Dissertation
submitted to the
Combined Faculties of the Natural Sciences and Mathematics
of the Ruperto-Carola-University of Heidelberg. Germany
for the degree of
Doctor of Natural Sciences

Put forward by
Konrad Briggl
born in Stuttgart
Oral examination: 10.1.2018

Silicon Photomultiplier Readout Electronics for Imaging Calorimetry Applications

Referees: Prof. Dr. Hans-Christian Schultz-Coulon
Prof. Dr. Norbert Herrmann

Abstract

Experiments at future linear collider experiments will allow to reach an unprecedented measurement resolution for standard model processes and the search for new physics. In order to exploit the full potential of the clean initial state in the electron-positron collisions, a jet energy resolution of 3-4% is required, which is not achievable with classical calorimetry approaches. The detectors will be optimized for the use of particle flow algorithms to achieve the required energy resolution, resulting in the need of highly granular (*imaging*) and compact calorimetry systems.
This work covers the development of specialized readout electronics for scintillator-based calorimeters read out by Silicon Photomultipliers. The readout electronics are required to provide a precise charge measurement capability over a large dynamic range, be fully integrated and self triggered. In order to allow a calibration of the calorimeters, the readout electronics must be capable of measuring the detector gain from the response to signals at the level of few photons. Noise contributions affecting the resolution for the gain calibration are discussed and used for the circuit optimization. Due to the high channel density, the power consumption of the front-end electronics is extremely limited, requiring to implement power pulsing techniques to minimize the power consumption.

Zusammenfassung

Experimente an zukünftigen Linearbeschleunigern werden eine nie dagewesene Auflösung für die Vermessung des Standardmodells und die Suche nach neuer Physik erreichen. Um die Möglichkeiten der Elektron-Positron Kollisionen mit wohldefiniertem Anfangszustand voll auszuschöpfen, müssen die Detektoren eine Jet-Energieauflösung von 3-4% erreichen können. Dies ist mit klassischen Ansätzen der Kalorimetrie nicht erreichbar, und die Detektoren werden für die Anwendung von *Particle Flow* Algorithmen optimiert. Dies setzt die Entwicklung hoch granularer, *bildgebender* Kalorimetriesysteme voraus.
Diese Arbeit beschäftigt sich mit der Entwicklung hochspezialisierter Ausleseelektronik für solch hoch granulare Kalorimeter, deren aktive Lagen mit Szintillatorkacheln bestückt sind und mit Silizium Photomultipliern ausgelesen werden. Die entwickelte Elektronik erlaubt die Ladungsmessung über einen weiten dynamischen Bereich und liefert ein voll integriertes Messsystem mit eigenständiger Triggerlogik. Zur Kalibrierung der Kalorimeter ist die Messung der Sensorverstärkung bei Signalen von wenigen Photonen notwendig. Rauschterme die diese Messung beeinflussen wurden untersucht um die Ausleseelektronik auch für die Messung kleinster Signale zu optimieren. Auf Grund der hohen Kanaldichte ist die Leistungsaufnahme für die Ausleseelektronik stark begrenzt. Um die diesbezüglichen Anforderungen einzuhalten wurden Funktionen implementiert, die die Leistungsaufnahme während der Zeit ohne zu vermessende Kollisionen minimieren.

Contents

1 Preface **1**

2 Calorimetry for future Linear Collider experiments **5**
2.1 Particle Flow Calorimetry . . . 8
2.2 The international linear collider . . . 11
2.3 The International Large Detector . . . 13
2.4 The Analog Hadronic Calorimeter concept . . . 15
2.5 Scintillator ECAL . . . 17
2.6 AHCAL Calibration . . . 17

3 Silicon Photomultipliers **19**
3.1 Avalanche photodiodes . . . 19
3.2 Silicon Photomultipliers . . . 20
3.3 Basic properties . . . 21
3.4 High dynamic range sensors . . . 24
3.5 Electrical model of Silicon Photomultipliers . . . 25

4 Introduction to Mixed signal ASIC design **31**
4.1 CMOS technology . . . 31
4.1.1 MOS transistors . . . 32
4.2 Analog signal processing and the Laplace transformation . . . 37

5 Noise contributions to SiPM single pixel charge measurements **43**
5.1 Electronic noise . . . 44
5.2 SiPM-intrinsic noise contributions . . . 46
5.3 Signal smearing from dark count rate pileup . . . 47
5.3.1 DCR pileup measurement setup . . . 49
5.3.2 Simulation . . . 49
5.3.3 Effective noise term . . . 49
5.4 Time measurement and the impact on voltage sampling . . . 52
5.4.1 Time jitter . . . 53
5.4.2 Distortions from timewalk effects . . . 55
5.5 ADC resolution . . . 57

6 KLauS: A low power SiPM charge readout ASIC **59**
6.1 Design considerations drawn from AHCAL requirements . . . 60
6.2 The KLauS4 channel . . . 61
6.3 Analog front-end . . . 62
6.3.1 Input stage . . . 62

6.3.2 Bias voltage tuning DAC . . . 65
6.3.3 Charge collection and noise filtering . . . 68
6.3.4 Noise analysis . . . 73
6.3.5 Comparator . . . 78
6.4 Analog to Digital Converter . . . 80
6.5 Channel control logic and mixed signal interfaces . . . 90
6.6 Digital part . . . 93
6.6.1 Event Data transmission . . . 94
6.6.2 Configuration interface . . . 96
6.6.3 On chip data reduction . . . 96
6.7 Power pulsing . . . 98
6.7.1 Analog Front-end . . . 98
6.7.2 ADC and digital part . . . 102
6.8 Design verification . . . 102

7 Characterization measurements 105
7.1 Test setup . . . 105
7.2 Analog front-end . . . 106
7.2.1 Input terminal DAC . . . 106
7.2.2 Dynamic range & charge noise . . . 107
7.2.3 Trigger comparator . . . 109
7.3 ADC performance . . . 112
7.4 Full-chain measurements . . . 115
7.5 Power pulsing . . . 119
7.6 Power consumption . . . 122
7.7 Test-beam . . . 122

8 Summary 127

A Description of additional sub-circuits 131

B Additional characterization measurements 137

C Lists 141
C.1 List of Figures . . . 141
C.2 List of Tables . . . 143

D Bibliography 145

1 Preface

Most modern particle physics detector are based on the measurement of particles produced in collisions of high-energetic particles. The standard model of particle physics describes the elementary particles and their interaction mediated by the exchange of vector bosons. With the discovery of the Higgs boson by the ATLAS and CMS detectors at LHC [1, 2] the standard model is complete and proves to be a very successful theoretical description of the observed fundamental particles and their interactions. Lacking an explanation of dark matter as well as a link to the gravitational force, it can however not describe all aspects of our universe. Precision measurements are performed to find differences to the standard model predictions which might guide the way to a more general theory.

The large hadron collider at CERN provides collisions with a center of mass energy of up to $\sqrt{s} = 13\,\text{TeV}$. While delivering highest collision energies to find new particles at the TeV scale, the possibilities of precision measurements tackling the standard model are limited due to the underlying QCD background generated in the collisions of protons. Lepton collider facilities may complement the searches for new physics performed at the LHC. The collision of leptons being fundamental particles provide a well-defined initial state with the full energy available in the collision, and clean underlying background from QED processes. This also allows to perform precision measurements of hypothetical particles not interacting with the detector, which leads to missing energy or momentum. A proposed lepton collider is the International Linear Collider (ILC), where electrons and positrons are collided with a center of mass energy of up to 500 GeV (1 TeV after a potential upgrade).

Prominent measurements to be conducted at such a collider would of course be the properties of the Higgs boson, for example it's couplings and total width. They are expected to be measurable with much better precision (or accessible at all) compared to the LHC experiments for most of the decay channels [3, 4].

In order to fully exploit the possibilities of the e^+e^--collisions, the detector are optimized to achieve the best momentum and energy resolution possible. Many processes involve jets coming from the hadronization of quarks in the final state. The detectors developed for the ILC aim to achieve a jet energy resolution of 3%-4% over a wide jet energy range, sufficient to distinguish the hadronic decay of W and Z bosons using solely the energy measurement informations. This translates into a required resolution of $30\%/\sqrt{E[\text{GeV}]}$ in the calorimeters, substantially better than any detector system developed before.

To reach the envisioned jet energy resolution, the detectors are optimized for so-called *Particle Flow Algorithms* (PFA). In PFAs, the jet energy measurement is obtained by measuring the energy of every single particle in the jet, in the subdetector which provides the best energy resolution for the given particle type. Typically, only neutral particles of the jet are measured in the calorimetry systems. For the ILC detectors, sampling calorimeters will be used for both the electromagnetic (ECAL) as well as the hadronic (HCAL) parts, with tungsten and steel used as the absorber material. The need to distinguish the differ-

ent particles from a single jet requires the calorimeters to have an unprecedented spacial resolution.

Within the CALICE (*CAlorimeter for a LInear Collider Experiment*) collaboration, different concepts for the calorimetry systems are being studied. Differing in the active material and detector technology placed between the absorber layers, the quantization resolution of the individual cells and the cell size, they all have an outstanding spacial granularity. One of the concepts studied for the hadronic calorimeter system is the Analog Hadronic Calorimeter (AHCAL), where scintillating tiles of $\approx 10\,\mathrm{cm}^2$ area are used as active cells. Each Scintillating tile is read out by a Silicon Photomultiplier, semiconducting devices which are sensitive to light levels as low as single photons, compact to allow the spacial segmentation, and insensitive to magnetic fields.

The readout electronics will be placed inside the active layers of the calorimeters. Application Specific Integrated Circuits (ASICs) need to be developed, providing a digitized charge and time information. This work describes the development of such custom readout electronics for the AHCAL as well as the Scintillator ECAL which is using scintillating strips read out by Silicon Photomultipliers, with even higher spacial granularity compared to the AHCAL. Based on the requirements on the measurement of charge signals from the SiPMs, covering the range of few photons for calibration purposes up to the response to 10-100 minimum-ionizing particles passing a single cell, the integrated readout electronics need to cope with an extremely low power budget driven by the high granularity and compactness of the calorimeters. As no external triggering scheme is foreseen for the ILC detectors, the readout electronics also need to be self triggered. Also the measurement of time stamps is required. The time information will allow to improve the energy resolution further, and may reduce the effect of limited jet energy resolution of high-energetic jets.

Structure of the thesis

The thesis is separated in 7 chapters, starting with the requirements on the front-end ASICs in terms of detector design, the use of Silicon Photomultipliers and it's capabilities allowing an intrinsic gain calibration. In Chapter 2, an introduction to calorimetry and detectors at the ILC will be given, focussing on the requirements of the calorimetry system to achieve the targeted jet energy resolution by the use of particle flow algorithms.

The AHCAL will be using Silicon Photomultipliers to measure the light generated by the scintillating tiles. An overview on the basic functionality of this type of sensor is given in Chapter 3, postulating basic requirements on the readout electronics developed for charge measurements.

For the calibration of the calorimeters consisting of millions of readout channels, the capability of the SiPMs resolving single photons will be used to extract the gain of each sensor. Resolving the single photon peaks in the measured charge spectra for small signals is therefore a key requirement of the readout electronics. Different noise sources contribute to the smearing of the spectra. They will be discussed in Chapter 5.

A main focus of this work is the development of the analog readout chain for the front-end ASIC. Chapter 4 will give an introduction in the development of mixed-signal ASICs, transistors in CMOS technologies and analog signal processing theory.

In Chapter 6, the ASIC developed in this work will be described. The ASIC consists of an analog front-end converting the charge signal to a voltage signal with well defined shape.

This allows to digitize the charge measurement using a integrated ADCs connected to each of the channels. The digital part combines the event data and provides the connection to the data acquisition system collecting the event data from the front-end ASICs. All of these blocks and their key characteristics will be discussed in this chapter.

The ASIC has been characterized in detailed laboratory measurements using single channels. The dependence of the measured charge and noise on the most important configuration settings and detector parameters were studied. In order to investigate the operation and stability of the chip in an environment closer to the final detector, a test setup consisting of several ASICs was build. All key aspects of the ASIC were evaluated in a measurement campaign at the DESY electron test-beam facility, from calibration of a multi-channel setup, to the measurement of the response of MIP-like particles passing the multi-layer setup. Results of the characterization measurements will be discussed in Chapter 7.

Contributions from the author

The development of a complex microchip consisting of analog, digital, and mixed-mode building blocks is a task usually performed by a group of designers. The KLauS ASIC is a development over several years, based on the development of a purely analog readout ASIC discussed in [5]. The front-end schematic and layout of the ASIC presented in this work, as well as the simulations of the front-end were carried out by the author. The development of the ADC was carried out by other members of the group. In the digital part, the author designed all blocks included in the single-channel block, such as the ADC control logic, interfaces between the front-end and digital part, as well as the interface to the following digital part combining the individual events. The blocks connected to the I^2C interface were also developed by the author. The physical implementation of the chip, such as floorplan, timing validation and final design checks were performed by other members of the group. The automatized digital verification environment allowing for an in-depth validation of the full ASIC functionality was also developed by the author. The development of the necessary PCBs to test the ASIC was shared between the electronics department of KIP and the author. The DAQ and slow control software, as well as the software framework used to perform the characterization measurements in an automatized fashion were written mostly by the author, starting with developments of a proceeding diploma thesis. The characterization measurements were shared between several group members and the author. Also the construction of the setup and measurements during the test beam were a joined effort between members of the detector development group and the author. The development of the new versions of the KLauS readout ASIC leading to this thesis was presented in several international conferences and published in proceeding papers [6–8], partially in peer reviewed journals.

2 Calorimetry for future Linear Collider experiments

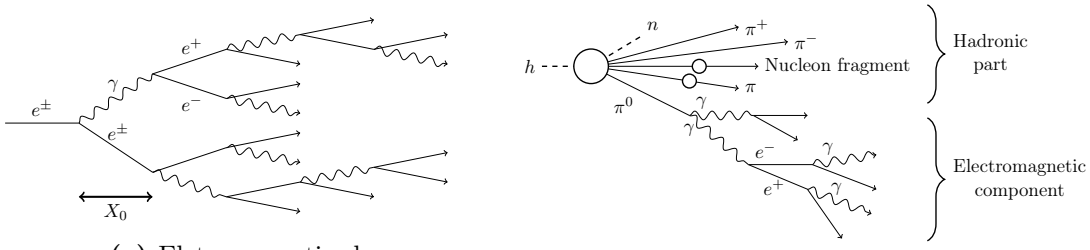
The measurement of particle and jet energies in elementary particle physics detectors is performed in calorimeters. By the interaction of the particles with matter, the kinetic energy is passed to an absorber material, ideally releasing all kinetic energy within the absorber. The energy measurement is therefore a destructive process. A fraction of the dissipated energy leads to the production of free charge carriers or scintillation light which is then measured to reconstruct the energy of the incident particle. For high incident particle energies, secondary particles are created in the absorbing material, carrying a fraction of the initial momentum. This leads to a cascade of secondary particles. Depending on the particle generating the cascade, one speaks of *electromagnetic* or *hadronic* showers. Once the energy of the secondary particles passes the production threshold energy for the generation of secondary particles, the shower stops as the particles loose all remaining energy by absorbing processes not sustaining the cascade.
The discussion on the development of showers will cover single particles entering the calorimetric volume. For jets consisting of many particles from a single initial quark, the energy deposited by several particles is summed up in a classical calorimeter. As it will be discussed in Section 2.1, future calorimeters will be able to separate the individual particles in the shower in order to improve the intrinsically limited energy resolution of the calorimeters.

Electromagnetic showers

At high energies, electrons and positrons interact with the absorber material predominantly by generation of bremsstrahlung photons with an average energy loss of

$$\left\langle \frac{\mathrm{d}E}{\mathrm{d}x} \right\rangle = -\frac{E}{X_0} \tag{2.1}$$

The characteristic length scale X_0 is called the radiation length, approximately proportional to A/Z^2 for large proton numbers Z of the absorbing material. For photons at energies above few MeV and high-Z materials, the dominant interaction process is the pair-production of electron-positron pairs in the vicinity of a nucleus. The mean free path for pair-production in the absorbing material is $\lambda_{pp} = 9/7X_0$. Based on these two processes, an electromagnetic shower forms as sketched in Figure 2.1a. This process is the same for high-energetic positrons. Excluding the first interaction, also the formation of an electromagnetic shower generated by a photon entering the calorimeter volume results in such an electromagnetic shower, however with a later shower starting point. At the critical energy $E_c \approx (\mathrm{d}E/\mathrm{d}V) \cdot X_0 \approx 610\,\mathrm{MeV}/(Z + 1.24)$, [9], the secondary particles start to loose most of their energy by processes such as direct ionization, annihilation for positrons and coulomb-scattering for photons. No additional secondary particles are generated, and the



(a) Eletromagnetic shower

(b) Hadronic shower including EM component

Figure 2.1: Formation of showers in the calorimeter

cascade is stopped. While the electrons and positrons are captured within one radiation length, the photons can travel longer distances. In order to contain the full shower, the calorimeters should have a thickness in the order of 20-30 X_0.

The lateral spreading of the shower is governed by multiple scattering of the charge particles. The size of the shower is characterized by the Moliere radius, containing 90% of the total deposited energy:

$$R_m = m_e c^2 \sqrt{\frac{4\pi}{\alpha}} \frac{X_0}{E_c} \tag{2.2}$$

Hadron showers

For hadrons, strong interactions with the absorber nuclei govern the shower development sketched in Figure 2.1b. The initial energy is carried by secondary particles generated in nuclear reactions, leading to breakup, spallation and excitation of the nuclei and the generation of pions of any charge. The characteristic length scale of the hadronic interaction λ_{nuc} depends on the absorber material (nucleon number) as well as the particle type. It is much longer than the radiation length X_0 for most absorber materials. In order to contain the hadronic shower, a total nuclear length of about $10 \cdot \lambda_{nuc}$ is required, significantly more material than for the containment of electromagnetic showers. With 30-40%, a significant amount of energy is transferred to binding energy of the generated lighter nuclei and is therefore invisible in the energy measurement. The response of a hadronic shower measured in the calorimeter is therefore lower than for electromagnetic showers. The response ratio depends on the particle type and energy. The prompt decay of neutral pions, $\pi^0 \to \gamma\gamma$, leads to a electromagnetic shower component within the hadronic shower. On average, the energy fraction of the electromagnetic component can be parameterized as

$$f_{em} = 1 - \left(\frac{E}{E_0}\right)^{(k-1)} \tag{2.3}$$

where E_0 is the pion production threshold energy, $\approx 0.7\,\mathrm{GeV}$ for iron, and $k \approx 0.8$ is a absorber-dependent parameter related to the π^0 production probability. From shower to shower, significant fluctuations of the electromagnetic shower components are observed. Due to the different response of the individual shower components, the total response to a hadronic shower is showing nonlinearities with respect to the particle energy. For the

same reason, the statistical fluctuations of the electromagnetic shower fraction introduce fluctuations of the measured signal. In order to reduce the effect on the linearity and energy resolution, the detector response must be *compensated*, equalizing the response in the relevant energy range. This can be accomplished by enhancing the hadronic response in the detector, for example by enhancing the neutron and γ emmision by the use of uranium as the absorber material. Similarly, the electromagnetic response can be reduced by shielding soft photons. As it will be discussed in the following chapters, the high granularity of future calorimeter systems will allow to perform the compensation offline on the bases on software algorithms. In this case, the different length scales of the hadronic and electromagnetic showers are exploited.

Sampling calorimeters

In order to measure the energy deposited in the shower, an active material is required. In homogeneous calorimeters, the absorbing material also acts as the active medium generating the measured signal, for example scintillation light produced in dense scintillating crystals. With such detectors, a high energy resolution can be acheived since all energy depositions may be sensed. Especially for hadronic calorimeters, this method is not feasible due to the large amount of material required. Sampling calorimeters use a segmented structure of dense but insensitive absorber material, interleaved with active layers providing a measurable signal. Figure 2.2 shows the sketch such of a sampling calorimeter. A fraction of the energy depositions in the shower are *sampled* in the active layers. This setup is much cheaper than homogeneous calorimeters since the absorber can be made from plates of relatively cheap material such as steel, copper or thungsten. For the active layers, for example using plastic scintillators or gas detectors are used. A drawback of the sampling calorimeter is the small fraction of energy deposits visible by the active material, which results in a lower energy resolution due to fluctuations of the sampled energy.

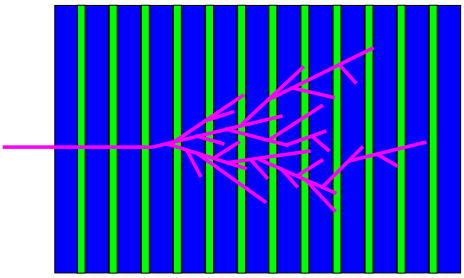


Figure 2.2: Structure of a sampling calorimeter [10]. The absorber is sketched in blue colour, the active layers are shown in green. The particle shower sampled by the calorimeter is indicated by the purple lines.

Energy resolution

The relative energy resolution of the sampled energy deposits can be decomposed in three terms added in quadrature:

$$\frac{\sigma_E}{E} = \frac{a}{\sqrt{E}} \oplus b \oplus \frac{c}{E} \tag{2.4}$$

The first term is called the stochastic term and is given by the statistical fluctuations of the sampled energy. It dominates the energy resolution at small energies. The constant term b is given by the energy losses due to leakage of the shower, calibration errors, and other detector effects such as the response differences between hadronic and electromagnetic shower components. The last term is a noise term from the activated absorber material and noise in the sensors and readout electronics. Due to the effects listed above, the hadronic energy resolution is intrinsically limited by the fluctuating visible energy resolution and further degraded if the calorimeter is not compensated.
Typically, the stochastic term of a hadronic calorimeter is in the order of $\approx 60\%/\sqrt{E\,[\mathrm{GeV}]}$.

2.1 Particle Flow Calorimetry

In classical high energy physics detectors, the energy of a jet is determined by the energy deposits of the full jet measured in the calorimeters. The energy measurement is therefore governed by the intrinsically poor energy resolution of the hadronic calorimeter. In an average jet, 30% of the total energy is carried by photons, and only 10% of the energy is carried by neutral hadrons. Both detectors planned for the international linear collider are optimized for the use of *Particle Flow* algorithms [11] for the event reconstruction. In general, only the neutral particles leaving no signal in the tracking detectors need to be measured in the calorimeters. The energy of all other charged particles may be obtained from momentum measurements in the tracker, which have a better energy resolution than the calorimeters especially at low energies. The idea of the particle flow calorimetry aims to combine the informations from tracking and calorimetry systems such that the detector with best resolution is used for each single particle of a jet. The energy of charged particles are obtained from the tracking detectors, and only energy deposits of neutral hadrons and photons are measured with the hadronic and electromagnetic calorimeters, respectively. Figure 2.3a shows the transition from classical calorimetry to the particle flow approach. The total reconstructed energy is determined from the sum of all particle energies taken from the subdetectors. In order to give correct results, the calorimeters must be capable to identify and separate energy deposits of neutral hadrons from the other energy deposits which are already accounted for using the tracking information. Thus, the capability to decompose the individual particle tracks and showers are key elements of detectors optimized for particle-flow algorithms. These are driving the requirements on all subdetectors, in particular the calorimetry systems which need to have a high spacial granularity. Figure 2.3b shows different sources of errors clustering the energy deposits in the calorimeters and connecting them to tracks. In order to minimize the energy measurement errors due to misidentification, precise knowledge of the structure of hadronic showers is required. Also the layout and material of alternating absorber and active layers forming the calorimeter structure must be optimized to provide a good separation power between the showers.

In addition to the calorimeter performance, an excellent tracking performance is needed

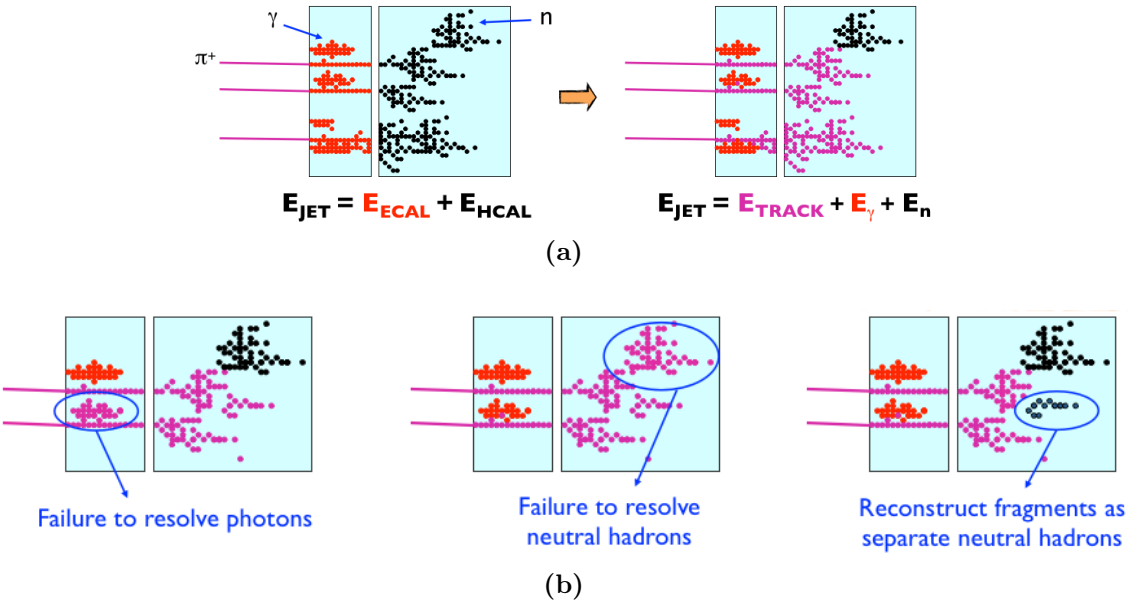


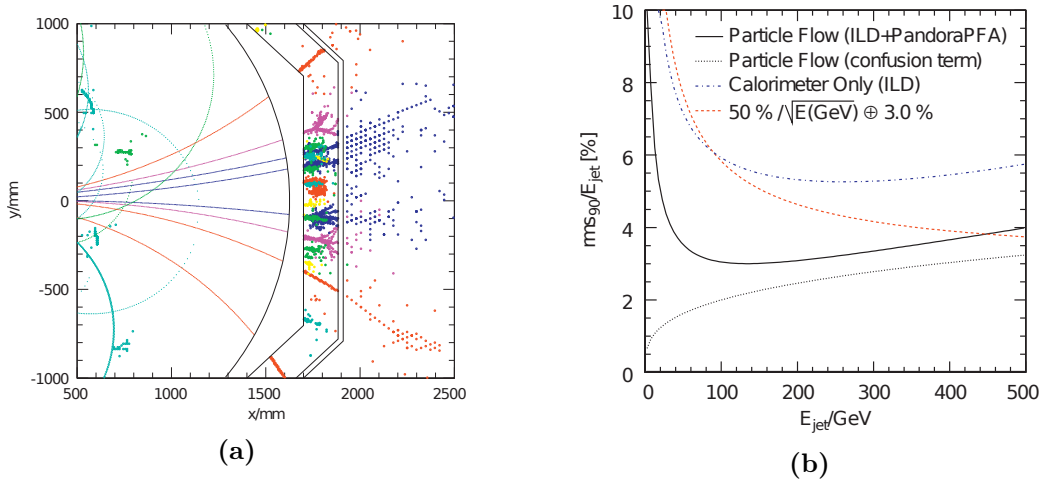
Figure 2.3: Particle flow calorimetry a) Transition from classical calorimetry to particle flow. b) Sources of confusion due to wrong pattern recognition. Figures taken from [12].

for the momentum measurement, and to be able to separate tracks and connect them to clusters of energy deposits in the calorimeter systems. Figure 2.4a shows an event view for a 100 GeV jet after event reconstruction and clustering into particle flow objects, using a realistic simulation model of a detector optimized in this way.

The jet energy resolution is determined by the resolution of the tracking system, ECAL and HCAL, weighted by the energy fraction of charged particles, photons and neutral hadrons:

$$\sigma_j = f_c \cdot \sigma_{tr} \oplus f_\gamma \cdot \sigma_{ECAL} \oplus f_{h0} \cdot \sigma_{HCAL} \oplus \sigma_{conf} \tag{2.5}$$

Only about 10% of the total jet energy is measured by the hadronic calorimeters. With a typical single-particle resolution of the hadronic calorimeter of $\sigma_E/E = 0.55\sqrt{E(\mathrm{GeV})}$, even the small fraction of neutral hadrons dominates the jet energy resolution [11]. Therefore, the calorimeters are still required to achieve a high single-particle resolution. The last term in (2.5) is called the confusion term. With increasing jet energies, the separation of the particles decreases. This worsens the discrimination power of deposits from neutral particles in the jet as the showers start to overlap, also due to the slightly increasing transversal dimensions. In the confusion term, errors in the assignment of energy depositions to neutral particles in the PFA shower clustering are accounted for. In Figure 2.4b, the relative jet energy resolution and it's contributions are shown in a parameterized form, obtained with the Pandora PFA. Using the particle flow approach, the jet energy resolution of 3%-4% envisioned for future linear collider experiments is achieved for a wide range of jet energies. Compared to the classical calorimetry approach, the algorithms always improve the resolution.



(a) **(b)**

Figure 2.4: a) Simulation of a 100GeV jet in the ILD detector after reconstruction with a particle-flow algorithm. b)Contributions to the total jet energy resolution using particle-flow algorithms. Figures taken from [13].

Software compensation

As discussed in Section 2, hadronic showers contain fluctuating electromagnetic components. The calorimeter response is different for these subshowers, deteriorating the single-particle resolution and linearity. In intrinsically non-compensating, but highly granular calorimeters optimized for particle flow, compensation can also be achieved in software by reweighing the measured visible energy depositions in the calorimeter cells. Thanks to the high granularity, electromagnetic shower components with a higher density of visible energy can be discriminated from shower components from hadronic interactions. Both the local hit energy density for each calorimeter cell as well as the energy distribution for the whole shower can be used as input data to calculate a weight, shown to yield improvements of the in the order of 20-25% [14].

Time information in calorimetry

Apart from using the time information to distinguish deposits from different bunch crossings, it may also be utilized to improve the jet energy resolution. In the T3B[1] experiments, the time development of hadronic and electromagnetic showers was studied [15]. Electromagnetic showers are dominated by prompt processes due to ionization from relativistic particles and the electromagnetic cascade. On the other hand, hadronic showers show a more complex time structure due to the additional processes involved. The time spectrum shows a prompt component ($<$1 ns) coming from direct ionization of charged hadrons and electromagnetic components. For hydrogen-rich active materials such as plastic scintillators, another relatively fast component with a time constant of about $7-8$ ns is observed and coming from neutron-proton scattering in the active medium [15, 16]. A slow component from spallation processes and capture of slow neutrons is depending on the absorber

[1] T3B: *Thungsten timing test beam experiment*, also carried out using steel as absorber material.

material, between $\approx 80\,\text{ns}$ for steel and $\approx 500\,\text{ns}$ for thungsten.

The electromagnetic cascades in the hadronic showers are governed by the prompt time constant. Thus, the time information may provide another handle to perform software compensation of the calorimeter response besides the energy density information. Simulations by Benaglia et. al. [17] performed for an idealistic sampling calorimeter have shown that time-based compensation allows to reduce both the stochastic and especially the constant term in Equation (2.4) significantly, comparable to the compensation-power of a dual-readout calorimeter.

The lateral size of the hadronic shower is not driven by multiple scattering as in the case of electromagnetic showers. In hadronic cascades, the transverse width is increased by large momentum transfers in nuclear interactions [9]. These have a longer time constant as they are transported by relatively low-energetic neutrons which travel further from the shower axis [15]. Thus, the time delay of the hit correlates with the distance from the shower core, and the average shower radius is reduced by limiting the integration window. Potentially, this may improve the shower decomposition in the particle flow algorithms, reducing the confusion term. Studies on this matter are ongoing.

2.2 The international linear collider

Hadron colliders such as the LHC can reach energies allowing a large variety of searches of physics bejond the standard model. The precision of measurements with hadrons is however limited. The precise initial state of the collision is unknown since the momentum of the proton is shared among the constituents. Also the large QCD background limits the measurement sensitivity.

Lepton colliders provide a well defined initial state including the spin of the colliding particles, since they are elementary particles. Due to the low mass of electrons, it is however more challenging to achieve high center-of-mass energies. In the predecessor of the LHC residing in the same circular tunnel, the Large Electron Positron Collider, the center of mass energy was limited to 209 GeV by energy losses due to synchrotron radiation emitted by the electrons on their circular track.

The International Linear Collider (ILC) is a proposed electron-positron collider with a center of mass energy of up to 500 GeV and a potential upgrade to 1 TeV in a second stage. For the planned site of construction, an area in the northeast region of Japans main island has been selected [18]. Figure 2.5 shows a schematic layout of the accelerator complex highlighting the key subsystems. A detailed description of the accelerator base-line design can be found in [19].

The electrons are generated by a laser-induced photo-cathode DC particle gun, where the polarized beam is brought into the bunch train structure of about 1ms duration and 5Hz repetition time. After pre-acceleration to 5 GeV, the electron beam is injected into a damping ring where the beam-emittance is reduced to achieve the required luminosity. The electrons are extracted from the damping rings and led into a transport tunnel of $\approx 13\,\text{km}$ length. After a 180° turn and passing several stages of collimation, bunch compression and spin rotation systems, the beam is injected into the main linear accelerator (LINAC).

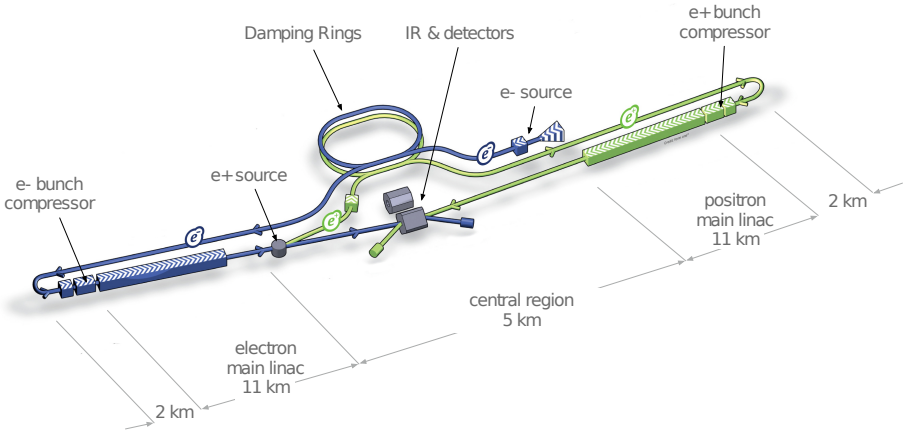


Figure 2.5: The international linear collider [20].

In the main LINAC consists of about 850 *cryomodules* operating at 2 K. The acceleration field is generated in superconducting niobium cavities operating at a frequency of 1.3 GHz. The average gradient in the approximately 1m long cavities is 31.5 MV/m. A very similar type of acceleration modules has been implemented at the *European X-ray free electron laser (XFEL)* facility based at DESY, Hamburg, where electrons are accelerated to 17.5GeV in a 3.4km tunnel. Figure 2.6 shows a photograph of the acceleration tunnel of the XFEL facility, giving a realistic impression of the main accelerator tunnel as it would be realized at the ILC.

The positrons are produced using the accelerated electron beam, passing a helical undulator to generate synchrotron radiation. The photons are led to a cooled metal target consisting of a thin rotating titanium disc, where electron-positron pairs are generated. The electrons and positrons are captured and pre-accelerated to 125MeV. Electrons, photons and positrons with largely differing energy are extracted and led to a beam dump.
Before injection into the positron damping ring, the positrons are accelerated to 5 GeV using a booster linear accelerator. The positron beam then follows the equivalent path as the electrons, a damping ring, long transfer line, and main linear acceleration stages.

The electron and positron beams are brought to collision in a single interaction point, where the beams are focussed to reach the required luminosity. For center of mass energies between 250 GeV and 500 GeV, the aimed delivered luminosity is between $0.56 \cdot 10^{34}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$ and $1.8 \cdot 10^{34}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$ [19]. Within the bunch train, the collision period in the baseline parameter-set is 554 ns. The bunch train structure, namely the bunch train repetition rate of 5Hz and the pulse length of 726 μs, is limited by the maximum pulse length allowed by the klystrons generating the RF power for the acceleration cavities. Other limitations come from the damping time of the pre-accelerated electron and positron beams in the damping rings, as well as the load on the cryogenic infrastructure. Two detectors will be sharing the common interaction point, using a push-pull approach to switch the position. While one detector is taking data, the other is parked in a maintenance position.

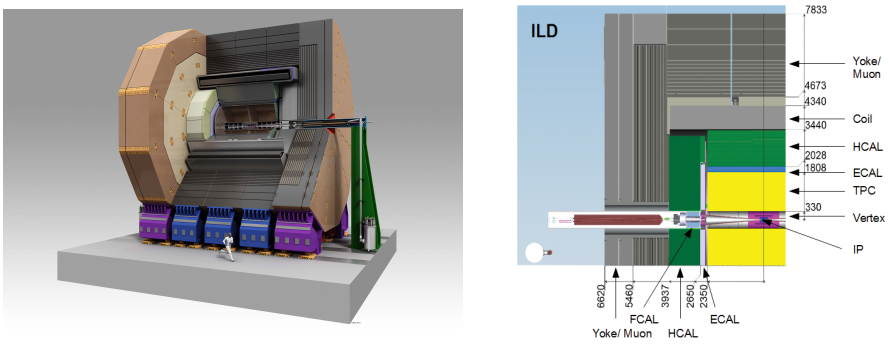
Figure 2.6: View of the acceleration tunnel at the European XFEL facility [21].

Detectors at the International Linear Collider

Two detector concepts have been selected for the ILC: The *Silicon Detector* (SiD) and the *International Large Detector* (ILD). Both detectors are designed as general-purpose experiments, allowing a wide variety of high-precision measurements. Also, both detectors are optimized for the use of event reconstruction algorithms based on the particle-flow idea. SiD is a compact design, with a tracking system entirely built from silicon detectors. The ILD is a slightly larger detector, using a hybrid tracking system consisting of silicon detectors and a time projection chamber. In both experiments, the tracking as well as the calorimetry systems are contained in the solenoidal magnetic field (5 T for SiD, 3.5 T for the ILD). The return-yoke is instrumented with position sensitive detectors to serve as a muon detector and tail catcher for the hadronic calorimeter.

2.3 The International Large Detector

In the following, an overview of the individual subdetectors of the ILD is given. Figure 2.7 gives an overview of the detector subsystems at ILD. To exploit the potential of the clean events with known initial state delivered by the ILC accelerator, the ILD is developed with emphasis on the high efficiency of a full event reconstructions following the needs of the particle flow algorithms. Thus, every charged particle should be reconstructed in the tracking system and matched to the showers in the calorimeters, requiring a good spacial resolution in all subdetectors. A high tracking efficiency including the matching to calorimeter clusters is achieved by redundant measurements of the trajectories. As described in Section 2.1, a electromagnetic and hadronic calorimeter systems with an unprecedented granularity are required to separate the individual showers from the particles in a jet. In order to allow matching of tracks with the calorimeter clusters, the non-sensitive material between tracking and calorimetry systems must be minimized. Therefore, most of the subdetectors of ILD are placed within the 3.5 T solenoidal magnetic field, and the muon system is placed inside the return yoke of the superconducting magnet.



(a) (b)

Figure 2.7: The ILD detector [20].

