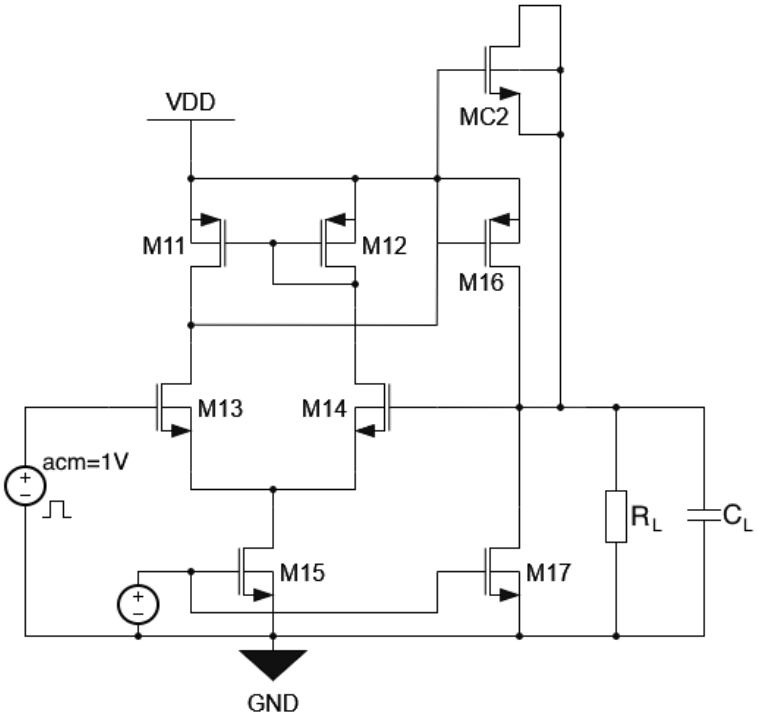


Figure 8.4: Schematic used for the simulation of the isolated buffer amplifier. Voltage biasing is provided by voltage sources at the gates of M13, M15 and M17. The voltage source at the gate of M13 has an AC magnitude (`acm`) of 1 V for AC analysis, and it also generates voltage pulses for transient analysis.

DC analysis is again used to check if the DC operating point satisfies the pre-calculated one. AC analysis is used via the voltage source at the gate of M13, which has an AC magnitude of 1 V to analyze the frequency response of the buffer amplifier and whether it provides a gain of 1. Finally, a transient analysis is also applied by checking the step response of the buffer amplifier. This helps to determine if the frequency compensation introduced by MC2 is enough to get rid of any ringing.
A DC analysis without R_L yielded the same DC operating point calculated above, up to negligible deviations under 1 %. The output resistance was determined by comparing the voltage drop at the output node of the buffer when the load resistor of $R_L = 1\,\mathrm{M\Omega}$ is applied. It amounted to $R_{\mathrm{out}} = 458.59\,\Omega$, which is slightly higher than the output

resistance calculated using the LUTs. With a voltage drop of 4.59 mV, the drain current of M16 increased to 112.6 µA when the load resistance was applied.
Using an AC analysis, the gain as a function of signal frequency was simulated using $C_L = 4\,\text{pF}$ and $C_L = 20\,\text{pF}$ for the load capacitance, which is shown in Fig. 8.5. For $C_L = 4\,\text{pF}$ the bandwidth of the circuit amounted to $f_{\text{BW}} = 591.68\,\text{kHz}$. For $C_L = 20\,\text{pF}$ it is $f_{\text{BW}} = 484.31\,\text{kHz}$. The bandwidth is, in both cases, almost 300 kHz removed from the core amplifier's bandwidth. The low-frequency gain amounted to $A_v = -0.03\,\text{dB}$ for both load capacitances, which corresponds to a small reduction of an input signal of 0.4 %.

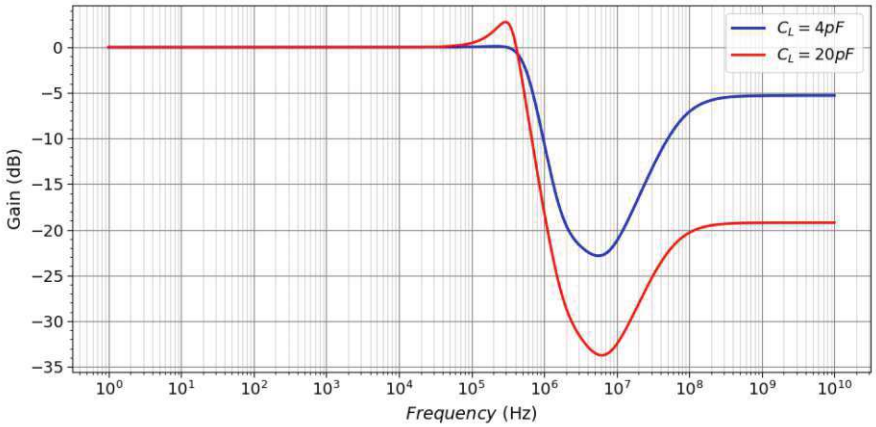


Figure 8.5: Gain as a function of frequency for the isolated unity-gain buffer stage.

For $C_L = 20\,\text{pF}$ shows gain peaking up to 2.73 dB at a frequency of 293.50 kHz, which hints at oscillatory behaviour. To check the extent of the oscillatory behavior at the higher load capacitance, a transient analysis was carried out using a 1 mV high step signal with a rise time of 1 ps as an input signal. Fig. 8.6 shows this input signal and the resulting output voltage signal.
It seems that the buffer suffers from ringing. The output signal shows a damped oscillation with $C_L = 20\,\text{pF}$ and the same resonant frequency at which the gain peaking for the red curve in Fig. 8.5 occurs. Note that the introduction of MC2 initially reduced the ringing by limiting the bandwidth of the buffer amplifier to cut off before the frequency, where the gain peaking occurs. However, the limited width did not allow for enough frequency compensation and could not completely remove the ringing.

8.3.4 Adding the Biasing Network and Final Adjustments

The biasing of the buffer amplifier was rather straightforward, as it could be integrated into the biasing network of the core amplifier. Since M15 and M17 are also NMOS transistors, they could simply be added to the current mirror network of the core amplifier. The ideal voltage sources at their gates were, therefore, simply replaced by a connection to the gate of M5. To also properly bias the input node of the buffer amplifier, the output node of the core amplifier can be connected to the gate of M13. Fig. 8.7 shows a schematic with a combined core and buffer amplifier with biasing network. Using this

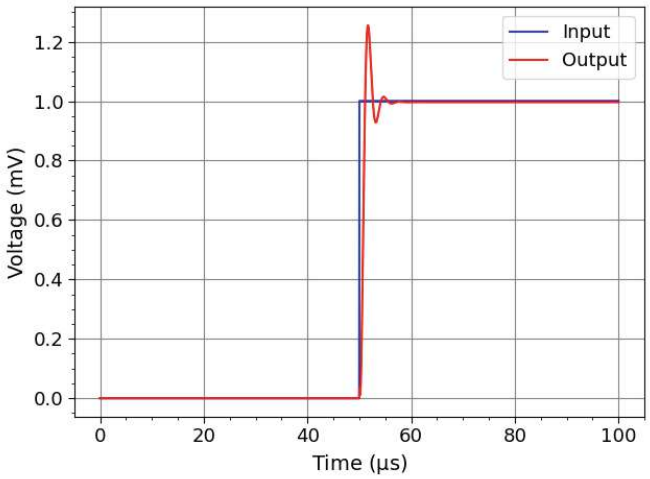


Figure 8.6: Step response of the unity-gain buffer stage with a load capacitance of $C_L = 20\,\mathrm{pF}$ - step input signal of $1\,\mathrm{mV}$ in blue and response output signal in red.

schematic, the combination of the core amplifier and buffer amplifier was again tested for the DC operating point with a DC analysis, the frequency response with an AC analysis and for ringing via the step response using a transient analysis.
A DC analysis without the load resistance showed that the DC operating point of both of the core and buffer amplifier did not change after their combination, up to another negligible deviation below 1 %. Adding a load resistance of $1\,\mathrm{M\Omega}$ decreased the potential at the output of the buffer amplifier by $4.58\,\mathrm{mV}$. This did not significantly change the performance of the buffer amplifier. The output resistance decreased slightly to

$$R_{\mathrm{out}} = 458.48\,\Omega.$$

The output resistance, therefore, satisfies the design requirements. Tab. 8.4 summarizes the MOSFET dimensions and the DC operating point of M11 to M17 and MC2, when the load resistance is applied.
An AC analysis was carried out again with the load capacitances of $C_L = 4\,\mathrm{pF}$ and $C_L = 20\,\mathrm{pF}$ for the combination of a combination of the core amplifier and buffer amplifier. The gain versus frequency response of the buffer amplifier itself stayed essentially unchanged i.e. the gain peaking at roughly $484\,\mathrm{kHz}$ persisted with $C_L = 20\,\mathrm{pF}$. The voltage signal amplitudes were compared at the input and output of the buffer amplifier to see that.
Fig. 8.8 shows the gain as a function of frequency for both load capacitances if the gain from the input of the core amplifier to the output of the buffer amplifier is analyzed. For $C_L = 20\,\mathrm{pF}$, a well-pronounced and rather sharp gain roll-off could be located at roughly the same frequency, at which the gain peak maximum of the buffer amplifier appeared. The gain peaking, however, seemed to be less pronounced, hinting at the possibility of reduced ringing. For $C_L = 4\,\mathrm{pF}$, no signs of gain peaking were visible.
The gain resulted in

$$A_v = 51.80\,\mathrm{dB}$$

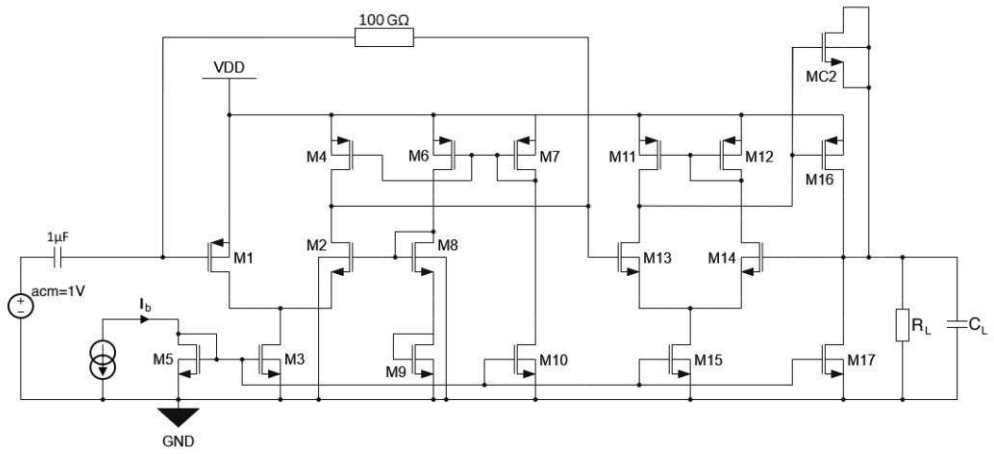


Figure 8.7: Schematic used for the simulation of the buffer amplifier with biasing network i.e. a combined core and buffer amplifier without the feedback network. Biasing is provided by a biasing network supplied by externally provided biasing current I_B, which is represented as a DC current source in the schematic. For AC analysis, a voltage source with AC magnitude (`acm`) of 1 V is AC coupled to the input of the core amplifier.

for both load capacitances. The bandwidth differed only slightly for both load capacitances. They amount to

$$f_{BW,4\,\mathrm{pF}} = 27.38\,\mathrm{kHz},$$
$$f_{BW,20\,\mathrm{pF}} = 27.83\,\mathrm{kHz}.$$

The combination of the core amplifier with the buffer amplifier shows a smaller gain when compared to the isolated core amplifier. This was expected as the buffer amplifier has a gain slightly below 1. A significant reduction in bandwidth could also be observed, which could be attributed to the parasitic capacitances that the buffer amplifier introduced to the output node of the core amplifier, where the dominant pole of the circuit is located. The most notable contribution to the dominant pole frequency is probably the Miller-capacitance of the gate-source capacitance of M13.
Lastly, a transient analysis again with a step input signal of 1 mV and a rise time of 1 ps was carried out to check if the circuit still suffered from ringing due to the buffer amplifier. The step signal was applied to the input of the core amplifier. Fig. 8.9 shows the output signal response, where the ringing is indeed removed.