Tracking and Vertex detectors

The vertex detector, the innermost part of the tracking system, consists of 5 or 6 layers of pixel sensors arranged in a barrel geometry around the interaction point. The innermost layer is placed at a radius of $15 - 16$mm from the interaction point to provide an excellent resolution for secondary vertices in order to resolve decays from D and B mesons[3] and allow flavour-tagging of jets from c or b quarks[22]. In order to minimize the effect of multiple scattering on the pointing resolution, the pixel sensors are thinned to a thickness of $\lesssim 50\,\mu$m. Also the power consumption is reduced such that cooling by air flow only is sufficient [22].
The main tracking system of ILD is relying on a large number of measurement points along the particle trajectory in order to achieve high momentum resolution and reconstruction efficiencies. The core part of the tracking system is a time projection chamber (TPC) continuously active throughout the length of one bunch train. For the amplification of the primary ionization electrons, either micromesh-gas detectors or GEM foils will be used, the readout will be implemented by pads or CMOS pixel sensors [3]. To improve momentum resolution and matching to the other subdetectors, the TPC is surrounded by silicon strip sensors in both the barrel and endcap region, as well as two double-layers of silicon strip detectors between the vertex detector and the TPC.

Electromagnetic calorimeter concepts

The electromagnetic calorimeter is a sampling calorimeter of 30 layers. Tungsten is used as the absorber material to reduce the thickness, and therefore the required size and cost of the magnet. As the active material, two different concepts are studied. Silicon sensors with a pixel size of $5 \cdot 5\,\mathrm{mm}^2$ read-out individually by integrated readout electronics are one option (SiECAL). An alternative solution resulting in a lower number of channels is the use of scintillating strips strips arranged orthogonally in adjacent layers, read out by Silicon Photomultipliers. The so called ScECAL option can provide similar spacial information as the SiECAL solution at the cost of more complicated reconstruction in dense showers [22].

Hadronic calorimeter concepts

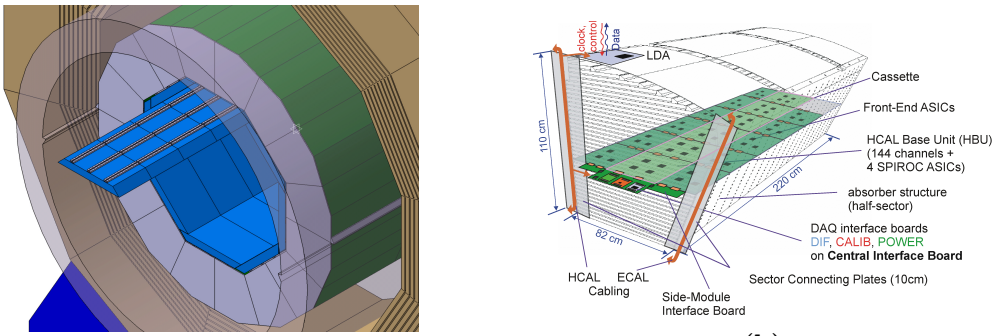
Essentially two different concepts are studied for the hadronic calorimeter system in ILD. In the digital hadronic calorimeter (DHCAL), small cells with a lateral segmentation of $1 \cdot 1\,\mathrm{cm}^2$ are used, with resistive plate chamber detectors giving a digital information. Semi-digital approaches using two or three different thresholds are also studied (SDHCAL concept). The analog hadronic calorimeter concept (AHCAL) is the main application for the ASIC developed in this thesis. It uses active cells with coarser segmentation built from scintillating tiles. The scintillation light is read out using Silicon Photomultipliers, giving an analog information of the energy deposited in the cell, allowing to relax the segmentation. Both scintillator-based calorimeter concepts (ScECAL and AHCAL) will be covered in Chapter 2.4.

Muon system

The superconducting coil generating the solenoidal magnetic field inside the detector volume is surrounded by an iron yoke to contain the magnetic field, providing a return path for the magnetic flux. The yoke is segmented in by iron-plates of 10 cm, and instrumented with position-sensitive detectors in order to serve as a muon detector. Also, it allows to achieve a better containment of hadronic showers, acting as a coarsely segmented tail catcher of the hadronic calorimeter. Detector simulations [3] as well as Measurements of prototype modules [23] have shown that the additional instrumented nuclear interaction lengths may significantly improve the single-particle energy resolution at higher energies.

2.4 The Analog Hadronic Calorimeter concept

The Analog hadronic calorimeter (AHCAL) concept is a sampling calorimeter using steel as the absorber material. Plastic scintillating tiles read out by Silicon Photomultipliers are used as the active medium. Since the light signal providing an analog information of the visible energy in the cell, the segmentation can be coarse compared to a digital readout solution. Scintillator tiles with a dimension of $30 \cdot 30 \cdot 3\,\mathrm{mm}^3$ are chosen as a compromise between channel count and particle flow performance [3, 13]. Overall, the hadronic calorimeter will still consist of about 8 million channels.
The barrel part foreseen for the AHCAL consists of 32 modules, segmented in two parts for the z-direction and eight octants in ϕ [3]. Figure 2.8b shows a drawing of one half of the AHCAL octant modules inside the ILD detector calorimetry system (Figure 2.8a). The structure consists of 48 absorber plates made from 16 mm stainless steel, interleaved with the active readout layers. The active layers include the scintillator tiles of 3mm thickness, photo-sensors, the readout ASICs and printed circuit boards. The thickness of the active layer is limited to 5.4mm, not including the cassettes made from $\approx 3\,\mathrm{mm}$ stainless steel plates. A single readout board (*HBU - HCAL Base Unit*) includes 144 scintillating tiles and the Photosensors, read out by 4 readout ASICs with 36 channels each. The digitized data, control signals and power connections are generated by dedicated PCBs at the side of the module, and distributed through a chain of 6 HBUs in the layer. In order to reduce the lateral shower size, the compactness inside the HCAL volume should be as high as possible. A cooling system for the readout ASICS should therefore be avoided. Only the components



(a) (b)

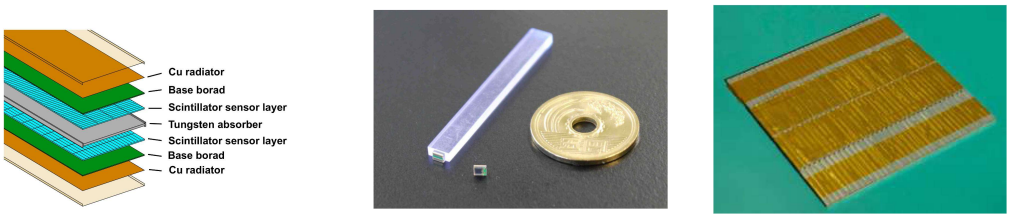
Figure 2.8: a) The ILD calorimetry system consisting of ECAL (blue) and HCAL (green). b) One half of the half-octant AHCAL module. Figures taken from [3].

on the auxiliary boards, in particular the FPGAs for used for data concentration and the voltage regulators for the ASICs, are actively cooled with a water cooling system. This puts stringent requirements on the power consumption of the front-end ASICs, limited to 25 μW per channel. This is only feasible by taking advantage of the ILC bunch train structure. During 99.5% of the time, no collisions are expected and most of the circuits in the readout ASICs can be switched off. The long time window also allows to transfer the data slowly out of the calorimeter, without the need of an external trigger system.

In order to validate the detector concept, a 1m^3 "Physics Prototype" with 7808 channels was built and operated in combination with modules of the ECAL and tail-catcher in various test-beam campaigns between 2006 and 2012 [24]. With the technical prototype, the capabilities of the scintillator-based calorimeter in terms of energy resolution [25], [14] could be demonstrated. By the use of software compensation techniques taking the energy density in the clusters into account, a resolution of $\sigma_E/E = 44.3\%/\sqrt{E[\text{GeV}]} \oplus 1.8\%$ [14] to single $\pi^{\pm}$ showers was achieved. Also the imaging capabilities required for particle flow have been studied [26], all showing that the AHCAL concept is not only capable of reaching the ambitious goals for the ILD detector, but also that stable conditions and calibration are feasible.

The physics prototype can however not be scaled to a full ILD design due to the missing integration level and complex production of the individual components. Current efforts in the CALICE AHCAL community concentrate on the preparation of a "Technical prototype" in order to show the feasibility of the AHCAL design proposed for the ILD. For this purpose, scintillating tile designs being easier to produce in large quantities, ASICs with the required power consumption and functionality, as well as the DAQ are being developed. Many aspects of the AHCAL design have been revisited in this process, simplifying production steps and implementation such that standard industrial solutions can be applied. As examples, the use of SMD[2] packaged SiPMs, air-coupled injection moulded scintillators tiles without the use of wavelength-shifting fibers and standard BGA[3] packages for the readout ASICs greatly simplify the production process and reduce the cost.

[2]Surface Mountable Device
[3]Ball Grid Array



(a) (b) (c)

Figure 2.9: Scintillator-based electromagnetic calorimeter option[29, 30]. a) ECAL active layer segmentation. b) Scintillating strip of $45 \cdot 5 \cdot 2\,\mathrm{mm}^3$. c) Technical prototype layer equipped with 144 scintillating strips and SiPMs.

2.5 Scintillator ECAL

Also for the Scintillator-based ECAL a physics prototype was built, achieving an energy resolution of $\sigma_E/E \approx 13.8\%/\sqrt{E[\mathrm{GeV}]} \oplus 3.5\%$ [27]. A technical prototype using the same readout hardware as the AHCAL is being built, however with a much higher segmentation. While for the physics prototype a segmentation of 10 mm was used, the aimed lateral segmentation for the technological prototype is 5 mm in order to achieve a better particle flow performance. In order to reduce the lateral shower radius, also the thickness of the active layers need to be reduced in the technological prototype. Figure 2.9a shows a sketch of the Scintillator ECAL. In total 30 active layers will be used, with a total absorber material of 24 radiation lengths. The scintillator strips are aligned orthogonally in neighboring layers, and the 2D information is extracted by special reconstruction algorithms [28]. The scintillator strips, shown in Figure 2.9b, are read out by Silicon Photomultipliers at the side. Currently, different scintillator geometries are under study to minimize the dead area while keeping a good uniformity along the strip and light yield [29]. Figure 2.9c shows a module of the ScECAL technical prototype. The EBU[4] module has a size of $18 \cdot 18\,\mathrm{cm}^2$ and consists of 144 scintillating strips individually wrapped in reflective foil. Due to the higher visible energy density in the ECAL, the sensors are required to have a high dynamic range in order to reduce the sensitivity to the nonlinear behaviour of the Silicon Photomultipliers.

2.6 AHCAL Calibration

The calibration of a calorimeter consisting of several million channels will be a difficult task. Given that the calorimeter can not be calibrated fully in beam-tests before installation, methods to achieve the necessary precision and long-term stability have been studied and applied in the AHCAL physics prototype. With some simplifications, the discussion of the calibration steps here follows the strategy used in the AHCAL physics prototype [10].
The relation between the incident particle energy and the AHCAL response is given by the sum of all measured signals in the shower,

$$E = w \cdot \sum_i E_i \tag{2.6}$$

[4] ECAL base unit

where E_i are the measured signals in each cell in units of a known reference signal. The scaling factor w relates the signal sum to a particle energy and can be determined from test-beam measurements with known energy [31]. The weighting factor should be independent on the particle type, energy and detector environment such as the temperature. Additional weighing factors from software compensation are not considered in this discussion. A common unit of measure for the cell signals E_i is the response to minimum-ionizing particles (MIP) such as muons traversing the calorimeter without generating a shower. The data measured by the readout ASICs is a charge signal digitized by an analog-to-digital converter, including an offset value from zero. The cell energy in units of MIP responses is then given by

$$E_i = (S_i - P_i) \cdot w_i \cdot f^{-1} \tag{2.7}$$

where S_i is the digitized charge signal, P_i is the offset (pedestal) determined in dedicated noise measurement runs and w_i is a reweighing constant for each channel. The inverse function f^{-1} is used to correct for the nonlinear behaviour of the Silicon Photomultiplier response. The function f depends on the number of scintillation photons produced, which requires the knowledge of the sensor gain. As it will be discussed in Chapter 3, Silicon Photomultipliers allow to extract the gain by the use of small light pulses, either obtained in dedicated calibration runs, or in-situ using the charge signals of cell hits with a small amplitude. At least two calibration constants are required for each channel.

The required precision of the calibration constants has been studied by the AHCAL group [31]. After an initial calibration using hadronic Z-decays obtained within three weeks of collisions at the Z-resonance, the calibration precision can be maintained by monitoring the "Environmental" aspects such as the temperature and bias voltages, but also the SiPM gain for each channel and the MIP response for groups of channels. For the MIP-scale calibration constants w_i, track segments within the hadronic showers as well as muons may be used to provide sufficient statistics in order to allow monitoring on the bases of submodules. While correlated variations over all detector cells (e.g. gain change over time) introduce a significant change of the calorimeter response, constant cell-to-cell variations introduce negligible changes in the jet energy resolution. Thus, the (ideally in-situ) monitoring of the sensor gain and the MIP scale provide sufficient precision to calibrate the AHCAL. While not studied for the Scintillator ECAL in the same detail, the required calibration precision and possibilities will be similar for this subdetector.

3 Silicon Photomultipliers

In experimental high energy physics, photo-detectors are mostly used to measure the light from cherenkov radiation, as well as scintillation light.

Especially for cherenkov photons, but also for scintillation light in detectors with a reduced cell size due to the increasing demands on the granularity, the amount of light produced is small. This requires sensors that are sensitive to the level of single photons. Semiconductor devices without an intrinsic gain, such as PIN diodes, can not achieve this. Photomultiplier tubes (PMT) have been used for decades for this purpose. However, the demand for more compact and highly granular detector systems does not allow to use such voluminous devices. They are also hardly operable in strong magnetic fields present in the detectors. Silicon Photomultipliers are solid-state detectors with a functionality similar to ordinary PMTs. With the maturation of this detector type in the last decade, strongly driven by their application in medical devices, they also reach the performance of PMTs in most aspects. Their compactness, lower cost, speed and insensitivity to magnetic fields make Silicon Photomultipliers a popular choice for applications measuring small light signals.
In the following, an introduction into the basic working principle will be given. In the last section of this chapter, an electrical model of the sensor will be discussed. Focussing on the application of energy measurements, basic requirements for the readout electronics are derived.

3.1 Avalanche photodiodes

The possibility measuring single-photon light signals is offered by avalanche pixel diodes (APD), providing an intrinsic amplification method based on the avalanche generation from a primary electron-hole pair. Figure 3.1 shows a sketch of the doping structure in an APD. By implanting a highly p+ doped volume underneath the n+ region, a region with high

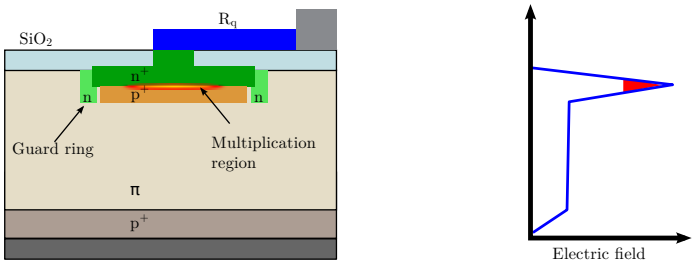


Figure 3.1: Sketch of the typical APD doping profile and electric field. Based on a Figure in [32].

electric field is formed when a reverse bias voltage is applied. In this high field region, secondary electron-hole pairs can be generated by impact ionization, yielding an amplified signal. To prevent undesired high field regions in the border area of the APD, guard ring structures are introduced. For moderate bias voltages, the field is high enough to generate free charge carriers by the impact ionization from electrons, but not holes. Thus, the avalanche generates a charge signal proportional to the charge produced in the primary photon-substrate interactions. This is then called the linear region of the APD, where a gain in the order of 10^3 is reached. In this operation mode however, the diode exhibits a large response to particles traversing the sensor and generating electron-hole pairs which are amplified, potentially resulting in a signal much larger than the signal generated by the scintillation light. This unwanted effect is eliminated by applying a higher bias voltage above the breakdown voltage. In this case, the field is sufficiently strong such that also the movement of holes generate secondary free charge carriers by impact ionization. This operation mode is called the Geiger operation region of the diode. The avalanche is self sustaining, and a constant current is flowing through the diode. In order to stop the avalanche, a *quenching* mechanism is required. In commercial sensors, a passive quenching mechanism is used, adding a sufficiently large resistance in front of the diode. After an avalanche has been triggered, the current through the avalanche is mainly sourced by the charge stored on the diode junction capacitance, and the voltage at the diode reduces exponentially. When the current reaches a certain threshold current I_q where the statistical recombination processes would outnumber the reduced impact ionization processes in particular for holes, the avalanche is quenched. The threshold current depends on the specific doping structure and is in the order of $\lesssim 20\,\mu$A [33]. After the quenching, the diode is then recharged to the bias voltage through the quenching resistor. For sufficiently large quenching resistors, typically in the order of $100 - 2000\,\mathrm{k}\Omega$, the current through the resistor is small compared to I_q and the quenching time is relatively well defined. Thus, the charge generated in one single avalanche is given by the product of the *overvoltage* and the total capacitance of the diode:

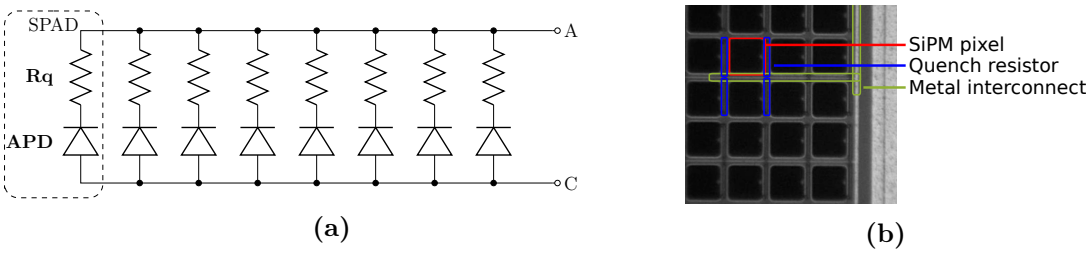
$$Q_{av} = G \cdot e = (V_{bias} - V_{br}) \cdot C_d = V_{over} \cdot C_d \tag{3.1}$$

where V_{bias} is the reverse bias voltage applied to the diode and V_{br} is the breakdown voltage, defined as the bias voltage where gain drops to 1. Typical sensors have a breakdown voltage in the range of 20 V-100 V. The diode then acts as a single photon "counting" device, however with very limited dynamic range (0 or 1). This device is therefore also called a *single photon avalance diode* (SPAD). The single photon gain G of the SPAD is in the order of $10^5 - 10^6$ for typical overvoltages of $3 - 6\,\mathrm{V}$ and depends on the area of the diode. It is considerably larger than the gain achieved with APDs in the linear operation mode due to the Geiger operation.

3.2 Silicon Photomultipliers

The limited dynamic range of the single SPAD is solved by Silicon Photomultipliers (SiPM)[1]. In order to achieve a dynamic range for multiple photons, a SiPM consists of an array of SPADs on the same sensor. The individual SPADs will be referred to as a (SiPM) pixel in

[1] Also multi-pixel-photon-counter (MPPC) and other abbreviations are common, depending mostly on the manufacturer.



(a) (b)

Figure 3.2: The connection scheme (a) and a microscopic picture (b) of an analog Silicon Photomultiplier. Shown is a sensor from Hamamatsu Photonics with a pixel size of 50 μm.

the following. The most common commercially available type of SiPM is called an analog Silicon Photomultiplier, where all pixels are connected in parallel as shown in Figure 3.2a. The pixels are placed in a matrix, most commonly covering a sensor area of $1 \times 1\,\mathrm{mm}^2$ or $3 \times 3\,\mathrm{mm}^2$. The pixel size of commercially available SiPMs reaches from $10 \cdot 10\,\mu\mathrm{m}^2$ to $100 \cdot 100\,\mu\mathrm{m}^2$.

3.3 Basic properties

When a Silicon Photomultiplier is exposed to small light pulses, some of the photons generate primary electron-hole pairs resulting in an avalanche in the pixels. Since the individual pixels act as binary photon counters, the generated charge is proportional to the number of fired pixels,

$$Q_{pe} = N_f \cdot G = N_f V_{over} C_d \tag{3.2}$$

The number of fired pixels triggered by detected photons is determined by the photon detection efficiency (PDE). For SiPMs, the PDE can be decomposed into four terms [34],

$$\mathrm{PDE} = (1 - P_R) P_{ff} \cdot P_{qe} \cdot P_{av} \tag{3.3}$$

where P_R gives the wavelength-dependent probability of the photon being reflected at the silicon surface and P_{ff} is the fill factor determining the geometrical fraction of the sensor area sensitive to light. The quantum efficiency P_{qe} is also a function of the wavelength and gives the probability for a photon to create an electron-hole pair in the silicon substrate. The Geiger efficiency P_{av} gives the probability of the primary electron-hole pair to trigger an avalanche and is a function of wavelength, temperature and bias voltage. The wavelength of the photon defines the average absorption depth in the silicon bulk. Depending on the absorption depth and doping profile of the sensor, either the primary electron or hole drifts into the high field region, eventually triggering an avalanche. Since the ionization coefficients are in general lower for holes as for electrons, the probability for a drifting hole not to cause an avalanche is higher. Therefore, two different doping schemes exist for sensors being sensitive in the visible blue spectral region (p-on-n structure) or to the green region (n-on-p structure as shown in Figure 3.1).
A typical charge spectrum for light pulses of small amplitude is shown in Figure 3.3. Since the charge is ideally quantized with the number of fired pixels, the spectrum shows distinct peaks corresponding to a certain number of fired pixels. Because of the proportionality

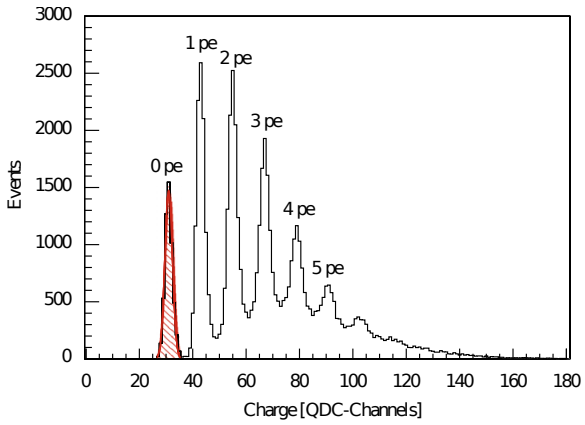


Figure 3.3: SiPM single photon spectrum [36]

of fired pixels to the number of photons, such a spectrum is also called a *single-photon-spectrum*. The distance of the peaks is determined by the single-pixel gain G of the sensor. This allows to measure and monitor the SiPM gain using LED pulses. The SiPM gain is then extracted by fitting a suitable fit function, for example a superposition of multiple Gaussian functions. The individual integrals under the peaks follow a compound Poisson Distribution [35]. Apart from the poisson-distributed number of avalanches from photons, additional statistical processes contribute, which will be discussed in the following paragraphs.

Dark count rate

An avalanche can not only be triggered by a photon reaching the SiPM pixel, but also any other process generating an initial electron-hole pair. Such single pixel events are indistinguishable from avalanches from photons, resulting in the *Dark count rate* (DCR) of the SiPM. Two processes are known to generate dark-count events. The thermal generation of free electron-hole pairs assisted by the presence of intermediate energy levels due to lattice defects can generate electron-hole pairs triggering an avalanche. Since the dark counts due to this effect are generated with a rate growing exponentially with Temperature, approximately by a factor 2 for a change of 8C° [37], the DCR can be reduced by cooling the sensor. This is often used for SiPMs exposed to radiation to compensate the increased lattice defect concentration in the silicon.
Due to the strong electric field in the avalanche region of the pixel, an electron can also tunnel through the band gap, potentially generating an avalanche. The dark counts generated by the field tunneling are only weakly dependent on the Temperature. It is determined by the bias voltage applied to the sensor and dominating at low temperatures.

The DCR in Silicon Photomultipliers is much higher compared to classical photomultiplier tubes. This poses the biggest disadvantages of these devices. Modern sensors reach a rate of such undesired avalanches of about $30\,\mathrm{kHz/mm^2}$ at room temperature ([38]).

Correlated Noise

The avalanche formation can also be a result of another avalanche in a neighboring pixel (*cross-talk*) or the same pixel earlier in time (*after-pulses*).

In the avalanche process above the breakdown voltage, hot carriers can generate photons with an energy sufficient to trigger secondary electron-hole pairs [39]. For photons above the band-gap energy of silicon, an generation efficiency of $2.9 \cdot 10^{-5}$ is reported [40]. While the photons absorbed within the pixel area contribute to the lateral development of the avalanche, the photons reaching neighboring pixels may also trigger an avalanche process there. The SiPM cross-talk can be reduced by adding trenches between the pixels to prevent the photons from travelling to the neighboring cells. For SiPMs with a small cell size and low gain, the average number of produced photons is smaller. Trenches are therefore less often used since they also increase the insensitive area. Most of the crosstalk events are prompt as the photons are absorbed in the high field region of the neighboring cell and directly trigger an avalanche. Also delayed crosstalk events are observed which seem to come from the drift of the released carriers [41]. The photons are absorbed in region of lower field strength in the neighboring cells. The generated electron or hole (depending on the doping structure) drifts towards the high field region and triggers an avalanche with a delay of few ns [42].

In contrast to the crosstalk avalanches, the after-pulse avalanches are generated in the same pixel. Due to the single pixel recovery time τ_{rec}, these events have a smaller amplitude depending on the delay between primary and secondary avalanche

$$Q(t) = Q_{pe} \cdot \left(1 - e^{-t/\tau_{rec}}\right) \tag{3.4}$$

where Q_{pe} denotes the signal for a fully recovered SiPM pixel (3.2) and t is the time difference to the previous event. During the avalanche, charge carriers are trapped in energy levels generated by impurities in the crystal. If the pixel has recovered to a sufficiently high overvoltage, a second avalanche can be triggered when the charge carrier is released again. In general, different characteristic time constants of the after-pulse events are observed [36]. Simulations show that a significant fraction of the fast component may also come from the same effect as the delayed crosstalk and not only from capture-release processes [41]. Figure 3.4 shows the processes of photon-assisted correlated noise for the prompt crosstalk (DiCT), delayed crosstalk through drift (DeCT) and the fast after-pulse contribution from photon-absorption and drift.

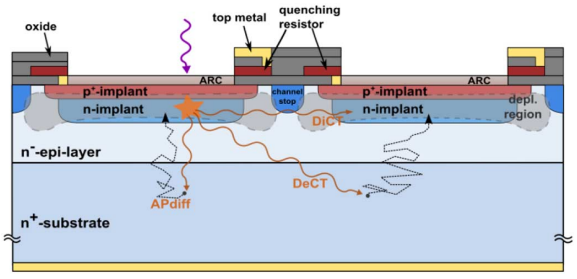


Figure 3.4: Correlated avalanche processes induced by photons from hot carriers [43]

Dynamic range

Neglecting the dark count events, the response of the SiPM is proportional to the light pulse amplitude for a small number of incident photons uniformly spread over the sensor area. For larger signals, the number of pixels available to trigger an avalanche limit the dynamic range of the Silicon photomultiplier. The average number of fired pixels is given by

$$N_f = N_{pix,eff} \cdot \left(1 - e^{-\mathrm{PDE} \cdot N_\gamma / N_{pix,eff}}\right) \tag{3.5}$$

The effective number of pixels $N_{pix,eff}$ is given by the number of pixels in the SiPM. This number is reduced by pixels insensitive to photons since they have been triggered before or give lower signal due to the pixel recovery. Thus, the dark count rate as well as the correlated noise processes reduce $N_{pix,eff}$ and the dynamic range of the sensor. The additional signal generated by crosstalk events is not included in (3.5). These processes add another factor of $1/(1 - P_{xt})$ to the numerator in the exponential term [44], where P_{xt} is the cross-talk probability. This also reduces the dynamic range of the sensor, and adds additional statistical fluctuations to the output signal. The effect on the signal-to-noise ratio is however only dominant for extreme saturation when more than 80% of the pixels fire, as shown in Figure 3.5. Otherwise, the statistical fluctuations of the number of fired pixels (enhanced by the correlated noise) and the number dark-count events in the considered time window dominate in this example.

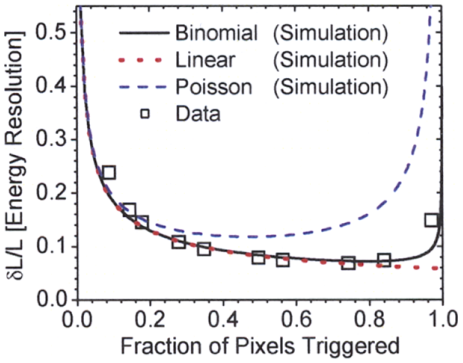
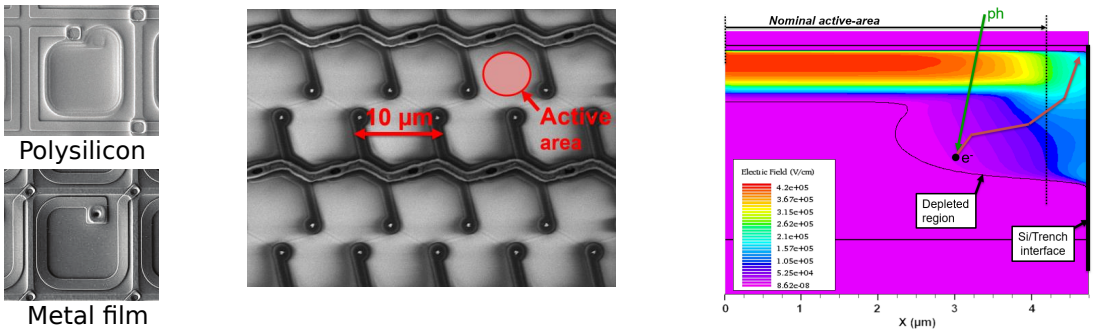


Figure 3.5: Photon counting resolution including DCR and correlated noise as a function of the fraction of fired pixels. The red dotted line ("linear") assumes a Poisson distribution without a limited upper bound and corresponds to the case of no saturation. The blue dashed line assumes response limited by the saturation function (3.5) with Poisson statistics, and the black line gives the response using binomial statistics including the number of available cells. Taken from [45].

3.4 High dynamic range sensors

In many applications the sensor area can not be increased to achieve a larger dynamic range for measurements of the amplitude of the light pulse. To reduce the effects of strong saturation, the light might be spread in a non-uniform manner. Finally however, the only solution is to increase the density of pixels on the silicon area, and therefore reduce the pixel



(a) **(b)** **(c)**

Figure 3.6: Techniques to improve the fill factor in high dynamic range SiPMs. a) Transparent metal film quenching resistors [46] b) Dense metal routing [43] and c) Field strength in a high density SiPM, the reduced size of the avalanche region close to the pixel border is indicated by the dotted line [43].

size. In general, the gain of such devices will be much lower since it scales with the pixel area. Commercial sensors with 10 μm pixel size typically have a gain of $(1 - 1.5) \cdot 10^5$. For small pixels, the area required for the quenching resistors, metal connection and guard ring structures degrade the photon detection efficiency because the fill factor is lower. Quenching resistors based on thin metal films instead of the commonly used polysilicon resistors are used in some Sensors from Hamamatsu. The metal resistors are transparent and can be placed on top of the active area of the sensor [46](Figure 3.6a). As mentioned previously, the cross-talk probability is reduced for sensors with small pixels due to the reduced gain. Therefore trenches are often removed for pixel sizes below 25 μm. Also the metal routing area can be reduced, for example by a honeycomb structure as visible in Figure 3.6b for a sensor with 10 μm pixel size developed by FBK. For the sensor shown, both trenches and polysilcon resistors are used. The main Problem in the reduction of the pixel size is however the guard ring structure. Figure 3.6c shows a simulation of the field strength in the avalanche region of a pixel of 5 μm size. Without modifications of the doping structure in the guard ring region, the active region is reduced due to the lower field towards the pixel edge.

3.5 Electrical model of Silicon Photomultipliers

In order to optimize the readout electronics for the use of a specific sensor configuration or the needs of a measurement (e.g. time or charge), an electrical model of the SiPM is required. This allows to simulate the response of the sensor in combination with the readout, and tailor the electronics to the needs of the application, giving the best measurement precision possible. Also the choice of the used sensor will have an impact on the optimization process.

For the single APD, an electrical model was introduced by Cova et al. [33] for both actively and passively quenched avalanche diodes. The electrical model for analog SiPMs using a passive and resistive quenching circuit has been established by Corsi et al. [47], extending the single pixel model to a combination of lumped passive (not in avalanche) and a single active SPAD. Many further additions to this model exist, for example including

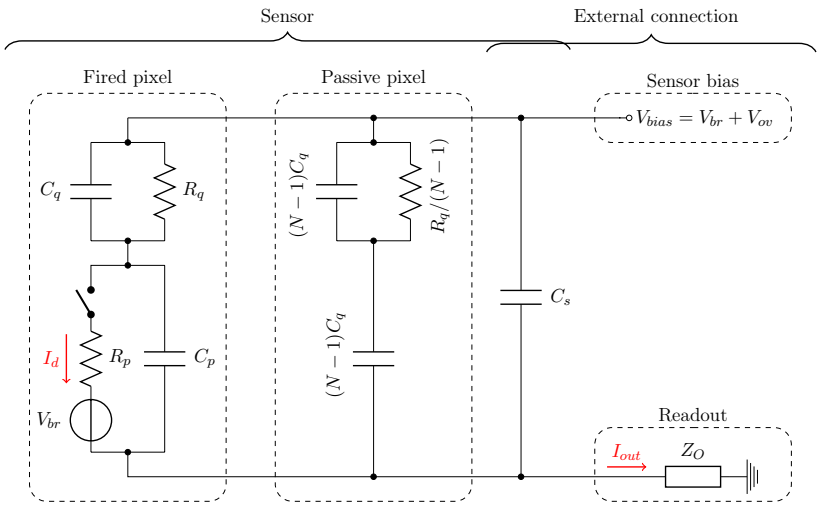
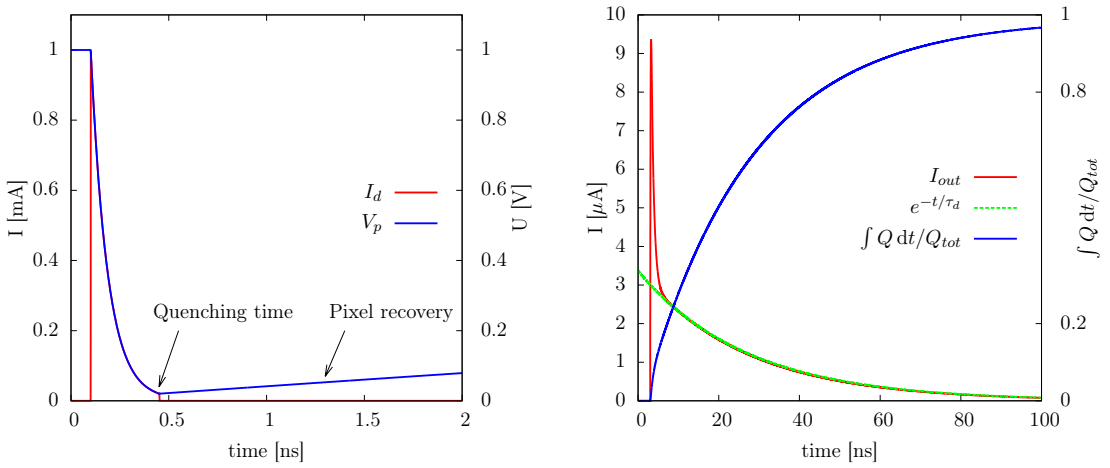


Figure 3.7: Electrical model of a SiPM including one firing pixel and indicating the bias voltage connection and readout.

bonding wire inductances[48] or the detailed simulations of the Geiger discharge formation[49]. Figure 3.7 shows the electrical model proposed by Corsi et al. which will be used in the following. To reduce the SPAD array of consisting many diodes to fewer components, the SiPM of N pixels is decomposed into one SPAD generating a signal, and $N-1$ pixels which are purely passive and can therefore be combined into parallel connected passive components. In many published analyses of this model, this is trivially generalized to the case of N_f pixels fired and $N-N_f$ passive pixels. This case will not be studied here because the difference in the pulse shape is marginal in case of charge measurement signals.

The active avalanche pixel diode is modelled by the junction capacitor C_p, initially charged to the bias voltage V_b. During the avalanche , the junction capacitance is discharged through the conductive channel, reaching a voltage level close to the breakdown voltage V_{br}. The diode current I_d is limited by the resistor R_p, representing the effective resistance of the Geiger discharge to the substrate bulk. The avalanche starting and quenching times are controlled by a switch in series to R_p. The quenching resistor R_q connected between the positive bias line of the SiPM and the APD is accompanied by a parasitic capacitance C_q since the metal connection is placed close to the pixel. Since all passive pixels behave the same, they are represented by the parallel connection of all remaining components, $R_{q,pass} = R_q/(N-1)$, $C_{p,pass} = (N-1)C_q$ and $C_{p,pass} = (N-1)C_p$. The parasitic capacitance introduced by the metal routing grid on the SiPM connecting all pixels is proportional to the number of pixels on the sensor. These parasitics are merged with the parasitics from the sensor package and routing on the printed circuit board, resulting in the parasitic C_s shown in the schematic.

Figure 3.8 shows the results of a numeric circuit simulation using the model. When an avalanche is triggered by a primary charge carrier drifting into the avalanche region, the switch in Figure 3.7 is closed, causing the pixel capacitance to discharge with a time constant of approximately R_pC_p, i.e. sub-nanosecond time scales. Once the current falls below a threshold current in the order of few μA, the avalanche is quenched and the switch is opened. The diode capacitance is then recharged through the quenching resistor with a



(a) **(b)**

Figure 3.8: SPICE simulation of the SiPM response using the electrical model with the parameters of a 50 μm pixel sensor and an ohmic readout impedance of 50 Ω. a) Current and voltage in the active diode b) Current sensed by the readout electronics and integrated charge relative to the generated total charge.

much slower recovery time constant of $R_q(C_p + C_q)$. The total charge released in one Geiger discharge is determined by the integral of the diode current I_d until the quenching time. It defines the gain of the single pixel, and thus the gain of the SiPM in the linear region. Since the quenching time is in the order of 100 ps and $R_p \ll R_q$, the charge flowing through the quenching resistor before the quenching time can be neglected. The single pixel charge signal then simplifies to

$$Q_{px} \approx V_{ov}\,(C_p + C_q) \tag{3.6}$$

For large gain SiPMs, the parasitic capacitance of the quenching resistor is much smaller than C_p and can be ignored. For small gain sensors with pixel sizes below 25 μA, this is no longer the case. Here, the junction capacitance decreases proportionally with the pixel area, while the quenching resistance is increased[2].

For the optimization of the read out electronics, the effective detector capacitance seen by the readout electronics, and the time constant of the signal sensed by the readout electronics represented by the frequency dependent input impedance Z_O are the relevant quantities. The effective capacitance is of high importance for the noise analysis of the electronics. For the measurement of the signal charge observable at the impedance Z_O, two readout schemes can be applied. In the *voltage sensitive* readout, the sensor current is converted to a voltage by the use of an ohmic resistance. This voltage is then sensed by the readout electronics. For a *current sensitive* scheme, the signal current is sensed directly by the impedance. In practice, the readout electronics will always have a bandwidth limitation. For the voltage sensitive readout scheme with a frequency independent impedance converting

[2]Based on and a comparative measurement of different sensors from Hamamatsu Photonics performed by the author, and [50] for a sensor from FBK

the current, the voltage across the resistor is measured with a limited bandwidth. Thus, signal components with a frequency beyond the bandwidth of the electronics are not sensed.

This is different for the case of a current sensitive readout, where the current is sensed with limited bandwidth (the impedance rises at the bandwidth of the first stage of the readout electronics). The signal current is integrated to obtain the charge information. If this bandwidth can be represented by an increased input impedance as it is the case for the current sensitive readout, no charge is lost, since the limited bandwidth will cause the charge to remain on the detector. In the impractical limit of an infinite charge integration time, the full signal is then sensed out by the electronics. This scheme will be used in the ASIC described in this work. In order to analyse the efficiency of the charge measurement, the high frequency components of the signal generated by the SiPM pixel discharge can be ignored, and the input impedance R_o at frequencies below the bandwidth of the readout electronics is considered.