8.4 Design of the Closed-Loop Structure

In order to complete the full circuit of the CSA, the feedback network presented in section 5.5 needed to be added. The design requirements for the feedback network were mostly to provide that the CSA has an SNR of at least 10 with a reasonably stable and not overdamped output signal and a phase margin in the range of 65°-90°. For optimizing the SNR, the circuit was simulated using the thin sensor configuration with a signal charge of $Q_{in} = 660\,\mathrm{e}^-$ and an input capacitance of $C_{in} = 1\,\mathrm{pF}$. This case

Table 8.4: Width W, length L, drain-source voltage V_{DS}, gate-source voltage V_{GS} and drain current I_D of M11 to M17 in the output buffer amplifier when combined with the core amplifier - a load resistance of $R_L = 1\,\mathrm{M\Omega}$ is applied at the output of the buffer amplifier.

	W (µm)	L (µm)	V_{DS} (V)	V_{GS} (V)	I_D (µA)
M11	97	7	10.13	9.99	10.02
M12	97	7	9.99	9.99	10.02
M13	447	5	3.88	4.01	10.02
M14	447	5	4.02	4.00	10.02
M15	167	5	5.99	6.00	20.04
M16	626	6	10.01	10.13	112.53
M17	844	5	9.99	6.00	102.54
MC2	1500	10	0.00	-0.13	0.00

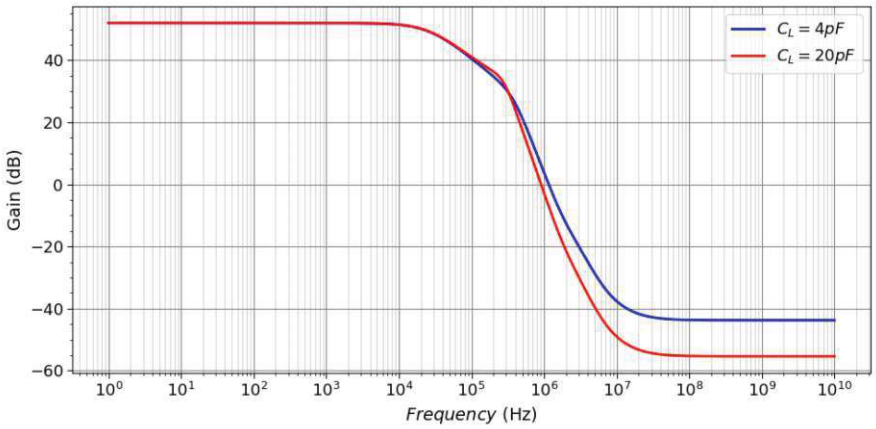


Figure 8.8: Gain as a function of frequency for the combination of the core amplifier and buffer amplifier.

was considered more critical than the thick sensor configuration due to the lower signal charge. For testing the stability of the circuit, a load capacitance of $C_L = 20\,\mathrm{pF}$ was considered a more critical case due to the gain peaking the buffer amplifier originally showed using this load.

The drain current of M18 and M19 flows into the output of the CSA. It needs to be low enough not to load the output of the buffer amplifier significantly. Considering the output resistance of roughly $R_{out} = 469\,\Omega$ of the buffer amplifier determined above, a current below $\leq 3\,\mu\mathrm{A}$ was determined to cause an acceptable voltage drop in the range of 1 mV and below.

Unlike the core amplifier and the buffer amplifier, the feedback network was not analyzed via the generation of LUTs. Instead, it was directly incorporated into the full closed-loop circuit schematic as shown in Fig. 8.10. M18 was biased via the PMOS current mirror network by connecting its gate with the gate and drain of M7. The potential applied to the source of M19 was fixed by connecting it to the output of the buffer

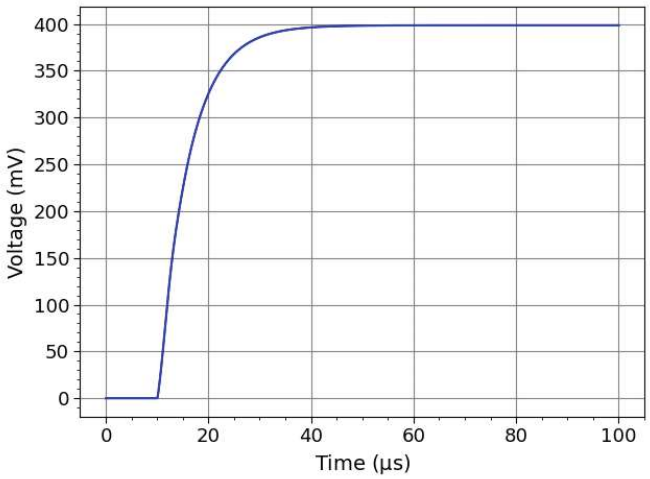


Figure 8.9: Voltage signal response of the combined core amplifier and unity-gain buffer to a step input signal of 1 mV.

amplifier, provided that the drain current of M19 does not load the output of the buffer amplifier. However, the biasing of M19 and also of MR still needed to be determined by the potential applied to the gate of both transistors. This voltage depends on the sizing of M19 and the drain current of M18, which could also be controlled via its sizing. Hence, a pre-definition of the voltage biasing, like for the core and buffer amplifier, was omitted and directly determined via the sizing of the MOSFETs. The MOSFET sizing was, in turn, determined by a pre-determination of the channel lengths and a subsequent parametric sweep of the channel widths. The desired set of widths was chosen according to the requirements of the feedback network.
Note that in Fig. 8.10 the equivalent circuit of the sensor consisting of a parallel input capacitance and current pulse source is included as well. To simulate the thin sensor configuration, the input capacitance of $C_{in} = 1\,\mathrm{pF}$ and a rectangular current pulse with a height of 1.05 µA, a width of 0.1 ns and a rise and fall time of 1 ps are used. The equivalent signal charge of the current pulse is roughly $Q_{in} = 662\,\mathrm{e}^-$.

8.4.1 MOSFET Sizing

Determine MOSFET Channel Lengths

The length of MR was chosen as long as possible with 10 µm to potentially maximize its effective resistance. The length of MC was chosen to be 10 µm to maximize the accuracy of the MOSFET models to the physical devices. The capacitance of MC1 is otherwise controlled by the width and not the length because the bulk terminal is grounded. The length of M19 was chosen to be $L_{19} = 5\,\mu\mathrm{m}$ to again increase the comparability of the MOSFET models to the physical MOSFETs.
The length of M18 should be determined by its inclusion in the PMOS current mirror network with M7 as the reference device, where $L_7 = 7\,\mu\mathrm{m}$. However, a length of $L_{18} = 6\,\mu\mathrm{m}$ was chosen instead to decrease its width for the same current and hence

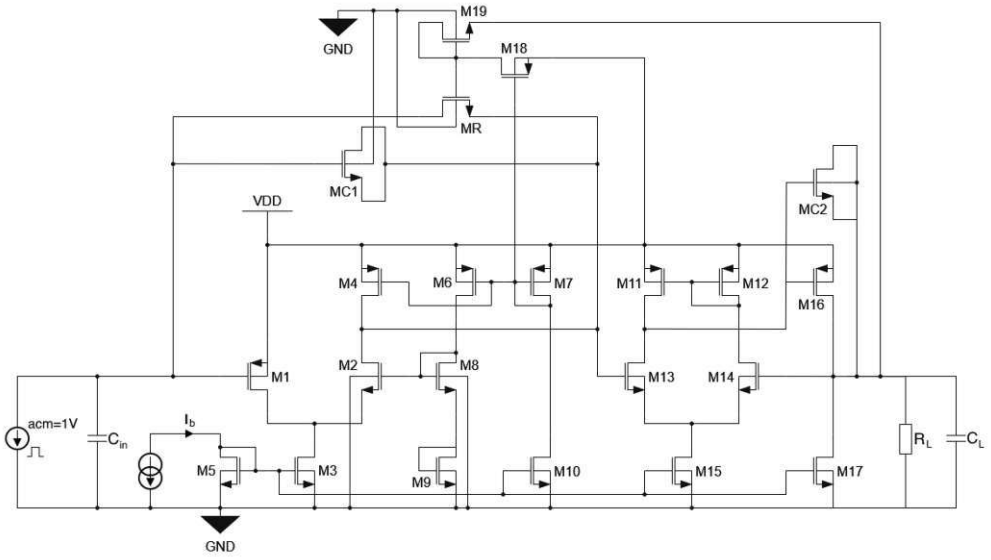


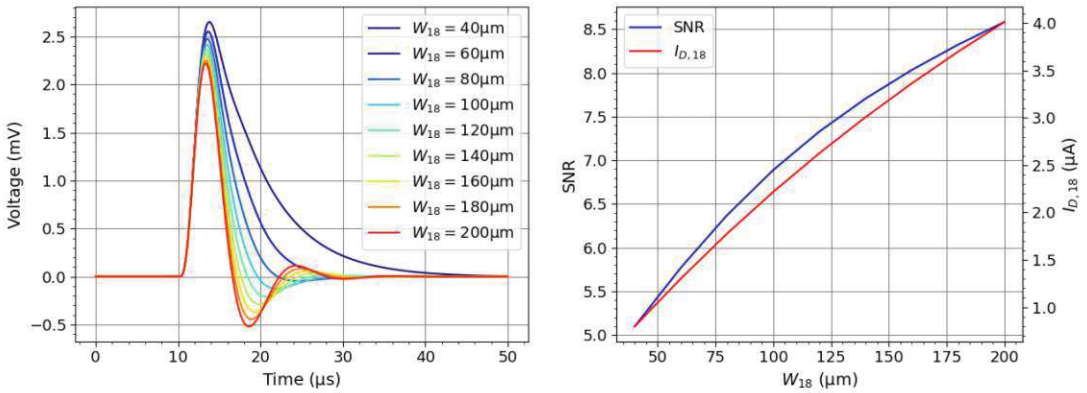
Figure 8.10: Schematic used for the simulation of the full CSA circuit consisting of core amplifier, buffer amplifier and feedback network. A parallel combination of a current source (AC magnitude of 1 V for AC analysis, current pulse for transient analysis) and the input capacitance C_{in} at the core amplifier input forms the equivalent circuit of a particle sensor. The load impedance consisting of the load resistance R_L and load capacitance C_L is added to the buffer amplifier output.

parasitic capacitances in the feedback network. This was at the cost of the drain current provided by M18 being more sensible on changes in its drain-source voltage and channel dimensions. This could prove to be disadvantageous for a MOSFET that is supposed to supply a constant current, but feedback stability was nonetheless of higher priority.

Determine MOSFET Channel Widths

The width of MR was chosen right away with the smallest possible width, i.e. $W = 3\,\mu\text{m}$, to increase its effective resistance. In order to find the widths of MC, M18, and M19, a sweep of different values was carried out, and at each combination of the sweep, the design requirements were considered as the figure of merit for an appropriate set of widths. One simulation per width combination entailed a DC analysis, an AC analysis, a stability analysis, a transient analysis and a noise analysis. The DC analysis was used to check the drain current of M18 and M19 as well as the gate-source voltage of MR. The AC analysis was used to extract the bandwidth of the CSA. For the stability analysis, the feedback loop was broken at the core amplifier output node to simulate the loop gain and phase as well as the phase margin of the circuit, including both feedback paths. The transient analysis was carried out over a time frame of 45 µs, with the input signal delayed by 10 µs to give the circuit time to settle into the DC operating point. The result of the transient simulation was a simulation of the output signal, which was used to extract the voltage amplitude and to check whether the output pulse was oscillating or over-damped. The noise analysis was used to extract the total root-mean-square noise

voltage of the CSA up to its bandwidth. By dividing the output signal amplitude by the total root-mean-square noise voltage, the SNR of the circuit was determined.
A sweep over a wide range of widths was carried out to narrow down the appropriate current range, where the output signal did not show any oscillations as determined by inspection via a transient analysis. Over this sweep, the highest possible SNRs could be found to be > 10. However, these cases could all be associated with under-damped, oscillating output signals. The SNR decreased the more the gate-source voltage of MR gradually decreased, which means an increase of the effective resistance of MR and the output signals gradually getting more damped. Along with a decreased SNR, over-damped signals also exhibit relatively long settling times, which reduce the rate at which the CSA can process incoming particle signals without any signals piling up.
Fig. 8.11 to 8.13 show the evolution of the output signal, SNR and drain current of M18 with a sweep over the widths of M18, M19 and MC1, respectively. Note that the range of widths, as well as the step size shown in these plots, are reduced compared to the ones that were used to find the optimal set of widths for the sake of visibility. Fig. 8.11a shows several output signals based on a signal charge of $662\,\mathrm{e}^-$ over a sweep of the width W_{18} of M18 in the range of $40\,\mu\mathrm{m}$ to $200\,\mu\mathrm{m}$. An increase in W_{18} seemed to cause the signals to get under-damped and oscillate, which also increased the SNR as shown in Fig. 8.11b. This directly correlated with the drain current of M18, i.e. the higher the width of M18, the higher its drain current and the drain current of M19. As a consequence, the drain-source voltage of M19, the potential at its drain and, therefore, the gate-source voltage of MR needed to increase. Hence, MR gradually opened up, reducing its effective resistance and reducing the damping of the feedback.

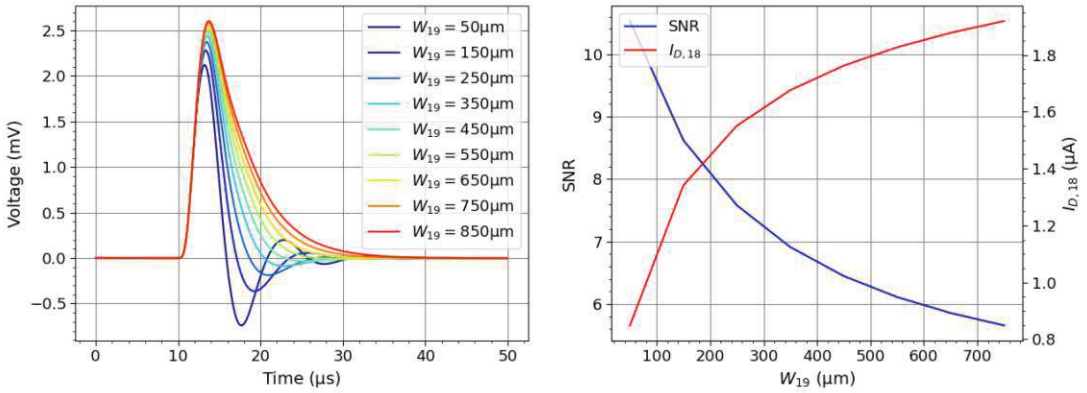


(a) (b)

Figure 8.11: (a) Simulated output signals and (b) SNR as well as drain current $I_{D,18}$ of M18 over a sweep of the width W_{18} of M18 – $W_{19} = 470\,\mu\mathrm{m}$, $W_{C1} = 13\,\mu\mathrm{m}$ were used as the widths for M19 and MC1.

Fig. 8.12a shows several output signals over a sweep of the width W_{19} of M19 in the range of $50\,\mu\mathrm{m}$ to $850\,\mu\mathrm{m}$. An increase in W_{18} seemed to cause the opposite effect as an increase of the width of M18, i.e. the signal starts to oscillate less and get more strongly damped. Increasingly damped signals also caused the SNR to reduce. This also seemed to be indirectly correlated to the drain current of M18 and, consequently, also of M19.

If the width of M19 increased, its drain-source voltage could decrease to carry the same amount of drain current. As a consequence, the potential at the drain and gate of M19 and the gate-source voltage of MR decreased. MR, therefore, gradually closed up and its effective resistance increased, damping the feedback.



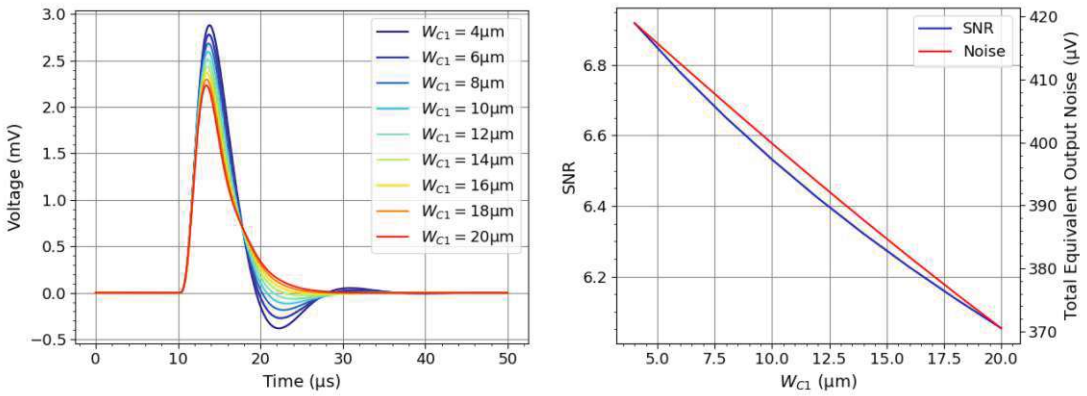
(a) (b)

Figure 8.12: (a) Simulated output signals and (b) SNR as well as drain current $I_{D,18}$ of M18 over a sweep of the width W_{19} of M19 – $W_{18} = 80\,\mu m$, $W_{C1} = 13\,\mu m$ were used as the widths for M18 and MC1.

Fig. 8.13a shows several output signals over a sweep of the width W_{C1} of MC1 in the range of 4 µm to 20 µm. An increase of W_{C1} increased the capacitance of MC1, which caused output signals to gradually oscillate less. This also reduced the SNR, even though the total equivalent output noise decreased as well, as shown in Fig. 8.13b. This was due to the decreasing amplitude of the voltage signal. Hence, any decrease of the SNR due to over-damping the feedback could be traced to a simultaneous decrease of the output signal amplitude at a higher rate as the noise decreased.
Appropriate signals were identified to be the ones with only a small undershoot after the signal settles to the baseline voltage for the first time. It was assumed that under changes in the DC operating point due to process variations, variations in the supply voltage, or variations in the biasing current, the feedback could potentially be pushed into the more over-damped or under-damped direction. Signals with only a small undershoot, therefore, seemed to be a good compromise, in which the feedback could be pushed in either direction without reaching a critical state of oscillation or too long settling time. The set of widths that was chosen is

$$W_{18} = 80\,\mu m \quad W_{19} = 470\,\mu m$$
$$W_{C1} = 13\,\mu m.$$

In Fig. 8.11 to 8.13, the subtle appearance of a bump in the decaying signal could be observed for signals with a longer decay time. These could be attributed to the fact that the feedback network introduced a second feedback path due to the connection of the source of M19 to the output of the buffer amplifier. The feedback path is presumably



(a) (b)

Figure 8.13: (a) Simulated output signals and (b) SNR as well as total equivalent output noise of the CSA over a sweep of the width W_{C1} of MC1 – $W_{18} = 80\,\mu m$, $W_{19} = 470\,\mu m$ were used as the widths for M18 and M19.

established via the parasitic capacitances of M19 and MR. The bump is created as soon as the output pulse is superimposed to the input pulse over this feedback path.
Tab. 8.5 summarizes the MOSFET dimensions and resulting DC operating point of M18, M19, MR and MC1. It needs to be noted, considering the threshold voltage of M18, that M18 could not be properly placed in saturation with $V_{DS18} = 1.78\,V$. While this makes the drain current of M18 more sensible to changes in its DC operating point, this issue was inevitable due to the grounded bulk terminal increasing its threshold voltage and, therefore, gate-source voltage to reach an appropriate effective resistance.

Table 8.5: Width W, length L, drain-source voltage V_{DS}, gate-source voltage V_{GS} and drain current I_D of M18, M19, MR and MC1 in the feedback network of the CSA.

	W (µm)	L (µm)	V_{DS} (V)	V_{GS} (V)	I_D (µA)
M18	109	6	1.78	9.00	1.77
M19	470	5	8.22	8.22	1.77
MR	3	10	0.00	8.22	0.00
MC1	13	10	0.00	0.00	0.00

Using the chosen set of widths for the feedback network, the resulting output signal, as well as loop gain and phase, can be seen in Fig. 8.14 and 8.15. The output signal in Fig. 8.14 does show a small undershoot but is otherwise free from ringing.
The SNR could be determined to be 6.37 with a total noise of 390.69 µV at a bandwidth of 28.54 kHz. The voltage gain of the voltage amplifier from the input of the core amplifier to the output of the buffer amplifier resulted in 51.54 dB. The loop gain shown in Fig. 8.15 seems to have one pole around the feedback path, which results in a phase margin of 90°.

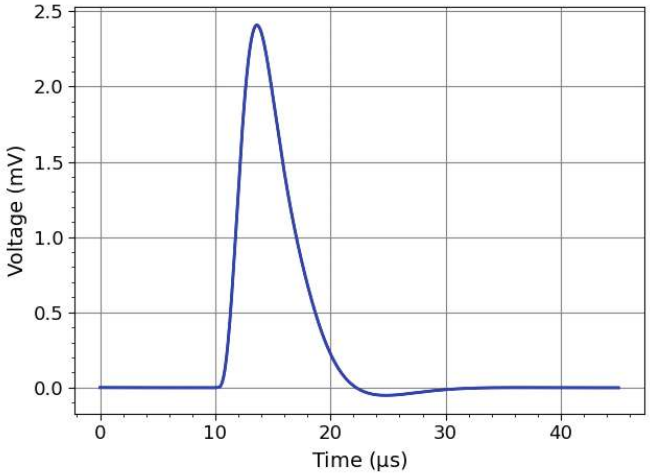


Figure 8.14: Simulated output signal of the closed-loop CSA circuit for a signal charge of $Q_{in} = 662\,\mathrm{e}^-$, input capacitance of $C_{in} = 1\,\mathrm{pF}$ and load capacitance of $C_L = 20\,\mathrm{pF}$.

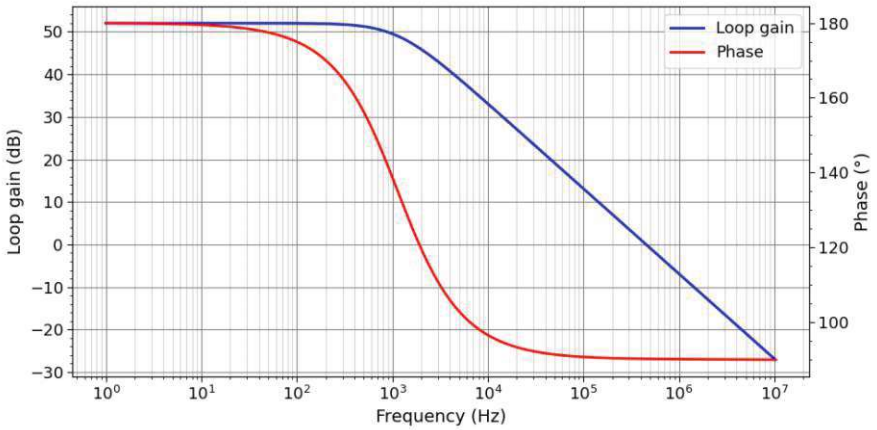
8.4.2 Conversion to Finger Structure

The Fraunhofer IISB SiC process does not support an automatized way to account for the distribution of the MOSFET width onto several fingers in the schematic or layout. Hence, this distribution was added manually. In the schematic, a MOSFET with a width of W consisting of n fingers was converted to n separate MOSFETs with a width of $\frac{W}{n}$. Before the conversion, the MOSFET widths were adjusted such that the finger widths have an integer number of µm i.e. the total width W is divisible by n without remainder. Then, the MOSFETs representing the single fingers corresponding to a single MOSFET in the schematic were connected by respectively shorting their gate, source, drain and bulk terminals. Fig. 8.16 shows a schematic of the conversion of a single MOSFET with a width of W to several fingers with a width of $\frac{W}{n}$. The converted schematic was then built by simply connecting the shorted terminals of a MOSFET according to how the single MOSFETs were connected before conversion.
Due to the conversion to the finger structure, a slight change in the DC operating was expected. Therefore a DC analysis was carried out to review, whether adjustments to the schematic were necessary. The highest observed deviation of the operating point voltages was the gate-source voltage of M2 with a deviation of roughly 150 mV. To reduce this deviation, the width of M9 was increased to

$$W_9 = 120\,\mu\mathrm{m}.$$

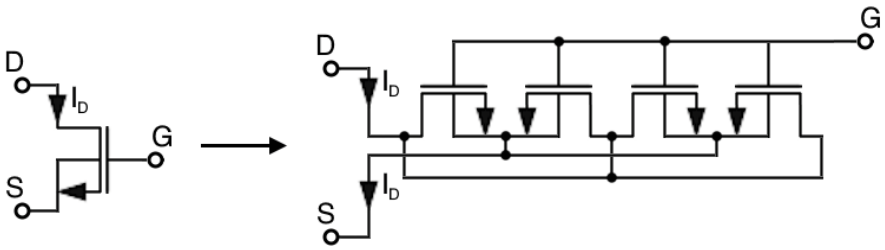
This did, however, not change the drain current of M2, which increased by roughly 0.6 µA, but rather decreased the potential at the folded node by roughly 130 mV. While not having a significant influence on the gain, bandwidth, or SNR, the change was unnecessarily kept anyway.
Additionally, the widths of M11 and M12 were reduced to

$$W_{11} = W_{12} = 90\,\mu\mathrm{m},$$



(a)

Figure 8.15: Simulated loop gain and phase of the closed-loop CSA circuit for an input capacitance of $C_{in} = 1\,\mathrm{pF}$.