The time behaviour in the sensor as well as the measurable signal current I_{out} was analysed by Marano et al. [51], showing that in total four time constants (reduced to three with some approximations) describe the pulse shape. Two fast rise and decay time constants below 1 ns, generated during the avalanche, and a fast and one slow recharge time constants after the SPAD has been quenched. These time constants have already been quoted in a strongly approximated manner in the previous discussion. For charge measurements, all fast time constants can be ignored and only the slow component of the current signal needs to be taken into account. In [51], one (inappropriate) conclusion is that this decay time constant τ_d is governed by the recharge time constant $R_q(C_p + C_q)$, i.e. independent on the input electronics input impedance. This finding is a consequence of some simplifications made in the analysis which are appropriate for applications requiring a good time resolution, where the high frequency components of the signal pulse are of main interest. This is not the case for applications mainly requiring the measurement of the charge signal, in particular when using SiPMs of small pixel size. In fact, for a large input impedance, one would expect a decay time constant of R_oC_d, where C_d is the effective capacitance of the detector seen by the readout electronics.

For the study of the effective capacitance at low frequencies, the parasitic capacitance C_q can be ignored as the complex impedance is much larger than R_q. The effective capacitance is then given by

$$C_d = NC_p + C_s \tag{3.7}$$

At frequencies above $\omega = 1/(R_qC_q)$, the parasitic can no longer be ignored, causing the effective capacitance falling to

$$N\frac{C_pC_q}{C_p + C_q} + C_s$$

For noise filtering, the integration time of the charge has to be limited in all practical implementations. Thus, the charge measurement efficiency is enhanced by minimizing the time constants of the signal current which dominate the charge integral. The relevant time constant of the charge signal τ_d is calculated including all parasitics while ignoring the signal generation by the active pixel. It is simplified to a current pulse, because the avalanche process is much faster and only contributes to the fast signal components directly. With

these approximations, the τ_d is calculated from Kirchhoff's current law equation, yielding

$$\tau_d = (NC_p + C_s)R_o + (C_p + C_q)R_q = C_d R_o + \tau_{rec} \tag{3.8}$$

It is limited by the recovery time constant of the SiPM at small R_o. As shown in Figure 3.8b, this time constant reproduces the low-frequency component of the signal well and dominates the charge integral. In order to keep the time constant small, the input impedance should then be small compared to

$$\hat{R} = \frac{(C_p + C_q)R_q}{C_p N + C_s} \tag{3.9}$$

Using the values listed in Table 3.1, the input impedance should be small compared to $\approx 790\,\Omega$ for high gain SiPMs with pixel dimension in the order of $50\,\mu$m, and $\approx 150\,\Omega$ for small gain sensors.

Using (3.8), also the bandwidth requirements posed on the first current-buffering stage of the readout electronics can be inferred. In order to keep the τ_d the dominant time constant on the slow end, the bandwidth should be kept above $1/(2\pi\tau_d)$. For a given input impedance of $50\,\Omega$ at low frequencies, the bandwidth should be large compared to 6 MHz and 14 MHz for high gain and low gain sensors, respectively.

	N	R_p	C_p	R_q	C_q	C_s
$50\,\mu$m SiPM [51]	400 (*)	$1\,\text{k}\Omega$	78 fF	$290\,\text{k}\Omega$	8 fF	$1\,\text{fF} \cdot N + 10\,\text{pF}$ (*)
$10\,\mu$m SiPM (typ.)	10^4	$1\,\text{k}\Omega$	4.6 fF	$1550\,\text{k}\Omega$	1 fF	$0.2\,\text{fF} \cdot N + 10\,\text{pF}$

Table 3.1: Component parameters for the SiPM electrical model. The values marked with a star are different from the cited reference. The values are adapted to a sensor size of $1\,\text{mm}^2$, and the parasitic capacitance on the PCB is included. Note that the values quoted for the $10\,\mu$m SiPM are only estimated values since to the knowledge of the author no electrical characterization measurement exists for this type of sensors. The value R_q is measured for a Sensor from Hamamatsu and consistent with [50] for a device from FBK. The other values are extracted from the datasheet [52] or based on best guessing. These should be considered as an order of magnitude estimation only.

4 Introduction to Mixed signal ASIC design

In modern chip production technologies, the small feature size and versatile library of components that can be produced on the same chip. This allows to develop custom electronics including both analog and digital circuits. In these so called mixed-signal designs, the analog information is digitized in the chip by the use of analog to digital converters, and the digital information is processed further in the digital part of the chip. Because these ASICs combine all functionalities required by the specific application in one chip, they are also called a System-On-Chip (SoC). While the design of the analog blocks is a mostly manual process in the view of both description (schematic) and physical implementation (physical layout), the digital part is designed using hardware description languages and is implemented with the help of automated software tools. At the interfaces between the analog and digital domains, additional information is required by the software analysing and optimizing the design in the digital domain. For the physical implementation, abstract views of analog blocks are generated from the detailed analog layout in an automated process, for example the pin positions and areas designated for routing the metal traces. The correct timing behaviour of the analog blocks connected by the digital part is ensured by timing libraries which define pin capacitances and maximum transition times[1] for input terminals, the maximum loading capacitance for outputs and the timing behaviour of the output with respect to the input pins. This allows the digital implementation tools to analyse and optimize the timing of the digital nets, and add buffers if necessary. Detailed descriptions of the digital implementation steps required to obtain a working digital design are given in [32] and [53], the digital design flow used for the ASIC described in chapter 6 are based on these works.

In this chapter, the basic principles of analog circuit design will be discussed. After a description of MOS-transistors as the most basic devices in the circuits, an introduction to the theory of analog signal processing will be given. This will be applied in the analysis of the analog performance of the ASIC described in chapter 6.

4.1 CMOS technology

In the complementary metal-oxide-semiconductor (CMOS) technology, MOS transistors are the most important building blocks of the analog circuits. In the following, the working principle of these transistors will be covered, leading to a small signal model which will be used extensively in the analysis of the circuits in chapter 6.

[1] The transition time defines the rise time of a digital signal, usually given as the time between 10% and 90% of the logic voltage levels.

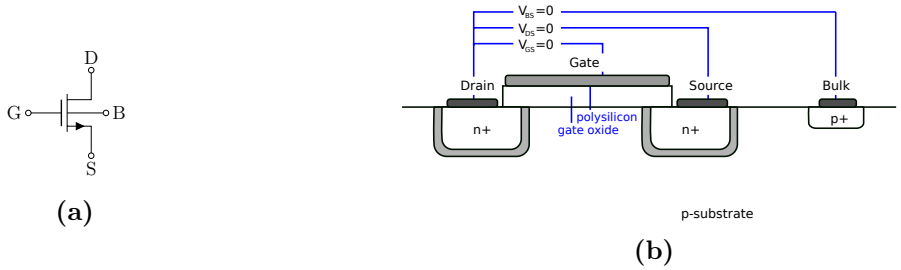


Figure 4.1: Schematic symbol and sketch of an n-MOS transistor

4.1.1 MOS transistors

Essentially, MOS transistors are three-terminal devices which behave as current sources between the drain and source terminal, controlled by a voltage at the gate terminal. Figure 4.1 shows the schematic symbol and cross-section through a standard n-MOS transistor. Embedded in a p-doped semiconductor substrate, the drain and source contacts are realized by strongly n-doped regions connected to a metal interface. The potential of the substrate is defined by the bulk contact in the vicinity of the transistor, connected to the ground potential. The gate contact is realized by a layer of grown poly-silicon on top of a thin SiO^2 insulator.

By convention, the control voltage of the transistor is given as the potential difference between the gate and the source terminal (V_{gs}). Once the potential difference between the gate and the source contact exceed a threshold voltage V_{th} defined by the p-type doping concentration below the gate, a conductive channel is formed underneath the gate. This channel is generated by the enhancement of n-type carriers due to the increased surface potential, leading to an inversion of the carrier concentration. This is then called the *strong inversion* operation region of the transistor. With the channel formed, a small voltage difference between the drain and source contacts (V_{ds}) will lead to a current I_d flowing through the transistor as sketched in Figure 4.2a. As V_{ds} increases, the potential difference between gate and drain decreases. Thus, the channel is pinched off when V_d exceeds $V_{\mathrm{ds,min}} = V_g - V_{th}$ as shown in Figure 4.2b. Until this point, the current through the channel is strongly dependent on the drain-source voltage difference, which is called the *triode mode*[2]. With the channel pinched off at the drain, I_d is only weakly dependent on V_{ds} and the transistor is operating in the *saturation mode*.
Figure 4.2c shows the current through the channel as a function of the drain-source voltage for different values of V_{gs} above threshold. The border between the triode and saturation mode is indicated by the dashed line. In the triode mode, i.e. $V_{ds} < V_{ds,min}$, the current through the channel is given by

$$I_d = \mu_n C_{ox} \frac{W}{L} \left(2\left(V_{gs} - V_{th}\right) V_{ds} - V_{ds}^2\right) \tag{4.1}$$

where μ_n is the mobility of the carriers (electrons in the case of the n-MOS transistor), C_{ox}

[2] Also called triode region. The differentiation is made to distinguish between strong and the weak inversion regions, which will be introduced at a later point.

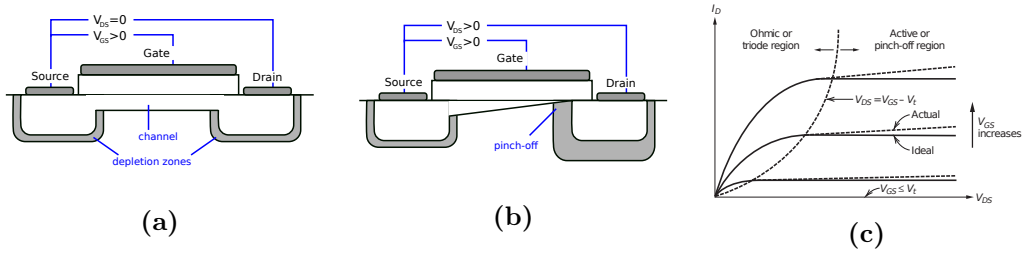


Figure 4.2: Operation regions of the n-MOS transistor in strong inversion. a) Triode mode. b) Beginning of saturation operating mode. c) Current through the n-MOS channel in strong inversion region, from [54].

is the capacitance per area between gate and semiconductor and W, L are the width and length of the transistor, respectively. While the strong dependence on V_{ds} is unfavoured to implement a voltage controlled current source, transistors operated in the triode mode can be utilized as voltage-controllable resistors, which is approximately the case for small V_{ds}, where the last term can be ignored *deep triode mode*.

For $V_{ds} > V_{ds,min}$, the current through the transistor channel is approximately given by the maximum of 4.1,

$$C_{ox}\frac{W}{L_{eff}}\left(V_{gs}-V_{th}\right)^2 \tag{4.2}$$

However, due to the pinch-off, the effective length L_{eff} of the channel is now reduced with larger V_{ds}, and I_d increases. This is called the channel length modulation effect. Taking only the channel length modulation into account, I_d now shows a linear dependence on the drain-source voltage as the effective channel length is reduced with $L' = L/(1 + \lambda V_{ds})$. Here, the channel length modulation parameter λ is a function of the transistor geometry increasing for short channel (small L) devices. In fact, the channel length modulation is not the only source of dependence on V_{ds}, for example the drain-induced barrier lowering (DIBL) is more dominant for short channel lengths[55]. At first order however, these effects may also be included in λ. In the saturated operation mode, the current through the transistor is then given by

$$I_d = \frac{1}{2}\mu_n C_{ox}\frac{W}{L}\left(V_{gs}-V_{th}\right)^2\left(1+\lambda V_{ds}\right) \tag{4.3}$$

For gate-source potential differences below the threshold voltage, the carrier concentration below the gate show no inversion and no channel is formed. A small current can still flow between the drain and source terminals, driven by the diffusion of minority carriers generated by thermal excitation. The concentration of the free charge carriers has a exponential dependence to V_{gs}, and so has the current through the channel. This operation mode of the transistor is called *weak inversion* or *sub-threshold* region. Here, the current through the channel is given by

$$I_d = \frac{W}{L}I_t \exp\left\{\frac{V_{gs}-V_{th}}{nV_T}\right\}\left[1-\exp\left\{\frac{-V_{ds}}{V_T}\right\}\right] \tag{4.4}$$

Where I_t is a technology and temperature dependent current in the order of $1\,\mu$A, $n \approx 1.5$ and $V_T = {}^{k_b T}/_{e} \approx 26\,\text{mV}$ is the thermal voltage. For $V_{ds} > 3V_T \approx 78\,\text{mV}$, the last term

in (4.4) is negligible, and the current is almost independent of V_{ds}. This, together with the small current in the gate, makes the sub-threshold operation region very attractive for low-voltage, low-power applications in modern CMOS technologies[56]. It will be used excessively in the ASIC developed in this work, targeting a low-power consumption. Due to the small current however, the relative current gain is low compared to the strong inversion region. It is therefore applicable mainly for circuits with a low bandwidth[3].

Small signal model of MOS-transistors

In order to facilitate the mathematical analysis of circuits built from nonlinear components such as MOS-transistors, linear small signal models are used to analyse the component's behaviour at a specific operation point. For transistors, this operation point is defined by the channel current or V_{gs} (i.e strong inversion or sub-threshold region) and the drain-source voltage difference (triode or saturated operation mode). Figure 4.3a shows a schematic representation of the n-MOS small signal model. The gate current is generated by an ideal current source of $g_m V_{gs}$ where g_m is called the *transconductance* with units of an inverse resistance,

$$g_m = \frac{\mathrm{d}I_d}{\mathrm{d}V_{gs}} \tag{4.5}$$

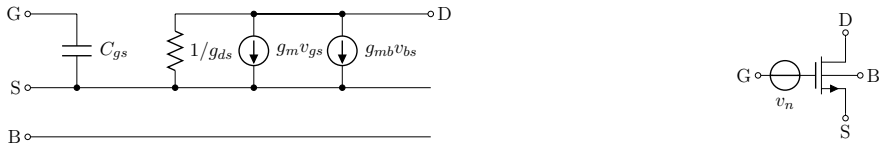
The dependence on V_{ds} is modelled by a resistance of $1/g_{ds}$ with

$$\frac{1}{g_{ds}} = \frac{\mathrm{d}I_d}{\mathrm{d}V_{ds}} \tag{4.6}$$

Because the threshold voltage of the transistor is defined for potential differences between the gate and the transistor bulk, it is dependent also on V_{bs}. This modulation of the current due to threshold voltage changes is called the *body* or *back-gate* effect, relevant only if the source and the bulk are not at the same potential. In the small signal model, it is represented by a second current source with a back-gate transconductance of g_{mb} Typically the transconductance of the back gate is about 3-10 times smaller than g_m. Therefore the body effect is of no practical concern for most of the circuits implemented in this work. In some cases the circuit would be affected, and the body effect is minimized by connecting the bulk potential to the source. Because the substrate silicon is normally shared by all nMOS-transistors and has a low impedance, this is possible only for transistors residing in a separate well.

Generated by the depletion regions and the gate oxide, a parasitic capacitance proportional to the transistor dimensions is present between each of the transistor terminals. For the most common operating case of the transistors, strong-inversion in saturation, the capacitance C_{gs} between the gate and the source is most dominant. This capacitance is also shown in the small signal model schematic, while the other capacitances are left out for clarity.

[3]For advanced nodes below 90nm, the sub-threshold operation is also used for circuits which are not slow at all: Since the gate capacitances are small in the sub-threshold region, propagation delays in the order of 100ps are obtained for digital circuits fully based on sub-threshold MOST. This family of digital cells have much lower power loss due to leakage currents, normally dominating the power consumption in these technology nodes.



(a) Small signal model of a MOS transistor

(b) Gate-referred noise sources

Figure 4.3: Small signal model of a MOS transistor. a) Small signal model schematic. b) Equivalent noise voltage at the gate terminal

Table 4.1 summarizes the relevant parameters of the small signal model for the sub-threshold region (with $V_{ds} > 3V_T$) and the strong inversion region in triode and saturated mode. In the case of strong inversion, the transconductance scales with $\sqrt{I_d}$, whereas in the sub-threshold region, the transconductance stays constant with respect to I_d . Thus, the transconductance is much smaller in sub-threshold due to the small channel current. More conveniently for the analysis, the transconductances are also rewritten as a function I_d at the operation point.

Operation mode	g_m	g_{ds}
sub-threshold	$\frac{I_d}{nV_T}$	0
strong inversion, triode	0	$\mu_n C_{ox} \frac{W}{L} (V_{gs} - V_{th})$
strong inversion, sat.	$\mu_n C_{ox} \frac{W}{L} (V_{gs} - V_{th}) (1 + \lambda V_{ds})$ $= \sqrt{2\mu_n C_{ox} \frac{W}{L} I_d}$	λV_{ds}

Table 4.1: Small signal model parameters for the n-MOS Transistor

Noise sources of MOS transistors

Transistors exhibit noise modelled equivalently as a noise current through the channel and parallel connected to the transistor, or as a noise voltage at the gate of the transistor which is then modulating the current, as illustrated in Figure 4.3b. For MOS transistors, two noise sources with different frequency behaviour exist. As for resistors, the *thermal noise* is generated by the random thermal movements of the charge carriers in the channel. The squared noise voltage density referred to the gate from thermal noise in the transistor channel is given by

$$v_{n,therm}^2 = \frac{8k_bT}{3 \cdot g_m} \Delta f \tag{4.7}$$

The second noise source is found in all active devices and called the flicker or 1/f noise. It is generated from trap and release processes due to lattice defects in the channel. Because the traps are associated to specific time constants given by their energy levels, the noise generated by these processes is frequency dependent. Effectively, this leads to a noise density

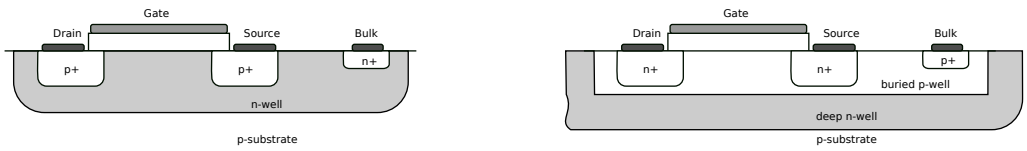
with a 1/f frequency dependence. Referred to the gate terminal of the transistor, the noise density from flicker-noise is given by

$$v_{n,flicker}^2 = \frac{K}{C_{ox} W L} \cdot \frac{1}{f} \Delta f \tag{4.8}$$

where K is a technology and transistor-type dependent constant. It is dominant at low frequencies and can be made small by increasing the transistor area $W \cdot L$, potentially sacrificing bandwidth due to the increased parasitic capacitances.

p-MOS and burried well n-MOS transistors

As the name suggests, the complementary version of the n-MOS transistor, the p-MOS, is also available in CMOS technologies. A cross section of the p-MOS doping structure is sketched in Figure 4.4a. The transistor resides in a weakly n-doped *well*, with the drain and source being $p+$. The bulk potential is defined by the contact inside the well, which allows to connect the bulk to a different potential, e.g. the source of the transistor to mitigate the body-effect. The p-MOS is fully complementary to the n-MOS counterpart in the sense that the source is at the highest potential. A decreasing voltage at the gate terminal will lead to an (absolute) increasing current through the drain. All formulas for the transistor currents, as well as the small signal model have the same form as for the n-MOS, with the difference that now holes are the majority carriers in the channel. The mobility for holes μ_p is 2-3 times smaller than for electrons. Therefore a p-MOS transistor of the same dimensions will have a correspondingly smaller transconductance than the n-MOS counterpart. Note that with the source-terminal being at the higher potential, V_{gs} is negative and so is I_d.
In some CMOS technologies optimized for mixed-mode applications, a special type of n-MOS transistor sitting in a separate well exists. A sketch of such a structure is shown in Figure 4.4b. By doping a *deep* or buried n-well into the substrate connected to a normal n-well contact, an isolated weakly p-doped region is formed. This is then called a buried p-well, allowing to manufacture n-MOS transistors with a bulk isolated from the substrate. As mentioned before, this type of transistor can be utilized to reduce the body-effect. More important is however that also the noise on the substrate potential, generated by switching activity of other circuit blocks such as the digital logic cells, is shielded effectively from the transistor. This allows to separate the noisy digital domain of the chip from the analog parts that require a clean environment for optimum noise performance. Such types of transistors have been used throughout the analog circuits of the design discussed in Chapter 6.



(a) p-type MOS transistor **(b)** Buried well n-MOS transistor

Figure 4.4: Other (MOS) transistor types in CMOS technologies

4.2 Analog signal processing and the Laplace transformation

The behaviour of time invariant analog circuits be analysed in the signal processing theory using linearized models of the circuit components. As discussed in the previous section, the linearized models of nonlinear circuit component such as MOS transistors are only an first-order approximation around the operating point. This operating point is defined by the component parameters and calculated by the stationary currents and voltages in the circuit. Thus, these models are only valid for small changes of the operating point. This can be ensured by designing the circuit to behave linearly use of feedback, often being a requirement on the electronics anyway. In the following paragraphs, an introduction to the relevant theory for the analysis of linear and time-invariant (LTI) systems will be discussed for the application of pulse signal processing. Details on this matter can be found in text books on signal processing such as [57, 58] and also in books on analog circuit design, e.g. [54, 59].

If a system is linear and time invariant, there is a linear operator $\mathfrak{H}$ independent on the absolute time, mapping an input $x(t)$ to the output signal $y(t) = \mathfrak{H}(x(t))$ of a signal processing block. In time domain, the output signal can then be written as a convolution of the input signal with the *impulse response* function h(t).

$$y(t) = \int_0^\infty x(\tau) \cdot h(t-\tau)\, d\tau \tag{4.9}$$

The impulse response $h(t)$ is defined as the response of the system to a delta function as input stimulus, $h(t) = \mathfrak{H}(\delta(t))$. Because it can be used to calculate the response to any input voltage stimulus put to the system, the impulse response function fully characterizes the linearized circuit.

The behaviour of the system is given by the current node and voltage loop equations following the Kirchhoff's laws, which provide a differential equation system describing the circuit. In time domain, this differential equation has the general form

$$\sum_{j=0}^{m} b_j \frac{\partial^j x(t)}{\partial t^j} = \sum_{i=0}^{n} a_i \frac{\partial^i y(t)}{\partial t^i} \tag{4.10}$$

where the coefficients a_i, b_i are given by the circuit topology and it's component parameters, describing the dependence on the input and output signal, respectively. The impulse response is then obtained by solving the differential equation with the input stimulus pulse. This most tedious part of the calculation can be avoided by a Laplace transformation to the s-domain, which is more suitable for the analysis of pulse-processing circuits. A signal x(t) in time domain is transformed to the s-domain using the Laplace-integral:

$$X(s) = \mathcal{L}\{x(t)\} = \int_0^\infty e^{-st} x(t)\, dt \tag{4.11}$$

The inverse Laplace transformation is given by

$$x(t) = \mathcal{L}^{-1}\{X(s)\} = \int_0^\infty e^{st} X(s)\, ds \tag{4.12}$$

Under the replacement $s \to i\omega$, the functions are transformed to frequency domain, and the similarity to the Fourier transformation is apparent. Therefore, the s-variable of the Laplace transformation can be understood as a complex frequency variable. The Laplace transformation is however more general as the integral in (4.11) may also converge for functions that are not convergent themselves, for example $\exp\{a\,t\}$ with $a > 0$. This is achieved by adding a convergence factor $\sigma \in \mathbb{R}^+$, with $s \to i\omega + \sigma$. For the above example, the integral then has a region of convergence of $a < \sigma$.

In the s-domain, the differential operators in (4.10) transform as

$$\frac{\partial^i}{\partial t^i} f(t) \leftrightharpoons s^i F(s), \tag{4.13}$$

and yield a polynomial with same coefficients a_i, b_i for the left and right side of the differential equation. The input and output functions in the s-domain $x(t) \to X(s)$, $y(t) \to Y(s)$ can be separated off, and (4.10) can be rewritten into the form

$$Y(s) = H(s)X(s) \tag{4.14}$$

with

$$H(s) = \frac{Y(s)}{X(s)} = \frac{\sum_{i=0}^{n} a_i s^i}{\sum_{j=0}^{m} b_j s^j} \tag{4.15}$$

The operator H(s) transforms the input signal X(s) to the output signal Y(s) and is called the *transfer function*. Under Laplace transformation, the delta function in time domain transforms to a constant value of 1. Thus, H(s) is the representation of the impulse response in the s-domain and $h(t) \leftrightharpoons H(s)$.

A comparison of (4.9) and (4.14) already shows an important transformation rule, the convolution in time domain. In the s-domain, the convolution integral transforms to a product. Thus, the connection of several sub-blocks ($h_1 \leftrightharpoons H_1$, $h_2 \leftrightharpoons H_2$) yields a common transfer function

$$h(t) = h_1 \circ h_2 \leftrightharpoons H(s) = H_1(s) \cdot H_2(s) \tag{4.16}$$

This allows to easily write down the transfer function of an analog processing chain by partitioning in sub-blocks, under the assumption that the blocks are not affecting (loading) each other.

Instead of applying an inverse Laplace transformation in order to analyse the circuit behaviour using the impulse response, the transfer function can also be rewritten to study the circuit in the s-domain. One form of the transfer function is the decomposition into products of complex roots in the nominator and denominator,

$$H(s) = k \cdot \frac{\prod_i (s - Z_i)}{\prod_j (s - P_j)} \qquad k \in \mathbb{R};\ \{Z_i\}, \{P_j\} \in \mathbb{C} \tag{4.17}$$

Z_i and P_j are called zeros and poles, respectively. While they are complex numbers, imaginary poles and zeros only appear in pairs of complex conjugates since the transfer function is a function mapping to $\mathbb{R}$. From the position of the poles in the complex plane, corresponding terms in the impulse response can be inferred. This gives some insight to the stability of the system already in the s-domain. A transfer function of a signal pole at $-1/\tau$ corresponds to an exponential function,

$$\frac{1}{s+1/\tau} \leftrightharpoons e^{-t/\tau} \tag{4.18}$$

Therefore, only poles in the left half of the complex plane will show a stable response. For a pair of complex conjugate poles with a negative real part, the system will show a damped oscillating behaviour,

$$\frac{\omega}{(s+1/\tau+i\omega)\cdot(s+1/\tau-i\omega)} = \frac{\omega}{(s+1/\tau)^2+\omega^2} \leftrightharpoons e^{-t/\tau}\cdot\sin(\omega\cdot t) \tag{4.19}$$

For the special cases of roots at 0, the nature of the Zeros or Poles becomes apparent. The differential operator transforms as in (4.13), corresponding to a Zero at s=0. The integration operator is transformed to a Pole at s=0:

$$\int x(t)\,\mathrm{d}t \leftrightharpoons \frac{1}{s}X(s) \tag{4.20}$$

Thus, the poles can be understood as terms with integrating nature, and the Zeros as such with differentiating behaviour. For roots not at zero, the position corresponds to the characteristic frequency where this behaviour sets in.

Figure 4.5a shows the passive implementation of the passive $\mathrm{CR}-\mathrm{RC}^2$ Filter, often used for the processing of pulse signals in nuclear instrumentation applications. It consists of one high-pass and two low pass filters connected together. The transfer function of this filter can be calculated by decomposition into one CR high-pass and 2 RC low-pass stages. The current node equations following the Kirchhoff's law give rise to differential equations in time domain for the CR and RC stages, respectively.

$$\frac{V_o(t)}{R} = C\cdot\frac{\mathrm{d}}{\mathrm{d}t}\left(V_o(t)-V_i(t)\right) \tag{4.21a}$$

$$\frac{V_o(t)-V_i(t)}{R} = C\cdot\frac{\mathrm{d}V_o(t)}{\mathrm{d}t} \tag{4.21b}$$

Under Laplace transformation, the terms $\frac{\mathrm{d}V(t)}{\mathrm{d}t}$ from the capacitors are transformed to $s\,V(s)$ using (4.13). Thus, the frequency dependent impedance of a capacitor shows up as $1/sC$ in the s-domain. The transfer functions of the CR high-pass and RC low-pass filter stages are then given by

$$H_{CR} = \frac{sCR}{1+sCR} \tag{4.22a}$$

$$H_{RC} = \frac{1}{1+sRC} \tag{4.22b}$$

The full transfer function of the filter is given by their product ($H_{CR} \cdot H_{RC}^2$). For the general case of an CR $-$ (RC)^n filter, the total transfer function is

$$H = \frac{s\tau}{(1+s\tau)^{n+1}} \qquad \text{with } \tau = CR \tag{4.23}$$

with the impulse response

$$h(t) = \frac{1}{(n+1)!}\left(\frac{1}{\tau}\right)^{n+1} t^{n-1} e^{-\frac{t}{\tau}}(n\tau - t) \tag{4.24}$$

The impulse response shows an undershoot with a zero crossing at $t = n\tau$ which is coming from the CR high-pass introducing a Zero at $s = 0$.
In comparison, a $(RC)^n$ filter with transfer function

$$H = \frac{1}{(1+s\tau)^n} \qquad \text{with } \tau = CR$$

has a impulse response peaking at $t = (n-1)\tau$, and the time derivative of this function has the same shape as the impulse response of the CR $-$ RC^{n-1} filter with it's additional Zero in the transfer function.
The same filter response can also be obtained using an active filter topology as shown in Figure 4.5b. Here, the RC-stages are implemented using a negative feedback loop around an operational amplifier, and the first CR-RC stages can be combined using one amplifier since it's input acts as an AC ground.

Figure 4.5c shows a charge sensitive readout circuit consisting of an active charge sensitive preamplifier followed by an active filter of second order. The filter is in this case also called a shaper, underlining the function of the block in terms of pulse processing. The charge sensitive preamplifier integrates the signal current on the Capacitor C_i, discharged by the resistor R_i. The transfer function of this block is then given by

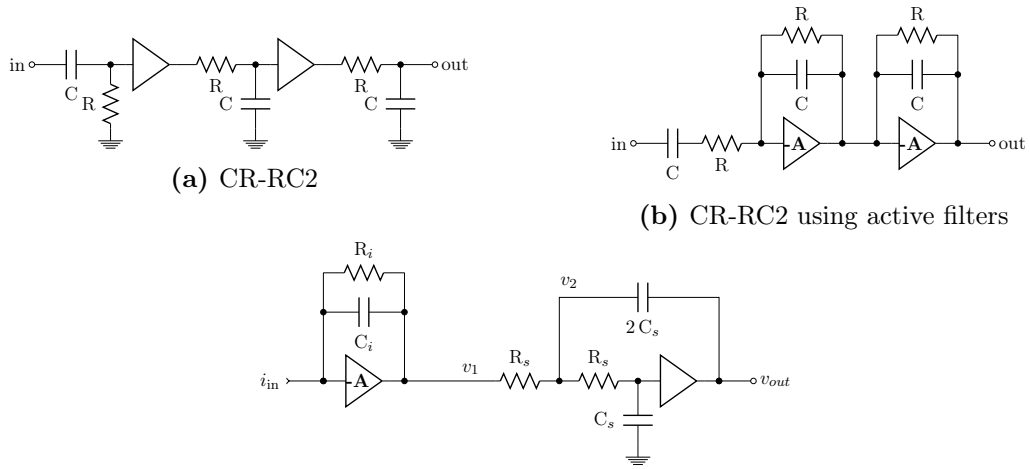
$$H_i = \frac{1}{C_i}\frac{1}{1+s\tau_i} \qquad \text{with } \tau_i = R_i C_i \tag{4.25}$$

The topology of the active filter following the preamplifier is named after the inventors, Sallen and Key [60]. In this configuration, the shaper implements a second-order low-pass filter. The response of the shaper is again calculated from the Kirchhoff's current law equations,

$$0 = 2C_s s\,(v_2 - v_o) + \frac{v_2}{\frac{1}{C_s s} + R_s} + \frac{v_2 - v_1}{R_s} \tag{4.26a}$$

$$0 = C_s s v_o + \frac{v_o - v_2}{R_s} \tag{4.26b}$$

which yield a transfer function of second order with a pair of two complex conjugate poles.



(a) CR-RC2

(b) CR-RC2 using active filters

(c) Charge sensitive preamplifier in combination with a Sallen-Key low-pass filter ("IP+CPP")

Figure 4.5: Illustration of commonly-used pulse shaping systems (filters)

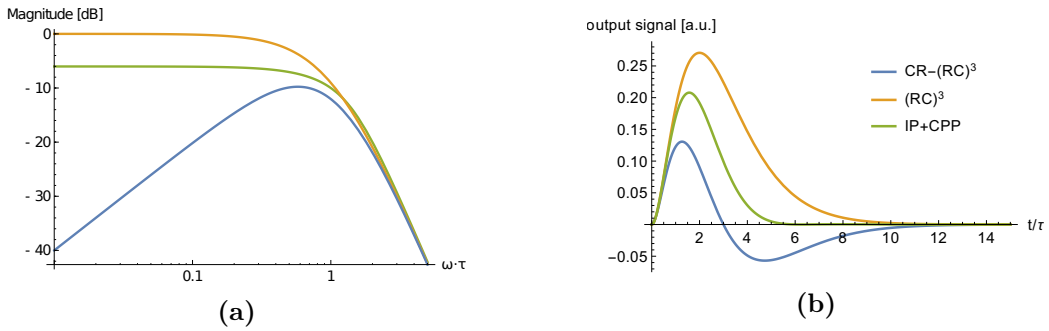
This would result in a damped oscillating impulse response for the shaper (4.19).

$$H_s(s) = \frac{2}{(s\tau_s + 1 + i)\,(s\tau_s + 1 - i)} \qquad \text{with } \tau_s = 2R_sC_s \tag{4.27}$$

For the full transfer function however, also the pole of the preamplifier is included. When the time constants of the preamplifier and shaper are made equal, the real-valued damping term is enhanced and the total impulse response is a semi-Gaussian pulse with little undershoot,

$$h(t) = \frac{2}{C} e^{-\frac{t}{\tau}} \left(1 - \cos\left(\frac{t}{\tau}\right)\right) \tag{4.28}$$

Figure 4.6 gives a comparison of the frequency response and impulse responses for the different shaper topologies. The Zero of the CR-RC3 shaper manifests as positive slope of 20dB per decade starting from $\omega = 0$. For each pole at $\omega = 1/\tau$ showing up in the transfer function, the slope decreases by 20bB per decade. Therefore, all filters have the same response at frequencies well above the pole position. The main purpose of the CR-stage in the CR-RC3 shaper is the filtering of low frequency noises, e.g. the 1/f noise from MOS transistors or the detector itself. As discussed previously, this causes a prominent undershoot in the impulse response function, which is unfavorable for the development of high dynamic range readout electronics. Since the low frequency noise components are small in the ASIC discussed in Chapter 6, it can also be implemented by other means not requiring 3 stages of operational amplifiers. Thus, only the RCn and the combination of one real pole and the pair of complex poles, denoted as "IP+CPP" in the following chapter, will be considered an option for the shaping stage.



(a) **(b)**

Figure 4.6: Comparison of the frequency behaviour (a) and impulse response (b) of the CR-RC3, RC3 and integration+complex pole shapers. The pole position is chosen the same for all cases.

5 Noise contributions to SiPM single pixel charge measurements

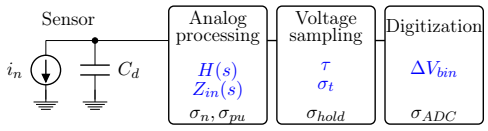
The charge measurement of signals from scintillation photons generated by traversing minimum ionizing particles is dominated by statistical fluctuations of the number of incoming photons and finally the number of fired pixels. Since the charge spectra of single minimum ionizing particles generating scintillation photons in the tile are only recorded with a low rate, a frequent calibration of the tile response is not feasible in this way.

Due to the photon counting nature of the Silicon Photomultiplier, the gain of the sensors may be extracted from single photon spectra. This information can be used to track changes of the tile response due to changing SiPM gain e.g. from temperature changes, and allows to adjust the bias voltage accordingly. Therefore, the measurement of small light-flux signals is critical for calibration of the detector. The calibration can be performed either in dedicated calibration runs using a pulsed light source integrated into the detector, or even during data taking using hits with small energy deposit.

For the gain estimation using single pixel information, a clear separation of the individual peaks is required. The sensor gain largely depends on the pixel size, which is related to the pixel count in the sensor and thus their dynamic range. For high dynamic range sensors with pixel sizes smaller than 25μm, this can pose a challenge on the noise performance of the readout electronics.

The small signal response for single to few fired pixels has several noise contributions. Some can be altered or are directly generated by the readout electronics. In the following chapter, the different noise terms contributing to the smearing of single photon peaks will be discussed. Where applicable, the influence of design parameters in the readout electronics will be described, with the goal to find an optimum set of parameters for the design of integrated readout electronics.

The noise can be separated in contributions from the sensor itself and in noise sources affected or generated by the readout electronics, following the individual building block (and "tasks" of the electronics) in the sketch below.



$$\sigma_n^2 = \sigma_{lk}^2 + \sigma_{APD}^2 + \sigma_{pu}^2 + \sigma_{el}^2 + \sigma_{hold}^2 + \sigma_{ADC}^2 \tag{5.1}$$

Here, σ_{lk} denotes the noise from the sensors leakage current, σ_{APD} are noise sources coming from statistical gain fluctuations of the APDs or non uniformities of the SiPM pixels, σ_{pu} is

the effective noise term from signal pile-up of dark count pulses, σ_{el} is the intrinsic noise of the readout electronics and σ_{hold} is coming from the uncertainties when sampling the peak voltage, which is then digitized by the ADC with a limited resolution (σ_{ADC}).

To study the separation of the individual photo-electron peaks, it is useful to define a single pixel signal to noise ratio,

$$\mathrm{pSNR} = V_{pe}/\sigma_n \tag{5.2}$$

The input referred noise for a charge sensing readout can also be described by an input referred noise charge, ENC (Equivalent noise charge). It is given by the integrated output voltage noise divided by the charge conversion factor, $G = \frac{\mathrm{d}v_{out}}{\mathrm{d}Q_{in}}$,

$$\mathrm{ENC} = \sigma_n/G \tag{5.3}$$

and can be understood as the charge that can be sensed by the readout electronics with a signal to noise ratio of one. The ENC allows to directly calculate the single pixel signal to noise ratio affected by electronics noise for a SiPM with known gain.

5.1 Electronic noise

The input referred noise generated by the readout electronics may be decomposed into two noise sources at the input, as sketched in Figure 5.1: A voltage noise source in series to the input, and a current noise source parallel to the input. These noise components are called "Series" and "Parallel" noise terms and fully characterize the noise behaviour of the analog front-end [61]. In addition, the bias voltage source together with the filtering network introduces additional noise components, neglected in the noise analysis since its characteristics including the noise are finally dictated by system aspects such as the quality of the voltage supply and the event rate if the same supply is shared between many channels. The sensor is represented by the total capacitance C_d (including the parasitic capacitances on the circuit board). As discussed in 3.5, the simplification to a single capacitance from the SiPM electrical model holds true for decent bandwidths of the analog processing chain, which is a reasonable assumption for the charge sensitive readout where time resolution is not of highest importance. The capacitance changes the frequency behaviour of the noise sources if converted to the current or voltage that is read out, depending on the type of the noise (current or voltage, series or parallel). For an idealistic current sensitive readout scheme, the terms can both be converted to a total noise current density present at the input,

$$i_{n,tot}^2 = i_{n,e}^2 + v_{n,e}^2\omega^2 C_d^2 \tag{5.4}$$

As the signal and noise currents are normally read out by the electronics with non-zero input impedance, they the charge sensed by the readout electronics is shared between the detector capacitance C_d and the resistor modelling the input stage of the electronics.