(a)

Figure 8.16: Simulated loop gain and phase of the closed-loop CSA circuit for an input capacitance of $C_{in} = 1\,\mathrm{pF}$.

reducing their current by roughly 0.35 µm. To account for that, the width of M15 was accordingly changed to

$$W_{15} = 160\,\mu\mathrm{m}.$$

This change, did not have significant changes to the performance of the circuit as well, however later on solved an issue in the layout involving M11 and M12.
After the conversion, the output signal of the CSA exhibited slightly stronger oscillation with a more pronounced undershoot and a slight overshoot after that. This was most likely due to a reduction of the effective feedback capacitance after the conversion to the finger structure of MOSFETs. At the cost of SNR, the width of MC1 was increased to

$$W_{C1} = 20\,\mu\mathrm{m},$$

in order to increase its effective capacitance so that the output signal can be stabilized. Tab. 8.6 summarizes the final MOSFET sizing, DC operating point and number of fingers under nominal operation of every MOSFET in the CSA.

Table 8.6: Summary of the widths W, lengths L, number of fingers n, drain-source voltage V_{DS}, gate-source voltage V_{GS} and drain current I_D of every MOSFET in the CSA with a load resistance of 1 MΩ.

	n	W (µm)	L (µm)	V_{DS} (V)	V_{GS} (V)	I_D (µA)
M1	6	552	6	14.05	14.05	93.38
M2	2	200	3	3.91	7.37	7.60
M3	9	846	5	5.95	5.99	101.00
M4	2	196	7	10.14	9.15	7.60
M5	9	846	5	5.99	5.99	101.00
M6	2	196	7	6.68	9.15	6.91
M7	7	700	7	9.15	9.15	26.68
M8	2	200	3	7.28	7.28	6.91
M9	2	120	10	6.04	6.04	6.91
M10	3	219	5	10.85	5.99	26.68
M11	1	90	7	10.26	10.03	9.58
M12	1	90	7	10.03	10.03	9.54
M13	5	445	5	3.84	3.96	9.58
M14	5	445	5	4.07	3.95	9.54
M15	2	160	5	5.90	5.99	19.12
M16	7	623	6	10.15	10.26	109.18
M17	9	837	5	9.85	5.99	101.56
M18	1	80	6	1.87	9.15	2.23
M19	5	470	5	8.28	8.28	2.23
MR	1	3	10	0	18.13	0
MC1	1	20	10	0	0	0
MC2	15	1500	10	0	-0.11	0

The simulation and results of the performance of the full circuit are presented in the following section.

9 Circuit Performance

The performance of the CSA was simulated using a supply voltage of $V_{DD} = 20\,\mathrm{V}$ and a biasing current of $I_b = 101\,\mu\mathrm{A}$. For the input and output of the circuit, four combinations of sensor configurations and load capacitance requirements were simulated. For the sensor configurations, the thin sensor configuration representing a $50\,\mu\mathrm{m} \times 50\,\mu\mathrm{m}$ pixel sensor designed in the Fraunhofer IISB SiC CMOS with a signal charge of $Q_{in} = 662\,\mathrm{e}^-$ and an input capacitance of $C_{in} = 1\,\mathrm{pF}$ and a thick sensor configuration representing a $50\,\mu\mathrm{m}$ strip sensor with a signal charge of $Q_{in} = 2755\,\mathrm{e}^-$ and an input capacitance of $C_{in} = 4.5\,\mathrm{pF}$ were used. For the load capacitance, the lower and upper design requirements of 4 pF, representing the input capacitance of another buffer amplifier, and 20 pF, representing a 10x oscilloscope probe, were used.
One simulation tested one combination of signal charge and capacitance, including several analyses to extract the output resistance, power consumption, phase margin, gain, bandwidth, noise, settling time, SNR and charge linearity of the circuit. The output resistance and power consumption were extracted using a DC analysis. A total power consumption P of

$$P = 6.1\,\mathrm{mW}$$

could be found as the product of the total current flowing from VDD to GND and the supply voltage of 20 V and adding the product of the biasing current with the drain-source voltage of M5. The output resistance R_{out} resulted in a value of

$$R_{out} = 457.45\,\Omega,$$

by determining the voltage drop at the output of the CSA by applying a load resistance of $R_L = 1\,\mathrm{M}\Omega$.
Fig. 9.1 shows the loop gain and phase as a function of frequency for the CSA obtained by a stability analysis with the input capacitances C_{in} of 1 pF and 4.5 pF. The gain around the feedback shows, in both cases, one pole, while for a higher input capacitance, the pole frequency is smaller. The phase margin PM in both cases, i.e. the phase angle found at the unity loop gain frequency, resulted in roughly

$$PM \approx 90^\circ,$$

which hints at sufficient feedback stability. The load capacitance had no influence on the loop gain and phase in the simulation.
The voltage gain and bandwidth of the amplifier were extracted using an AC analysis. The total equivalent output noise (root-mean-square noise) was extracted using a noise analysis by integrating over the simulated noise versus frequency spectrum up to the bandwidth frequency. Using a transient analysis, the output voltage signal was simulated and then analyzed for its 1 % settling time. By comparing the total equivalent output noise with the amplitude of the signal, the SNR was calculated. The ENC of the circuit was calculated by scaling the input signal charge with the SNR. Tab. 9.1 summarizes the performance values obtained from the AC analysis, transient analysis and noise analysis. Fig. 9.2 shows the voltage gain of the CSA as a function of frequency obtained by an AC analysis with the load capacitances C_L of 4 pF and 20 pF. Similar to before, the

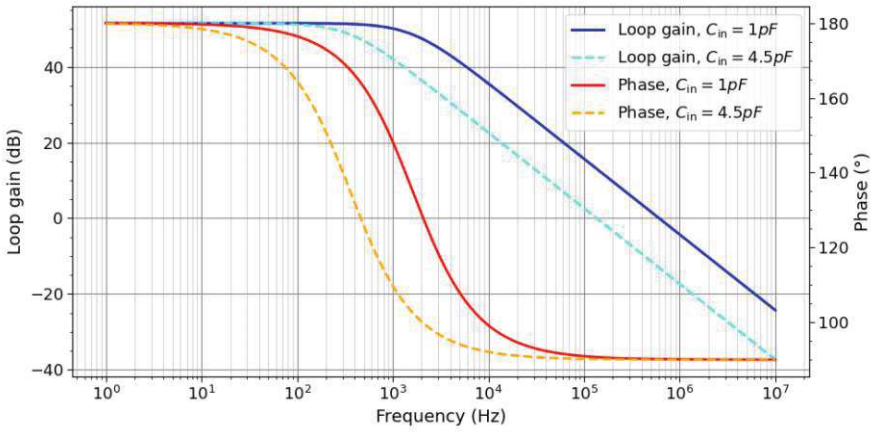


Figure 9.1: Loop gain and phase as a function of the frequency of the full CSA for an input capacitance C_{in} of 1 pF and 4.5 pF.

Table 9.1: Summary of the simulation results of the CSA for all combinations of the sensor configurations and load capacitance C_L including the voltage gain A_v, bandwidth frequency f_{BW}, 1 % settling time t_s, total noise voltage u_n, ENC and SNR – thin sensor configuration: signal charge $Q_{in} = 662\,e^-$ and input capacitance $C_{in} = 1$ pF; thick sensor configuration: signal charge $Q_{in} = 2755\,e^-$ and input capacitance $C_{in} = 4.5$ pF.

Sensor	C_L (pF)	A_v (dB)	BW (kHz)	t_{set} (µs)	u_n (µV)	ENC (e^-)	SNR
thin	4	50.97	30.19	13.76	347.2	119	5.58
thin	20	50.97	30.79	13.72	351.1	107	6.18
thick	4	50.97	30.19	23.57	464.8	217	12.7
thick	20	50.97	30.79	23.09	472.1	205	13.4

circuit had been converted to the MOSFETs in finger structure, and the higher load capacitance caused the buffer amplifier to show gain peaking. This is evident by the second gain roll-off at a frequency of approximately 300 kHz shown in the red curve in Fig. 9.2, which is again quite abrupt. The gain peaking, however, seems suppressed when observing the full circuit, as the core amplifier provides enough filtering of the critical frequencies.

The changes in the sensor and load capacitances had no influence on the low-frequency voltage gain A_v, which stayed the same at roughly $A_v = 51$ dB for all capacitance combinations as seen in Tab. 9.1. While the input capacitance had no influence on the bandwidth, the load capacitance caused a slight difference of 0.6 kHz. This slight difference in bandwidth caused by the load capacitance can also be observed in the total output noise voltage u_n. Fig. 9.3 shows the noise versus frequency spectrum for all combinations of the sensor and load capacitances, where the curves for the same load capacitance do not significantly differ. Hence, the difference in noise of a few µV, for both input capacitances respectively, originate from the slight difference in bandwidth. The significant change in total noise between the cases of different input capacitance is evident from Fig. 9.3 as well. While the noise spectra at low frequencies are pretty

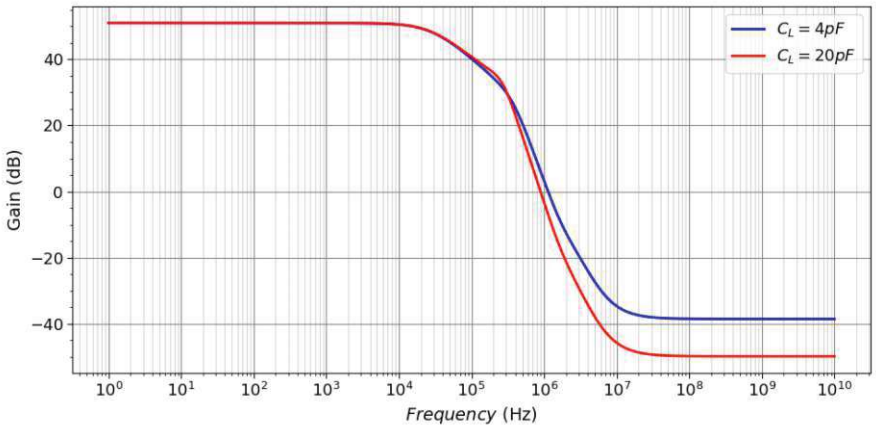


Figure 9.2: Gain as a function of the frequency of the full CSA for a load capacitance C_L of 4 pF and 20 pF.

much identical, the cases with higher input capacitance rise in noise at high frequencies. This hints at an increased noise gain towards the bandwidth frequency caused by the increase in input capacitance. The rise in noise towards low frequencies is caused by the $1/f$-noise contributions created in the circuit's electronics.

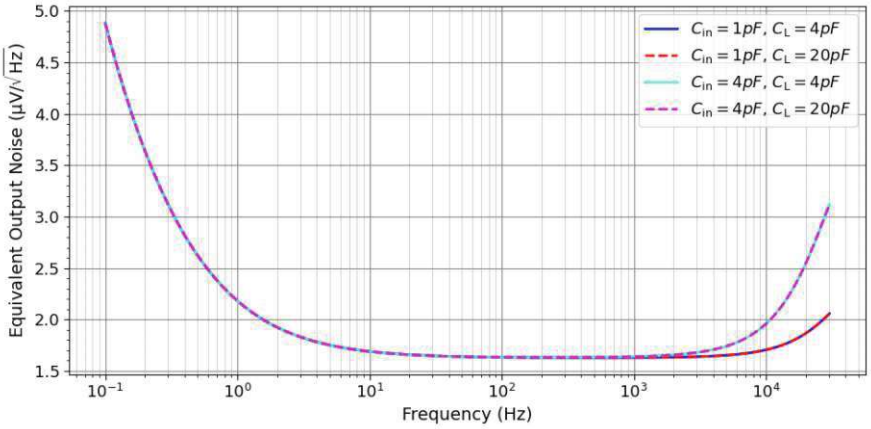
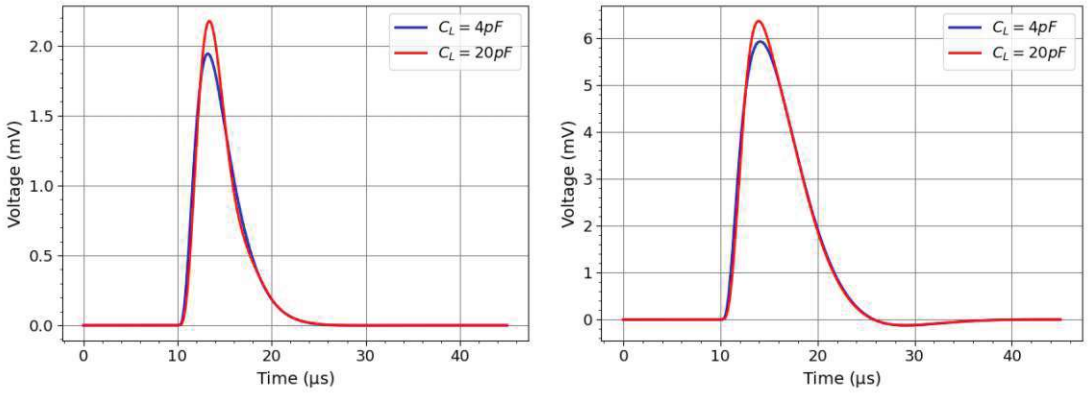


Figure 9.3: Equivalent output noise voltage as a function of the frequency of the full CSA for four combinations of input capacitance C_{in} and load capacitance C_L.