Under the assumption that the input impedance is not frequency dependent ($\tau_{el} \ll Z_{in}C_{det}$), the current (noise as well as signal) sensed by the electronics is given by

$$i_{in,el}(s) = \frac{i_{in}}{1 + sZ_{in}^2 C_d^2} \tag{5.5}$$

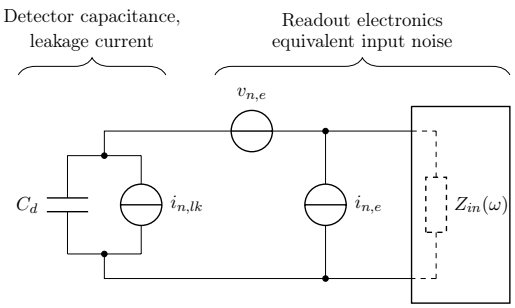


Figure 5.1: Electronic noise sources

Thus, also the parallel noise component is made frequency dependent with a time constant $Z_{in}C_d$. The total sensed noise current density is given by

$$i_{n,tot}^2 = \frac{1}{1+\omega^2 Z_{in}^2 C_d^2}\left(i_{n,e}^2 + v_{n,e}^2\omega^2 C_d^2\right) \tag{5.6}$$

While not noted explicitly, the two noise contributions are a function of frequency since they are generated by transistors which exhibit both broad-band thermal noise as well as low frequency 1/f noise (see Chapter 4).
The total input referred noise current is filtered and converted to a noise voltage observable at the output of the analog part. As discussed in the previous chapter, the behaviour of the analog processing system can be described by transfer functions in the s-domain. For stable circuits, the transfer function can always be transferred to the frequency domain, and the noise current densities at the input are propagated to the output giving a noise voltage density. Since the noise is uncorrelated, the total noise variance at the output is obtained by integration over frequency, adding the noise current densities in quadrature.

$$\sigma_v^2 = \int_0^\infty |H(i\cdot 2\pi f)|^2 \cdot i_n^2(f)\mathrm{d}f \tag{5.7}$$

$H(i\cdot 2\pi f)$ is the transfer function of the system in frequency domain and i_n is the input noise density.

For the optimized readout electronics, a transfer function needs to be found that effectively limits the noise including all contributors while keeping the signal response high. The dominant poles and zeros in the transfer function generated by the filter circuits can also be identified with frequency range which the front-end is sensitive to. To some extend, these represent the starting and cut-off frequencies in the noise integral. With increasing filter order, the noise filtering is more effective as the roll-off gets steeper. Since the power consumption increases as the number of stages does with the filter order, a compromise has to be found in the design and topological choices for the analog processing chain. The intrinsic time constants of the processing stage are selected depending on the contribution of the two noise terms. The series noise shows up at higher frequencies, favoring a lower time constant in $H(s)$. On the other hand for dominating parallel noise, a smaller processing time constant is favoured.

5.2 SiPM-intrinsic noise contributions

The majority of free charge carriers generated by tunneling and thermal excitation in the vicinity of the guard ring structure of the SiPM pixel diode do not trigger an avalance but generate a constant leakage current. This results in a shot noise current which is sensed by the readout electronics. The current noise density is given by

$$i_n^2 \,\mathrm{d}f = 2qI_{lk} \,\mathrm{d}f \,. \tag{5.8}$$

For a charge sensitive readout system with transfer function $H(s)$ or impulse response $h(t)$, this converts to a voltage noise at the output using the noise integral (5.7),

$$v_n^2 = 2qI_{lk} \int_0^\infty |H(i2\pi f)|^2 \,\mathrm{d}f = qI_{lk} \int_0^\infty h(t) \,\mathrm{d}t \tag{5.9}$$

The leakage current is typically in the range of 100pA, yielding a noise voltage which is negligible compared to the other noise sources. Sensors operated in a high radiation environment can however have a much higher leakage current [62]. Since the dark count rate also increases significantly at the same time, the noise from the leakage current will usually still be small in comparison.

Avalanche charge fluctuations

By the use of the passively quenching ohmic resistors connected in series to every Geiger-mode APD in the SiPM, the avalanche is stopped as the current drops below a certain threshold current. Since the ionization and recombination are statistical processes, the quenching time is also subject to statistical fluctuations, and so is the charge generated during an avalanche. This is particularly true for SiPMs operated at high overvoltages, where the current through the quenching resistor can no longer be neglected as it becomes comparable to the quenching threshold current.
Due to production tolerances, the average gain of the individual pixels given by the average quenching time, pixel capacitance and the resistance of the triggered pixel are not exactly the same. As a sensor typically consists of several hundreds to thousands of pixels, this variation of the gain also leads to a uncertainty of the single pixel gain if the light onto the sensor is spread over the SiPM area. For multiple pixels triggered simultaneously, the sensor response is a sum of the individual signals. Including the statistical gain variations and the distribution of the individual pixel gains, the variance of the signal rises with the number of pixels,

$$Q(N_f) = N_f G_{px} \tag{5.10a}$$

$$\sigma_Q^2(N_f) = N_f \sigma_G^2 = N_f(\sigma_{tq}^2 + \sigma_{G,apd}^2) \tag{5.10b}$$

where N_f denotes the number of fired pixels, G is the average gain of all APDs, σ_{tq} are the gain fluctuations due to uncertainties of the quenching time and $\sigma_{G,apd}$ are the fluctuations of the APD gain due to pixel-to-pixel fluctuations.

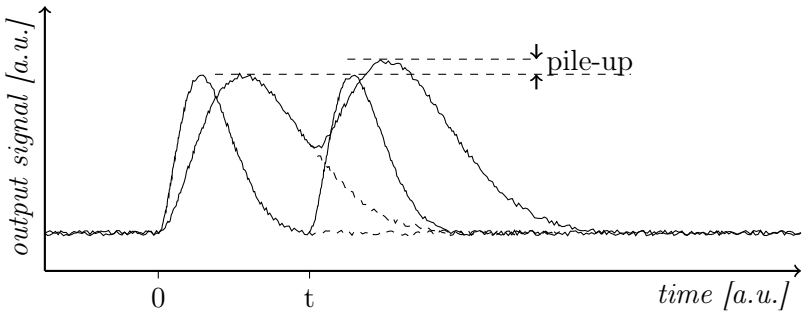


Figure 5.2: Illustration of signal pileup from dark-count pulses.

5.3 Signal smearing from dark count rate pileup

One noise source in the SiPM readout comes from the random triggering avalanches in the gAPD pixels due to thermal excitation or tunnel effects, which generates random pulses of single pixel amplitude[1] on the analog output.
The mathematical theorem of Campbell [63] gives the variance of the output signal for such point processes with a rate $N_{\mathrm{DCR}} = {}^{1}/_{\tau_{\mathrm{DCR}}}$

$$\sigma^2_{\mathrm{pu}} = N_{\mathrm{DCR}} V^2_{pe} \int_0^\infty h(t)^2 \, dt \tag{5.11}$$

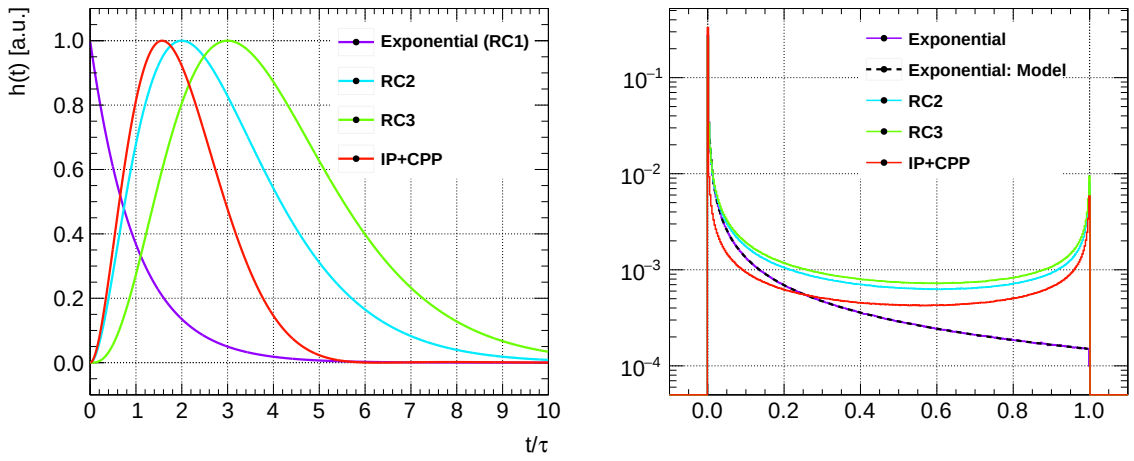
where $h(t)$ denotes the normalized response of the readout electronics to a single pixel hit, having a characteristic time constant τ_p for the following discussion, and V_{pe} amplitude of the single pixel signal. The theorem is very convenient in it's simplicity and can be directly calculated in the s-domain without the need for an inverse transfer function using Parcevalls theorem. For the single pixel signal noise analysis however, it is not applicable in the general case: The theorem gives the standard deviation of the full spectrum rather than the smearing of the peaks in the single photon spectra. Depending on the actual shape of the dark rate charge spectra, the individual pixel discharge peaks may still be distinguished even if the variance of the spectrum given by the campbell theorem is comparable with the single pixel signal. The theorem is therefore only applicable for small variances predominantly coming from the smearing of the photoelectric peak, such as for small dark count rates ($\tau_{\mathrm{DCR}} \gg \tau_p$), or for components of the pulse which are much slower than τ_{DCR} such that many avalance signals are averaged. Since the analog system is linear and time invariant, the different components may be considered separately in the noise analysis.

To estimate an effective noise term due to DCR pileup, the pedestal spectrum including the DCR pulses is needed. The probability density for the time delay between a dark pulse and some fixed observation time is given by

$$P(t) = \frac{e^{-\frac{t}{\tau_{\mathrm{DCR}}}}}{\tau_{\mathrm{DCR}}} \tag{5.12}$$

If the pulse is short, the probability to have more than one pulse contributing to the voltage

[1] Neglecting the correlated noise



(a) **(b)**

Figure 5.3: Simulated DCR pileup considering one preceeding DCR pulse for different impulse response functions with time constant τ a) Pulse response. b) Pedestal spectrum

at the observation time is small and only the last pulse needs to be considered. Summing over all solutions of $h(t) = v/V_{pe} \equiv \hat{v}$, the shape of the charge spectrum is given by

$$Q(\hat{v}) = \sum_t P(t) \left| \frac{\mathrm{d}h}{\mathrm{d}t} \right|^{-1} \quad \text{for } t = h^{(-1)}(\hat{v}) \tag{5.13}$$

Figure 5.3a shows the pulse response of the Filters with different order which have been discussed in chapter 4. The single pixel response (amplitude) has been normalized to one, and the characteristic time constant is set to 1 for all cases. Using a toy monte carlo model, pedestal spectra are obtained as shown in Figure 5.3b.

For a exponentially decaying pulse response $h(t) = V_{pe} e^{-t/\tau_v}$, equation 5.13 gives a concise expression for the charge spectrum

$$Q(\hat{v}) = \frac{\tau_p}{\tau_{\mathrm{DCR}}} \hat{v}^{\left(\frac{\tau_p}{\tau_{\mathrm{DCR}}} - 1\right)} \qquad \text{with } \hat{v} = \frac{v}{/} V_{pe}$$

Since the inverse function in 5.13 does not exist for the full definition range in the other cases, no analytical expression for the pedestal spectra exists. For such pulses, the spectra also show a peak at the single photon level ($\hat{v} = 1$) since the peak regions of $h(t)$ are smooth (*peak pileup*).

To overcome this limitation, and to find a range where the theorem is applicable for pSNR analysis, a simulation was developed and validated in measurements. The simulation will then be used to obtain an effective noise term for the dark count pileup, and to compare the different filter topologies and time constants beyond the possibilities available in measurement.

5.3.1 DCR pileup measurement setup

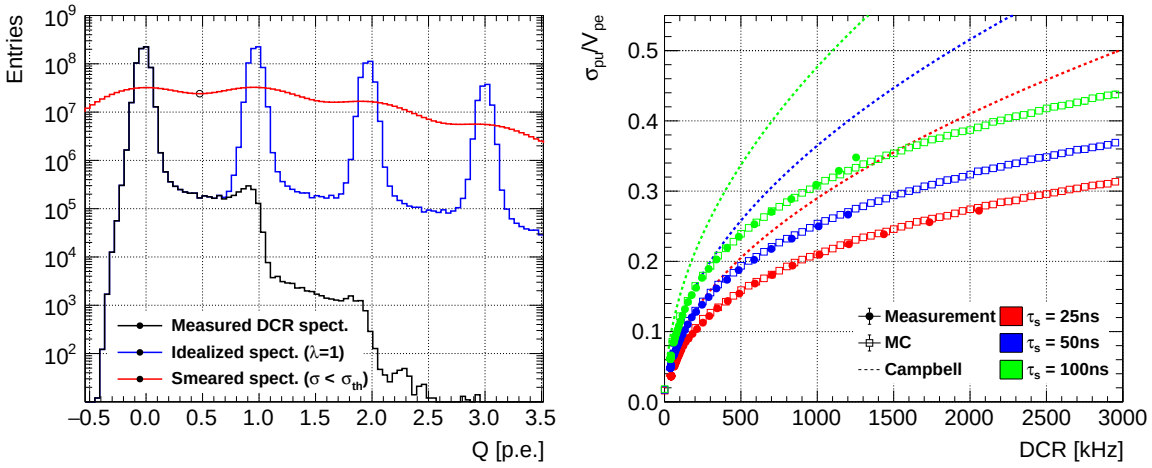
Because the dark count pulses are indistinguishable from pulses generated by photons, the DCR can be variated by applying a continuous light signal to the SiPM. The SiPM is read out using the fully analog KLauS2 ASIC [64]. The analog output of the chip is connected to a digital storage oscilloscope storing the waveforms on a PC. A green LED controlled by a programmable DC source is used to generate a constant light to model the DCR variations without changing other SiPM parameters. In addition, a fast laser is used to generate a pulsed light signal in order to measure the SiPM gain. To reduce the intrinsic dark counts of the SiPM and to keep the sensor in a stable operating condition, the setup is kept in a temperature controlled box and cooled down to 10°C. The LED voltage is varied in small steps, and the transient output voltage is recorded. To obtain the rate of pulses, i.e. the mimicked dark count rate, the oscilloscope traces are analysed using a peak finding algorithm [34], and the time delays between the hits are recorded. By fitting the time-difference spectra with an exponential function , the rate of pulses from the real dark counts and the light from the LED is extracted (5.12). The sensor gain is obtained from the peak distances in the recorded single photon spectra. With the laser switched off, the dark-rate charge spectrum of about $5 \cdot 10^8$ samples is recorded. The black line in Figure 5.4a shows an example of a recorded DCR spectrum.

5.3.2 Simulation

To simulate the SiPM output signal digitized by the scope, a SiPM simulation framework [65] was used. The software generates a list of SiPM pixel avalanches with time and amplitude informations, using basic sensor parameters such as DCR, PDE, crosstalk and afterpulsing probabilities, etc. This list is then used to generate a transient voltage signal from an impulse response function. As in the measurement case, the voltage samples are collected in voltage spectra for every simulated DCR setting. All sensor and setup parameters, most importantly the sensor gain and binning of the resulting histograms, are matched to the measurement results and the oscilloscope's resolution, respectively. The impulse response function parameters are extracted from a measurement of the average pulse sampled by the oscilloscope, requiring that no other dark count pulse that would change the recorded response function is captured before. The average pulse is fitted to the model impulse response function of the KLauS2 analog front-end's transfer function [5]. The pulse response model function has five free parameters: The single pixel gain, magnitude of a low frequency component due to a low frequency feedback in the chip, and three time constants for integration, shaping, and the low frequency feedback.

5.3.3 Effective noise term

To extract an effective noise term from the measured and simulated spectra, the separation power of pedestal and first photo electron peak is observed. The separation of the two peaks heavily depends on the ratio of the two, given by the average number of pixels fired at the same time. To simplify the analysis, the peaks are expected to have the same magnitude, i.e. the average number of fired pixels is 1. Controlling the light level accurately is challenging in the measurement setup, therefore the dark rate spectra are used to estimate the noise term, and an idealized spectrum is generated. The idealized spectrum is constructed by



(a) **(b)**

Figure 5.4: Effective DCR pileup noise measurement and simultaion. a) Measurement and simulation for all shaping time settings in the KLauS2 ASIC. b) Noise term components and comparison to ideal response for the 50ns shaping time setting of KLauS2

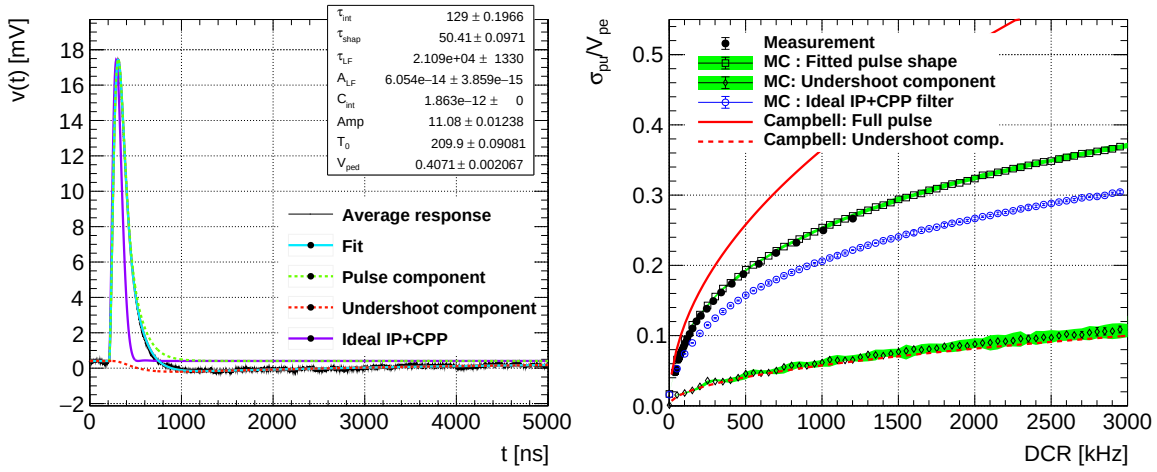
summing dark rate spectra shifted by the gain and weighted by the Poisson distribution. The parameter of the distribution is chosen to be $\lambda = 1$ such that the first two peaks have the same magnitude (blue line in Figure 5.4a). The spectra are then smeared by convolution with a normal distribution $N(0, \sigma)$ to extract an effective noise term,

$$\sigma_{pu} = \sqrt{\sigma_0^2 - \sigma_{th}^2}, \tag{5.14}$$

Where σ_0 is the constant noise of the system and σ_{th} is the smallest noise artificially added by convolution where the folded spectrum $Q_f(v) = Q(v) * N(0, \sigma)$ looses the minimum between the pedestal and first photo electron peak (i.e. the peaks can not be distinguished). For a system without additional noise contributions and infinite binning, the term σ_0 is half of the single photoelectron gain $(0.5 \cdot V_{pe})$.[2] For the measurement case, σ_0 is calculated from a noise measurement below the breakdown voltage of the sensor. For the noise free simulation case, only binning effects contribute and $\sigma_0 \approx 0.5$. Figure 5.4a shows one of the measured spectra (black), the idealized spectrum (blue), and a smeared spectra (red) generated to find σ_{th} where the local minimum at 0.5 pe disappears. A binary search algorithm is used to find σ_{th}. In Figure 5.4b, a comparison of measured and simulated effective noise terms is shown for all shaping time configurations available in the KLauS2 ASIC. The noise term calculated using the campbells theorem is plotted as dashed lines. As expected, it is comparable to the measurement results for small dark rates and is highly overestimated for large count rates.

Because the simulation is based on a mathematical model of the impulse response well describing the measured impulse response, the different pulse components can be separated

[2] This can be proven solving $\frac{\mathrm{d}^2}{\mathrm{d}x^2}\left(N(0,\sigma) + N(1,\sigma)\right)\Big|_{0.5} = 0$ for σ, i.e. the curvature of the spectrum between the peaks vanishes at the threshold width.



(a) (b)

Figure 5.5: Pulse contributions of the effective DCR pileup noise. a) Fitted pulse used for simulation and separated components b) Noise term components and comparison to ideal response for the 50ns shaping time setting of KLauS2

as shown in Figure 5.5a. This allows to simulate their contributions separately as shown in Figure 5.5b for the 50ns shaping time configuration. The slow undershoot component of the pulse has a negligible contribution since the amplitude is much smaller than the actual pulse. Due to the longer time scale, the undershoot pulse is averaged over many dark pulses and the simulation is in agreement with the expectations given by Campbells theorem. In the signal processing stage of the KLauS2 ASIC, the integration and shaping time constants are not exactly matched, causing a slightly longer tail compared to the ideal IP+CPP shaper. The blue data points in Figure 5.5b show the pileup noise assuming a response function with matched time constants $\tau_{shap} \approx 50$ns (red function in Figure 5.5a).

The simulation can also be used to investigate the effective noise term for other shaping topologies, as well as the shaping time constant itself. As a typical value for SiPMs with 1mm^2 area, the dark count rate is fixed to 100kHz, and the simulation is performed as a function of the shaping time constant. Since the contribution from the low frequency feedback turns out to be negligible, it is not included in the simulation. With only two time constants present, the effective noise term is dependent only on the ratio of the τ_p and τ_{DCR}. Figure 5.6 shows the simulation results as a function of the shaping time constant for the same filter topologies discussed previously. While the exponentially shaped signal has the lowest noise due to the fast decay, also it's filter order is the lowest, resulting in higher electronic noise. For the CR-RCn filters, the pileup noise term increases with order due to the increasingly long pulse. The IP+CPP shaper has an order of 3, but faster decay compared to the CR-RC2 and is therefore preferred. Because the effective noise contribution is proportional to the SiPM gain, it is typically the biggest contributor for large gain SiPMs, favoring small time constants of the electronics. For small gain sensors, the contributions are smaller and the terms independent on the sensor gain, such as the electronics noise, will have a bigger influence. The noise optimization for the KLauS front-end in the following chapter will reveal however that the pileup noise is still significant and has an influence on

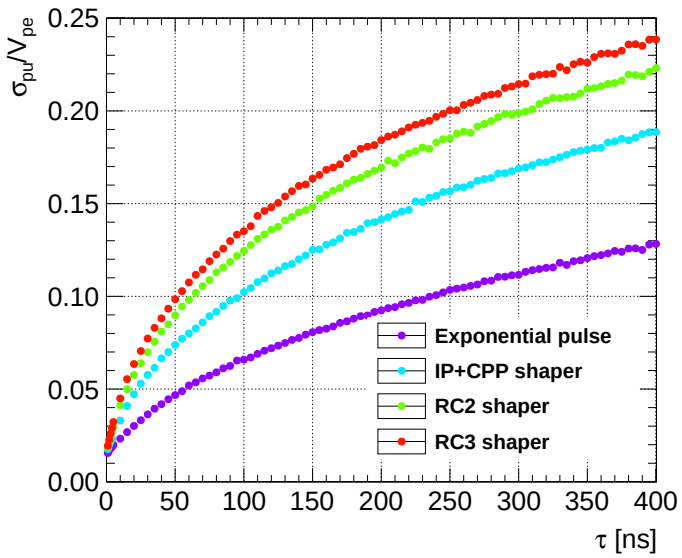


Figure 5.6: Simulation of pileup noise as a function of the shaping time for different filter topologies

the optimum shaping time.

5.4 Time measurement and the impact on voltage sampling

When a number of pixels in the SiPM fire after an incoming light pulse generated by the scintillator, the charge measurement branches generate an analog signal which is digitized in case the signal exceeds a certain threshold and the trigger comparator fires. Because the hit rate is small, it is sufficient to digitize the pulse maximum rather than the full waveform. Consequently, the peaking time of the analog signal should be well defined by implementing a proper shaping scheme, and the comparator generating the sampling signal should fire at a fixed time relative to the peaking time. Since the trigger decision will happen at a time t_{trig} before the analog signal reaches it's maximum, the trigger signal is used to generate a digital *sample* signal. Rising with the comparator decision at t_{trig} and falling at the peak maximum (t_{hold}) generated by shifting the comparator decision by a fixed delay, this sampling signal is used by the ADC to sample the analog voltage and hold it at the pulse maximum as illustrated in Figure 5.7a. The charge pulse generating from the detector is subject to variations due to the statistical nature of the scintillation, avalanche formation and generation and other SiPM related processes [49, 66, 67]. The impact of these processes on the time measurement accuracy have been studied in detail in the scope of applications requiring a very high timing resolution, for example the time of flight positron emission tomography (TOF-PET). While in these applications the long time constants of the scintillation processes play an important role, the organic scintillators used in calorimetry or other HEP experiments have much shorter time constants and the electronics performance may be a dominating term to the time resolution [34], particularly if the power consumption allowed to spend on the timing measurement is limited.

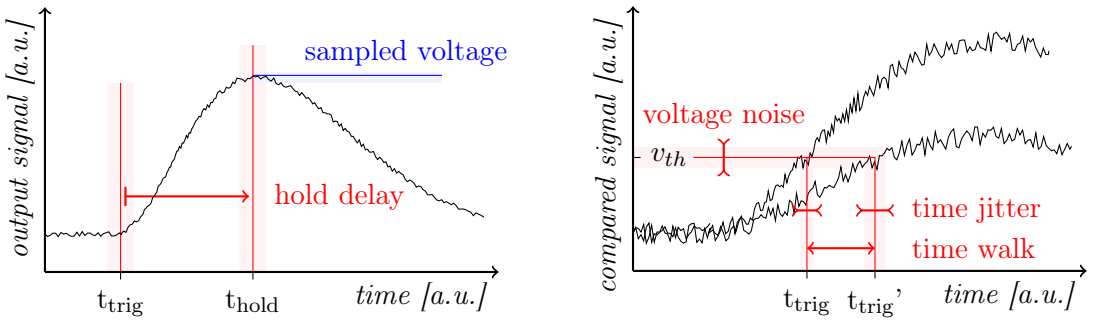


Figure 5.7: a) Illustration of the peak voltage sampling using an internal trigger comparator. b) Sources of trigger descision time uncertainties and distortions.

One commonly used and simple scheme to generate a trigger decision is the *leading edge discrimination*. As indicated in Figure 5.7b, the trigger condition is met once the signal in a separate comparator branch passes a fixed threshold. The time constants for this decision process are usually much shorter than the shaping time of charge measurement circuits, and the leading edge discrimination scheme can be used to generate the sampling signal for the charge measurement. To study effects of the comparator decision time on the voltage sampling, the pulse response of the compared signal is modelled by two time constants. The current signal from the sensor has a time constant τ_i, equivalent to $R_{in}C_{det}$ for a large input impedance of the electronics or detectors with a large detector capacitance, and otherwise dominated by the frequency response of the electronics in front of the comparator circuit. The parasitics at the compared node introduce another, longer time constant τ_c. The transfer function and pulse response of this system is simply

$$H_{\mathrm{tot}} = H_i \times H_{\mathrm{c}} = \frac{1}{s\tau_i + 1} \times \frac{R_c}{s\tau_c + 1} \tag{5.15a}$$

$$h(t) = \frac{R_c}{\tau_c - \tau_i}\left(e^{-\frac{t}{\tau_c}} - e^{-\frac{t}{\tau_i}}\right) \tag{5.15b}$$

where R_c is the impedance at the comparing node which converts the input current to a voltage signal which is then discriminated by the comparator circuit.

5.4.1 Time jitter

The contribution of the readout electronics to the time measurement uncertainties is called the *time jitter* and is given by the noise and limited rise time of the signal used for comparison.

$$\sigma_{t,el}^2 = \frac{\sigma_v^2}{\left(\frac{\mathrm{d}}{\mathrm{d}t}v(t)\right)^2} \qquad \text{for } t:\ v(t) = v_{th} \tag{5.16}$$

In order to minimize the jitter, voltage signal amplitude of the compared signal (5.15b) is made large by maximizing R_c by the use of cascode transistors, and the parasitic capacitance at the comparing node is minimized to obtain a fast rising component of the pulse limited

by τ_i. Because the slope of the signal increases with larger signal amplitude, the time jitter decreases monotonically and is then limited by the constant jitter introduced in the following digital cells and the time to digital converter. For a given charge, the total uncertainty of the hold time including the time jitter introduced by the comparator and the delay circuit $\sigma_{t,h}$ leads to a noise term of the sampled voltage, σ_{hold}. As the derivative of the analog output signal h(t) vanishes at the pulse maximum, the impact of uncertainties in the hold-time are calculated up to second order and the pulse response h(t) of the sampled analog voltage is approximated at some hold time t_{hold} as:

$$h(t) \approx h(t_{\mathrm{hold}}) + h'(t_{\mathrm{hold}})\ (t - t_{\mathrm{hold}}) + \frac{1}{2} h''(t_{\mathrm{hold}})\ (t - t_{\mathrm{hold}})^2 \tag{5.17}$$

For the probability distribution of the hold time T with the density function P(t), the mean sampled voltage is given by

$$\begin{aligned} \mu_v = E\left[h(T)\right] &= \int_{-\infty}^{+\infty} h(t)\, P(t)\, \mathrm{d}t \\ &\approx h(\mu_T) + \frac{1}{2}\, h''\left(\mu_T\right)\ \mu_{T,2} \end{aligned} \tag{5.18}$$

and the variance of the sampled voltage is

$$\sigma_{hold}^2 \approx \left(h'\left(\mu_T\right)\right)^2 \mu_{T,2} + \left(h'\left(\mu_T\right)\ h''\left(\mu_T\right)\right)\ \mu_{T,3} + \frac{1}{4}\left(h''\left(\mu_T\right)\right)^2 \left(\mu_{T,4} - {\mu_{T,2}}^2\right) \tag{5.19}$$

Here, μ_T is the average hold time and $\mu_{T,k}$ are the k-th central moments of T:

$$\mu_T = \mathrm{E}\left[(T)\right] = \int_{-\infty}^{+\infty} t\, P(t)\, \mathrm{d}t \tag{5.20}$$

$$\mu_{T,k} = \mathrm{E}\left[\left(T - \mathrm{E}[T]\right)^k\right] = \int_{-\infty}^{+\infty} (t - \mu_T)^k\, P(t)\, \mathrm{d}t. \tag{5.21}$$

Using the pulse response of the complex pole shaping stage introduced in Chapter 4 normalized to an amplitude U_{pk},

$$h(t) = V_{pk} e^{-\frac{t}{\tau_{\mathrm{int}}} + \frac{\pi}{2}} \left(1 - \mathrm{Cos}\left[\frac{t}{\tau_{\mathrm{int}}}\right]\right) \tag{5.22}$$

the voltage noise due to hold time variations for a normal-distributed hold time distribution with mean $\mu_T \equiv t_{pk} = \frac{\pi \tau_{\mathrm{int}}}{2}$ and width $\sigma_{t,h}$ can be calculated from (5.19) and simplifies to

$$\sigma_{hold} = V_{pk} \frac{\sigma_{t,h}^2}{\sqrt{2}\tau_{\mathrm{int}}^2} \tag{5.23}$$

For typical values, $\tau_{int} = 50\,\mathrm{ns}$ and $\sigma_{t,h} = 1\,\mathrm{ns}$, the introduced noise from the hold time jitter is $\approx 2.8 \cdot 10^{-4} U_{pk}$, which is again a small term compared to other noise sources.

In reality, the hold time can not be placed exactly at the pulse maximum as the adjustment resolution of the delay will finally be limited. This introduces additional terms to (5.23)

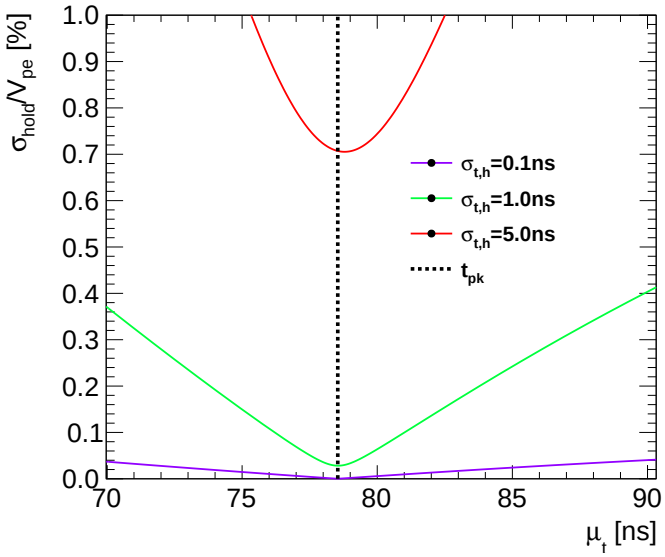


Figure 5.8: Noise introduced by sampling time jitter as a function of the average sampling time. The ideal value of t_{pk} is indicated by the dashed line.

as the terms with $h'(t)$ in (5.19) do not vanish any more. Figure 5.8 shows the introduced sampling noise for an input signal with amplitude 1 as a function of the hold time for $\tau_{int} = 50\,ns$ and normal distributed hold time distributions with different values of $\sigma_{t,h}$. The reduction of μ_v due to the sampling time jitter (Eq. 5.18) is always less than 1% at the peaking time. In summary, the noise peak sampling introduced by the time jitter is negligible compared to other sources for $\sigma_{t,h} < 0.1\sigma_{int}$, a constraint that is already covered by the requirements on the time resolution in the AHCAL application, assuming shaping times in the order of 50 ns. To avoid degradation of the average amplitude and a significant increase of the sampling noise, the hold time should be adjustable with a resolution of better than 2 ns such that these effects stay negligible.

5.4.2 Distortions from timewalk effects

While the leading edge comparison scheme can produce a low-jitter trigger signal to create a time-stamp, a drawback is the effect of time-walk as the input signal amplitude changes with respect to the trigger threshold as shown in Figure 5.7b With increasing amplitude, the comparator decision happens earlier in time, also moving the hold time with respect to the peak maximum when the comparator decision is used to generate the sampling gate. Consequently, the hold time changes with amplitude. As discussed in the previous section, this would increase the noise introduced by sampling time jitter. The main concern however is that the sampled peak amplitude changes, introducing nonlinearities in the peak voltage. The effect can be mitigated by using a different comparison scheme such as the zero crossing timing or constant fraction discrimination [61]. Here, the trigger threshold is effectively varied proportional to the signal amplitude, leading to a comparison decision time independent on the signal amplitude. This however comes with the cost of increased power consumption due to the additional signal shaping, and slightly worse time jitter.

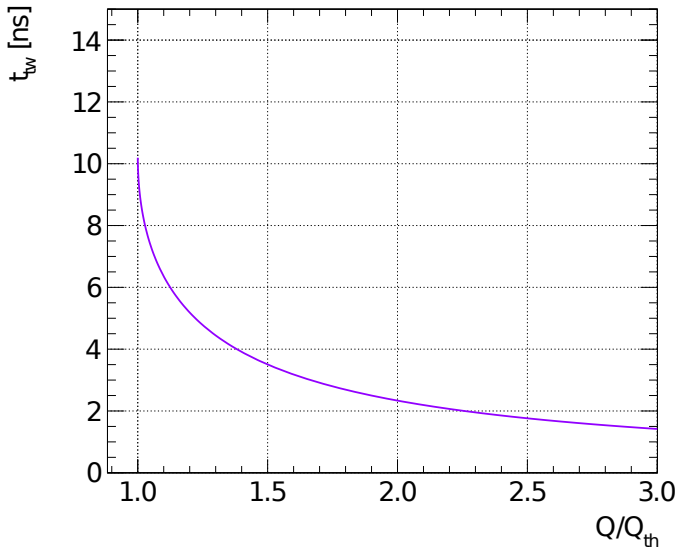


Figure 5.9: Comparator decision time (time walk) as a function of the input charge.

For the time response of the compared voltage signal normalized to the comparator threshold Q_{th}, $V(t) = Q/Q_{th}\,\hat{h}(t)$, the time walk for the leading edge discrimination is given by the time when $V(t)$ crosses $Q_{th} \equiv 1$

$$t_{tw} = V^{-1}(Q_{th}) = \hat{h}^{-1}(Q_{th}/Q) \tag{5.24}$$

The hat ($\hat{h}$) in the above equation denotes the impulse response (5.15b) normalized to the threshold charge Q_{th}. The time-walk effect is vanishing in the limit of $Q \gg Q_{th}$, and reaches a maximum given by the peaking time of (5.15b) for small signals close to the threshold.

$$t_{tw,max} = t_{pk} = \frac{\tau_c \tau_i}{\tau_{\mathrm{isg}} - \tau_c} \log\left(\frac{\tau_{\mathrm{isg}}}{\tau_c}\right) \tag{5.25}$$

For typical values reachable in the design, $\tau_i = 7\,\mathrm{ns}$ and $\tau_c = 16\,\mathrm{ns}$, the peaking time is $\approx 10\,\mathrm{ns}$. In Figure 5.9, the time walk as a function of the input charge divided by the comparator threshold is shown. Due to the time walk effect, the sampling time $t_s = t_{pk} + t_{tw}(Q) + \delta$ is shifted from later to earlier times with increasing input amplitude. Here, t_{pk} denotes the peaking time of the sampled analog pulse, $t_{tw}(Q)$ is the time walk and δ is a time offset. The sampling voltage error

$$\delta_v(Q) = Qh(t_pk + t_{tw}(Q) + \delta) - Qh(t_{pk} + t_{tw}(\infty) + \delta) \tag{5.26}$$

is plotted for different values of δ in figure 5.10.

If the hold time is placed exactly at the peak maximum at the limit of large charges, the introduced nonlinearity is monotonic and reaches a maximum of about 2% at the comparator threshold. When the hold time is placed slightly before, the nonlinearities get nonmonotonic, but decrease for the low charge regions since the absolute changes in the hold time relative to the peak get smaller. Since the threshold is assumed to be put just above the DCR noise

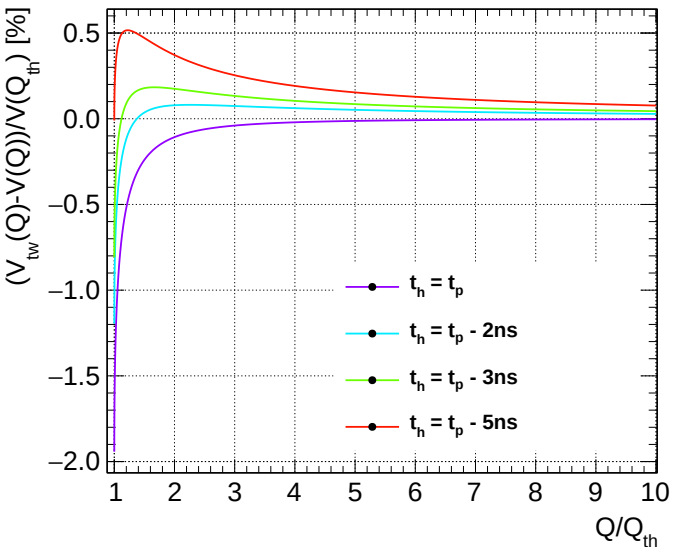


Figure 5.10: Voltage sampling errors as a function of the input amplitude for different offsets of the shaping time

threshold at around 3 PE, the relative nonlinearities given the large dynamic range of the analog processing stage are always small, and distortions in the converted charge close to the threshold are still acceptable.

5.5 ADC resolution

After processing the analog signal in the front-end, the peak voltage is digitized by the analog to digital converter. The limited quantization resolution gives rise to another noise term

$$\sigma_{\mathrm{ADC}} = \frac{1}{\sqrt{12}} V_{LSB} \tag{5.27}$$

where V_{LSB} is the bin size of the ADC, given by the number of quantization bits and the input voltage range.

$$V_{LSB} = \Delta V_{in} \cdot 2^{-N_{bits}} \tag{5.28}$$

While for an ideal ADC the bin sizes are constant, a realistic ADC will show nonlinearities and the bin size varies for the different codes and an average bin size differing from (5.28).