Fig. 9.4 shows the output signals for all four combinations of sensor configurations and load capacitance. Fig. 9.4a shows the signal for the thin sensor configuration and both load capacitances, while Fig. 9.4b shows the signals for the thick sensor configuration and both load capacitances. While the increase in load capacitance increased the signal amplitude, the increase in input capacitance more significantly increased the 1 % settling time of the output signal. As can be seen in Tab. 9.1, the settling times for the same input capacitances do differ slightly up to tens of ns, which might be explained by a slight alteration of the effective feedback capacitance due to the load capacitance and hence the

time constant $R_f C_f$ of the circuit. The exact determination of the feedback capacitance of the CSA is not trivial, due to the complex structure of the feedback network and was hence omitted. This, however, prevents a calculation of the time constant and, therefore, of the exact possible measurable rate of particle pulses. From the inverse of settling times, an approximate rate of roughly 70 kHz for the thin sensor configuration and roughly 40 kHz for the thick sensor configuration could be estimated.

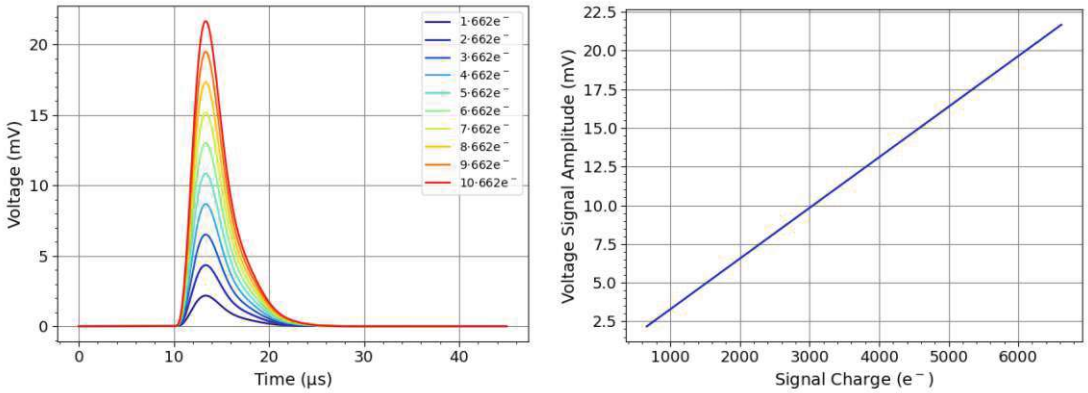


(a) Thin sensor (b) Thick sensor

Figure 9.4: Output signal of the full CSA for four combinations of sensor configurations and load capacitances C_L – (a) thin sensor configuration: signal charge $Q_{in} = 662\,\mathrm{e}^-$ and input capacitance $C_{in} = 1\,\mathrm{pF}$; (b) thick sensor configuration: signal charge $Q_{in} = 2755\,\mathrm{e}^-$ and input capacitance $C_{in} = 4.5\,\mathrm{pF}$.

The increase in signal amplitude due to the change in load capacitance likely also has to do with the time constant of the circuit. A higher load capacitance causes the feedback capacitance to discharge slower, letting it store up more charge and hence cause a higher voltage signal amplitude. Note that even though the signal charge was roughly four times higher for the thick sensor configuration than for the thin sensor, the outgoing signal amplitude only increased by a factor of approximately 3. This is only surprising, assuming that the amplifier output depends linearly on the signal charge. Assuming so for now, this could most likely be explained due to the higher input capacitance.
The SNR could be obtained by dividing the signal amplitudes by the total noise voltage as shown in Tab. 9.1. While the noise did not change significantly due to a change in the load capacitance, the change in signal amplitude caused a noticeable difference in SNR by roughly 0.6 for the thin sensor configuration and 0.7 for the thick sensor configuration. However, only for the thick sensor configuration, the design goal of an SNR > 10 could be achieved.
Knowing the signal charge for a given sensor configuration, as well as the resulting SNR, the ENC could be calculated by dividing the signal charge by the SNR. The resulting values are presented in Tab. 9.1. For the thin sensor configuration, an ENC of almost only $100\,\mathrm{e}^-$ could be reached without dedicated noise optimization. However, the signal charge produced by a MIP seems not enough to reach a sufficient SNR nonetheless. For the thick sensor configuration, the ENC is approximately doubled, but with the roughly 4 times higher signal charge, the SNR could be sufficiently increased despite the higher

input capacitance.
To get information about the linearity of the circuit, the transient analysis was repeated using the thin sensor configuration and a load capacitance of $C_L = 20\,\mathrm{pF}$. However the signal charge of $Q_{in} = 662\,\mathrm{e}^-$ was multiplied by a factor with a starting value of 1 and gradually increased by 1 over 10 transient analyses. Fig. 9.5 shows the result of these transient analyses. In Fig. 9.5a the resulting output signals can be seen, which did not differ significantly in the 1 % settling time. However, the signal amplitude increased linearly with the increase of the signal charge, as presented in Fig. 9.5b, where the output signal amplitude is plotted against the signal charge.



(a) (b)

Figure 9.5: (a) 10 output signals and (b) corresponding signal amplitudes for linearly increasing signal charge – an input capacitance of $C_{in} = 1\,\mathrm{pF}$ and load capacitance of $C_L = 20\,\mathrm{pF}$ were used.

Note that the performance parameters presented in this section represent the nominal circuit performance. This does not include changes in temperature or power supply, device mismatch, variations in the manufacturing process of the MOSFETs, and parasitic impedances extracted from the layout.

10 Layout

The circuit layout is the physical implementation of a circuit on the semiconductor chip. When designing a CMOS layout, several factors must be considered, including minimizing the space that the circuit takes up and reducing parasitic impedances such as capacitance and resistance introduced by devices and wires, which can degrade performance. In doing so, it must be considered that the spacing and overlap of specific layers cannot be done arbitrarily but must obey the design rules of the process.
In this work, the layout of the CSA was designed by the layout editor in Cadence Virtuoso. This section, first, gives an overview of the design considerations regarding the routing, the layout of single MOSFETs, as well as the placement of the bonding pads. This is followed by a description of the design process.

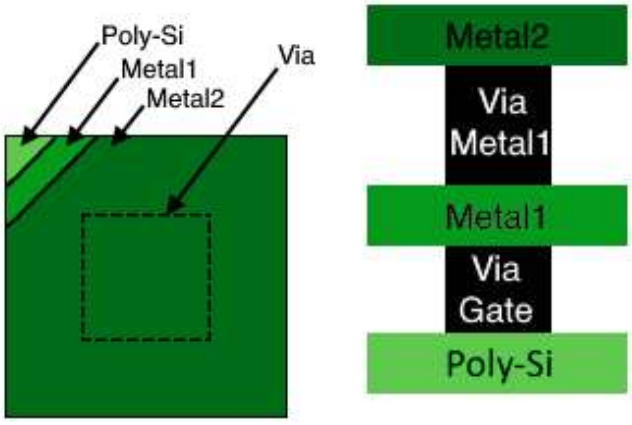
10.1 Layout Design Considerations and Methodology

10.1.1 Routing

As discussed in section 6.3.4, the resistance of metal wires with the smallest possible width according to the design rules could be neglected given that these wires were not longer than 1 mm. Hence, except explicitly necessary, metal wires were placed using the minimal allowed widths. The use of Poly-Si wires was prevented as much as possible and only used if Poly-Si connections between MOSFET gates needed to be extended over short distances. Between wires on the same layer, a space of at least 10 µm was considered, if possible, to ensure isolation. Vias were generally placed with the smallest possible feature sizes allowed by the design rules, as the parasitics of the vias at these sizes could be neglected.
For routing from the Poly-Si layer to the upper-most Metal2 layer, a via-stack was used as visualized in Fig. 10.1. Fig. 10.1a shows the stack design from a top-down view, while Fig. 10.1b shows it from a side view. The stack was also designed using the smallest possible dimensions allowed by the design rules. From the Poly-Si layer on the bottom, a ViaGate1 was connected to a Metal1 layer. From Metal1, a ViaMetal1 was used to connect to a Metal2 layer. The edges of the metal squares needed to have a distance of a few µm to the edges of the via squares as dictated by the design rules.

10.1.2 MOSFET Layout

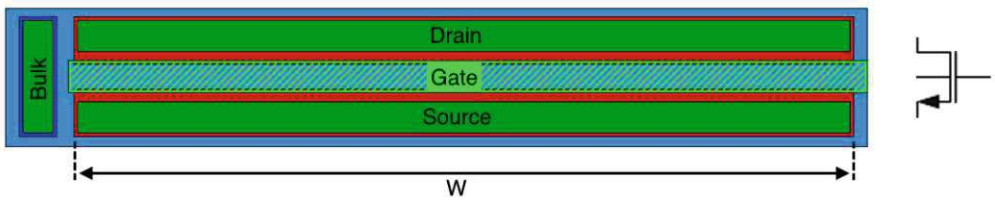
The ready-to-use layout cells of the MOSFETs included in the Fraunhofer IISB SiC process PDK version 1.2 do not have an automatic distribution of the MOSFET width onto several fingers (see section 3). Hence, this distribution had to be done manually using separate MOSFET cells. As discussed in section 6.3.4, the maximum width per finger could set to 100 µm. This is the highest width for which the MOSFET SPICE models were experimentally validated.
Fig. 10.2 shows two versions of the layout of an NMOS transistor with an arbitrary width of W from a top-down view as well as the equivalent circuit schematics. 10.2a shows the NMOS with a single channel and 10.2b with the channel distributed onto four fingers of $\frac{W}{4}$ channel width. For Fig. 10.2b, two neighboring MOSFETs share the same source or drain terminals, such that the label of the terminals alternates between source



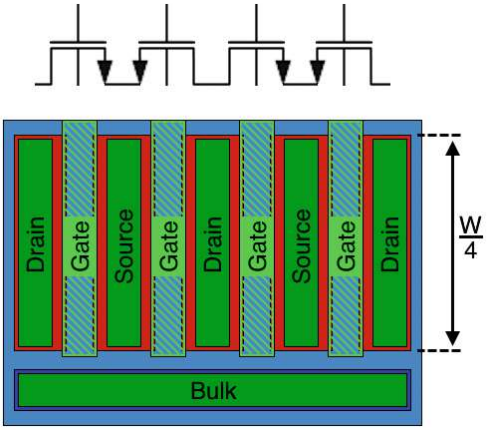
(a) Top-down view (b) Side view

Figure 10.1: Via stack consisting of a $12 \cdot 12\,\mu m^2$ Poly-Si layer on the bottom, a ViaGate1 with dimensions of $4 \cdot 4\,\mu m^2$ connecting to a $12 \cdot 12\,\mu m^2$ Metal1 layer and a $4 \cdot 4\,\mu m^2$ ViaMetal1 connecting $12 \cdot 12\,\mu m^2$ Metal2 layer - the dashed lines in (a) enclose the cross-sectional area of both vias stacked over each other.

and drain when going from left to right (not counting the gate terminals). The bulk terminal is shared by all the fingers via one single Metal1 strip perpendicular to the source and drain terminals.
Using this finger structure, routing could be simplified while also reducing the size of one MOSFET when compared to a structure where the neighboring terminals are not shared. The gate, source and drain terminals still needed to have their respective short connections in order to arrange the MOSFET channels in parallel. The short connection for the bulk was automatically given by the presented finger structure. If a MOSFET has the source and bulk connected to each other, the source Metal1 terminals just needed to be extended towards the bulk Metal1, as shown in Fig. 10.3.
For the gate terminals, the Poly-Si layers were extended at the top edge (the edge opposite of the bulk terminals) and could be connected using Poly-Si wires or metal wires. The latter option was used to minimize the capacitance to the substrate and the resistance between the single gates of a MOSFET. For this to work, the gate poly-Si needed to be connected to the metal layers using a via stack, as presented in section 10.1.1.
The short connections of the source terminals (if they were not connected to bulk) and drain terminals could be done in two ways: by extension of the metal connection to the side opposite of the bulk terminal or by routing the metal connections over the MOSFET channels. Fig. 10.3 shows a visualization of both options as well as the equivalent circuit schematic for the drain terminals. Depending on the metal layer that was used to connect the gates, the extension to the top edge could be done either by directly extending the Metal1 rectangles and routing below the Metal2 gate connections as shown in Fig. 10.3a or by routing each terminal to Metal2 using a via and routing over the Metal1 gate connections – which would be analogous to Fig. 10.3a if only the Metal layers of drain and gate had been switched and the drain would overlap the gate wire in the center.



(a) Single channel width W



(b) 4 channels with width $\frac{W}{4}$

Figure 10.2: Layout of an NMOS with (a) a single channel of a width of W and (b) a finger structure i.e. a width of W divided onto 4 channels of width $\frac{W}{4}$ as well as the equivalent circuit schematics – blue: p-Well, dark blue: p-Implant, red: n-Implant, lime (hatched and solid): Poly-Si, green: Metal1.

Routing over the channels was generally done by routing the Metal1 strips to Metal2 using vias and connect them with perpendicular Metal2 wires. A big advantage of routing over the channels is that it was more efficient in terms of layout space since the extensions outside of the MOSFET area take up additional space budget.
Both options contribute parasitic capacitances between the gate and the source or drain. Routing over the channel lengths contributes slightly higher capacitances for high channel lengths because the overlap between the metal layer and the gate Poly-Si is larger than for the extensions on the top of the MOSFET. The extension on the top edge, however, contributes more capacitance between the substrate and the source or drain.
Whichever option was chosen depended either on which parasitic capacitance needed to be reduced at a specific node in the circuit or if there were space constraints. If the capacitance between the gate and the source or drain could be neglected, routing over the channels was preferred due to the more efficient use of layout space.

10.1.3 Placement of the Pads

For the biasing of the circuit, 3 pads are necessary to provide the VDD potential, GND potential, and biasing current I_b. Two pads are needed for the signal transmission of the ingoing and outgoing signals. Since the signal comes from a sensor pixel on its own

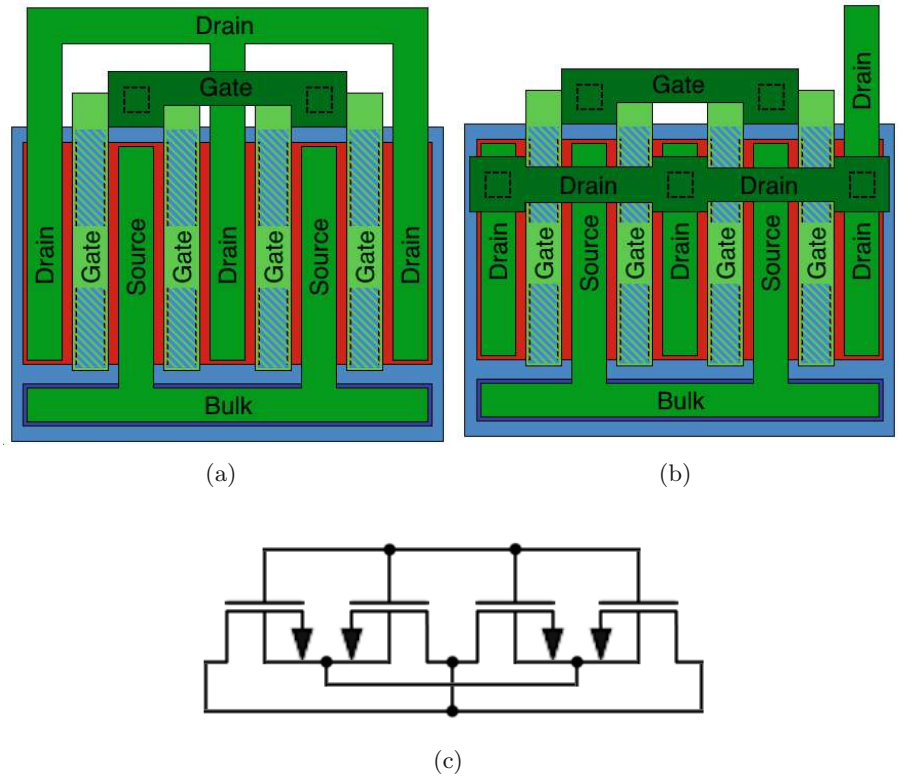


Figure 10.3: Layout of an NMOS finger structure with four channels and the corresponding routing to parallelize the four channels as well as the equivalent circuit schematic (c). The source is shorted with the bulk by extending the source metal1 terminals to the bulk metal1 terminal. The gate is shorted by shorting two neighbouring Poly-Si routing these to metal2 using a via stack and interconnecting all gates in metal2. Shorting of the drain is done either by (a) extending the metal1 terminals in parallel to the edge opposite of the bulk terminal or (b) routing the metal1 terminals to metal2 using a ViaMetal1 and interconnecting them over the gates – blue: p-Well, dark blue: p-Implant, red: n-Implant, lime (hatched and solid): Poly-Si, green: Metal1, dark green: Metal2, dashed line square: vias below metal layer.

chip, it was decided that it would be simpler to have the pad of the ingoing signal on the opposite side of the power supply pads and the output signal pad. This way, no complicated routing on the PCB and around the sensor pixel chip would be needed to connect the external supplies. This is especially true if several CSA chips are placed in parallel to read out multiple sensor pixels.
Consequently, on one side of the CSA chip, the sensor pixel can be placed and connected to the CSA via wire bonds to an IN pad. On the opposite side, a VDD and GND pad would connect to VDD and GND rails on the CSA chip and connect to the PCB board via wire bonds to the side opposite of where the IN pad and sensor pixel would be located. Similarly, also an IBIAS pad would be placed among the VDD and GND pads and be routed away from the chip in the same direction onto the PCB, while on-chip the pad connects to the gate of M5. For the outgoing signal, an OUT pad connects to the output node of the CSA and is bonded from the pad onto the PCB in the direction

away from the IN pad.
A sixth pad was considered to be included, which directly connects to the gate of M2 and provides an externally supplied biasing voltage. This consideration comes from the fact that the operating point of M2 is quite delicate, as it is placed on the lower edge of saturation due to its low drain-source voltage compared to its gate-source voltage. Further, this adds another degree of freedom if the DC operating point needs to be adjusted on the physical chip if it deviates from the simulated DC operating point or the biasing network does not work as intended. The optional voltage supply over this pad could be seen as equivalent to the red-colored voltage supply added to the CSA schematic in Fig. 10.4. A voltage applied to this pad would, therefore, bypass the biasing network of M6, M8 and M10.

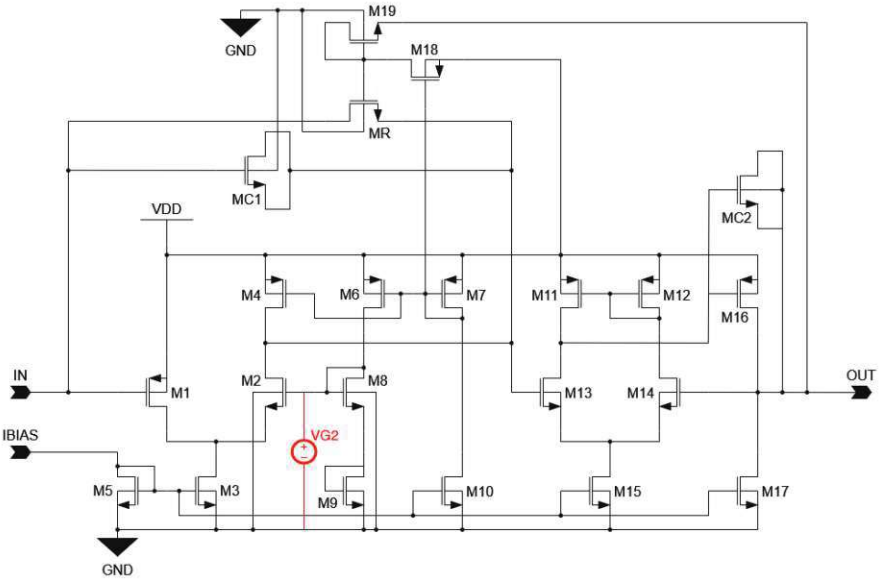


Figure 10.4: Equivalent circuit schematic of the CSA if an additional external voltage supply VG2 (red) is introduced to the gate of M2 via an additional bonding pad on the layout.

The pad got the label VG2. As a supply pad, its placement was also considered on the opposite side of the IN pad and among the VDD, GND, IBIAS and OUT pads. To connect it to the PCB, a wire bond could be routed away from the chip in the same direction as the wire bonds for the VDD, GND, IBIAS and OUT pads.

10.2 Design Process

The design workflow of the physical layout of the circuit could be divided into the following steps: floorplanning, MOSFET placement, routing and addition of the pads.

Floorplanning

The first step was to divide the available space on the chip into several regions in which certain layout structures could be organized. The rough regions that needed to be

considered were rails for the supply voltage and ground, routing wires and MOSFETs could be placed. The idea was to base the floorplan roughly on the schematic to simplify the floorplanning process. For the MOSFETs, several regions could then be considered based on the circuit block and the vicinity of the supply voltage and ground lines. The routing could then also be done mostly analogously to the schematic, which helped create the first draft of the layout.
Fig. 10.5 shows the CSA circuit schematic but with different structures highlighted in different colored boxes, where structures in the same colored boxes were considered to be organized in the same regions of the floorplan. The VDD nodes are highlighted in the light blue box and were considered to be organized in a VDD rail structure. The MOSFETs highlighted by the dark blue boxes are all PMOS transistors that have their source and bulk terminals connected to VDD. They could, therefore, be arranged next to each other and below the VDD rail, with their bulk Metal1 connections overlapping with the VDD rail on the Metal1 layer. The sources then simply needed to be connected by extending their Metal1 connections towards the bulk Metal1 connections as analogously shown in Fig. 10.3 for an NMOS. Note that it was also assumed that the input of the CSA and, therefore, M1 would be located on the left side of the floorplan.

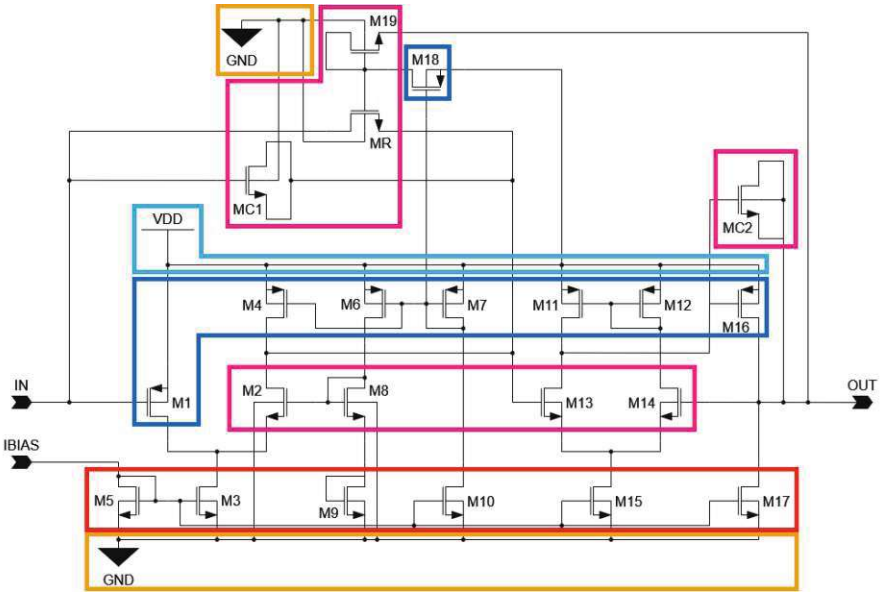


Figure 10.5: Grouping of MOSFET structures and circuit nodes for the organization of placement regions in the layout floorplan – Light blue: VDD rail, orange: GND rail, dark blue: PMOS transistors with source and bulk connected to the VDD rail, red: NMOS transistors with source and bulk connected to the GND rail, magenta: MOSFETs with gate, source and drain connected to nodes other than VDD and GND.

The GND nodes are highlighted by orange-colored boxes and were considered to be organized in a GND rail similar to the VDD rail. The NMOS transistors highlighted by red boxes are all connected to GND with their source and bulk terminals. Hence, they could be arranged next to each other and above the GND rail, overlapping the bulk Metal1 connections with the GND rail on the Metal1 layer. The source Metal1

connection could then be shorted to the bulk connection as shown in Fig. 10.3.
The MOSFETs highlighted by the magenta-colored boxes are MOSFETs that have their gate, source and drain terminals not connected to VDD or GND. They were considered to be arranged next to each other, but in between the rows of MOSFETs highlighted by the dark blue and red-colored boxes. M2, M8, M13 and M14 are NMOS transistors that were identified to vertically interconnect the row of PMOS transistors to the row of NMOS transistors below on the schematic. M19, MR, and MC1 of the feedback network, as well as MC2 of the buffer amplifier, were identified to rather interconnect nodes horizontally in the space between the rows of NMOS and PMOS transistors in the schematic. Hence, to simplify the routing, the MOSFET space in between the rails was divided into 3 zones, one each for the feedback network, MC2 and the remaining transistors i.e. M2, M8, M13 and M14. The feedback network was planned to be placed on the left side i.e. closer to the input of the CSA, to minimize the length of wires around it and, hence, minimize the parasitic capacitance between the wires and the substrate below. MC2 was decided to be placed closer to the output of the CSA and, therefore, on the right side. M2, M8, M13, and M14 were therefore planned to be placed in the center.
Fig. 10.6 shows the rough floorplan that was determined from these considerations. It shows the several regions where the MOSFETs could be placed in the way described above. It additionally shows space for the routing between the rows of MOSFETs, which is highlighted in green. It should be noted that the placement of the feedback network on the left side of the floorplan meant that the wire connecting the source of M19 and the output of the CSA would be quite long and needed to be routed horizontally across the majority of the layout. Therefore, it needed to be checked explicitly whether the resistance of this wire could be seen as negligible or not.