The differential nonlinearity (DNL) gives the relative difference of the bin size,

$$\mathrm{DNL}_i = \frac{w_i}{< w >} - 1 \tag{5.29}$$

where w_i is the bin size for the code i determined from the transition edges of the ADC, and $< w >$ is the average bin size given by (5.28) or determined from a linear fit. With the normalization, the DNL is given in units of least significant bits (LSB).

The integral nonlinearity (INL) gives the absolute deviation to the expected code

$$\mathrm{INL}_i = \sum_{j=0}^{i} \mathrm{DNL}_i \tag{5.30}$$

The nonlinearities of the ADC can be measured using code density tests from a known input signal, for example a slow ramp or sine wave or by scanning the input voltage in small steps [68, 69]. The measured DNL pattern can be used to correct for the ADC nonlinearities in software, or online in the digital part (see e.g. [70] for a description of different calibration approaches). However, the differing bin sizes effectively contribute to the noise as the bins with larger size have larger quantization error and potentially contribute more to the average error since a larger portion of the input spectrum will be captured in these bins. Apart from the nonlinearities themselves which can be corrected if precisely known, nonlinearities should also be kept small from the perspective of noise minimization. While both the DNL and INL have been introduced here to describe the nonlinearities of an ADC, they can also be applied to the characterization of other quantizing blocks, for example the linearity of digital to analog converters (DACs) where the bin size is given by the change of the DAC output per LSB.

6 KLauS: A low power SiPM charge readout ASIC

The KLauS (**K**anäle für die **La**dungsauslese von **S**ilizium-Photomultipliern) ASIC is developed to be used in the field of the highly granular calorimeters at future linear colliders. The design goal of the chip is to reach a high precision charge measurement with an emphasize on a low power consumption and power pulsing capabilities. The first versions of the KLauS ASIC have been designed in the 350nm AMS SiGe technology until 2010 [5] and were, except for the configuration interface, fully analog. In these ASICs, different readout topologies of the analog front-end were studied. In the framework of the CALICE collaboration, the effort was continued to design a mixed-mode, multi-channel readout ASIC which can be used in the CALICE AHCAL prototypes and the final ILD detector. The focus of the KLauS ASIC development is to provide a readout solution for small gain, high dynamic range sensors which can not be used with the current AHCAL electronics.
Since the noise coupling from the digital part into the analog circuits was expected to be significant and the area of the digital part was expected to be too large in the AMS technology, the ASIC was redesigned in the more advanced UMC CMOS technology with a smaller minimum transistor length of 180nm. In this technology, buried well n-MOS transistors are available which allow to decouple the digital noise activity on the substrate from the analog blocks effectively. While the overall readout topology of the KLauS2 ASIC, in particular the input stage, was identified to work also in the UMC180 technology, several modifications had to be implemented to keep the dynamic range of the charge measurement at the same level even at the reduced supply voltage in 180nm (1.8V instead of 3.3V). In parallel, a low power analog to digital converter was developed within the group. The front-end and ADC building blocks were produced in two iterations of single channel prototypes (KLauS3) [6–8], used to identify flaws in the new design before the production of a larger and more complex mixed signal ASIC. In summer 2016, the first multi-channel, mixed signal prototype ASIC KLauS4 was produced. The building blocks and key characteristics included in this ASIC will be described in the following chapter. Characterization results of the ASIC will be presented in Chapter 7.

Figure 6.1 shows the layout of the KLauS4 ASIC. It consists of 7 channels, each with an analog front-end, ADC and digital control logic. The digital part was designed to be scalable to the full chip of 36 channels and already includes all the functionalities needed to operate the ASIC in the AHCAL prototypes. In summer 2017, a first version of the 36 channel ASIC has been produced. While all of the ASICs produced in the UMC180 technology were developed during and as part of this thesis, the description of the ASIC following will only cover the seven channel prototype ASIC.

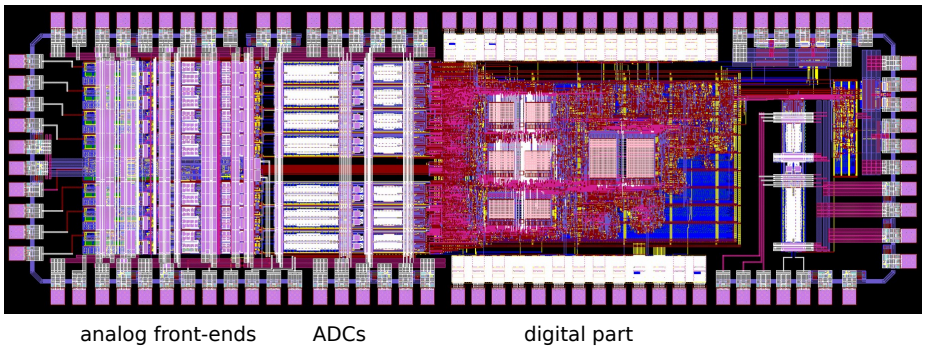


Figure 6.1: Layout of the produced seven channel KLauS4 ASIC.
The dimensions of the chip are $1.5 \times 4.5\text{mm}^2$.

6.1 Design considerations drawn from AHCAL requirements

For the analog hadronic calorimeter being developed for the ILD detector, the light pulses from the plastic scintillator tiles are read out by Silicon Photomultipliers. Each sensor is connected to the readout ASICs integrated into the layers of the calorimeter, measuring the charge and time-stamp of the events. For the dense electromagnetic showers in the ECAL and subshowers in the HCAL, the energy deposits in a single cell can be large, and sensors with a high dynamic range are desirable. For the ECAL, a dynamic range of more than 10^4 is required. The dynamic range of the SiPMs given by a large number of pixels in one sensor comes with the cost of a lower single-pixel gain. For currently available high dynamic range sensors with a pixel pitch in the order of 10μm and $O(10k)$ pixels in one sensor, the single pixel gain approaches 10^5. For the gain calibration of the sensors, these charge signals should be resolved by the readout electronics. The required equivalent noise charge needed to distinguish the individual pixels is then at the level of 5fC. At the same time, the electronics need to be able to measure also the maximum charge that can be delivered by the sensors, not affected significantly by the reduced pixel size and in the order of 150pC. The dynamic range range required from the readout electronic circuit is about 90dB, which would necessitate an analog to digital converter with an effective number of bits of about 15. The full resolution is however only required for the calibration of the sensor gain, i.e. the small charge region up to about 10 fired pixels. For the larger signals, the signal fluctuations due to gaus-landau fluctuation and poisson smearing due to photon statistics are always dominating the measurement uncertainty accuracy and the single pixel signals do not need to be resolved. In order to allow an in situ calibration from tracks in the calorimeter, both the small and large charge range need to be accessible at the same time , and the electronics are required to provide the small signal measurement with high resolution as well as the large signal range without changing the configuration.

The ILD detector has no external trigger system deciding which of the events should be read out. The readout electronics are required to operate in an auto-triggered fashion, storing the charge information only if the charge signal has passed a certain threshold. The internal trigger in each of the channels is also used to record the time stamp, which might be utilized to improve the particle flow algorithms. For the time stamps, a timing resolution in the order of $\lesssim 1\,\text{ns}$ for single minimum ionizing particles is required. In addition, the

trigger is used to initiate the peak voltage sampling of the charge signal, which requires the trigger to have a short delay compared to the timescale of the charge output signal and small time-walk.
The dead time of the channel after a trigger decision should be in the order of the bunch crossing rate[19] of about 550ns.

Since the Silicon Photomultipliers are subject to variations of the breakdown voltage due to production tolerances and temperature changes, a method needs to be provided to make the response throughout the detector uniform. This is accomplished by tuning the DC voltage at each input in a small range of the expected breakdown voltage fluctuations.

A key role choice of the design topologies is dictated by the requirement of the very dense and highly granular calorimetry system. The readout ASICs are integrated in the layers, and in order to obtain the best sampling fraction and shower containment the amount of material which is not absorber nor scintillator should be minimized. To avoid active cooling in the layers, the readout electronics are required to consume only a minimum of power. The only devices being cooled are the power regulators for the chip and the DAQ hardware combining the data from the ASICs, residing at the side of the calorimeter barrel. Each readout channel in the calorimeter is required to consume less than 25μW, not including the power consumption of the Silicon Photomultipliers. To meet this constraint the readout-electronics are switched to a low power mode when there are no collisions expected from the ILC accelerator. The beams are only colliding in bunch trains of less than 1ms followed by a gap of 200ms. In this time, the detector is read out and the power consuming blocks of the front-end readout ASICs are switched off. These power pulsing techniques are an important aspect of the design and must be implemented into the chip.
As the digital readout of the ASICs will be performed using a slow link, all events need to be stored on the ASIC until read out. Initial simulations carried out without taking the full detector occupancy into account suggest that a buffer capable of storing 16 events per channel suffices in order to avoid overflow conditions and ensure efficient data taking [71].

6.2 The KLauS4 channel

A block level schematic diagram of the KLauS channel is shown in Figure 6.2. It is comprised of an analog front-end generating the analog charge information and trigger decisions from the current pulse generated by the sensor, and an ADC which digitizes the analog information. After registering the trigger signal from the front-end, a digital control circuit initiates and controls the analog to digital conversion and passes the digital data of the hit to the following digital part responsible to combine, buffer and send the digital data to the DAQ hardware. The bias currents and voltages required by the front-end blocks are generated in a bias block shared by all channels. In the following section, the analog part of the channel generating the charge and time informations will be presented. In Section 6.4, the analog to digital converter design is discussed, followed by a description of the digital parts of each channel and the full ASIC. The necessary additions required to implement the power pulsing capabilities of the chip, i.e. both the analog and mixed signal blocks, are discussed in Sections 6.7.

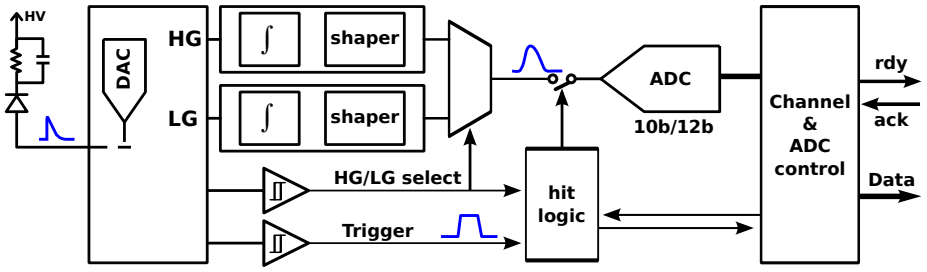


Figure 6.2: Block diagram of the analog front-end of a KLauS channel.

6.3 Analog front-end

The analog front-end is the part of the channel processing the charge and time information in the analog domain. It consists of several blocks also shown in Figure 6.2. An input stage of the channel buffers the signal current and distributes it to the charge measurement and comparator branches. It also implements the tuning of the SiPM bias voltage required to compensate the SiPM gain fluctuations. The charge measurement is performed in two separate branches ("high gain","low gain") with different dynamic range to cover the full required range of charge amplitudes. Two comparator branches are implemented in the front-end. The time comparator is used to record the time stamp of the hit and initiates the analog to digital conversion. The threshold of this comparator is placed at the level of few fired pixels, high enough to minimize the triggering rate due to dark pulses while keeping a high trigger efficiency for single particles traversing the scintillating tile. The second comparator is placed at a higher threshold close to the maximum charge of the high gain branch. It is used to select which of the analog signals from the two branches is sampled and digitized by the ADC. The correct communication with the digital part controlling the ADC sampling and conversion phases is ensured by a custom digital part in the front-end. It will be discussed in the scope of the digital control logic of the channel in Chapter 6.5.

6.3.1 Input stage

Figure 6.3 shows the schematic diagram of the KLauS input stage. The block is designed to provide a low input impedance such that the charge signal from the sensor is collected in the electronics in small timescales even for large area sensors. The DC input terminal voltage can be tuned in a small range in order to adjust the sensor gain and to compensate for different breakdown voltages between the channels. The signal current buffered by the input stage is copied to all the later processing stages: the two branches devoted to charge measurements, and the two comparator branches to implement the trigger and automatic gain selection functionalities. The chosen topology is a current conveyor based on a common-gate structure with voltage feedback, which allows to obtain the small input impedance, the SiPM bias tuning, a low power consumption and the power pulsing functionalities. The signal current is buffered by the input transistor M_1 is copied to the feedback path and to the charge processing branches using an NMOS current mirror. The mirror ratios are different for each of the output connected to the high gain, low gain, and trigger branches, as well as the feedback path in the input stage. The current mirror based on transistors M_2, M_3 and the triode M_t copies the signal to generate a feedback voltage which is used to lower the

input impedance. The magnitude of the voltage swing is determined by the mirror ratio and the impedance to ground at the feedback node, given by the inverse transconductance of M_4. When the channel length modulation resistors of the input stage transistors are neglected, the small signal impedance of the input stage at low frequencies can be calculated from the current flowing through the input terminal, and the resulting voltage swing at the feedback node

$$i_{in} = g_{m1}\left(v_{in} - v_{fb}\right) \tag{6.1a}$$

$$v_{fb} = i_{in} N \cdot \frac{1}{g_{m4}} \tag{6.1b}$$

where $i_{in} = i_{out}$ is the current through the input, copied to the feedback branch an effective scaling factor of N using the NMOS current mirror. The finite input impedance generates a voltage swing v_{in} at the input when a signal current is injected into the channel. The voltage swing is compensated by the feedback as the voltage at the source of M_1 (the input voltage) follows the gate of M_1 (v_{fb}). The voltage swing v_{fb} in the feedback loop is given by the fraction of signal current copied by the mirror and the impedance to AC ground given by the inverse transconductance of the source follower M_4. From (6.1b), the small signal input impedance at low frequencies is obtained as

$$R_{in} = \frac{v_{in}}{i_{in}} = \frac{1}{g_{\mathrm{m1}}} - N\frac{1}{g_{\mathrm{m4}}} \tag{6.2}$$

Since all the parasitic capacitors of the transistors as well as the large detector capacitance have been ignored, the frequency response of the input impedance and the current transfer to the output $H_{isg} = i_{out}/i_{in}$ are not analysed here. For sensors with an area of $\approx 1\mathrm{mm}^2$, the frequency response in the s-domain shows a set of two zeros and two complex poles which are dominantly generated by the C_{gs} and C_{gd} capacitances of M_1. For large area sensors, the time constant is given by the detector capacitance and the input impedance of the channel ($R_{in}C_d$). As a consequence of the limited bandwidth, the time constants of the integration and shaping stages can not be chosen too small in order to assure a good charge collection efficiency. The introduced limit on the shaping time constants that can be used

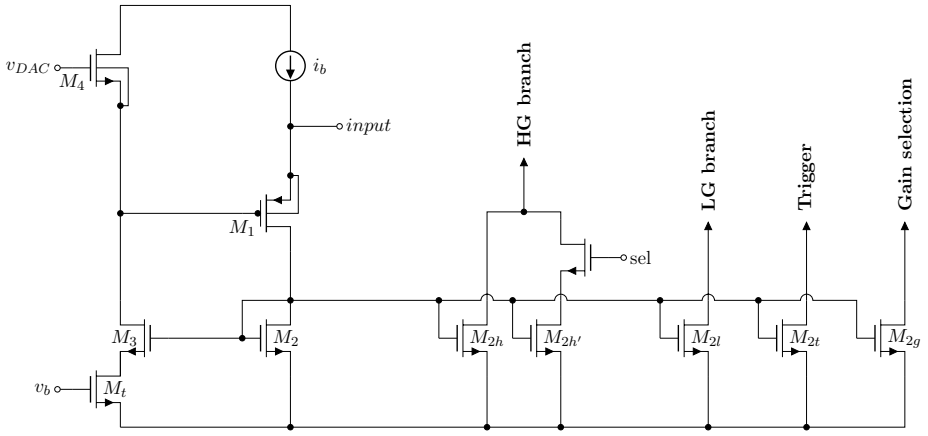


Figure 6.3: Simplified schematic diagram of the KLauS input stage.

in the charge measurement branches will have some impact on the noise performance of the input stage, as it will be discussed in Section 6.3.4.

An analysis of the frequency response including the dominant parasitics was conducted for the KLauS2 ASIC [5] and is also applicable to the KLauS4 input stage since the topology is very similar. The result can however not give much insight into the design optimization since it only includes parasitics from one transistor, and the results can not be used in the noise analysis because the complex noise integrals can not be solved with the additional poles included.

The input stage topology allows to set the DC voltage at the input terminal to adjust the SiPM bias voltage for each channel. Since M_4 and M_1 act as source followers, the voltage at the DAC terminal is reflected at the feedback node and finally the input terminal. To assure a good linearity of the bias tuning voltage with respect to the DAC input voltage, the bulk terminal of the source followers (M_1, M_4) are connected to their source terminals to suppress the body effect of these transistors. The upper and lower boundaries of the tuning voltage are given by the voltages where the transistors of the current source, and the NMOS-mirror enter the triode region.

$$(R_t I_f) + V_{dsmin,M3} + V_{gs1} < V_{in} < V_{cc} - V_{dsmin,SRC} \tag{6.3}$$

For the normal transistors and supply voltage of the UMC 180nm technology, this would allow a tuning range of only 0.5-0.7V. To extend the boundary to the maximum range, a transistor with a lower threshold voltage is used for M_1, and transistors with thicker gate-oxide are used for the input stage, allowing this block to be powered with a 3.3V supply. With these modifications, a tuning range of 2V is achievable. An important feature of the input stage topology is that the DAC terminal used to define the input terminal voltage is not loaded by the input stage. The voltage DAC may therefore be designed with very low power consumption. This is one requirement for the power pulsing capabilities to work, since the input terminal voltage should be stable also during power pulsing, which requires the DAC to be operational at all times.

In the calculation of the input impedance (6.2), the channel length modulation and body effects of the transistors have been neglected. When the voltage at the DAC terminal is changed and adjust the voltage at the input node, the input impedance may also change since the bias points of the transistors change. This results in the unwanted effect that the charge conversion factor changes as a function of the bias voltage, and the input impedance may get negative causing oscillations in some process corners. To compensate this effect, the triode transistor Mt is added to increase the output impedance of the current mirror, with a resistance of R_t.
This gives an effective transconductance of the mirror output transistor M_3:

$$g_{\mathrm{m3,eff}} = \frac{g_{\mathrm{m3}}}{g_{\mathrm{m3}} R_t + 1} \tag{6.4}$$

with the mirror scaling ratio and output impedance

$$N = \frac{g_{\mathrm{m3}}}{g_{\mathrm{m2}}} \frac{1}{1 + g_{\mathrm{m3}} R_t} \tag{6.5a}$$

$$\begin{aligned} r_o &= \frac{1}{g_{\mathrm{ds3}}} + R_t(1 + \frac{g_{\mathrm{m3}}}{g_{\mathrm{ds3}}}) \\ &\approx R_t \frac{g_{m3}}{g_{\mathrm{ds3}}} \end{aligned} \tag{6.5b}$$

While R_t is small and only increases the output impedance by about 30%, the value is tuned to counteract the input impedance change. With this method, R_{in} is stable within $1\,\Omega$ around the nominal value of about 60Ω.[1]

The higher supply voltage used in the input stage is used only to allow a larger tuning range for the input terminal voltage. This is not required for the later processing stages and would consume additional current in these blocks. To be able to use a lower supply voltage in the following chain of processing stages, the current is copied by the NMOS-mirror which has a common reference (ground) to the other supply voltage domains. As a drawback of this scheme, the input transistor M_1 has to be a PMOS device, which has a lower transconductance compared to the NMOS case due to the lower mobility of the holes. As a result, the size of M_1 is increased to keep the input impedance low, and the bandwidth is reduced.

The power pulsing capabilities and necessary additions to the input stage will be discussed in Chapter 6.7. Still the requirement to implement this feature plays a key role in the choice of the input stage topology, and some drawbacks of the input stage topology have to be accepted in order to compete with the stringent requirements on the power consumption.

6.3.2 Bias voltage tuning DAC

Since the voltage DAC connected to the input stage is always enabled even during power pulsing to keep the voltage at the input terminal stable, it must be designed with minimum power consumption. The voltage range of the DAC should be matched to the range of the voltage at the input terminal to use the whole range of output voltages provided by the DAC most efficiently. Since the voltage at the DAC terminal of the input stage is translated to the input terminal by the two source followers M_4&M_1, the tuning range at the input (6.3) translates into the required range of the DAC with $V_{in} \approx V_{DAC} + V_{GS4} + V_{GS1}$. This requires the block to use the 3.3V supply.

A schematic drawing of the DAC block is shown in Figure 6.4. An 8-bit current DAC operated in the subthreshold region generates a current which is converted to a DC voltage using a subthreshold operational amplifier with a fixed resistance in the feedback loop. The current DAC consists of 8 binary weighted and switchable current sources. The current is summed up and fed through the feedback resistor, resulting in a voltage at the output of

[1] The bias voltage v_b defining the triode resistance is derived from the input stage bias current generator, which reduces the effect of process variations on the input impedance due to the triode resistance value. For testing purposes, the triode resistance can also be changed by setting the bias voltage v_b manually with a voltage DAC. This allows to adjust the input impedance of the front-end.

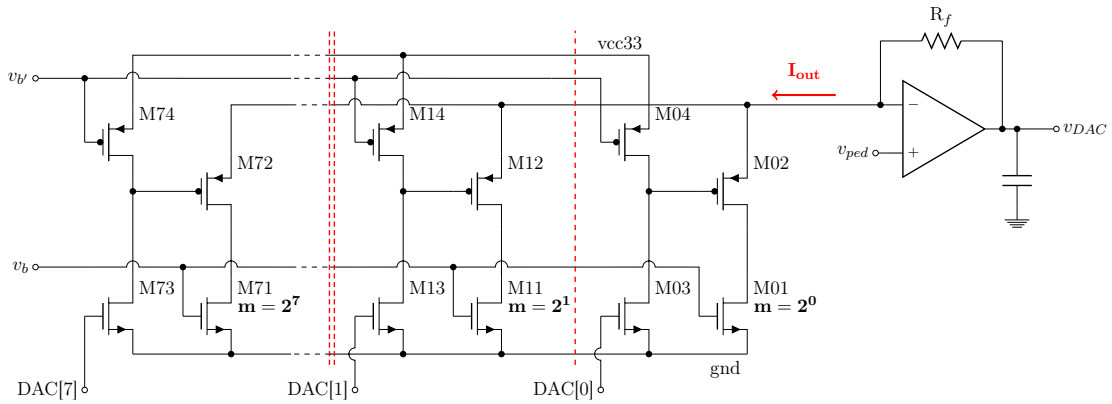


Figure 6.4: Schematic diagram of the low power voltage DAC.

the DAC given by

$$V_{out} = v_{ped} + R_f I_{out} = v_{ped} + R_f \sum_{i=0}^{7} b_i \cdot 2^i \, I_{\mathrm{LSB}} \tag{6.6}$$

where v_{ped} is the pedestal voltage of the operational amplifier defining the lower voltage range of the DAC, R_f is the feedback resistor with a nominal value of about 3MΩ, b_i is the i'th bit of the digital code and I_{LSB} is the unit current defining the voltage difference of the least significant bit. The subthreshold current is set to about 2.7nA/LSB to provide a minimum power consumption and the 2V tuning range, which requires the transistors to be large ($8\mu\mathrm{m}^2$ for each of the 255 transistors in the array). The current sources, implemented by the NMOS transistors $MX1$ and biased by a global generator circuit are switched by the transistors $MX2$. $MX3$ and $MX4$ act as level shifter to provide the correct voltage for the switch from the 1.8V digital control signals. The operational amplifier used to convert the current into a voltage output is also operated in the subthreshold region. As the pedestal voltage connected to the positive input of the operational amplifier is about 1.2V, it's dynamic range is designed to cover voltages between 1.2V and 3.3V. It consists of two stages, an NMOS input differential pair and a PMOS common-source output stage which can source the full current for the feedback. To ensure the stability of the amplifier in the full output dynamic range, the compensation the two stage amplifier carried out by miller compensation with a combination of metal-in-metal and MOS gate capacitances since the latter are not effective at output voltages close to the VCC rail where the MOS compensation capacitors enter the subthreshold region and do not provide sufficient capacitance. To filter the noise from the large feedback resistor and to stabilize the DAC against charge injections from the input stage which are only slowly recovered due to the low bandwidth of the operational amplifier, a capacitor of about 30pF is placed at the output of the DAC.

Bias current generation

The bias currents and voltages for the DAC are generated by a block commonly for all channels in the chip. The dynamic range of the DAC should be well defined and only show small differences from chip to chip to simplify the commisioning procedure of the full

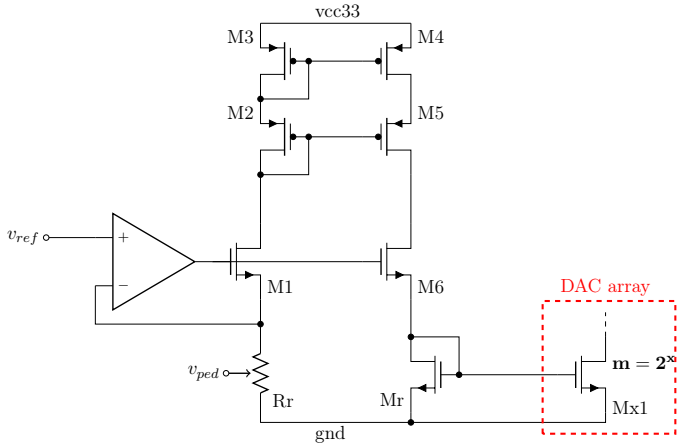


Figure 6.5: Circuit to generate the bias current for the low power DAC

detector. From (6.6) the LSB current required to obtain a 2V tuning range is calculated:

$$I_{\mathrm{LSB}} = \frac{2\mathrm{V}}{(2^8 - 1)\, R_f} \tag{6.7}$$

The LSB current should therefore be matched to the nmos device parameters and the resistor corner. These can vary significantly from wafer to wafer, and are subject to mismatches due to subthreshold operation. While there are several current reference circuits presented in literature that do or can be adapted to fulfill the matching against the process variations, e.g. [72–74], their mismatch would still be too large to obtain a well defined bias current, and thus require tuning after the chip production. Instead, the bias current of the DAC is generated based on an external voltage source from a reference voltage circuit [2]. Figure 6.5 shows the schematic diagram of the reference generator circuit. The reference current is defined by the voltage drop across the resistor Rr, which is a replica of the resistors in the DAC circuits. The voltage drop is fixed to the external reference voltage $v_{ref} = 1.2\mathrm{V}$ by the operational amplifier and M1. This generated current is copied through a cascoded mirror (M2-M5) to the transistor Mr which acts as the half mirror of the DAC array transistors which are sketched in the red box. This makes the current in the voltage DAC array independent of process, supply voltage and temperature variations (not considering the PVT variations of the reference voltage source on the board). Due to the subthreshold operation of the bias generator, the operational amplifier and mirror will show some mismatch and offset. Also the reference voltage will have some variations at the mV level. Since the replica resistor is composed of many unit-size resistors connected in series, the variations can be compensated on the chip level by shorting some of the resistors in the ladder.

The bias current is also used for other blocks requiring a subthreshold current reference[3]. Also the reference voltages required by the front-end blocks are generated by this block,

[2] These reference circuits are commercially available with very small voltage spread, temperature coefficient and supply current, which makes them a better solution than the use of an internal band-gap circuit.

[3] e.g. the operational amplifier of the voltage DAC, the subthreshold DACs for the pedestal holders and bias currents for the input stage used in power pulsed operation

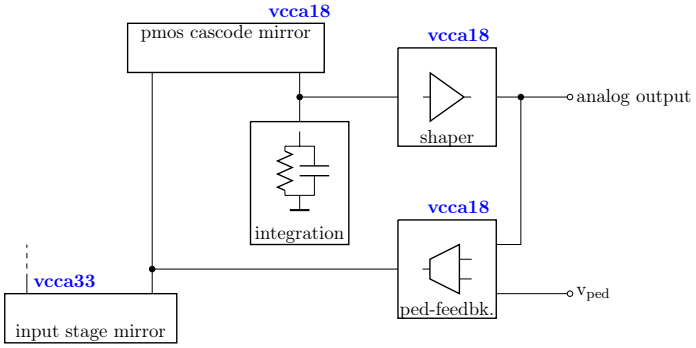


Figure 6.6: Block diagram of the charge measurement branches in the KLauS4 front-end channel.

such as the 0.6V pedestal voltage for the analog branches and the 1.2V reference voltage of the voltage DAC. They are tapped from the replica resistor and distributed to the channels. Since even the operational amplifier in the reference circuit is biased by the same current it is supposed to generate, a start-up circuit (see e.g.[54], Ch. 4) is implemented to ensure the feedback loop enters the correct stability point after enabling the supply voltages to the chip.

6.3.3 Charge collection and noise filtering

The signal current pulses from the sensor are buffered by the input stage and copied to the later processing stages by the nmos mirrors in Figure 6.3. Two of these branches are devoted to the charge measurement, with the distinction of different scaling factors of the charge to cover different signal ranges. The two stages are called *high gain stage* for the small charges, and *low gain stage* for the large charge region spanning the full dynamic range of the sensors. The dynamic range of the high gain stage is selected such that the signal of single pixel hit for 10μm pixel-pitch SiPMs can be digitized without having an effect from the limited resolution of a 12 bit ADC with a dynamic range of $\pm$1.8V and bin size of about 800μV. The scaling factor of the low gain stage is selected such that the full dynamic range of such sensors, 150pC for an overvoltage of 2.5V and the unlikely situation of all pixels firing at the same time, is still covered.

In Figure 6.6, the blocks used in each of the charge measurement branches are shown. The signal current from the input stage is copied and integrated on a capacitance and thus converted to a signal voltage. As explained in Chapter 6.3.1, the signal current must be copied using an NMOS mirror to the later stages to be able to switch to the 1.8V supply rail in these blocks to save power. The current is copied once more using a PMOS mirror connected to the 1.8V supply rail, which allows to cascode the mirrors in order to achieve a good linearity and large dynamic range in the charge measurement. The voltage signal after the charge integration is passed to a pulse shaper to provide a well defined pulse shape with smooth peak such that the peak voltage can be sampled by the ADC connected to the output of the front-end. Given the two different branches which are both connected to the same ADC and dynamically switched with the same sampling time, it is important

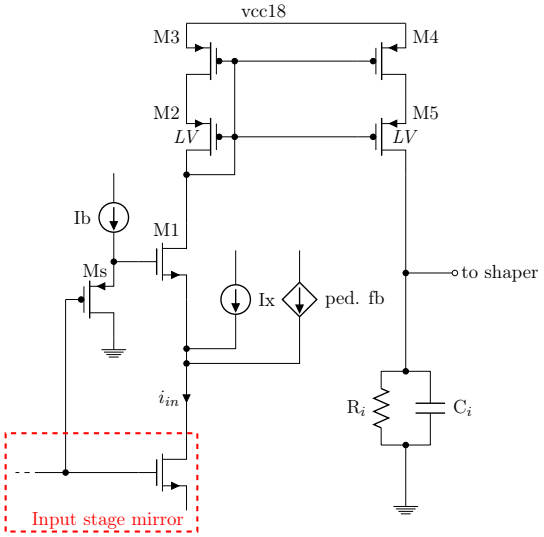


Figure 6.7: Schematic sketch of the charge integration block used in the high gain stage.

that the peaking time for both branches is matched. The pedestal voltage also needs to be well defined, which is provided by a low frequency pedestal feedback circuit in each of the branches. In the following paragraphs, a more detailed description of the individual blocks and their contribution to the overall transfer function will be given.

Charge integration and shaping

A schematic diagram of the charge integration circuit used in the high gain branch, including the half mirror of the input stage and the pmos mirror, is shown in Figure 6.7. Both mirrors are cascoded to acheive a good charge measurement linearity. The cascode for the input stage half-mirror M1 is biased dynamically by shifting the gate voltage of the mirror using a source follower (Ms and the current souce Ib). In this way, the V_{ds} of the half mirror tracks the drain-source voltage difference of the mirror input transistor.

The pmos mirror (M3,M4) is cascoded utilizing transistors with different threshold voltages available in the UMC180 technology. The cascodes M2 and M5 have a lower threshold voltage compared to M3, M4. $V_{ds}(M3, M4)$ is therefore given by the difference of the gate-source voltages of the cascode and mirror transistors. In the implemented circuit, the cascode and mirror pairs have almost the same size, and the drain-source voltage is approximately given by the threshold voltage difference in the absence of a signal. With a current pulse copied by the mirror, the $V_{ds}(M3, M4)$ get smaller and the integrator output voltage at the drain terminal of M5 rises. By properly fine tuning the transistor ratios of the mirror and cascode pairs, a good linearity is achieved.

Neglecting the input stage frequency behaviour, the signal shape exhibits an exponential decay with time constant $\tau_i = R_i C_i$, where C_i is the integrating capacitor defining the charge dynamic range, and R_i is the resistance discharging the capacitance.

Figure 6.8 shows the schematic diagram of the active filter used for shaping the pulse. It is based on the Sallen-Key topology introduced in Chapter 4.2, providing a second order transfer function with two complex-conjugated poles. With $\tau_s = 2C_1R_1$, the transfer

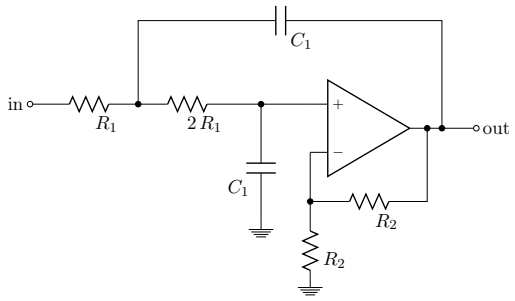


Figure 6.8: Schematic sketch of the active filter used for pulse shaping

function of the integration stage and shaper is given by

$$H_{i,s}(s) = \underbrace{\frac{\tau_i}{C_i\,(1+s\tau_i)}}_{H_i(s)} \cdot \underbrace{\frac{2}{(s\tau_s+1+i)\,(s\tau_s+1-i)}}_{H_s(s)} \tag{6.8}$$

The transfer function also includes the gain of 2 in the shaper, which is given by the resistors R_2 in the negative feedback loop of the operational amplifier. The peak amplitude seen at the integration stage is reduced in the active filter, which would result in a reduced linear range due to the voltage clipping at the integration node before the shaper reaches it's maximum output voltage. The gain factor is introduced to ensure the shaper amplifier is always saturating first, which provides the best linearity over the full output range of the analog front-end.

If the time constants of the shaping and integration stage are made equal, the pair of complex poles introduced in the shaper in combination of the integration pole yield an output pulse response which is flat at the peak, mandatory for the peak sampling by the ADC, and has a short tail to reduce the effective noise term from the DCR pile up events as discussed in Chapter 5.3. The inverse Lapace transform of the shaping system including the gain factor it is given by

$$h(t) = \frac{2}{C_1} e^{-\frac{t}{\tau_s}} \left(1 - \cos\left(\frac{t}{\tau_s}\right)\right) \tag{6.9}$$

The peak voltage of $V_{pk} = h\,(\pi\tau_i/2) = \frac{2}{C_1} e^{-\pi/2}$ is only dependent on the gain factor and integration capacitance. Thus, the dynamic range of the channel is given by the integration capacitance C_1 and the signal current scaling factor of the branch given by the mirror ratio (1:1 for the high gain branch). The amplitude is also independent on the shaping time, at the assumption that the input pulse from the input stage is a delta function. Of course this is not true, and the peak voltage is reduced if the bandwidth of the input stage comparable with the bandwidth of the shaper. This effect is called the ballistic deficit and poses a lower limit on the shaping time that can be used without degradations due to the input stage response. For $\tau_s \approx 10 \cdot \tau_{isg}$, the input stage poles can be neglected in the calculation. Since the shaping time of the system plays a key role in the noise performance of the channel, also the ballistic deficit should therefore be considered to reach the optimum signal to noise ratio. The noise analysis of the channel will be conducted in Chapter 6.3.4.

Pedestal voltage stabilization

The pedestal voltage at the integration stage and shaper output is not well defined because the DC current copied from the input stage is subject to variations from the process and transistor mismatch in the bias circuit and mirrors. It is fixed using an additional regulation circuitry with a small bandwidth compared to the integration and shaping time constant. The circuit block used for the pedestal voltage regulation is labeled *ped-feedbk.* in Figure 6.6. Since the shaper circuit is DC coupled, the feedback current is generated from the shaper output, mitigating the offset and the 1/f noise from the shaper amplifier. The pedestal voltage is compared with respect to a reference pedestal voltage, and a feedback current is added to the input stage DC current before the PMOS mirror. Since the DC voltage at the integrating node is defined by the input DC current and the resistance of the integration circuit, the pedestal is stabilized and low noise components are filtered. The transfer function of the full chain including the pedestal feedback is given by

$$H = \frac{H_{i,s}}{1 - H_{LF} H_{i,s}} \tag{6.10}$$

where $H_{i,s}$ is the transfer function for the integration and shaping stages and

$$H_{LF} = \frac{A}{1 + s/\omega_0} \tag{6.11}$$

is the transfer function of the pedestal holder amplifier with DC transconductance A and 3dB bandwidth ω_0. The output response including the pedestal feedback is not shown here due to it's length and the limited insight it can give to the circuit. The feedback introduces an undershoot to the pulse with an amplitude much smaller than the peak voltage and with a long decay time constant. It has been used in Chapter 5.3 to separate the different parts of the signal adding to the DCR pileup noise term, see Figure 5.5. As it has been discussed in the same chapter, this undershoot component has no significant impact on the pileup noise term due to the dark count pulses as long as the undershoot is small compared to the pulse amplitude. The parameters of the feedback are thus selected to obtain the best signal to noise ratio including the electronic noise only. Some further adjustments are performed in order to match the peaking time of the two charge measurement branches and to optimize the pedestal voltage settling during power pulsing.

Low gain branch and scaling options

As discussed in the beginning of this chapter, two charge measurement branches are implemented in each channel to cover the full dynamic range of the sensors. For the low gain stage to cover the large charge region, the signal current is copied to the integration capacitor with a scaling factor of 6:1 in the input stage mirror (Transistors M_2, M_{2L} in Figure 6.3), and a scaling factor of 6:1 in the PMOS-mirror of the integration block (Figure 6.6). Including the larger integration capacitor for the low gain stage, a total scaling factor of about 60 is reached. The different scaling factor results in very different DC currents in the PMOS current mirror for the low gain branch. Therefore, the polarity of the feedback current is different for the two stages: For the high gain stage, the current is sourced to the input stage mirror transistors for the purpose of power saving, minimizing the DC current

copied to the integration branch. For the low gain stage, the current through the PMOS mirror output would be too small using this approach, keeping the PMOS mirror from working as intended. The feedback current is therefore added to the PMOS mirror input.

While the DC currents are smaller in the low gain branch, the signal currents that need to be covered are much larger. Since also the gate source voltage of the input stage mirror transistors sees a much larger swing for this charge range, the cascode transistor M1 shown in Figure 6.6 would no longer be effective since it can not follow the swing of the input stage transistor. An NMOS cascode is therefore not used in the low gain charge measurement branch.

In order to properly merge the data after processing in the two separate stages in the later data analysis, the inter-calibration factor needs to be known precisely. As a result of component mismatch during the chip production, this parameter can vary for every channel and has to be extracted by measurement. This requires an overlapping signal region where the high gain stage is not saturated and the low gain stage already provides a measurable response. Since the scaling factor between the two stages is quite large, this requirement is not met and an additional switch is added to the input stage mirror to provide a scaling factor of 7:1 compared to the high gain response. This setting will be denoted by *MG* in Chapter 7, where the characterization measurements will be discussed. Two examples for possible inter-calibration methods methods will be presented in chapters 7.4 and 7.7. In both cases, the response at the HG/LG crossover point given by the threshold of the gain selection comparator will be used. Since this method is not assuming or inferring linearity for the inter-calibration, it allows to actually confirm linearity over the full matched range, but requires the overlapping region with sufficient lever arm to extract the response.