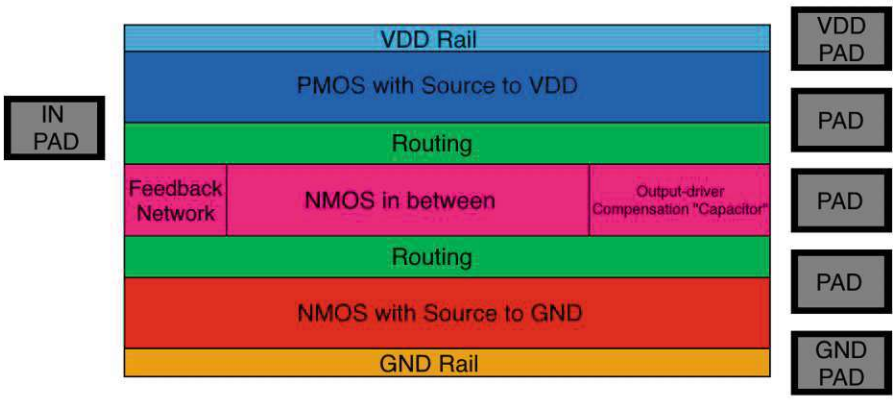


Figure 10.6: Rough layout floorplan of the CSA based on the circuit schematic.

The placement of the pads for the input and output signal, as well as for the power supply, also needed to be considered on the floorplan, as shown in Fig. 10.6. As explained in section 10.1.3, the pad for the input signal would be placed on the opposite side of the other 5 pads. It was placed on the left side, where the gate of M1 could potentially be placed in the vicinity.
The pads for the VDD and GND potentials were considered to be placed near the VDD

and GND rails on top and on the bottom of the layout. The IBIAS, OUT and VG2 pads were added between the VDD and GND pads. They were, however, not explicitly labeled because it was not clear where the optimal placement of each pad would be before the placement of the MOSFETs, as well as their routing, was completed. Instead, PAD was used as a placeholder in Fig. 10.6.

MOSFET Placement

The MOSFETs were then placed, in finger structures if necessary, in their corresponding zones dictated by the floorplan found in the previous design step. Their relative location was based mostly on the relative position of the MOSFET symbols in the schematic up to a few exceptions. M5 was placed on the right side of the layout as it is the reference device of the first NMOS current mirror and has, therefore, its gate and drain directly connected to the IBIAS pad, which is located on the right side of the layout as well.
M18 was integrated into the upper row of PMOS transistors between M1 and M4 in order to bring it closer to the other MOSFETs in the feedback network. Placing M19, MR and MC horizontally next to each other created an unnecessary empty space below MR and MC1, as their widths are relatively small. Instead, they were placed above M19 and filled out space on the left side of the upper routing region of the floorplan. This had to be accounted for in the next step but turned out to be beneficial to the compactness of the feedback path as well.
Due to the high width of MC2, placing its full finger structure in the central MOSFET row of the floorplan horizontally extended it beyond the upper and lower rows, which would have left a lot of unnecessary empty space between the upper and lower rows and the pads on the right side. Hence, 3 fingers of MC2 were removed from the full structure and integrated into the upper row of MOSFETs, such that they could then be connected in parallel to the other 12 remaining fingers.
Between the MOSFET rows, a rough space of 100 µm was left for routing. However, this space was flexibly adjusted during the routing process to make the layout as compact as possible or make space for more wires.

Routing

The layout of the electronics was then finalized by interconnecting the MOSFET terminals according to the schematic and using the considerations explained in section 10.1.1. It was taken into account that the routing of the feedback path between the CSA input node and core amplifier output needed to be as compact as possible to prevent parasitic capacitances between the wires and the substrate below. To minimize these capacitances further, wires on Metal2 were used as much as possible due to their higher distance from the substrate. Also, overlaps between different layers needed to be prevented to reduce parasitic capacitances between nodes.
For current mirrors with M5 and M7 as reference devices, the gates from the reference device to the gates of the mirroring devices were connected using a single wire along the corresponding row of MOSFETs. For the NMOS current mirror with M5 as a reference device, this wire could be done solely on Metal2. For the PMOS current mirror with M7 as a reference device, Metal1 had to be used for single gates.

To simplify the routing, M11 and M12 switched their relative position as dictated by the schematic. The 3 fingers that were removed from MC2 were then placed between M11 and M16. The gates of these 3 fingers are shared with the gates of 3 other fingers of MC2 below. The drain, sources and bulks were connected by extending all the drains and sources to the bulk connections. For the 3 removed fingers, the sources and drains were additionally extended to adjacent sources and drains of the fingers below between the gate connections.
The source of M19 had to be routed from the left of the layout across the width of four transistors to the bulk connection of MC2. Metal2 was used to ensure the capacitance of the wire and the substrate below was reduced. The length of the wire was possible to be kept reasonably short with roughly 600 µm. With a current of 2.23 µA flowing through the wire, the voltage drop should be below mV.
The VDD and GND rails were added using 25 µm thick wires to make the wire resistance as small as possible. These wires were then aligned with the edges of the bulk connections of the MOSFETs in the upper and lower MOSFET rows. The wiring to the pads was then applied in the subsequent step after the addition of the pads.

Addition of the Pads

As soon as the MOSFETs were placed and routed, the bonding pads with a pitch of 90 µm × 80 µm were connected to the corresponding circuit nodes in the layout. The IN pad for the transmission of the ingoing signal was placed on the left side of the electronics. Its shorter edge was vertically centralized on the Metal2 wire that extended from the gate of M1 to the pad. A horizontal space 10 µm was left between the pad and the Metal2 wire that horizontally extended from the bulk of MR.
The VDD, GND, IBIAS, VG2 and OUT pads on the opposite side of the electronics were placed with their shorter edge toward the electronics and with equal vertical space of 34 µm between them. The equal space was chosen such that a single wire of 12 µm thickness could pass between the pads with additional vertical space above and below the wire. The GND pad was then placed, aligning its lower edge with the lower edge of the Metal1 connection of the GND rail that was placed in the previous step. The space to the electronics was dictated by the space the ESD diodes on their left side needed, plus the thickness of the Metal2 wire connecting these diodes.
The VDD and GND pads were simply connected by extending the corresponding rails towards the pads. The IBIAS pad was added above the GND pad so that the gate of M5 could be routed to the lower edge of the pad. One pad above, the OUT pad was added, which was connected to the drain of M16, M17 and bulk of MC2 via the lower edge of the pad.
For the VG2 pad, above the OUT pad, a wire from the gate of M2 and M8 had to be routed above the VDD rail via the drain terminal of M6. The relative positions of M4 and M6 on the CSA circuit schematic had to be switched to simplify the routing. Then, using a ViaMetal1 from the Metal1 connection of the drain of M6 to the Metal2 layer, the wire could be routed over the VDD rail and horizontally routed to the upper edge of the VG2 pad. The length of the wire connecting the drain of M6 and the VG2 pad was found to be roughly 900 µm. A DC operating point simulation including a voltage source with a voltage of 13.3 V representing VG2 as in Fig. 10.4 did not significantly

change the DC operating point or the performance of the circuit. The simulated current flowing from the supply was roughly 35 nampere. With a wire length of roughly 900 µm, the voltage drop along the wire is less than 1 mV.
Next, the ESD diodes were added to the pads. On the IN pad, they were added to the longer edges of the pads. The diode based on the n-Well was connected to the VDD rail above, while the diode based on the p-Well was connected to the GND rail below on the Metal1 layer. For the VG2, OUT, and IBIAS pads, the p-well-based diodes were added to the left, and the n-well diodes were added on the right of the pads, i.e. on the shorter edges. The left diodes were connected to the GND pad below using a 25 µm wide Metal2 wire. Similarly, the diodes on the right side of the pads were connected to the VDD pad above using a Metal2 wire of the same width. Connecting to this wire a 50 µm × 50 µm Metal2 square with a cross-shaped passivation opening was added as an alignment marker to aid the wire bonding process as a reference point.
Fig. 10.7 shows the finished layout of CSA, including the labeling of every MOSFET and pad. The yellow wires are on the Metal1 layers and the orange wires are on the Metal2 layer. The layout takes up a space of 1293 µm × 508 µm.

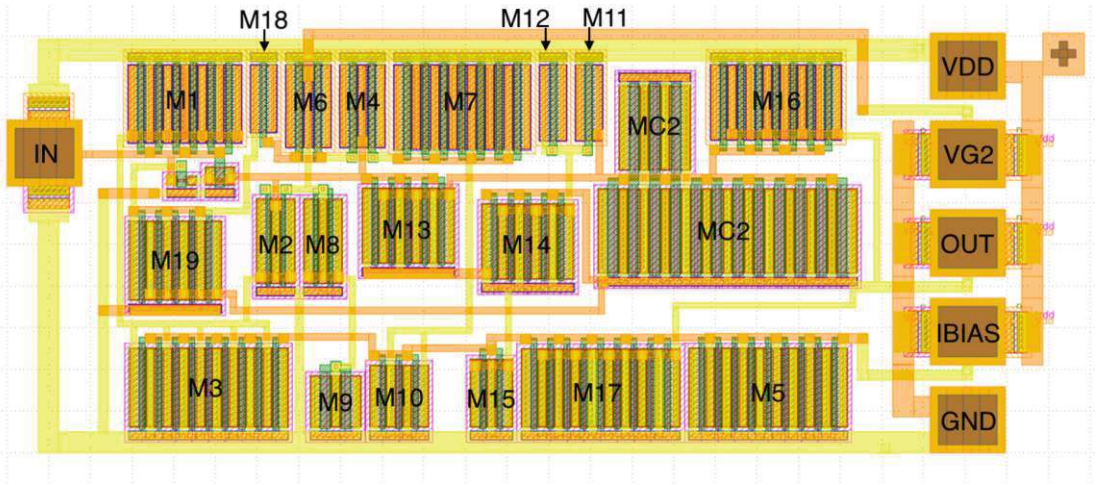


Figure 10.7: Finished layout of the CSA including labelling of all MOSFETs and Pads – yellow wires are on the Metal1 layers and the orange wires are on the Metal2 layer.

11 Discussion

The used 2 µm CMOS process comes with relatively large MOSFET devices compared to modern CMOS technologies. Consequently, only rather slow electronics in the scale of 10 kHz to 100 kHz were possible to develop in a first design run. Apart from that, the feature size of the devices potentially fills out tremendous space on a chip, which limits the remaining space needed to place an array of pixel sensors in a potential MAPS application. The lack of resistors and capacitors in the used PDK also introduced some complications in the design of the CSA. These complications could, however, be met with the use of parasitic capacitances of a MOSFET as a capacitor, while for a resistor, a MOSFET driven in the linear region with a regulated effective resistance was implemented.

The design process using the generation of look-up tables of MOSFET parameters proved to be quite convenient. However, this design process could be better streamlined by consistently using equal MOSFET channel lengths across the circuit. Additionally, a more efficient scheme for the determination of the voltage biasing should be worked out for a future design that includes considerations from the full circuit topology instead of doing it circuit block by circuit block. In this work, this led to M18 not being placed deep enough in saturation. This might subject its drain current to significant changes in the DC operating point due to fluctuations in the power supply, for example. This could turn out to be critical as M18 is supposed to provide a constant current and, therefore, degrade the functionality of the regulation of the feedback resistance.
During the design process of the feedback network, it was observed that it introduces an additional feedback path from the output of the buffer back to the input of the core amplifier via parasitic capacitances of M19 and MR. While this was initially not anticipated, it affected the performance of the CSA rather positively, as it caused a reduction of the noise created in the buffer amplifier by negative feedback.

Albeit the noise of the CSA could be found to be sufficiently low with enough signal charge, a more detailed noise analysis [40] of the circuit should be carried out in the next iteration. By modeling the noise behavior of the MOSFET more closely, this could be implemented into the design process via look-up tables. In combination with a noise requirement, a better setup with a signal-to-noise performance could be found.