Interstage loading effects

In the above discussion of the overall transfer function given by (6.8), it has been assumed that the integration and shaper blocks are separated and not affecting each other. The shaper however has a finite input impedance which is seen by and loads the integration stage. The input impedance of the shaper is given by

$$Z_{i,s} = \frac{R_1 \left(s^2\tau_s^2 + 2s\tau_s + 2\right)}{s^2\tau_s^2 + s\tau_s + 1} \tag{6.12}$$

The transfer function of the integrator and shaper system are modified by the loading term, effectively the impedance of the R-C integrator is parallel connected to the input impedance of the shaper, yielding

$$H_{tot} = \left(\frac{\tau_i}{C_i\left(1 + s\tau_i\right)} \parallel Z_{i,s}\right) \cdot \frac{2}{\left(s\tau_s + 1 + i\right)\left(s\tau_s + 1 - i\right)} \tag{6.13}$$

The loading term is proportional to $1/R_1$ and adds two poles and zeros to the system, effectively increasing the integration time constant τ_i and reducing the signal amplitude at the shaper output. The resistors in the shaper also introduce noise and can not be made too large to reduce the loading effects. They have been selected to yield the best signal to noise

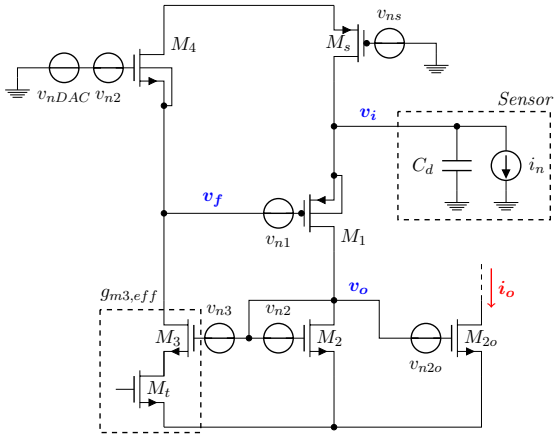


Figure 6.9: Schematic diagram of the input stage including the noise sources

ratio. Since the integration capacitor is different for the two stages, also the change in the pulse shape is different for the two. This effect is compensated by adjusting the integration time constants and pedestal holder transconductances.

6.3.4 Noise analysis

After the full processing chain has been discussed in the previous sections, the noise analysis of the front-end can be conducted. The noise seen at the analog output is generated by all of the noisy components in the processing chain and can equivalently be described by noise current and voltage sources at the inputs of the building blocks of the front-end. Here, the noise analysis will be performed only for the high gain branch, where the noise performance is of particular interest and the dominant noise comes from the input stage. Only for the low gain branch where the signal current of the input stage is scaled to achieve the large dynamic range, the noise from the integration and shaping stage become relevant. Thus, the noise from the integration stage, shaping and feedback circuits will be ignored in the noise analysis, and a scaling factor of 1 between the input and the integration stage is assumed as it is the case for the high gain branch[4]. The noise current is then generated by all transistors in the input stage including the bias current generator (a pmos transistor, M_s), the voltage noise of the DAC propagated through the input stage to the detector capacitance, and the mirror output transistor $M_{2\mathrm{o}}$ of the input stage ($M_{2\mathrm{h}}$ in parallel to $M_{2\mathrm{h}}$, in Figure 6.3) with a transconductance of $g_{\mathrm{m}2}$.

Based on their frequency behaviour and dependence on the detector capacitance, three different noise terms can be separated: The parallel noise, modelled by a current source parallel to the input terminal of the channel, the series noise term described equivalently as a voltage noise source connected in series to the input, and an internal noise current added to the signal current before the integration stage.

For the noise analysis of the input stage, the system of small-signal equations used to derive the input impedance (6.1b) is extended adding the gate referred noise voltages to every transistor of the input stage. Figure 6.9 shows the schematic diagram of the input

[4]In the normal 1:1 scaling configuration, i.e. the switch selecting the high gain dynamic range is closed in figure 6.3

stage including the noise voltage sources and the sensor modeled by the lumped total capacitance. The equation system composed of the input node, feedback node and output node and current equations is given by

$$0 = i_n + sC_d v_i + g_{\mathrm{ms}} v_{ns} - g_{\mathrm{m1}} \left(v_f + v_{n1} - v_i\right) \tag{6.14a}$$

$$0 = g_{\mathrm{m3,eff}} \left(v_{n3} + v_o\right) - g_{\mathrm{m4}} \left(v_{n4} + v_{n,\mathrm{DAC}} - v_f\right) \tag{6.14b}$$

$$0 = g_{\mathrm{m1}} \left(\left(v_f + v_{n1}\right) - v_i\right) + g_{\mathrm{m2}} \left(v_{n2} + v_o\right) \tag{6.14c}$$

$$i_o = g_{\mathrm{m2}} \left(v_{n2o} + v_o\right) \tag{6.14d}$$

where v_{nX} are the noise voltages of the input stage transistors with transconductance g_{mX} (the subscript $*_{ns}$ stands for the bias current source parameters), $v_{\mathrm{n,DAC}}$ denotes the noise of the voltage DAC, $g_{m3,eff}$ is the effective transconductance of M3 including the triode (6.4) and i_o is the output current copied to the high gain branch.

Solving the equation system for i_o yields the noise current density depending on all noise (voltage and current) sources seen at the output of the mirror. They can be separated into the tree classes of noise transfer functions by taking the partial derivatives $\frac{\partial i_o}{\partial v_{nX}}$ and $\frac{\partial i_o}{\partial i_n}$. Using the definition of the input impedance (6.2), the transfer functions of the noise components simplify to

$$\frac{\partial i_o}{\partial v_{\mathrm{n1}}} = \frac{\partial i_o}{\partial v_{\mathrm{n4}}} = \frac{\partial i_o}{\partial v_{\mathrm{n,DAC}}} = -H_{ser} \tag{6.15a}$$

$$\frac{\partial i_o}{\partial v_{\mathrm{n2}}} = -g_{\mathrm{m2}} H_{par} - \frac{g_{\mathrm{m2}}}{g_{\mathrm{m1}}} H_{ser} \tag{6.15b}$$

$$\frac{\partial i_o}{\partial v_{\mathrm{n3}}} = \frac{g_{\mathrm{m3,eff}}}{g_{\mathrm{m4}}} H_{ser} \tag{6.15c}$$

$$\frac{\partial i_o}{\partial i_n} = \frac{\partial i_o}{\partial v_{ns}} \frac{1}{g_{\mathrm{ms}}} = H_{par} \tag{6.15d}$$

$$\frac{\partial i_o}{\partial v_{n2o}} = g_{\mathrm{m2}} \tag{6.15e}$$

with

$$H_{ser} = \frac{sC_d}{1 + sR_{in}C_d} \tag{6.16a}$$

$$H_{par} = \frac{1}{1 + sR_{in}C_d} \tag{6.16b}$$

The different noise components can be understood as follows: The series noise components generate a noise voltage on the feedback node which generates the same noise voltage at the input node. This noise voltage is converted to a noise current by the sensor capacitance, which has an impedance of $1/(sC_d)$. The parallel noise currents are parallel to the current-sensitive input and thus not affected in this way. The total noise current is sensed by the input stage with a finite input impedance. It is shared by the parallel connected R_{in} and C_d, which yields the pole at $-1/(R_{in}C_d)$ in both noise transfer function terms.

The total noise voltage seen at the output of the front-end is calculated using the noise

integral (5.7) and the noise voltage densities generated by the transistors (4.7, 4.8). In order to be able to calculate the noise integrals for the different noise components, further simplifications need to be made.

In the noise voltage densities generated by the transistors, the low frequency 1/f noise will be neglected. For the transistors contributing only to the series noise, the low frequencies are already suppressed by the current sensitive readout scheme which adds a zero at $f = 0$. The most significant contributor of 1/f noise for the parallel noise is the current noise source. Since this transistor is connected to the input terminal where the total capacitance connected to the net is already large from the detector itself, the 1/f noise can be reduced by increasing the transistor area (compare Equation 4.8) without sacrificing bandwidth. Further, the voltage noise generated by the DAC is not broad band, but has a cut off frequency in the order of 10kHz and only shows up as a series noise term. Since the pedestal holder circuit is sensitive to these frequencies, this low frequency noise is filtered to a large extend. Thus, it is acceptable to neglect this noise component in the calculation. On the other hand, the thermal noise is broad band. The pedestal holder bandwidth is placed much lower the bandwidth of the integration and shaping system. The low frequencies filtered by the pedestal feedback are not contributing significantly to the noise integral and the pedestal feedback can be neglected. The overall noise transfer functions are then given by $H_{ser} \times H_{s,i}$, $H_{par} \times H_{s,i}$, $H_{s,i}$ for the tree noise species, simple enough to allow the calculation of the noise integral.

For the series noise term, the integral is given by

$$\begin{aligned} \sigma^2_{v,ser} &= D_s \cdot \int_0^\infty |H_{ser}(i \cdot 2\pi f) H_{i,s}(i \cdot 2\pi f)|^2 \cdot \mathrm{d}f \\ &= D_s \cdot \frac{2C_d^2 \tau \left(3R_{in}C_d + \tau\right)}{5C_{\mathrm{int}}^2 \left(R_{in}C_d + \tau\right)\left(2R_{in}C_d\tau + 2R_{in}^2C_d^2 + \tau^2\right)} \end{aligned} \tag{6.17}$$

with the series noise voltage densities D_s generated by the transistors M1-M4 and the voltage DAC, using (4.7) for all the series noise components collected from (6.15):

$$D_s = \frac{8k_bT}{3}\left(\frac{1}{g_{\mathrm{m1}}} + \frac{1}{g_{\mathrm{m4}}} + \left(\frac{g_{\mathrm{m2}}}{g_{\mathrm{m1}}}\right)^2 \frac{1}{g_{\mathrm{m2}}} + \left(\frac{g_{\mathrm{m3,eff}}}{g_{\mathrm{m4}}}\right)^2 \frac{1}{g_{\mathrm{m3,eff}}}\right) \tag{6.18}$$

The integrated parallel noise term is calculated in the same way collecting all parallel noise terms,

$$\begin{aligned} \sigma^2_{v,par} &= D_p \cdot \int_0^\infty |H_{par}(i \cdot 2\pi f) H_{i,s}(i \cdot 2\pi f)|^2 \cdot \mathrm{d}f \\ &= D_p \cdot \frac{\tau^2 \left(9R_{in}C_d\tau + 10R_{in}^2C_d^2 + 3\tau^2\right)}{5C_{\mathrm{int}}^2 \left(R_{in}C_d + \tau\right)\left(2R_{in}C_d\tau + 2R_{in}^2C_d^2 + \tau^2\right)} \end{aligned} \tag{6.19}$$

with the total parallel noise current density

$$D_p = \frac{8k_bT}{3}\left(g_{\mathrm{m2}} + g_{\mathrm{ms}}\right) \tag{6.20}$$

The internal noise is generated by the output transistor of the mirror only. Here, only

the frequency response of the processing stages have to be considered.

$$\begin{aligned}\sigma_{v,int}^2 &= \frac{8k_bT}{3}g_{\mathrm{m2}} \cdot \int_0^\infty |H_{i,s}(i \cdot 2\pi f)|^2 \cdot \mathrm{d}f \\ &= \frac{8k_bTg_{\mathrm{mi}}\tau}{5C_{\mathrm{int}}^2}\end{aligned} \tag{6.21}$$

To study the optimum parameters of the readout chain, the total noise seen at the output is obtained by adding all noise terms of the front-end, the effective noise term from DCR pileup, and the ADC quantization resolution (5.27) in quadrature:

$$\sigma_n = \sqrt{\sigma_{v,ser}^2 + \sigma_{v,par}^2 + \sigma_{v,int}^2 + \sigma_{pu}^2 + \sigma_{ADC}^2}$$

Apart from the transistor transconductances, the function has two free parameters: The shaping time constant and the detector capacitance. The latter can be estimated pessimistically as 75pF for typical sensors of 1.3mm^2 plus the routing parasitics on the circuit board in the mathematical calculations. The dependency on the detector capacitance is however relevant for the characterization. By measuring the noise as a function of the capacitance at the input, the different noise terms can be separated. Figure 6.10a shows the noise components of the front-end as a function of the shaping time constant. The series noise term is generated by the detector capacitance and therefore only dominant at high frequencies (small shaping time). On the other hand, the parallel noise increases with the shaping time constant as the timescale of noise integration gets longer. The optimum shaping time to obtain the lowest noise depends on the detector capacitance and dark count level. For small detector capacitances the parallel noise term always dominates the total noise, assuming reasonable shaping time constants in view of the power budget ($C_d = 75$pC and $\tau \approx 50$ns).

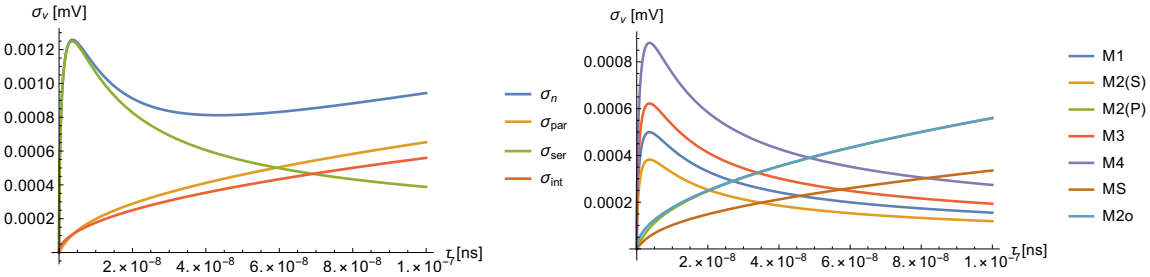
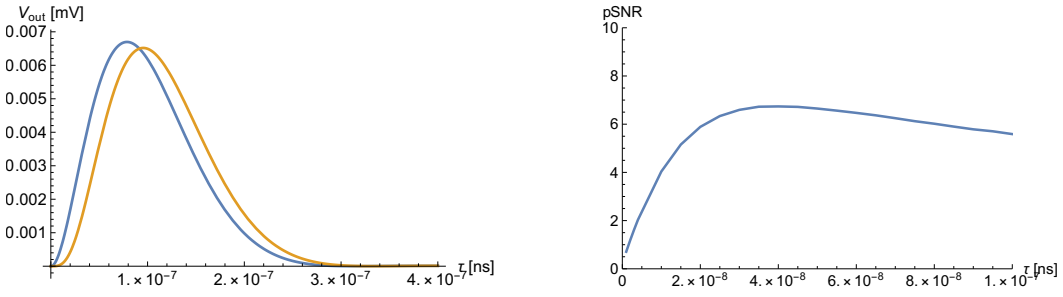


Figure 6.10: Noise components of the KLauS4 front-end. a) Sum of all electronic noise terms.
b) Individual contributions of the input stage transistors.

In Figure 6.10b, the contributions for the individual noise terms are shown separately. The input transistor has a small contribution to the noise as the transconductance is large to reduce the input impedance. MS in the Figure shows the contributions of the bias current source transistor adding to the parallel noise. A small transconductance g_{ms} limits the input voltage tuning range (6.3) as $V_{dsmin,SRC}$ is increased, and increases the effect of



(a) **(b)**

Figure 6.11: a) Ballistic deficit due to the limited input stage bandwidth. The blue line gives the output response for an unlimited bandwidth and zero input impedance of the input stage, whereas the yellow line gives the pulse response for an input stage bandwidth of 20 MHz. b) Signal to noise ratio for a single pixel signal of 15fC under consideration of electronic noise, pileup due to dark count pulses, limited ADC resolution and ballistic deficit.

power supply noise on the bias current. The most important parameter is the choice of g_{m2}. The transistor contributes to both the series and parallel terms and is one of the biggest contributors to the overall noise. M2 also fixes the dimensions of mirror output transistor, which is also a significant term. In fact, the transistor needs to have a relatively large transconductance to obtain the dynamic range delivered by the sensors. For large signals, the current fed through the input and copied by M2 is causing a swing at the gate of M2. For too small g_{m2}, the voltage swing is too large for the connected cascodes to follow, as they require some voltage headrooms in order to work as intended. This would introduce significant nonlinearities in the charge measurement. Thus, a compromise between noise performance and the dynamic range needs to be found for the size of the mirror transistors.

For the optimization of the shaping time constant given by the optimal single pixel signal to noise ratio (5.2), the input stage frequency response must also be considered. If the dominant poles of the input stage current transfer function are placed not much higher than the pole of the integration stage, the signal amplitude is decreased as the time constant of the current pulse is comparable to the discharge time constant of the RC-integrator circuit. This effect is known as the ballistic deficit, and yields a reduced signal amplitude for smaller time constants. As a rough estimation, the input stage is modelled by two poles at a frequency of 20MHz, a result obtained from circuit simulations[5]. Figure 6.11a shows a comparison of the output signal shape without and with the input stage bandwidth limitation included for a shaping time of $50ns$. The reduced signal amplitude affects the pile-up noise contributions since it is proportional to the signal height and deteriorates the signal to noise ratio for integration time constants approaching the characteristic time constant of the input stage. Also the contribution from the limited ADC quantization resolution is increasing. Figure 6.11b summarizes the signal to noise ratio analysis for a charge of 15fC (i.e. very low gain SiPMs), given as a function of the integration time constant. In the calculation, the electronic noise, DCR pileup, ADC resolution and ballistic deficit are included.

[5] A more detailed calculation of the ballistic deficit for the KLauS input stage was carried out in [5], including the most relevant parasitics in this design

The contributions from the DCR pileup are taken from the monte-carlo simulations shown in Figure 5.6, assuming a dark-count rate of 100kHz. This noise term turns out to have almost the same magnitude as the noise from the mirror transistors. The quantization resolution is given by the bin size of 800μV, as obtained from a differential ADC with a dynamic range of $\pm$1.8V and 12 bit resolution. The optimum value of about 40ns has been chosen for the design. Because the calculation only includes the noise sources most important for the design optimization, the calculated signal to noise ratio can not be acheived in reality. Even though the calculation has undergone significant simplifications and all the parasitics in the input stage as well as other noise sources have been ignored, numeric circuit simulations yield a very similar result for the optimum shaping time, and an equivalent noise charge of about 4 fC. From the above calculations, a general contradiction valid for most common-gate input stage topologies is apparent. To be able to use a small shaping time to reduce noise contributions from the DCR pile-up, the input impedance should be made small to minimize the ballistic deficit. This on the other hand generates large currents that have to be processed by the input stage, requiring that the mirror transistors are large to obtain a good linearity. This generates larger parallel noise contributions, pushing the ideal shaping time and thus bandwidth further. This ultimately leads to a higher power consumption, and a compromise between power, signal-to-noise ratio and linearity has to be made.

6.3.5 Comparator

For the generation of time stamps of the charge pulses, a comparator circuit is implemented in the chip. Since the channels shall be operating in an autotriggered fashion, the ADC conversion is also started only if the charge pulse exceeds the comparator threshold. Figure 6.12 shows a simplified[6] schematic diagram of the comparator block. Since the requirements on the time resolution of less than 1 ns for the AHCAL are relatively easy to achieve, the signal current copied from the input stage is scaled in order to reduce the power consumption. The copied current is compared against a digitally controllable threshold current i_{th}. The threshold current is the sum of a current generated by a 6-bit digital to analog converter common to all channels to set the coarse threshold globally, and a 4-bit DAC for each channel used for fine adjustment of the threshold. This allows to mitigate different SiPM operating conditions or parameters, and allows to remove differences in the channels due to mismatch. If $i_{in} < i_{th}$, the voltage at the comparing node *comp* in Figure 6.12 is at a logic high level. Once the input current is larger than i_{th} during an incoming charge pulse, the comparing node reaches a logic low state, and the comparator fires. The input stage and threshold current mirrors transistor (M_{2t} in 6.3 and M_3, M_0 in the comparator block) are cascoded to achieve a large impedance at the comparing node, which increases the voltage swing. The cascode transistors M_1 and M_2 are kept small to minimize the parasitic capacitance at the comparison node, which increases the comparison speed and thus reduces the timewalk effect.

The comparator core composed of the decision gate (M_9, M_{10}) and the "push-pull" gate (M_5-M_8) is based on the current comparator of Traff [75], which has been extended by

[6] For the simplified schematic, some logic inverters between the digital latch and the comparator core have been omitted, leaving the actual logic the same. Also the power connections of the digital gates are left out; They are supplied by the digital power supply while the comparator core is supplied by the analog 1.8V supply rail as indicated in the figure.

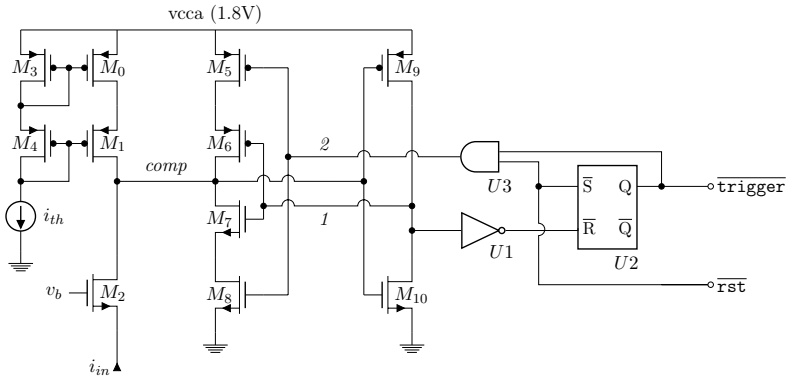


Figure 6.12: Schematic sketch of the KLauS comparator branches

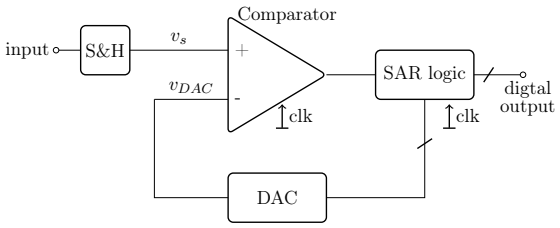
additional digital circuit to avoid glitches at the comparator output. The voltage at the comparison node is discriminated by a simple digital inverter optimized for minimum input capacitance and threshold voltage. For small charge pulses, the comparison node passes the threshold voltage slowly, which increases the decision time and eventually generates glitches at the output of the inverter (node *1*). To enhance the speed of the comparator, the output of this decision gate is used in a positive feedback configuration: In the idle state of the comparator, node *2* is at a logic high level causing transistor M8 to conduct and M5 to be open. The output of the gate M5-M9 will sink current from the comparing node if *1* is at logic high level and is at high impedance otherwise. Once the voltage of the comparing node falls below the threshold voltage of the comparing gate a logic high level appears at the gate of M7, and the push-pull gate rapidly pulls the comparison node to a proper logic zero. To further process the comparator output, the discrimination pulse is buffered and stored in an RS-latch.The output of the latch is sent to the later digital circuits to generate timestamps and initiate the ADC conversion. In addition, it is used to generate the direction signal for the push-pull gate, ensuring the recovery of the comparator after the trigger is registered in the latch. After storing a trigger hit, the comparator core is masked and not generating new triggers, which reduces the noise on the power lines due to digital switching activity. After a low going pulse on the $\overline{rst}$ pin, the comparator is ready to generate new triggers. If permanently pulled low, the pin also serves as a mask flag controlled by the configuration interface in the digital part of the chip, allowing the user to disable activity from this channel. The second comparator implemented in the front-end is used for the automatic gain selection. The threshold of this comparator is set much higher , $\approx$ 2pC close to the maximum dynamic range of the high gain branch, and does not need to be defined as accurate, and a separate DAC for fine-adjusting the threshold is not required. While the comparator core is the same as for the main comparator, the digital control logic is designed different to allow trigger signals also after the hit has been registered in the RS-Latch. This feature is used to allow the registration of a single bit information on large signals sensed while the ADC is busy from a preceding trigger.

6.4 Analog to Digital Converter

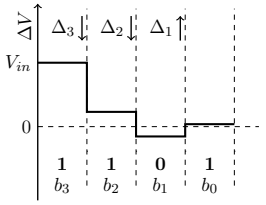
After the charge signal has been processed by the analog front-end, the charge information is available as a pulse with well defined shape and amplitude proportional to the charge at the output. The peak voltage is sampled and then converted to a digital code word by the ADC. The peak voltage sampling implemented in the chip will be discussed in section 6.5. As discussed in the beginning of this chapter, a quantization resolution of up to 12 bit is required for low gain SiPMs during calibration. In other cases, a 10 bit resolution is sufficient. In order to limit the dead time of the channel after a hit, the ADC block should allow conversion speeds in the order of 400 ns excluding the sampling time. The power consumption can be minimized further by disabling the ADC blocks with a static power consumption, since the event rate is small and the ADC will be in an idle state most of the time. Given the large number of channels in the detector system, the calibration of the ADC with respect to the correction of nonlinearities should be straight forward and ideally static after the chip production such that the calibration measurements only need to be performed once during detector commissioning.

ADC architecture

From the different ADC topologies known in literature (see e.g. [76]), the pipelined ADC and the Successive Approximation Registration (SAR) ADC fit the requirements on the sampling rate and resolution best [77]. A survey comparing a large number of published ADC designs from different conferences [78] shows excellent power efficiencies for ADCs in both pipeline, SAR and combinations of both. For the KLauS ASIC, the ADC will convert the analog signal only after a trigger from the comparator which allows to fully switch off the ADC to significantly save power given that the occupancy per channel is small. Another speciality is given by the fact that the quantization resolution required for the SiPM digitization is dependent on the gain of the used sensor and the mode of operation since in the standard calibration procedure of the AHCAL, the single photon spectra are only recorded in calibration runs.
While the survey can give hints to designs suiting the application best, it can not give a definite answer on the ideal topology since most of the converters presented in the publications are designed in smaller technologies and operated in continuous sampling mode, which should be considered the standard case of most ADCs in other applications. Given the special needs for the ADC in KLauS, the block should always be tailored to the requirements.
The block diagram in Figure 6.13a shows the basic architecture of a SAR ADC. After the analog voltage is sampled and held on an internal node of the ADC using the sampling switch (S&H block in the figure), the digital code word, i.e. the digital representation of the analog voltage, is searched by successively reconstructing the stored analog voltage using a voltage DAC. Figure 6.13b gives an example of the approximation process for a 4-bit SAR ADC. The x-axis refers to the conversion step i where the bit b_i of the ADC code is sampled. On the y-axis, the residual voltage difference between the sampled input voltage and the output of the DAC is plotted. The voltage generated by the DAC is fed back to the comparator, which gives the necessary information for the search algorithm. The DAC is controlled by a digital logic block which is implementing the search algorithm. After each



(a) Building blocks of a SAR ADC.



(b) Example of a 4 bit SAR conversion.

Figure 6.13

comparison, the DAC code is changed according to the comparator decision and the range of possible code words matching the analog voltage is reduced by 1/2 using a binary search algorithm. For a conversion with N bits of quantization resolution, a voltage DAC with a resolution of N-1 bit is required, and the conversion is finished after N comparison steps.

Since a clock much faster than the desired sampling speed is required, the conversion speed of such SAR ADCs is typically limited up to orders of 10MHz for mixed-mode technologies above 130nm. The quantization resolution is dependent on the quality of the DAC which will govern the nonlinearities introduced by the ADC. This limits the achievable resolution depending on the technology and implementation of the DAC, typically to a maximum dynamic range of 12 bit.
For the targeted application of the KLauS ASIC, where only the peak voltage needs to be digitized rather than the full waveform, the average rate of events to be converted is small, but the conversion speed should be high enough to avoid long dead-times beyond the collision period of the ILC. A distinct feature of the SAR ADC is that the power consumption scales with the conversion rate, which makes it a good choice for KLauS. Since the nonlinearities introduced by the ADC are governed by production tolerances and the art of circuit layout, the nonlinearities are fixed after the chip production, and the calibration only needs to be performed once. For CALICE detectors with millions of channels in the system that need to be calibrated this is an important aspect. The SAR ADC structure is chosen for the KLauS because of the calibration stability and lower power consumption at small conversion rates.

10 bit SAR ADC

The main part of the ADC implemented in the chip is shown in Figure 6.14. It is designed in a differential topology which allows to also track changes in the reference ground used by the front-end, and ensures a symmetric layout. At the positive input of the ADC, the analog output of the front-end is sampled, whereas the local ground potential of the channel is sampled at the negative input. This reduces the effect of ground potential changes along the supply lines, especially for channels placed in the middle of the ASIC. The voltage at both inputs is sampled and held using bootstrap switches based on the design in [79] [80]. The bootstrapped operation allows to minimize nonlinearities from charge injection or the switch impedance, which would vary with the signal amplitude using a normal single-transistor switch. The schematic and operation of the implemented switch is discussed in

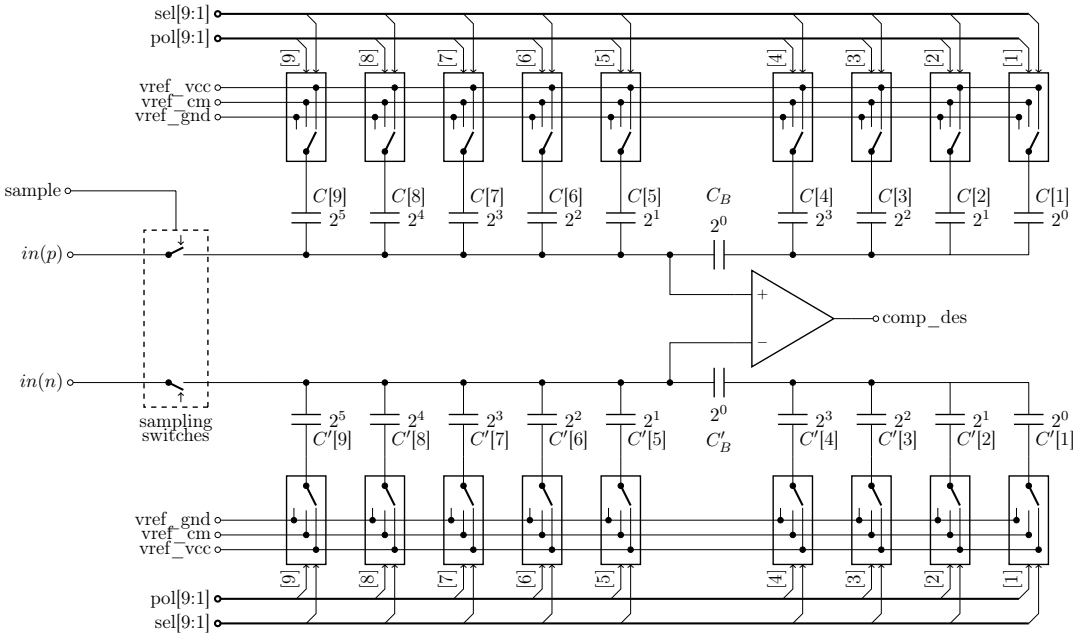


Figure 6.14: Block level schematic of the 10-bit SAR-ADC

Appendix A.

The sampled voltage is stored on an array of metal-in-metal (MIM) capacitances available as a special metal layer stack in the UMC180 technology. The differential pair of capacitors also implement the voltage DAC. After sampling the input voltage, the sampling switches are opened, leaving the nodes connected to the comparator floating. By changing the initially fixed reference terminal of the capacitors in every step of the approximation procedure, the voltage difference at the compared node is changed and compared. In this sense, the ADC is minimizing the difference of the sampled input voltage and the voltage change introduced by the DAC. The capacitor DAC generating the voltage changes are implemented using individual capacitor blocks $C[i]$ and $C'[i]$ for the positive and negative comparison nodes. Each capacitor is composed of a number of unit size capacitors with $C_u \approx 33\text{fF}$, indicated by the capacitor values in Figure 6.14. For a binary DAC with a resolution of N bits, $2^{N+1} - 1$ unit size capacitors would be needed. To reduce the chip area and total capacitance at the input of the ADC which needs to be driven by the shaper amplifier in the front-end, the capacitor array is split into two sub-arrays, the MSB and LSB array. They consist of capacitors $C[i]$ with $i = L \ldots N-1$ and $i = 1 \ldots L-1$ for the MSB and LSB array, respectively. The two sub-arrays are connected by a bridge capacitor of unit size. By the series connection of the bridge and LSB array capacitances, the effective capacitance of the LSB array capacitors seen at the comparator input is attenuated to achieve a linear switching behaviour. At the beginning of the conversion, all capacitor reference planes are connected to the *vref_cm* voltage at 0.9V. During the approximation procedure, the capacitor starting with the highest order are switched to either *vref_vcc* at 1.8V or *vref_gnd* at 0V depending on the last comparator decision. This changes the potential at the comparing nodes which initially hold the sampled voltage. The lower (negative) side of the differential capacitor array is switched to the opposite potential (*vref_gnd* instead of

vref_ vcc and vice versa). Together with the appropriate digital control logic, this *monotonic capacitor switching* implementation can reduce the power consumption from the charging and discharging of the capacitor banks and is one of the most efficient implementations presented in literature [81]. The differential voltage change at the comparator inputs when switching the back-plate potential of capacitor i is given by

$$\Delta_i = C_{eff}[i] \cdot V_{ref} \cdot \frac{C_B + C_L}{C_M \cdot (C_B + C_L) + C_B \cdot C_L} \tag{6.22}$$

Where C_B is the bridge capacitor of unit size, V_{ref} is the voltage change of the reference potential ($\pm$1.8V including the differential topology of the DAC), C_M and C_L are the total capacitances of the MSB and LSB arrays and $C_{eff}[i]$ is the effective capacitance of the switched capacitor pair $C[i]$,$C'[i]$ seen at the comparator input:

$$C_L = \sum_{i=1}^{L-1} C[i] = \sum_{i=1}^{L-1} 2^i \cdot C_u \tag{6.23a}$$

$$C_M = \sum_{i=L}^{N-1} C[i] = \sum_{i=L}^{N-1} 2^{i-L} \cdot C_u \tag{6.23b}$$

$$C_{eff}[i] = \begin{cases} C[i] \\ C[i] \cdot \frac{C_B}{C_B+C_L} \end{cases} = \begin{cases} C_u \cdot 2^{i-L} & \text{for } i = N-1, \ldots, L \\ C_u \cdot 2^i \cdot \frac{C_B}{C_B+C_L} & \text{for } i = L-1, \ldots, 1 \end{cases} \tag{6.23c}$$

For the MSB array capacitors ($i \geq L$), the effective is the same as the actual capacitance C_i. In case of the LSB array capacitors ($i < L$), $C[i]$ is attenuated by the bridge. While the total number of capacitors $M + L$ defines the quantization resolution minus the sign bit of the DAC, the splitting fraction L is a design parameter that needs to be selected. A splitting fraction of 5:4 (i.e. $L = 5$) is chosen in the design as a compromise between low input capacitance and matching requirements of the bridge capacitor. As the conversion proceeds, the residue voltage difference sensed by the comparator is minimized by the binary search algorithm. The LSB bin size of the ADC in absence of additional parasitic capacitances is given by

$$V_{LSB} = \frac{V_{ref}}{2^{M+L} - 1} \tag{6.24}$$

Apparently there is a difference to the ideal bin size (5.28) which reduces the dynamic range of the ADC even in the ideal case without parasitics. Since the difference is small, this is of no practical concern here. It will become relevant when the parasitics from the layout will be added, which will be done in the linearity analysis of the DAC.

The voltage comparator implemented in the ADC is discussed in more detail in Appendix A. It consists of a dynamic comparator clocked by the digital part to give a logic decision after each falling edge of the clock, and a amplifier connected before the comparator. The preamplifier stage is used to reduce the comparator offset, and to reduce the digital kick-back noise from the output stage of the comparator. The comparator is already designed to consume only a small amount of power. To minimize the power consumption at low

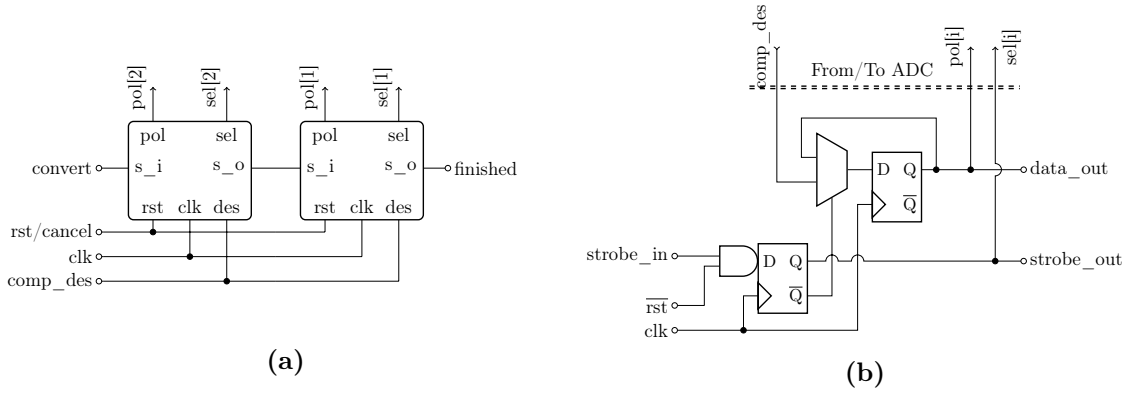


Figure 6.15: Core part of the SAR ADC control logic. a) Illustration of the cell connection giving a 2-bit logic example. b) Single bit logic

event rates, the comparator clock is switched off, and the control logic can be configured to disable the preamplifier bias current when there is no conversion in progress. This is possible since the time between the trigger signal and the start of the ADC conversion after sampling the peak voltage is sufficiently long to allow the settling of the comparator block.

ADC control logic

The SAR control logic realizing the binary search algorithm is implemented as a digital circuit in the synthesized digital part of the chip. Implemented as a chain of single-bit stages as illustrated in the example in Figure 6.15a, it handles the registration and approximation process of the ADC. The single bit stage schematic is shown in Figure 6.15b. It consists of a d-flipflop implementing a shift register to store the current step of the conversion, also used to select if the corresponding capacitor should be connected to the common mode potential or the vcc/gnd potentials. At every rising edge of the conversion clock, the *strobe* signal is passed through the shift register, propagating a logic one through the chain of flip-flops. As the strobe output of the single-bit cell rises to logic one, the current comparator decision (input *comp_ des*) is stored in a second flip-flop. The strobe output *sel[i]* and the stored data signal *pol[i]* are fed back to the capacitor array changing the reference switches as listed in Table 6.1. To perform a conversion cycle using a chain of these cells, a logic '1' is asserted at the strobe input of the first cell. When the logic one has propagated through the chain and appears at the strobe output of the last cell, the conversion is finished and the logic can be put back to the idle state by asserting a high-pulse to the rst signal. After the conversion is finished, the registered comparator decisions represent the converted digital code of the conversion which are then processed in the later digital stages and finally sent off-chip. Figure 6.16 shows the simulation of a single analog conversion using the 10 bit ADC implemented in the chip. The higher level digital block controlling the chain will be discussed in section 6.5, where the interface to the front-end and digital part further processing the digital data are also included.

sel[i]	pol[i]	p-side	n-side
0	X	$vref\_cm$	$vref\_cm$
1	0	$vref\_vcc$	$vref\_gnd$
1	1	$vref\_gnd$	$vref\_vcc$

Table 6.1: Potential of the capacitor back-planes connected to the charge steering DAC.