The circuit could unfortunately not be simulated under the influence of process variations, as the used PDK did not offer any process corners for the process. The same is true for device mismatch. However, this could somewhat be manually simulated by setting up a Monte-Carlo simulation, where random variations of the MOSFET channel dimensions of every MOSFET are randomly pulled from a random distribution. Combined with a sweep over different temperatures, supply voltages and bias currents, a simulation under environmental fluctuations could provide a closer look at the circuit performance under a more realistic environment.
Additionally, also parasitic impedances introduced by the layout geometry should be included into the analysis. However, parasitic extraction was not included in the PDK. A manual extraction could be carried out, but this could prove to be prone to errors.

12 Conclusion and Outlook

Throughout this thesis, the design of a CSA circuit with a 2 µm CMOS process in SiC developed at Fraunhofer IISB was presented. The presented work is a first step towards the potential development of MAPS using SiC, by designing the first block in the readout chain of a detector for the detection of high-energetic particles. Doing so allows to test the capabilities of CMOS electronics in SiC for their application in particle detection. While the presented work only entails the design simulation, a first glimpse into the performance of the electronics and the potential of the used SiC CMOS process could be obtained.
The performance of the circuit yields promising results for the confident detection of particles with enough signal charge. Assuming a thick SiC sensor with a higher input capacitance of 4.5 pF produced a maximum ENC of 217 e^- and SNR of 12.7. Only with a thin, 12 µm thick SiC sensor pixel using the epi-layer of a chip developed in the used SiC CMOS process showed to produce too less charge to meet the requirement of an SNR of at least ten. With an expected signal charge for a MIP of 660 e^-, an input capacitance of 1 pF and an ENC of 107 e^- a signal to noise ratio of 6.18 could be achieved. It can, however, be argued that with an ENC of 107 e^- for a 1 pF and a thicker sensor having a similar capacitance, a much better and sufficient SNR could be achievable.

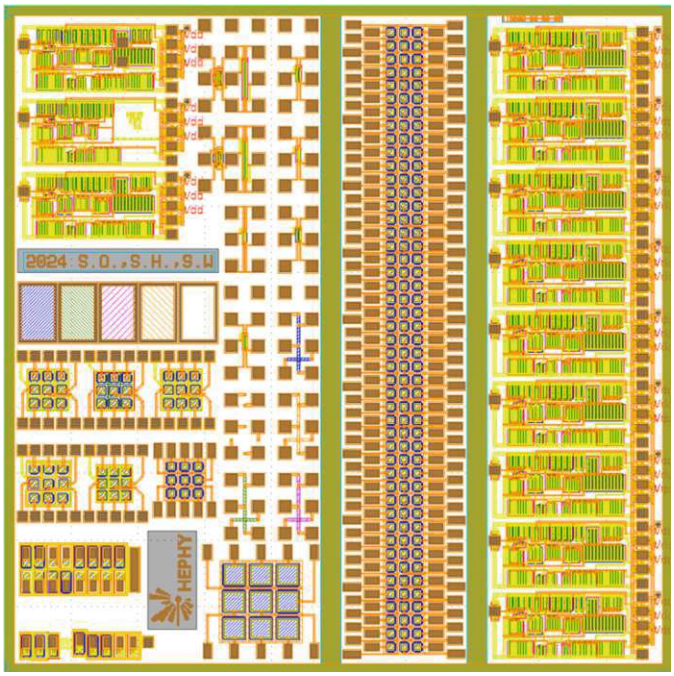


Figure 12.1: Chip layout design for the MPW production run via Europractice and Fraunhofer IISB, including the presented CSA circuit layout. Several copies of the CSA layout are implemented along the right edge of the die as well as in the upper-left corner.

Given the big feature sizes of the electronic devices a rather low bandwidth of 30.79 kHz in the best-case could be reached. Considering the simulated settling times, possible

particle rates of roughly 40 kHz could be estimated with the thick sensor setup. For testing in a lab setup, this, however, seems to be sufficient.
The circuit layout design was submitted for fabrication as part of a multi-project wafer production run via Europractice and Fraunhofer IISB. Fig. 12.1 shows the layout design of a $5 \times 5\,\mathrm{mm}^2$ chip die, including several test structures for layer material and MOSFET characterization, pixel sensor arrays and the CSA circuit layout. The design was submitted in February 2024. To the knowledge of the author at the time of publication of this thesis, delivery of the chip is scheduled for an unspecified date in 2025.
For the development of a SiC MAPS there is still a long way to go. The work presented in this thesis is, to the best knowledge of the author, the first work on SiC-based readout electronics for particle detectors in high-energy physics. There are still many parts of the readout chain that need to be designed and tested using SiC CMOS technology.
In the future, the simultaneous implementation of the electronics and sensors also needs to be figured out. The version of the Fraunhofer IISB SiC CMOS process used in this work would require modification, introducing the proper structure of wells and implants needed to shield the electronics from the high voltage needed to deplete the active sensor volume. Further, a reduction of the size of the electronics is essential to create the required space needed for a sensor pixel array on a chip. While these issues need to be solved in the long term, their solution is up to the capabilities of the manufacturer to meet the requirements of the development of a SiC MAPS.

References

[1] M. Garcia-Sciveres and N. Wermes, "A review of advances in pixel detectors for experiments with high rate and radiation," *Reports on Progress in Physics*, vol. 81, p. 066101, May 2018.

[2] N. Cartiglia, , *et al.*, "LGAD designs for Future Particle Trackers," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 979, p. 164383, Nov. 2020.

[3] E. Sicking, "Detector requirements for future high-energy collider experiments." TREDI 2020 [Online; accessed 30.08.2024], 2020. URL: `https://indico.cern.ch/event/813597/contributions/3727952/attachments/1988376/3314100/EvaSicking_DetectorRequirements.pdf`.

[4] E.D.R.R.P. Group, "The 2021 ECFA detector research and developement roadmap," 2021.

[5] F. Nava *et al.*, "Silicon carbide and its use as a radiation detector material," *Measurement Science and Technology*, vol. 19, p. 102001, Aug. 2008.

[6] J. W. Palmour, "Silicon carbide power device development for industrial markets," in *2014 IEEE International Electron Devices Meeting*, IEEE, Dec. 2014.

[7] C. Langpoklakpam, "Review of Silicon Carbide Processing for Power MOSFET," *Crystals*, vol. 12, p. 245, Feb. 2022.

[8] M. De Napoli, "SiC detectors: A review on the use of silicon carbide as radiation detection material," *Frontiers in Physics*, vol. 10, Oct. 2022.

[9] J. M. Rafi *et al.*, "Electron, Neutron, and Proton Irradiation Effects on SiC Radiation Detectors," *IEEE Transactions on Nuclear Science*, vol. 67, pp. 2481–2489, Dec. 2020.

[10] M. Mager, "ALPIDE, the Monolithic Active Pixel Sensor for the ALICE ITS upgrade," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 824, pp. 434–438, July 2016.

[11] G. Contin *et al.*, "The STAR MAPS-based PiXeL detector," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 907, pp. 60–80, Nov. 2018.

[12] J. Romijn *et al.*, "Integrated Digital and Analog Circuit Blocks in a Scalable Silicon Carbide CMOS Technology," *IEEE Transactions on Electron Devices*, vol. 69, pp. 4–10, Jan. 2022.

[13] N. Rinaldi *et al.*, "A 4H-SiC CMOS Oscillator-Based Temperature Sensor Operating from 298 K up to 573 K," *Sensors*, vol. 23, p. 9653, Dec. 2023.

[14] J. Romijn *et al.*, "Integrated 64 pixel UV image sensor and readout in a silicon carbide CMOS technology," *Microsystems & Nanoengineering*, vol. 8, Oct. 2022.

[15] C. W. Fabjan and H. Schopper, *Particle Physics Reference Library*. Springer International Publishing, 2020.

[16] W. Shockley, "Currents to Conductors Induced by a Moving Point Charge," *Journal of Applied Physics*, vol. 9, no. 10, pp. 635–636, 1938.

[17] S. Ramo, "Currents Induced by Electron Motion," *Proceedings of the IRE*, vol. 27, no. 9, pp. 584–585, 1939.

[18] R. L. Workman *et al.*, "Review of Particle Physics," *Progress of Theoretical and Experimental Physics*, vol. 2022, Aug. 2022.

[19] P. Barletta *et al.*, "Fast Timing With Silicon Carbide Low Gain Avalanche Detectors," 2022.

[20] M. Christanell *et al.*, "4H-silicon carbide as particle detector for high-intensity ion beams," *Journal of Instrumentation*, vol. 17, p. C01060, Jan. 2022.

[21] T. Kimoto and J. A. Cooper, *Fundamentals of Silicon Carbide Technology: Growth, Characterization, Devices, and Applications*. Wiley, Sept. 2014.

[22] M. Barbero *et al.*, "Radiation hard DMAPS pixel sensors in 150 nm CMOS technology for operation at LHC," *Journal of Instrumentation*, vol. 15, pp. P05013–P05013, May 2020.

[23] D. V. Berlea *et al.*, "Radiation Hardness of MALTA2, a Monolithic Active Pixel Sensor for Tracking Applications," *IEEE Transactions on Nuclear Science*, vol. 70, pp. 2303–2309, Oct. 2023.

[24] Wikimedia Commons, "File:MOSFET functioning.svg." [Online; accessed 21-August-2024], 2020. URL: `https://commons.wikimedia.org/wiki/File:MOSFET_functioning.svg`.

[25] G. Bertuccio *et al.*, "A CMOS Charge Sensitive Amplifier with sub-electron equivalent noise charge," in *2014 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC)*, IEEE, Nov. 2014.

[26] C. Gao *et al.*, "A Low-Noise Charge-Sensitive Amplifier for Gainless Charge Readout in High-Pressure Gas TPC," in *Proceedings of Topical Workshop on Electronics for Particle Physics — PoS(TWEPP2018)*, TWEPP2018, Sissa Medialab, May 2019.

[27] R. Kleczek *et al.*, "Charge sensitive amplifier for nanoseconds pulse processing time in CMOS 40 nm technology," in *2015 22nd International Conference Mixed Design of Integrated Circuits & Systems (MIXDES)*, IEEE, June 2015.

[28] B. Razavi, *Design of analog CMOS integrated circuits*. New York, NY: McGraw-Hill Education, 2 ed., 2017.

[29] W. Jendernalik *et al.*, "Unity-Gain Zero-Offset CMOS Buffer with Improved Feedforward Path," *Electronics*, vol. 10, p. 1613, July 2021.

[30] H. Camenzind, *Designing Analog Chips*. [Online; accessed 29.08.2024], 2005. URL: `http://www.designinganalogchips.com/`.

[31] M. Saukoski *et al.*, "Fully Integrated Charge Sensitive Amplifier for Readout of Micromechanical Capacitive Sensors," in *2005 IEEE International Symposium on Circuits and Systems*, IEEE, 2005.

[32] S. Hablas, "Passive SiC Sensor Design." Project Thesis, Vienna, 2023.

[33] R. Sattari *et al.*, "Design, Fabrication, and Characterization of a 4H-SiC CMOS Readout Circuit for Monolithic Integration with SiC Sensors," in *2023 24th European Microelectronics and Packaging Conference & Exhibition (EMPC)*, IEEE, Sept. 2023.

[34] Cadence Design System, "Cadence Virtuoso." [Online; accessed on 28.08.2024]. URL: `https://www.cadence.com/en_US/home.html`.

[35] S. Ahmed *et al.*, "DC Modeling and Geometry Scaling of SiC Low-Voltage MOSFETs for Integrated Circuit Design," *IEEE Journal of Emerging and Selected Topics in Power Electronics*, vol. 7, pp. 1574–1583, Sept. 2019.

[36] A. May *et al.*, "A 4H-SiC CMOS technology enabling smart sensor integration and circuit operation above 500 °C." accepted for publication.

[37] P. G. A. Jespers and B. Murmann, *Systematic Design of Analog CMOS Circuits: Using Pre-Computed Lookup Tables*. Cambridge University Press, Sept. 2017.

[38] A. Ortiz-Conde *et al.*, "A review of recent MOSFET threshold voltage extraction methods," *Microelectronics Reliability*, vol. 42, pp. 583–596, Apr. 2002.

[39] S. Waid *et al.*, "Detector development for particle physics," *e+i Elektrotechnik und Informationstechnik*, vol. 141, pp. 20–28, Jan. 2024.

[40] G. Bertuccio and S. Caccia, "Noise Minimization of MOSFET Input Charge Amplifiers Based on $\Delta\mu$ and $\Delta N1/f$ Models," *IEEE Transactions on Nuclear Science*, vol. 56, pp. 1511–1520, June 2009.