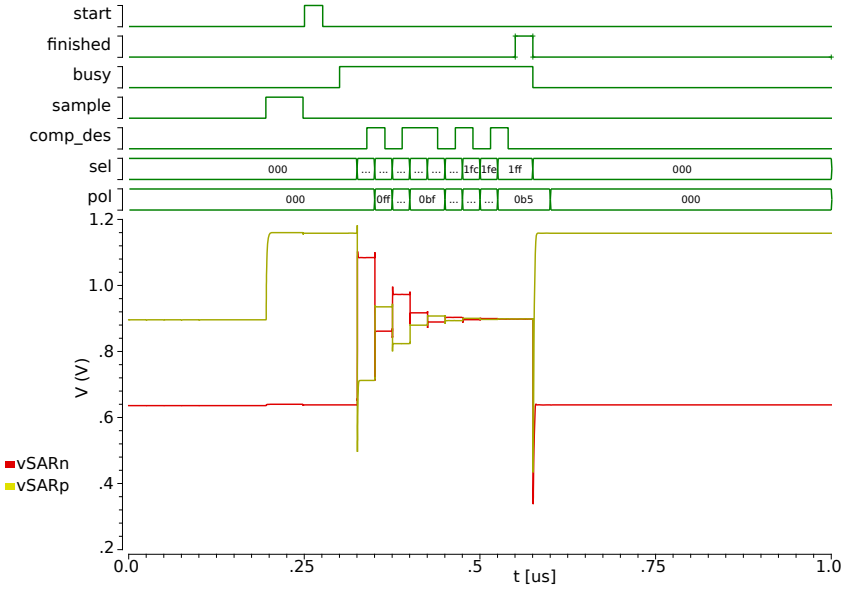


Figure 6.16: Simulation of a 10 bit ADC conversion cycle

ADC linearity

The static linearity of the voltage measurement is mainly affected by the capacitor DAC. At high sampling rates, also the dynamic performance of the sampling switch, comparator and the settling time of the DAC determined by the unit size capacitance and the switch resistances are reducing the effective resolution of the ADC as the sampling frequency increases. Since the conversion speed is still comparably small for the ADC implemented in KLauS, these effects will not be discussed here. From (6.22), it is apparent that every parasitic capacitance inevitably added to the DAC due to routing and placing the capacitances in the array will change the bin size of the DAC and affect the ADC linearity and dynamic range. Ideally, the linearity criterion for the binary DAC

$$\frac{\Delta_i}{\Delta_{i-1}} = 2 \tag{6.25}$$

is assured by the specific choice of $C[i]$ and the bridge capacitor. Figure 6.17 shows the most relevant parasitic capacitances that may occur in the layout, potentially turning into nonlinearities violating the linearity criterion of the DAC. By adding the parasitics to (6.23), the voltage changes including parasitics (6.22) is easily calculated and can be used to obtain the ADC linearity given by the bin sizes. Without analyzing the introduced errors quantitatively, the impact of the different parasitic capacitances shown in the figure will be

discussed in the following.

For each capacitor $C[i]$ in the array, a parallel connected parasitic $C_p[i]$ is introduced by the routing and parasitic coupling between the capacitances in the array. In addition, imperfections during the chip fabrication yield small fluctuations (mismatch) of the actual capacitance. While these parasitics can be reduced by increasing C_u, the increased routing parasitics also affect the linearity and an optimum unit size capacitance needs to be found during the layout optimization process. Also the chip area, loading capacitance at the ADC input must be considered in this process. To cancel gradual changes of the unit capacitors over the chip area at first order, a common centroid layout technique [82] is used. In order to satisfy the linearity criterion in view of the systematic part of $C_p[i]$, also the parasitics should be scaled in the same way as the $C[i]$ themselves, which is possible by a proper distribution the individual capacitors in the DAC and carefully adding metal traces to balance the parasitics. Since the metal trace parasitics are affected by mask misalignments in the production process, they should be made invariant under small mask misalignments and minimized to reduce the overall sensitivity on production tolerances such as etching effects.

The design rules posed by the chip manufacturer require a minimum coverage for all metal layers, the layers below the capacitance also need to be filled and should be connected to an AC ground potential. The result is a parasitic to ground for the signal lines connected to the sampling switches and comparator, as well as the signal line of the LSB array (C_{pM},C_{pL}). Both parasitics do not directly affect the linearity of the DAC, but the LSB size changes as C_M (6.23b) showing up in the denominator of (6.22) is modified. With increasing C_{pM}, the bin size decreases as the charge on the comparing node is shared between the parasitic capacitance. While small changes are certainly acceptable, large parasitics reduce the dynamic range of the ADC and should be avoided.

The bridge capacitance defines the attenuation factor of the LSB array in (6.23c), and thus the Δ_i's. The attenuation factor of the bridge capacitance is also reduced by the parasitic between the LSB array signal line and ground (C_{pL}). This requires the bridge capacitance parasitics to be adjusted accordingly. Indirectly also parasitics added to C_M have an effect on the nonlinearities due to the changed bin size, requiring C_pL and C_pB to be tuned. Since C_B is only of unit size, the linearity is quite sensitive to C_{pB}.

Finally, if the two sub-arrays are not separated well enough, cross parasitics between

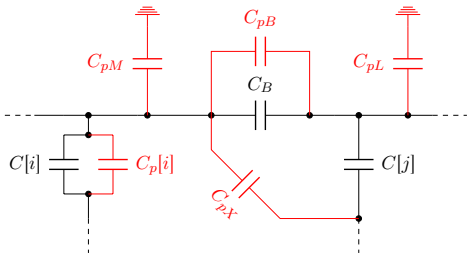


Figure 6.17: Illustration of CDAC parasitics (red) affecting the linearity of the SAR ADC

the reference nodes and the MSB or LSB node occur. The parasitic C_{pX} is particularly important. Here, a reference potential on the LSB side is directly coupled to the MSB array, "bypassing" the attenuation factor from the bridge capacitor. Thus, the voltage change at the MSB side is larger. Even for very small parasitics compared to the unit size, this can have a visible effect, and the arrays are separated by adding dummy capacitors in the area of the bridge where the MSB and LSB array are the closest.

The nonlinearity of the DAC manifest themselves in varying ADC bin sizes, quantified by the INL and DNL of the ADC (see chapter 5.5). The size of a bin is given by it's upper and lower voltage boundaries dictated by the voltage changes of the DAC. The size of the bin with code $X = \{b_{M+L-1}, \dots, b_0\}$ consisting of bits (comparator decisions) b_i is given by

$$W(X) = \Delta_j - \sum_{i=1}^{j-1} \Delta_i, \tag{6.26}$$

where j is the lowest bit differing from the previous:

$$b_{j-1} = b_{j-2} = \dots = b_0 \quad \text{and } b_j \neq b_{j-1}$$

As an example, the code X=$\{0, \dots, 0, 1, 1, 1\}$ gives j=4 and $W(X) = \Delta_4 - \Delta_3 - \Delta_2 - \Delta_1$. The DNL and INL are easily calculated using (5.29), (5.30) and the bin size (6.24), modified by the parasitics. In the linearity analysis performed previously, the capacitors C[i] and C'[i] were assumed to be the same. In fact, the difference between the differential pairs is only given by mismatch due to etching or local variances since the systematic parasitics are generated the same way and are invariant under mask misalignments if the layout is designed properly. The effect of the differential mismatch is negligible, since they only generate a change in the common mode voltage between the "n" and "p" side of the differential DAC. The ADC output code is then only affected via the common mode response of the comparator. Thus, the DNL pattern is symmetric around the central bin transition $\{0, 1, 1, \dots\} \rightarrow \{1, 0, 0 \dots\}$.

Since the DNL is a linear combination of the physical Δ_i errors, the measured DNL pattern can be used to reconstruct the parasitics. Clearly, the number of degrees of freedom is smaller than the $2^N - 2$ DNL informations. A repetitive pattern is observed in the DNL distribution, and only a subset of bin sizes need to be known in order to reconstruct the full DNL pattern. Since W(X) is only dependent on j, it is useful to define DNL_j which holds all DNL informations.

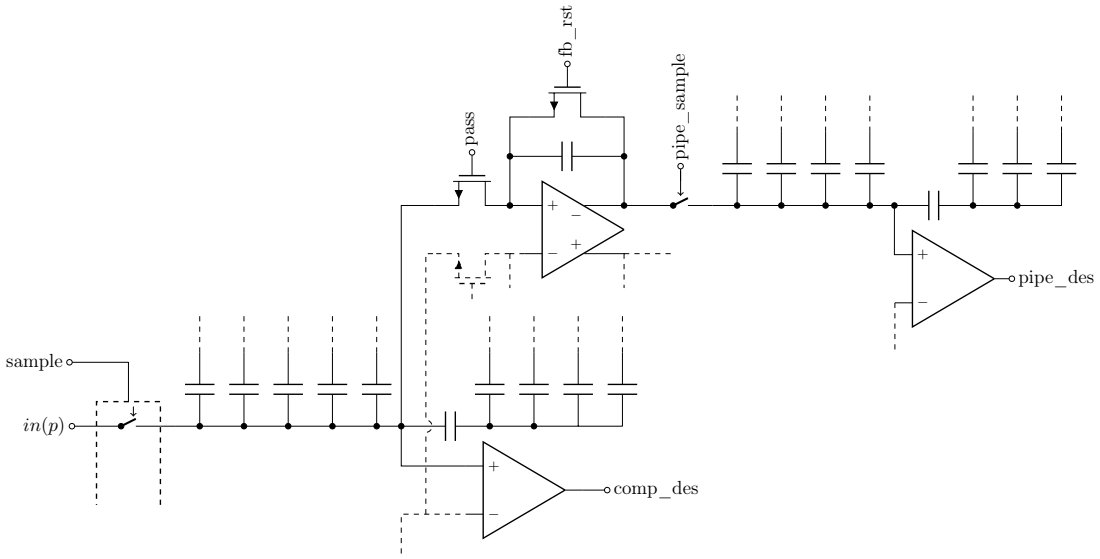


Figure 6.18: Schematic sketch of the 12-bit SAR-ADC

12 bit conversion mode

For SiPMs with a very low gain in the order of 10^5, the quantization resolution of the ADC presented in the previous paragraph is not sufficient to distinguish the single pixel peaks. When extending the DAC by two bit to the required resolution of 12 bit, the matching requirements for the capacitor array including the parasitics are much more stringent. Because a high production yield is preferred, the design of a 12 bit ADC achieving good linearity in a single stage is hardly possible in the CMOS technology used for the KLauS ASIC.

To achieve 12 bit quantization resolution, an additional pipelined stage is used to extend the quantization resolution. In this operation mode, the analog voltage is still sampled by the 10 bit SAR ADC, but the SAR operation in this converter stage is only carried out until the 5'th conversion step. The remaining residue voltage is amplified by a factor of 16 and stored in a second capacitor array included in the pipelined ADC stage, where the amplified residual voltage is further digitized with 8 bit resolution. For the largest expected residual seen at the first ADC stage, the voltage after amplification spans half the dynamic range of ± 1.8V of the second stage. This ensures that the output dynamic range of the residue amplification circuit is also not limiting the residual voltage in the case of an offset voltage of the amplifier which is propagated to the amplified residual. Figure 6.18 shows the ADC extended with the pipelined stage, for one side of the differential structure. Again, the pipelined stage is composed of a split capacitor array, in this case with a splitting ratio of 4:3, and a separate comparator used for this stage. Before the amplifier is tracking the residue voltage from the main ADC, the capacitor feedback is kept in a reset state by closing the switches connected to the *rst* pin. During amplification, the switch is opened and the *pass* and *hold* switches are closed and opened again to first amplify the residual voltage of the main 10 bit ADC and then hold the voltage on the capacitor array of the second stage in order to proceed with the conversion. The residual amplification performed to increase

the quantization resolution will inevitably add nonlinearities to the conversion, which can be corrected for if the distortions of the amplifier are known [83]. To avoid this additional calibration step, the amplifier's open loop gain is made extremely large by using a two-stage fully differential folded cascode amplifier structure. With the long settling time of 75 ns allowed for the residue amplification, the amplification error is not contributing to the overall nonlinearity significantly. Figure 6.19 shows the simulation result of a conversion cycle in 12 bit mode. During the first phase (1) the voltage is sampled and held at the input of the main ADC where the first 5 bits are registered (2). The residue is then amplified and allowed to settle for 3 clock cycles (3) and then converted with the second stage (4). The separate ADC data of the two stages is sent off chip separately and needs to be combined to a common code in software. For the code of the main SAR, $D_{\mathrm{SAR[9:5]}}$ and the code of the pipelined stage $D_{\mathrm{PIPE[7:0]}}$ in decimal representation, the merged code is given by

$$D_{12B} = D_{\mathrm{SAR[9:5]}} \cdot G + D_{\mathrm{PIPE[7:0]}} + C \tag{6.27}$$

where G denotes the inter-stage gain factor (ideally 16*8 in this representation) and C is an offset including the comparator offset (ideally 127, i.e. half of the dynamic range of the pipelined stage). The gain factor G is derived from the closed-loop gain of the residue amplifier. Due to capacitor mismatch and additional parasitics, the ratio will differ from the design value and can be adjusted by changing the feedback capacitor in a small range. Ideally, the gain should be tuned to yield an integer value of G such that all bins are covered uniformly after merging the information of the two stages.

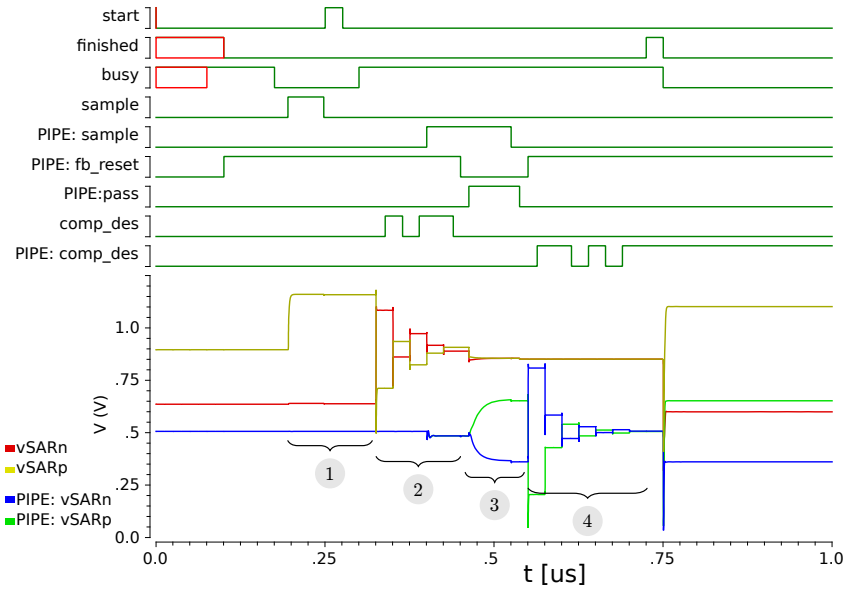


Figure 6.19: Simulation of a 12 bit ADC conversion cycle indicating the different conversion steps as described in the text

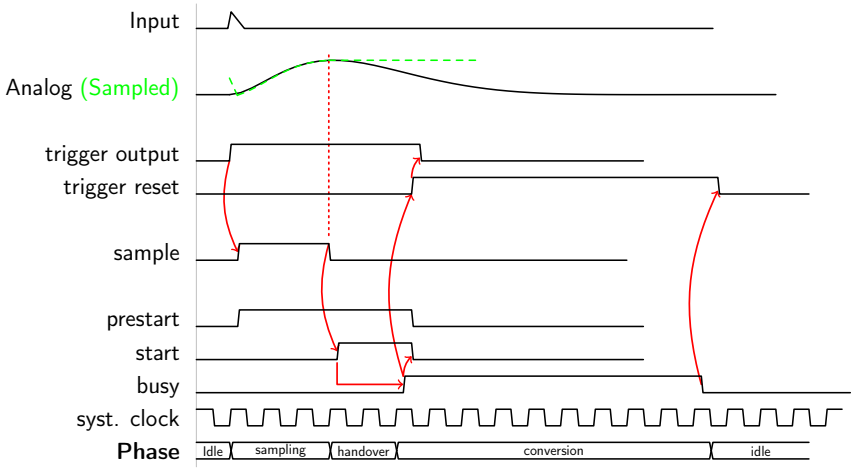


Figure 6.20: Timing diagram illustrating the analog-digital handshake between the hitlogic and the channel control logic.

6.5 Channel control logic and mixed signal interfaces

After a charge pulse from the sensor has triggered the comparator in the front-end, the peak voltage is sampled by the ADC, a conversion cycle is initiated and the gain selection flag and time stamp is stored. Finally, the digital data containing all hit information are committed to the digital part of the chip combining the data from the different channels. These processes are controlled in two separate digital blocks: The asynchronous *hit-logic* residing in the front-end channel, and the *channel control logic* which is synchronous to the system clock. Hand-shake processes are implemented to pass the boundary between the asynchronous domain of the front-end and the synchronous control logic, as well as the data interface between the control logic and the later processing stages. This ensures the hit is always passed to the next stage and blocks further events until the hit has been taken over correctly.

Front-end to channel control logic interface

Figure 6.20 shows the timing diagram of the handshake process at the interface between the front-end hit-logic and the channel control logic. By adding a fixed delay to the comparator output, the sampling signal is generated and directly connected to the ADC's sampling switches. After the sampling phase is finished, the hit-logic asserts the *start* signal. After the flag has been synchronized to the system clock in the channel control block using a two flip-flop synchronizer, the busy signal is asserted to notify the hit-logic that the hit has been taken over by the ADC control logic, which starts the conversion. The busy flag clears the start signal, finishing the handover process between the two blocks. The busy flag is kept high until the end of the conversion to block all hits that might be generated by the comparator during the time the ADC can not sample a new event (see also chapter 6.3.5). The *prestart* signal is also generated from the comparator and is deasserted with the busy flag from the ADC control logic. It is used to switch on the ADC comparator already during the hold phase, which allows the block to settle before the SAR operation is

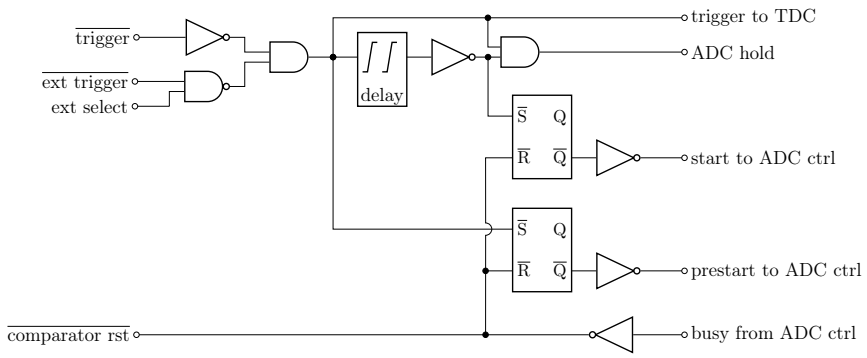


Figure 6.21: Simplified schematic of the hit-logic block. Left side: Connection to the comparators. Right side: Interface to the channel control logic.

started. Figure 6.21 shows a simplified schematic diagram of the hit logic block. Residing in the front-end, the custom logic cells are designed to isolate the digital switching activity from the surrounding analog part. The trigger signal can be configured to come from the comparator, or from an external pad of the chip for characterization purposes and cases where the internal comparator can not be used. The ADC sampling signal is generated from logic AND of the trigger and an inverted and delayed version. The delay cell is designed based on a CMOS-thyristor cell [84] [85] and described in more detail in Appendix A. Since the delay needs to be defined accurately to sample the peak voltage, the bias current of the cell defining the coarse delay is generated globally for all channels, and a fine tuning is implemented in the delay cell to allow independent adjustments for every channel. The jitter of the hold time also includes the jitter introduced by the delay cell. It is designed to not contribute more than 1 ns, including noise from the power lines during the comparator decision time which can affect the delay time significantly. The start and pre-start signals connected to the channel control logic are set by the trigger and delayed trigger signal and stored in RS-latches. Both latches are reset by the busy signal from the synchronous channel control block, implementing the handshake between these two blocks.

Channel control block

The channel control block shown in Figure 6.22 is implemented in the data path after the front-end's hit logic. It is designed using the hardware description language VHDL, and synthesized using the Faraday digital standard cell library [86]. It is comprised of the ADC control logic, registers to store the automatic gain selection information and the timestamp, a single-event buffer for the digitized event data and a block steering all the conversion and implementing the handshake to the front-end and digital buffer. All analog and digital blocks shown in the diagram are placed in the same hierarchical entity, which allows to validate the full channel block in mixed mode simulations.

The conversion control module processes the flag signals from the hit logic as described in the previous paragraph. After the sampling phase is finished and the start signal of the front-end is synchronized to the system clock, it starts the ADC conversion and issues the gain selection bit to be sampled.

The gain selection bit determines the branch that will be digitized. It is saved by the

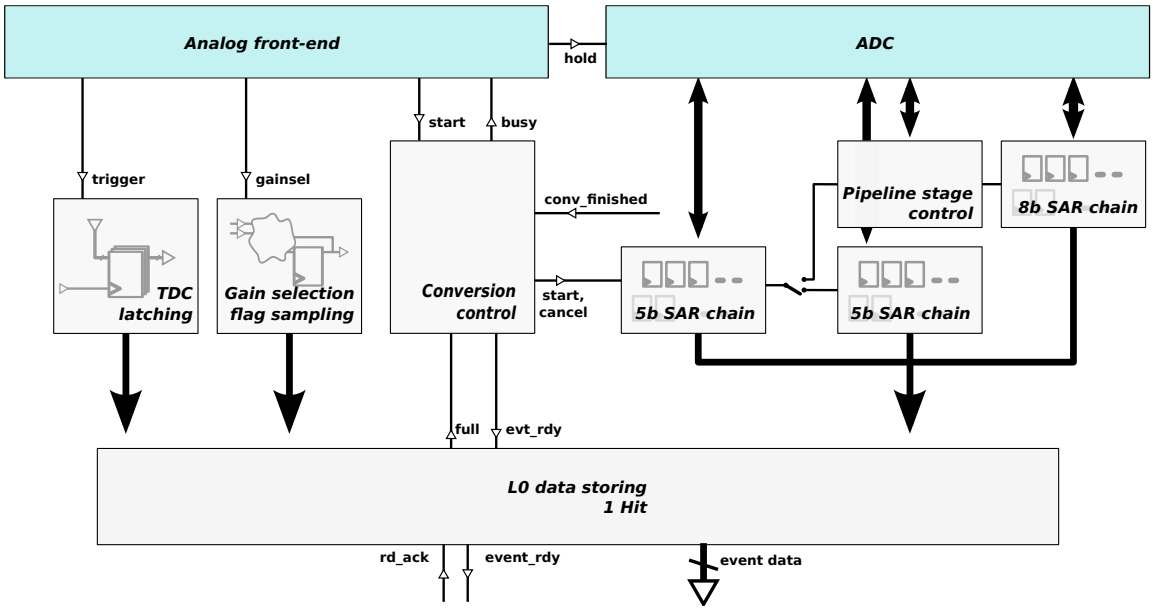


Figure 6.22: Block diagram of the channel control digital part

channel control logic after the start has been sampled. Since the comparators are reset while the busy signal are asserted, i.e. after the start is sampled, it is ensured that also the gain selection bit is kept stable until stored. While the ADC conversion is in progress and the channel is blind to additional pulses, the state of the gain selection comparator is sampled to get a single-bit information of hits with large charge happening during that time.

To implement the two operation modes of the ADC, three chains of the control logic discussed in Chapter 6.4 are used. In the 10 bit mode, the two chains of 5 bit in length each are connected to the main SAR array are loaded with the strobe signal consecutively. For the 12 bit mode, the output of the first (MSB) chain controlling the main ADC instead propagated to a block steering the residual amplification switches of the pipeline stage, followed by an 8 bit SAR chain for the pipelined ADC stage. All informations - the ADC results, gain selection bits and time stamp - are combined and stored in a single event buffer (*L0 buffer*). After the data is stored in the buffer, the channel is ready to digitize a new event. In the case the L0 buffer is still full when the conversion is finished, the channel is kept in a busy state blocking triggers from the front-end until the event data can be stored. The data from the L0 buffer is then presented to the later digital part, again a handshake procedure is implemented to ensure no event data is lost.

Time to Digital Converter

As summarized in chapter 6.1, a time resolution in the order of 1 ns is required by the AHCAL. Since no time to digital converter module compatible with the power budget was available, the development of the necessary blocks, namely a low power phase-locked-loop, counters and dynamic latches were postponed to later versions of the chip. Instead, only the coarse counter part of the TDC was implemented in the synthesized digital part. The TDC can provide a time stamp with bins corresponding to the system clock period of 25 ns.

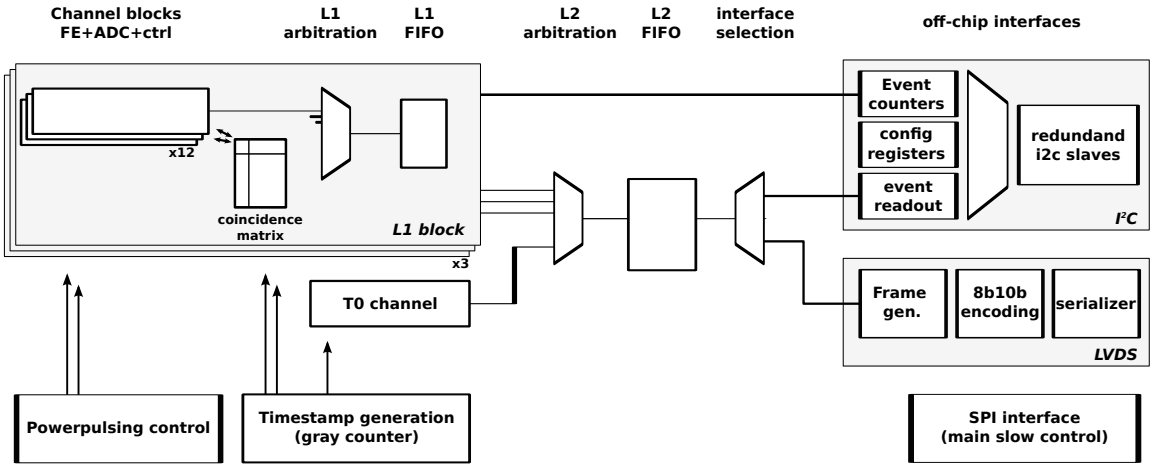


Figure 6.23: Block diagram of the ASIC. The numbers in the Figure correspond to a 36 channel ASIC

A counter running at this clock frequency is generating a time stamp, which is distributed to the control logic of all channels. The time stamp is gray encoded[87], meaning that at every clock cycle, only one bit of the time stamp changes. If the time stamp is latched with the rising edge of the front-end's trigger output signal happening close to the clock edge when the counter value changes, the recorded value can only be wrong by one count, in contrast to other encodings like the binary or PRBS counters where many bits change at the same time and a single wrongly sampled bit can introduce a tremendous error. The comparator output signal from the front-end is buffered and fed directly to the channel control logic, where the time stamp is stored at the rising edge. While the busy signal is asserted to the front-end's hit logic, the comparator is inhibited from generating new triggers. Thus, the time information latched at the rising edge is not overridden until the end of the conversion and can safely be registered by the clock-synchronous digital part of the control logic. The dynamic range of the counter is 12 bits ($\approx 100\mu$s @ 40MHz system clock) for the seven channel prototype. In the 36 channel ASIC, the dynamic range was extended to 16 bits ($\approx$ 1.63ms @ 40MHz) to in accordance with the ILC bunch structure. The coarse counter will be extended with a fine counter which will then provide the additional resolution required by the experiment.

6.6 Digital part

The main digital part of the chip processes the events after they have been digitized in the channels. The data from the channels are combined to a common datastream, buffered and sent off-chip over a serial data interface. A diagram of the individual blocks is shown in Figure 6.23. The channel blocks, consisting of a Front-End module, ADC and control logic as shown in Figure 6.22 are combined in a group of 12 channels in the Level 1 block. The data from the 12 channels is merged using a round-robin arbiter, and buffered in the L1 FIFO which can store 64 events.
The data from the three L1 groups is again combined by another arbiter and buffered in the

L2 FIFO with a capacity of up to 128 hits. The capacity of the L2 FIFO was extended to 512 hits in the 36 channel version of the ASIC. In the process of the L1 and L2 arbitration, channel and group number fields are added to the event data to fully identify the channel number. Besides the three L1 groups of channels connected to the L2 arbiter, a separate *T0* channel is implemented. This channel stores a time stamp at the rising edge of a signal from an external pad, which can be used to add timing informations of an external signal, for example from a trigger scintillator or a slow clock in order to synchronize the timestamps from several ASICs. As discussed in the previous paragraph, the time stamp used by all channels is generated by a global gray counter running at the system clock frequency, included in the toplevel hierarchy of the chip. The number of channels in the L1 block was selected to avoid a bottleneck in the instantaneous rate of hits that can be processed by a chip of 36 channels while keeping the number of FIFOs minimal. The dead time of the channel, i.e. the maximum event rate a single channel can produce data, is 12 clock cycles plus the hold time. With a L1 group of 12 channels connected to an arbiter which can process one event per clock cycle, it is assured that no bottleneck can occur in this state. At the same time the number of L1 FIFOs is kept small. While the overall rate limitation for the full chip will be governed by the L2 FIFO throughput or transmission speed, the instantaneous rate the channels can process is still maximized in the design.

To control the power pulsing functionality of the chip, a block generating all necessary control signals is added in the toplevel. It samples the state of a digital pad and enables the blocks in the channels in the correct time and order. The power pulsing function will be discussed in Chapters 6.7.1 and 6.7.2.

In the commissioning of a detector system consisting of thousands of channels running in auto-triggered mode, one key point is the selection of the correct trigger threshold for every channel. To simplify this procedure, an 8 bit hit counter is implemented in every channel. This allows to monitor the hit rate without reading the full event information, which would saturate the data link while tuning the thresholds, where the comparator threshold for some of the channels might be configured to be at the noise level.

6.6.1 Event Data transmission

The event data buffered in the digital part is sent off-chip using either a slow I^2C interface [88] with a data rate up to few Mbit/s, or a faster LVDS link providing a data rate of 160 Mbit/s. The selected interface will then read the data from the L2 FIFO and send it to the DAQ hardware.
The interface selection module connects the data and clock line at the reading side of the L2 FIFO to the selected transmission interface. Each event consists of 38 bit, padded to 40 bit (5 bytes) before transmission. It holds the channel number, gain selection flags, TDC and ADC informations.

I^2C Interface

For the CALICE calorimeters, the average hit rate per channel is small and the number of signal lines needed to transmit the data should to be minimized to reduce the routing area on the long PCBs in one calorimeter layer hosting many chips. This can be done by sharing the data lines by several ASICs in one calorimeter layer. In the I^2C transmission standard,

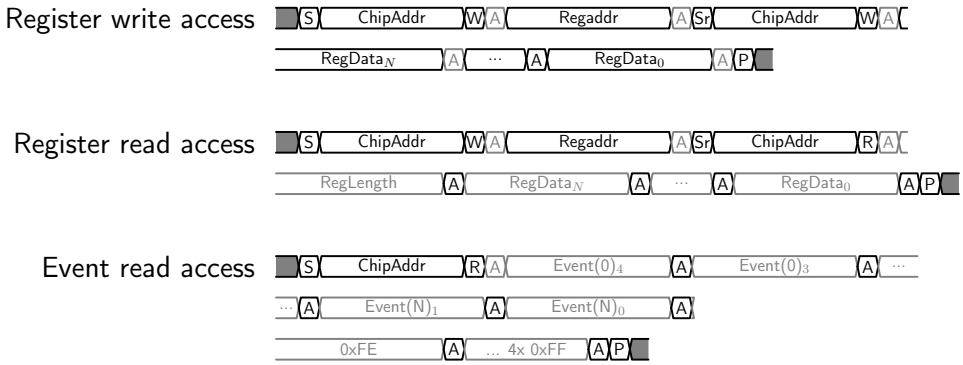


Figure 6.24: I^{2}C transaction types used to communicate with the ASIC. The *S*,*Sr* and *P* bits correspond to the I^{2}C start, restart and stop conditions. R/W is the write/read bit and *A* is the acknowledge bit as described in the I^{2}C specification. The data in light grey corresponds to data sent by the I^{2}C slave (KLauS ASIC).

multiple slave devices (ASICs) can be connected to a common pair of data and clock lines, and the individual ASICs are accessed by their 7 bit address. The I^{2}C specification gives a speed limit of 5Mbit/s in the ultra fast mode. This speed is already sufficient for the rates expected in the CALICE detectors, with the requirement to read out all ASICs within the time of no collisions (≈200ms). Depending on the expected noise level of the detector causing more hits to be read out, the speed can also be enhanced by reducing the capacitive load (i.e. number of ASICs) connected to the bus. The I^{2}C interface implemented in the chip consists of two separate slave instances which can be connected to two different buses. In the case of connection failures along the PCB, the two redundant slave instances allow to maintain a communication to the connected ASICs. The two buses can also be used to increase the readout speed as two ASICs can be read out at the same time. The I^{2}C interface is used to read out event data from the L2 FIFO, to read the channel event counter information, and to read and write control registers configuring the I^{2}C slave interfaces, e.g. the driving strength of the pads. The I^{2}C interface allows tree different access modes shown in the timing diagrams in Figure 6.24: Read and Write of the configuration and channel event counter registers, and a read transaction of the event data. The readout protocol is based on the SMBUS specification [89], which defines the read and write access to the registers (channel event counter and slow control registers of the I^{2}C part). For the read out of event data, normal I^{2}C block read transactions are used to reduce the communication overhead.

In all cases the ASIC is first addressed with it's 7bit I^{2}C address and a bit determining the direction of the transfer. For the SMBUS-based register transactions, a register address is written to the ASIC. After selecting the register, it can be read or written to. For the most frequently used event readout transaction, no register address is written to the chip and the data is directly transmitted by reading blocks of data from the I^{2}C slave. To inform the I^{2}C master implemented in the DAQ hardware that there is no more event to read, the ASIC will transmit an "empty event" indicated by an unused channel number.

Fast LVDS interface

For many other applications of the chip beyond the scope of the CALICE detectors, the rate limitation imposed by the shared connection of the I^2C bus will not be adequate. To make the chip more versatile, a second and much faster serial interface was implemented which can be used instead of the I^2C interface at the cost of higher power consumption and the necessity of individual data lines between chip and DAQ hardware. The interface uses a serial data link with a rate of 160Mbit/s. The data stream is 8b/10b encoded to balance the differential signal. The additional code words available in the datastream are used for clock recovery and data word alignment on the receiver side, as well as to identify header and trailer fields enclosing the event data. Every 1024 clock cycles, a header including preamble, packet number and number of events in the packet is transmitted, followed by the events of 5 bytes each. The packet is finished with a trailer field including a 16-bit CRC field used to identify transmission errors.

6.6.2 Configuration interface

For compatibility with the current DAQ firmware, the main configuration interface is implemented as an SPI chain. The slow control is used to configure the front-end blocks, ADCs and functionalities of the digital block. The interface is implemented as a chain of flip-flops which can also be connected through several chips. After filling the chain with the configuration data, is stored in a separate set of data latches connected to the blocks in the chip. For the SPI chain, three signal lines are needed to configure the chips connected in the chain. A drawback of this scheme is a lack of readback functionalities without changing the configuration data, while with increasing amount of configuration data, the chance of transmission errors rises. Future versions of the KLauS ASIC are planned to use registers connected to the I^2C bus in favor of the SPI interface.

6.6.3 On chip data reduction

In test beam campaigns, the particle rate and spill structure usually different from the ILC scheme with it's short collision time followed by a long time without collisions, in which the data from the ASICs is read using a slow link. For efficient data taking, this requires a longer acquisition time increasing the data volume that needs to be transmitted. For the electromagnetic ECAL, a scintillator option is under discussion. In this approach, scintillator strips are read out using SiPMs, and the response is much smaller compared to the tiles used in the analog hadronic calorimeter. Since the threshold needs to be reduced in order to maintain a good trigger efficiency, also the rate of noise triggers from the dark count pulses of the SiPMs increases. This can be mitigated by using a dual side readout scheme, where the amount of data to be transmitted can be reduced by on-chip coincidence logic.

Both of these cases require a validation logic to select which of the recorded events should be kept and transmitted, and which should be dropped in the chip. For the first situation, external scintillator panels are used to select the time frames of interest. In the second case, the detector is still self-triggered by the coincidence logic.
Two different methods to reduce the data are implemented in the chip. The span different time scales of event validation, i.e. the maximum time until the decision if the event should

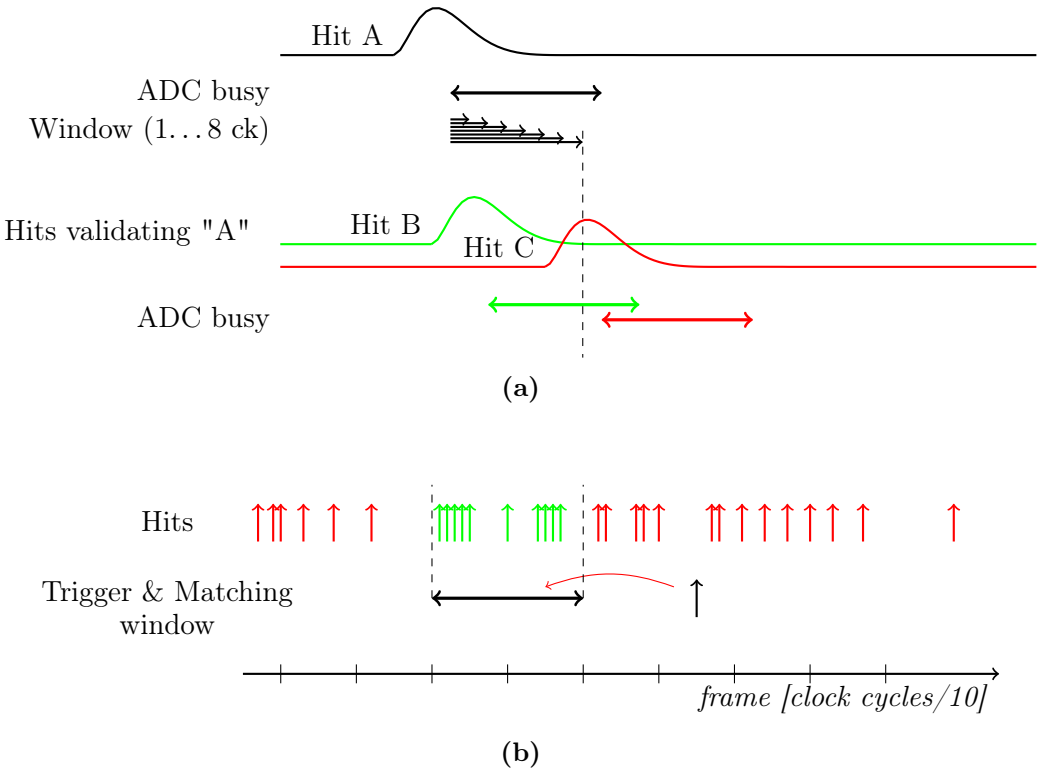


Figure 6.25: On-chip event validation features implemented in the chip. a) L0 validation implementing a coincidence logic. b) L1 validation from an external source with a configured offset and window of 1μs and 400ns, respectively.

be kept or not needs to be made. The first method is implemented using the channel control logic. While the ADC conversion is in progress and before the data has been stored in the L0 event buffer, the event can easily be dropped by cancelling the conversion by resetting the SAR conversion chain and conversion control circuit. For this purpose, the channel control logic has an input which allows to validate (or otherwise cancel) an ADC conversion. The validation signal can be configured to come from an external pin, or it is derived from the busy signal of any OR-combination of the 12 channels in the same L1 group, implementing a simple coincidence logic between any of the channels in this group. The validation time window can be configured to be up to $\pm$8 clock cycles. An example of the operation mode is given in Figure 6.25a. The first channel with hit A is configured to be validated by another channel (i.e. a coincidence logic). In the first case, hit B, the busy signal of the second channel is asserted within the validation window of hit A, validating A. In the second case, hit C, A is not validated and would not be stored in the L0 buffer.

The limitation on the validation window, i.e. the time until a validation signal needs to be present in the method described before can be a too tight constraint in some cases, for example when the trigger signal is coming from an external source propagated through a complex triggering scheme with longer latency. To cope with these requirements, a second event validation option is implemented which allows a longer trigger latency. The event

data can not be kept in the channel until validated to avoid an increase of the dead-time of the channel. The second level of validation is therefore implemented in the L1 FIFO and triggered by an external source [90] in order to minimize the dead time. The trigger decision is based on the time frame when the event is stored in the L1 FIFO, the rising edge of the trigger signal, and a configurable time window. All events stored in a specific window are validated on a positive trigger decision. The width and offset of the window can be configured in frames of 10 clock cycles, with a maximum offset of 4μs from the first validated event to the trigger condition and a width between 400ns (10 clock cycles) and 8μs. Figure 6.25b shows an example of the L1 validation. Here, the validation window is configured to an offset of 1μs and a width of 400ns.

6.7 Power pulsing

The ILC bunch structure allows to reduce the power consumption of the readout ASICs integrated in the detector. For nominal operating conditions of the accelerator, a bunch train with a length of about 730μs is followed by a time without collisions of about 200ms [19]. While there are no collisions most of the chip is put in a standby state, consuming only a minimum of power. The readout electronics will inevitably need some settling time, so the ASICs should be turned on before the collisions are expected. A key point is therefore to speed up the electronics such that the duty cycle can be reduced towards the bunch train length. This allows a switching duty cycle of the electronics in the order of 0.5%.
Another requirement on the electronics is to keep the sensors connected to the ASICs in a stable operating condition. Since the bias voltage is tuned by the input stage, also the offset voltage generated by the input stage should be stabilized during power pulsing. The corresponding DAC should therefore be enabled at all times and was designed to achieve a low power consumption (see Chapter 6.3.2). Also the input stage should traverse from low power mode to active mode without introducing distortions or offsets of the input terminal voltage.

While enabling the analog front-end block, noise triggers generated by the internal comparator should be masked to avoid unwanted hits to be digitized by the ADC. As mentioned in the description of the digital part, different control signals are needed for the power gating operation. They are generated by a digital module in the synthesized digital part. It is controlled by a single pad on the chip which determines if the chip should be in the low power or active mode. During the switch-on procedure, the different signals connected to the input stage, feedback, bias modules and the channel control logic blocks also steering the ADC are switched with configurable delay. Figure 6.26 gives an illustration of the switch-on procedure. The behaviour of the different blocks, as well as additions to the circuits necessary to allow the power pulsing, will be described in the following.

6.7.1 Analog Front-end

All blocks that do not need special care are switched by enabling or disabling the bias current in the global bias generator. The digital activity that might cause unwanted triggers and thus ADC conversions it excluded by the digital part. The individual mask signals connected to the channel control logic blocks are overridden with a power pulsing control signal to

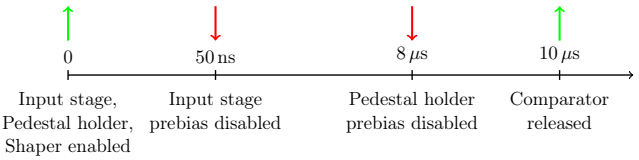


Figure 6.26: Illustration of the switching procedure in power-pulsed operation (time axis not to scale)

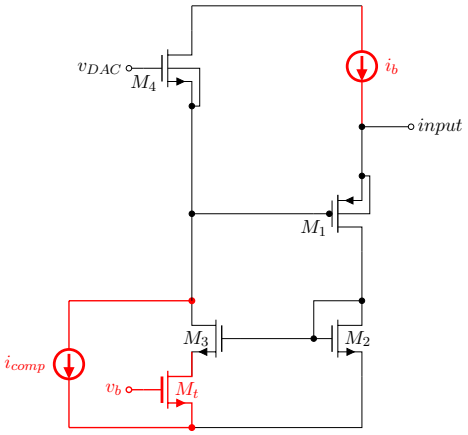


Figure 6.27: Input stage schematic including circuits for powergated operation. Parts that are actively changed or added compared to the schematic in Figure 6.3 are highlighted in red.

mask all channels while the front-end is enabled. This results in a constant reset of the comparators, masking any triggers during this time.

Input stage

The main part of the input stage including the additional components for the powergated operation is shown in Figure 6.27. In the low power mode, the bias current i_b of the input stage is reduced from about 150μA to a standby current of about 200nA. Since this can change the input impedance towards the instability region, the feedback path of the input stage is disabled by switching the bias voltage v_b of the triode transistor M_t to ground. A constant compensation current of about 100nA is added to the feedback branch, tunable using a 3bit current DAC. This allows to minimize the voltage differences of the input terminal voltage with respect to the active state. Since the current is very small compared to the on-state bias current, it is always enabled without having an effect in the normal operation mode.

While small changes in the input terminal are acceptable during the transition between low power and active state, large excursions especially in the turn-on procedure should be avoided. When the input stage bias current rises during switching to the on-state, the input stage mirror (M_3,M_2) will take some time to settle, causing a relatively large positive swing on the input terminal voltage. This is compensated by injecting a larger current of about 20μA through i_{comp} in the first 50-100 ns after enabling the input stage.

Pedestal stabilization

Switching the input stage bias current also changes the DC current copied to the processing stages such as the two charge measurement branches. Since these branches include a slow DC feedback circuit stabilizing the pedestal voltage they would, if not handled, take a long time to adapt to the new DC current. While this is no problem for the on-to-off transition, clearly the switch-on procedure should be as fast as possible to avoid switching on the front-end significantly before there are collisions. The pedestal feedback circuit has been shown in a simplified manner as a building block in Figure 6.6 and a controlled current source in Figure 6.7, which was sufficient to describe the signal generation for the charge measurement. Figure 6.28 gives the schematic of this part, including the necessary components for power gating. A differential amplifier (M1-M4) is tracking the difference between the expected pedestal voltage and the voltage at the output of the channel. In the active operation mode, the amplifier is biased in the sub-threshold regime which enhances the gain. The 3dB bandwidth is put to O(1Hz) using a filter capacitance. The output of the amplifier is normally connected to the gate of transistor Mf, which generates the current fed back to stabilize the DC voltage at the integration stage and thus the front-end's output.

In the low-power mode, the feedback current is disabled by setting the *enable* pin to a logic low level, which disconnects the amplifier output and shorts the gate of Mf to the vcca18 supply. This is necessary to avoid additional power consumption in the low power state caused by copying signal currents coming from the input stage. In the turn-on process, the settling time is mainly determined by the time it takes to recharge the capacitor back to the correct voltage, which is given by the maximum output current the amplifier can deliver since it is slewing during that time. Thus, the settling time is reduced by increasing the bias current of the amplifier until the pedestal is close to the expected value, which is done by switching the $\overline{pwrup}$ pin to a logic low level. Since the feedback loop is working also after the boosting is finished, the time the speed is enhanced in this way does not need to be defined accurately and should be put at the time just after the pedestal voltage reaches the expected value. Figure 6.29 shows a simulation of the pedestal voltage and the relevant digital control signals. After enabling the front-end, the feedback amplifier bias is increased until the pedestal voltage has reached the expected voltage as shown for the green lines. For comparison, a simulation result where the additional current is switched off too early is shown in the red lines. After switching off the additional current, the pedestal is slowly approaching the pedestal voltage. As a second example, the case where the current is decreased too late is also shown. In this case, the pedestal voltage is recovered fast, but with some distortions after the switching point. Because the settling behaviour is still fast, it is possible to configure the delay without requiring precise tuning. The typical configuration value will be placed at a delay slightly longer than the ideal value without deteriorating the settling time significantly. For the ideal configuration setting, the simulated recovery time of the pedestal in the high gain branch is about 10μs, with the low gain branch recovery being slightly faster due to different bias points because of the lower DC current copied from the front-end. While additional time is needed to settle the pedestal in the mV level, the overall settling time is small enough to allow the power pulsing operation of 0.5% in the nominal ILC beam parameters. From the simulations, a power consumption of 26μW per channel for a 36 channel ASIC is expected under the nominal configuration settings of the front-end.

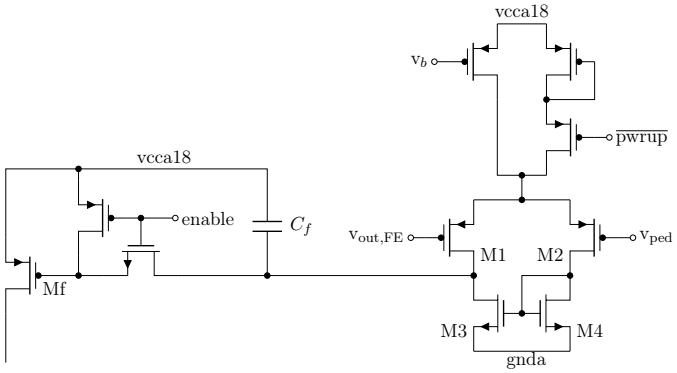


Figure 6.28: Schematic sketch of the low-frequency feedback circuit used for pedestal stabilization including circuits to reduce the settling time in power gated operation.

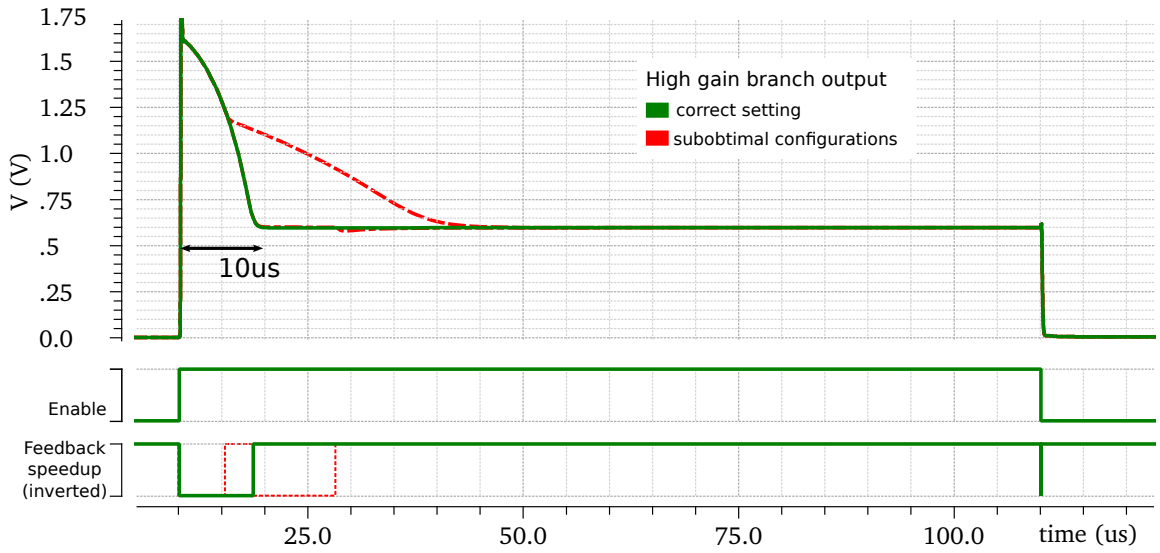


Figure 6.29: Simulation of the pedestal voltage settling behaviour in power pulsing mode (high gain branch). The lower digital traces show the logic levels of the relevant power pulsing control signals (compare Figure 6.28, the Feedback speedup signal refers to the $\overline{\text{pwrup}}$ signal). For comparison, the simulation is performed also for suboptimal settings of the control signal delays, shown as red lines. Note that the front-end is switched off again after 100μs in this simulation, shorter than the nominal ILC operation scheme.

6.7.2 ADC and digital part

While the front-end is consuming the majority of power in the on-state of the chip compared to the ADC and digital part, these blocks should still be gated since even a small power consumption not being power pulsed will contribute significantly to the overall power consumption of the chip.

As discussed in chapter 6.4, the ADC is already self gated on an event-to-event bases, since the capacitor array itself is only consuming power while a conversion is active. The comparators and residue amplifier do consume power while inactive, but are switched off while there is no conversion. The event-based switching of these blocks is possible because their settling time is shorter than the peaking time of the front-end's output pulse, so the blocks can be enabled just before the conversion without adding any dead-time. Clearly, this approach is even better than the power pulsing with the bunch train structure since the rate per channel is low and the conversion time is short.

For the digital part of the chip, two sources contribute to the power consumption and can be tackled with different methods. The static leakage power is coming from the transistor leakage through the gate and channels of the very short transistors in the digital cells. The leakage current rises drastically for smaller technologies[91] and can only be reduced by switching off the supply voltage of the hierarchical modules that are not in use. In the 180nm technology used for the KLauS ASIC, the leakage power in the digital part is in the order of nW and can be ignored. A larger contributor in this case is the dynamic switching power due to charging and discharging of the parasitic capacitances in the digital part while there is activity, including the clock distribution network itself. The natural solution is to reduce the activity in the digital part by switching off the clock. A simple method would be to do this on the DAQ side, which also reduces the power consumed by the clock distribution chips on the PCBs, but this can only be done when all data in the chip is transmitted[7]. To reduce more power, the clock can be switched off in parts of the chip that are inactive. These clock gating techniques[92] are well known strategies to save power in digital circuits. This functionality will be implemented in future versions of the chip, since the power consumption and thus saving potential are not modelled realistically by the digital library, in particular the SRAM cells used in the FIFO buffers. In principle, an event-based clock-gating can be implemented to save the maximum of power in the digital part.

6.8 Design verification

All blocks and the final design were simulated extensively before it was sent for production. Since a full analog simulation of the chip is not feasible within a reasonable time, different simulators (i.e. abstraction levels) are used to simulate blocks in the different domains.

The front-end and the shared bias generation block were simulated using the analog simulator SPECTRE, first using abstract models of the building blocks for feasibility studies, then on the full schematic and finally including the parasitics from the extracted layout. The impact of component mismatch and process variations were studied using automatized

[7] For the I^2C readout option, the clock may be switched off as soon as all data resides in the L2 FIFO (i.e. the L0 and L1 buffers are empty). The I^2C interface provides it's own clock signal which is sufficient to read the L2 FIFO.

simulations of the key specifications of the chip, for example the integrated noise and charge measurement linearity. The ADC, it's control logic and finally the full channel including block in the analog and digital domain (FE,ADC and control logic, see Figure 6.22), have been simulated using the AMS Designer mixed mode simulation framework provided in the Cadence software package.
For the verification of the digital part, specific simulations of the analog modules were carried out characterizing the logic and time behaviour of these blocks. Based on these simulations, abstract digital models of the front-end and ADC were developed and put into the digital simulation which can run much faster given the simplified models. This allows to cover a larger amount of cases in the digital part to exclude special conditions that may cause the digital part to work differently as intended.

The verification of the correct behaviour was carried out using input data randomized both in the data contents and time of generation. A digital verification environment was implemented, based on the Cadence Universal Verification Methodology (UVM) standard. Written in the proprietary language "e"[93] specifically covering the need of automatized digital verification, the framework allows to develop reusable verification modules and randomizing stimulus drivers. The stimulus drivers generate events with randomized data for every channel and implement the abstract models of the analog modules (mainly the front-end's hitlogic). As the events propagate through the digital part and are "read out" in the test bench, monitor modules sample the events passing the different stages in the design. Such monitor instances have been implemented for the L0, L1 and L2 buffer outputs, as well as the deserializer or I^2C master instance included in the testbench. The contents of the sampled events are compared against the produced ones, expecting the events are buffered and transmitted in correct order and no event is lost in the process. By verifying the data at all critical stages, potential bugs are found reliably and can be identified quickly. Simulations for different scenarios of event rates for single channels, groups of channels and the full chip have been performed to exclude critical errors in the data path.

7 Characterization measurements

The functionality of the chip was validated in characterization measurements of the analog front-end, the ADC and the full readout chain. The results of these measurements will be presented in the following chapter and have been obtained using the 7-channel prototype ASIC. The individual building blocks and the performance of the full readout chain were characterized in measurements of single channels in the laboratory. To evaluate the chip performance under realistic conditions, a setup consisting of three ASICs was assembled with Silicon Photomultipliers and scintillating tiles. It was successfully used at the DESY test beam facility.

7.1 Test setup

In order to perform the characterization measurements with reproducible results, a test setup including all the necessary components to operate the ASICs was developed and is shown in Figure 7.1a. It consists of three printed circuit boards (PCBs):
The *test-board* hosting the chip, reference voltage generation circuits for the front-end and ADC blocks, as well as buffer amplifiers for the analog monitoring signals. The *test-board* is connected to an interface board including the voltage regulators for the supply voltages, the clock oscillator and fanout buffers for the control signals. The chip readout and configuration, as well as the control of the power regulators and the oscillator, is carried out by a commercial ARM processor board (Raspberry Pi-3b) running a standard linux operating system. Since the ARM processor provides all peripherals required to operate the chip (GPIO pins, an SPI and I^2C interface), the DAQ can be written conveniently in C++ code. For the DAQ and slow control software, a server-client scheme was applied. The ARM processor handles all communication over the physical interfaces in one thread, and sends the event data using a separate server thread over the network interface when requested by client DAQ and slow control software running on one or several separate PCs. This allows very flexible data quality monitoring solutions in parallel to the data taking. For the serialization and transmission of the event data, the network and message classes of the ROOT data analysis framework[94] are used.

The test system PCBs and the software framework was developed also allowing the possibility of measurements in test-beams using several ASICs connected to the same DAQ hardware. The test-board allows to solder surface-mounted SiPMs on backside of the PCB which is kept empty otherwise. The sensors are distributed to fit the dimensions of the CALICE scintillating tiles, and a UV LED-system developed by the DESY FLC group and the university of Wuppertal[95] is integrated. The light emitting diodes generate light pulses triggering scintillation light, which are read out by the sensors. This allows to calibrate the SiPM gain. The total height of the test board including the scintillating tiles has been limited to 7 mm, allowing the integration of the board into the CALICE absorber stack.

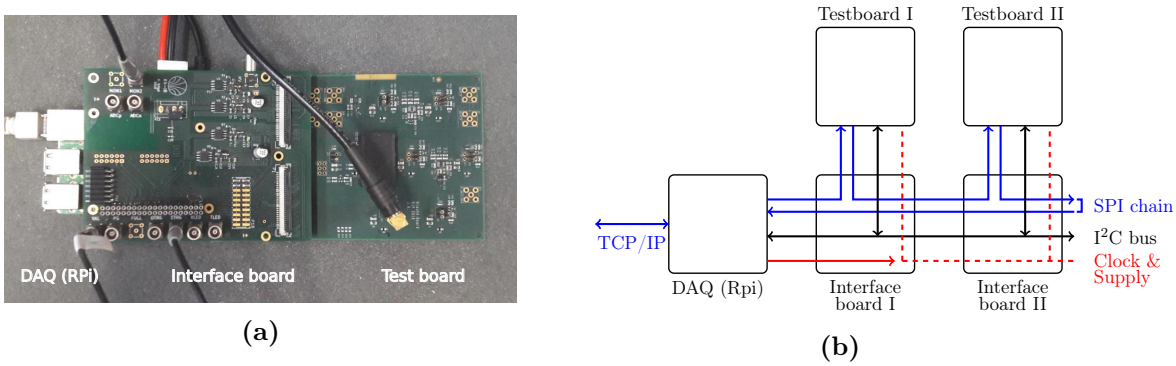


Figure 7.1: The KLauS4 test setup.
a) Picture of the test setup including the DAQ processor board, one interface and one test board. b) Overview of the test system components and their interconnections for a system comprised of multiple ASICs.

For the operation of multiple ASICs read out by the same data-acquisition board, several ASICs and interface boards can be connected as shown in Figure 7.1b. The SPI slow control interface is chained, and all ASICs are connected to the same I^2C bus. The supply voltages for the chip are generated on the first interface board and distributed to the other boards.

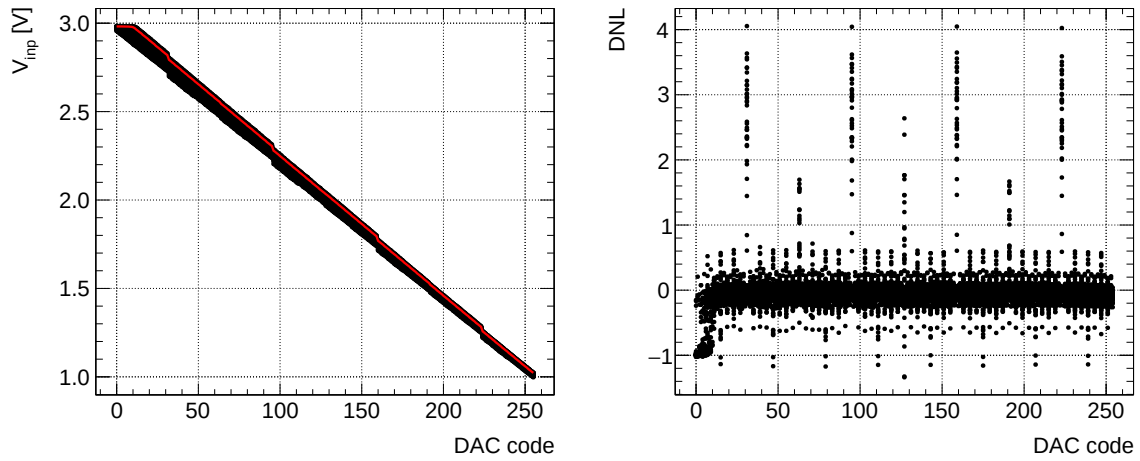
7.2 Analog front-end

The performance of the analog front-end blocks have been studied without using the integrated ADC in order to separate the effects of these blocks. The common bias-block connected to the front-ends includes a set of voltage buffers which allow to measure the analog output voltage of the front-end after the shaper amplifiers. Since the signal used to monitor the analog output of the front-end shows pollutions from the digital domain, the clock to the digital part was switched off for the front-end measurements. An exception are the studies of the comparators, since these blocks require a running clock to be functional. The analog performance on the chip is not affected by the digital noise. The pollution is only visible on the board level and comes from noise coupling of the digital ground potential to the connector used to monitor the analog output.

7.2.1 Input terminal DAC

The tuning range of the DC voltage at the input terminal of the channel was measured with a multimeter[1] as a function of the configured DAC value. The measurement was performed for 21 channels from three ASICs in order to evaluate the spread of the nonlinearities and the voltage slope. Figure 7.2a shows the measured voltage at the input terminal, with one channel highlighted to visualize the nonlinearities introduced by the DAC and input stage. For a linearity requirement of better than 1.5% of the full scale range, a tuning range of more than 1.8V is obtained for all channels. A visible nonlinearity is observed at small DAC values, limiting the maximum voltage to about 3V. This is an expected consequence of the

[1] Agilent/Keysight 34401A $6^1/_2$digits digital multimeter



(a) **(b)**

Figure 7.2: Voltage at the input terminal as a function of the voltage DAC code a) Measured voltage for 21 channels from 3 ASICs, showing the dispersion of the slope and DNL effects. The highlighted measurement has been selected arbitrarily. b) DNL distribution of all measured channels

input stage bias current source transitor entering the triode region. From a linear fit in the value range between 15 and 255, the voltage slope and DNL pattern is extracted. Without trimming the bias current in the generator circuit, the RMS dispersion of the voltage change per LSB code is 1.24% RMS for all chips. For the individual ASICs, an average of 0.92% RMS is obtained, showing that the mismatch of the resistors and the current sources in the digital to analog converters has almost the same contribution as the bias generator circuit and the voltage reference. Most of the nonlinearities visible over the full range are generated by transistor mismatch of the current DAC, which manifests in a DNL pattern visible for the highlighted measurement and in Figure 7.2b, where all DNL patterns of the measurements are shown. From Monte-Carlo simulations, the maximum expected DNL from statistical mismatch is 3 LSB at a 3σ confidence level. The DNL patterns show a larger mismatch with a systematic behaviour most pronounced at the transitions where the 5^{th} and 6^{th} bits are flipping. Since the transistors of the low-power DAC are very sensitive to doping gradients or voltage drops on the gate potential due to the subthreshold operation, they are strongly affected by the mismatch. In order to minimize the random mismatch between the transistors, the unit transistor size is chosen large, and a transistor width to length ratio of $1\mu\text{m}/8\mu\text{m}$ is used. From the analysis of the DNL pattern in correlation with the layout, the mismatch is expected to come mainly from the transistors at the border of the DAC array, because dummy transistors can not be placed due to space limitations.

7.2.2 Dynamic range & charge noise

The linearity and noise performance of the charge measurement branches was evaluated by charge injection measurements. The output of a function generator generating a voltage step function is AC-coupled to the channel input terminal with a capacitor value similar to

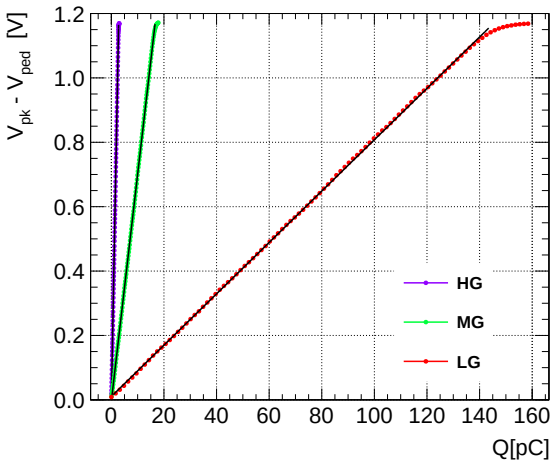


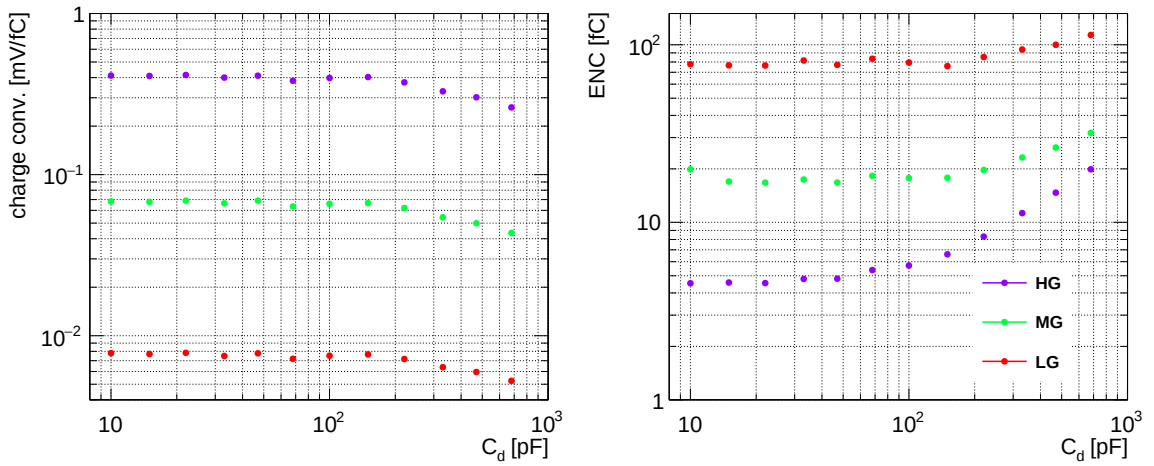
Figure 7.3: Analog output voltage as a function of the injected charge for all charge ranges of the KLauS4 front-end. The pedestal voltage of 0.6V is substracted. The black lines show the result of the linear fit used to obtain the linear range and charge conversion factors.

the detector capacitance to emulate the current pulses by a detector. If not stated otherwise, the capacitor values chosen are $C_d = 33\text{pF}$. For a voltage step with an amplitude ΔV and fast rise time, the charge injected into the channel is given by

$$Q_i = C_d \cdot \Delta V \tag{7.1}$$

The peak voltage generated by the analog front-end is buffered by a monitoring amplifier on the chip and measured by a digital sampling scope to obtain the response function of the different charge measurement ranges available in the chip. Figure 7.3 shows the measured peak voltages as a function of the injected charge for the high gain branch (HG), the high gain branch with scaled current (MG) and the low gain branch (LG). From a linear fit, the charge conversion factor ${}^{\mathrm{d}V_{pk}}/_{\mathrm{d}Q}$ and the linear range is extracted for each of the curves. To obtain comparable results, the considered charge range is adjusted in the fitting algorithm such that the maximal integrated nonlinearity is better than 1% of the fitted voltage range. The individual measurements including the nonlinearities are shown in Appendix B. The resulting charge conversion factors and linear ranges obtained from the fits are summarized in Table 7.1. As discussed in Chapter 6.3, the effect of the SiPM bias voltage tuning DAC on the input impedance, also affecting the charge conversion factor due to the ballistic deficit, is compensated by adding the triode transistor in the KLauS input stage. From a measurement of the charge conversion factor as a function of the configuration setting of this DAC, a distortion of less than ±0.2% is seen over the linear range of the DAC (see Figure B.2 in Appendix B). Also the linear range is not affected by the setting. Both are significant improvements[96] to the previous KLauS2 input stage topology[5].

The electronic noise generated by the KLauS front-end was measured as a function of the detector capacitance in order allow a study of the different contributing terms. Also the charge conversion factor is dependent on the detector capacitance as a consequence of the ballistic deficit.



(a) **(b)**

Figure 7.4: Charge measurement response as function of the detector capacitance. a) Charge response. b) Equivalent noise charge

From the noise divided by the charge conversion factor, the equivalent noise charge (5.3) is obtained. Figure 7.4 shows the measured charge conversion factor and the equivalent noise charge. With increasing detector capacitance, the charge conversion factor drops as the time constant given by the product of the input impedance and the detector capacitance rises, affecting the peak voltage as the ballistic deficit is increased. Also the noise voltage generated by the front-end rises due to the C_d dependence of the series noise term. For detector capacitances below 100 pF, the parallel and internal noise terms are approximately constant and dominating the total noise. This is the case for the $\approx 1\,\mathrm{mm}^2$ sensors used in the CALICE prototypes. Also the parasitics from longer routing tracks on the HBU PCB should not affect the total noise significantly.

Range	Charge conversion factor	Linear range (1% FSR)	ENC @ 50pF
HG	$424\,^{\mathrm{mV}}/_{\mathrm{pC}}$	2.77pC	4.55 fC
MG	$70.5\,^{\mathrm{mV}}/_{\mathrm{pC}}$	16.7pC	16.9 fC
LG	$7.98\,^{\mathrm{mV}}/_{\mathrm{pC}}$	144pC	77.8 fC

Table 7.1: Charge conversion factors, linear range and equivalent noise charge at small detector capacitances for all charge ranges of the front-end. All measurement results are dominated by the systematic capacitor size uncertainties of 1%.

7.2.3 Trigger comparator

The outputs of the comparator used for time-stamping and hold signal generation can be connected to common digital output pad which gives an or-combination of all comparator signals. In order to extract the comparator threshold, jitter and timewalk effect, charge injection measurements are conducted, measuring the trigger efficiencies as a function of the injected charge.

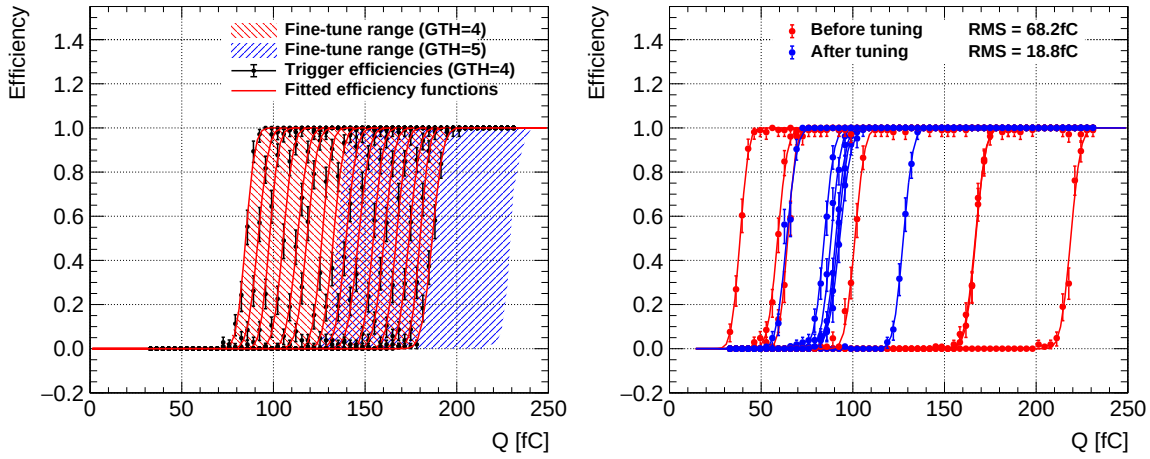


Figure 7.5: Threshold adjustment of the time trigger circuit a) Efficiency curves for different global and fine-tuning DAC values. The ranges tunable on a per-channel bases using the fine DAC are indicated by the filled areas. b) Mismatch of the thresholds for 7 channels within one ASIC in the case of untuned (equal) fine threshold DAC settings (red data points and fitted functions), and after adjusting the fine tuning DACs (blue data points)

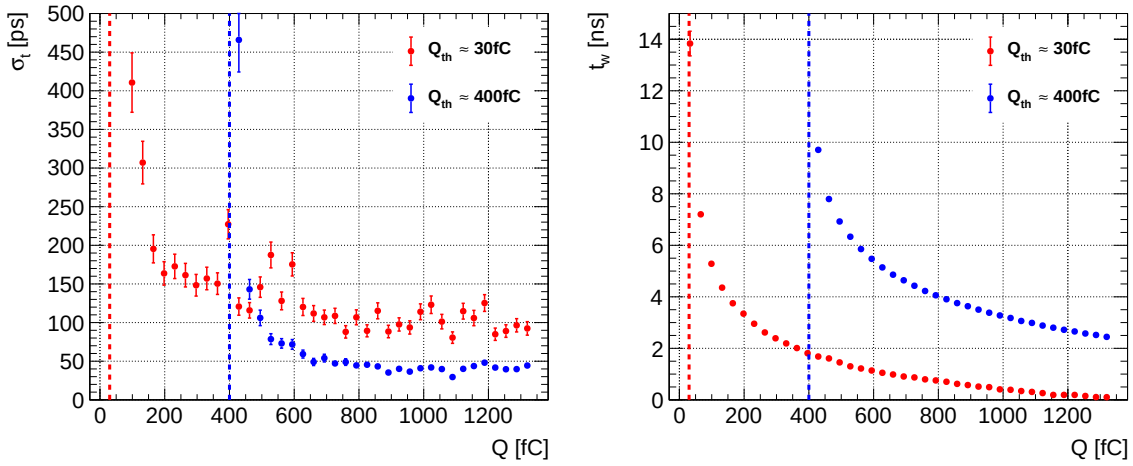
Threshold setting

The comparator efficiency curves were measured for the global 6-bit and the cannelwise 4-bit DAC settings to extract the comparator thresholds and their adjustment range and resolution. Figure 7.5a shows the trigger efficiency as a function of the injected charge for two global DAC and all 15 fine tune DAC values. The thresholds are obtained by fitting a Gauss error function to the measured data. Using the 4-bit DAC included in the channels, the threshold can be adjusted with a resolution of 6.8fC. The global threshold DAC allows to set the threshold in steps of 50fC.

The main purpose of the 4-bit DAC in each channel is to mitigate threshold differences between the channels. These differences mainly come from mismatch in the threshold generation current and the mismatch of the signal current copied by the input stage mirror. Figure 7.5b shows the trigger efficiency curves for all channels in one ASIC. Before tuning the threshold (i.e. the 4-bit DAC configurations are the same), the RMS dispersion between the channels is 63fC. After adjusting each of the 4-bit configuration values, this dispersion is reduced to about 18fC. The remaining differences given by outliers which can not be adjusted further due to the limited range of the 4-bit DAC. As a consequence, the range of the channel-wise DAC has been extended in the 36 channel version of the chip.

Comparator jitter and time-walk

The electronic time jitter and time-walk introduced by the timing comparator was studied by measuring the variation of the trigger time with respect to the synchronization output of the pulse generator. Figure 7.6a shows the jitter as a function of the injected charge for different threshold settings. The jitter appoaches a minimum value for larger charges which



(a) (b)

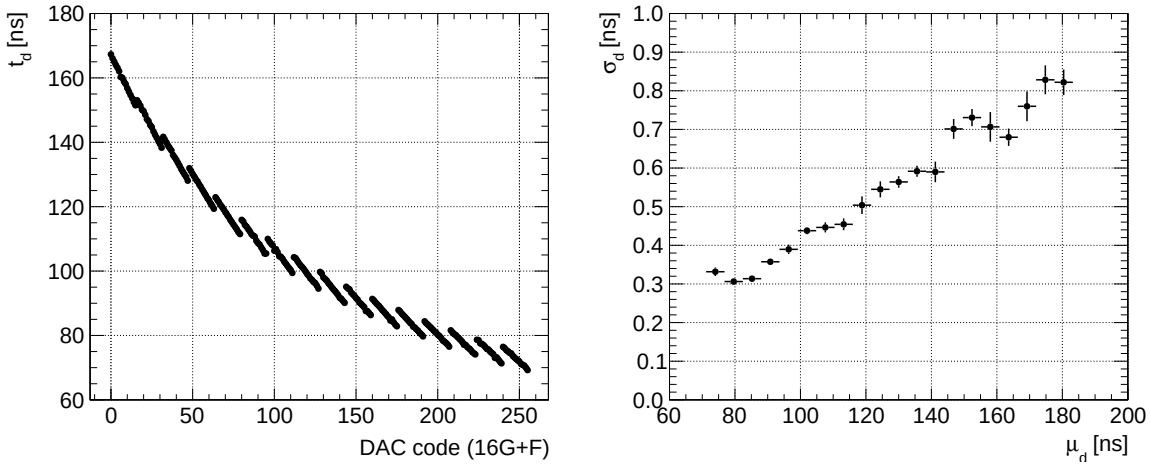
Figure 7.6: a) Jitter as a function of the injected charge for two comparator threshold settings. The threshold positions are indicated by the dashed lines. b) Timewalk effect for the same thresholds.

is dependent on the threshold value. For small gain SiPMs coupled to scintillators with small light yield, a threshold configuration close to the noise level is preferred to obtain a good trigger efficieny for single particles passing the scintillator. In this case, the jitter approaches values of about 100ps at signals above the threshold position. If larger thresholds are acceptable, the threshold can be put higher, resulting in a minimum jitter in the order of 60ps. Under all conditions, the requirements on the timing resolution of 1ns are met, also ensuring that the peak sampling is not deteriorated.

The measured average time of the comparison allows to study the time-walk as a function of the charge. As discussed in Chapter 5.4.2, this will have an impact on the linearity of the sampled signal. As shown in Figure 7.6b, the time-walk is always smaller than 10ns, which should reduce the sampled signal at the threshold position by about 2%. Given that the dynamic range is much larger than the comparator threshold this nonlinearity will have a small effect on the overall integrated nonlinearity. The time walk effect also motivates to set the threshold low to minimize the impact of the time walk on the charge measurement to a small charge range close to the threshold. A measurement of the introduced nonlinearities will be presented in Chapter 7.4.

Hit logic delay box

In order to set the hold time accurately to the peak position of the analog output pulse, a delay box is implemented which allows to set the delay in coarse time steps for all channels in common, and fine steps for each of the channels individually in the same way as the threshold setting of the comparator. This allows to precisely adjust the delay and to mitigate channel to channel differences in the delay box and peaking time of the analog signal. Figure 7.7a shows a measurement of the sampling window width generated by the delay cell. The x-axis gives the configuration value combining the global and fine-tuning settings. The delay



(a) **(b)**

Figure 7.7: Characterization measurements of the delay box used to adjust the hold time a) Tuning range. b) Introduced jitter as a function of the average delay.

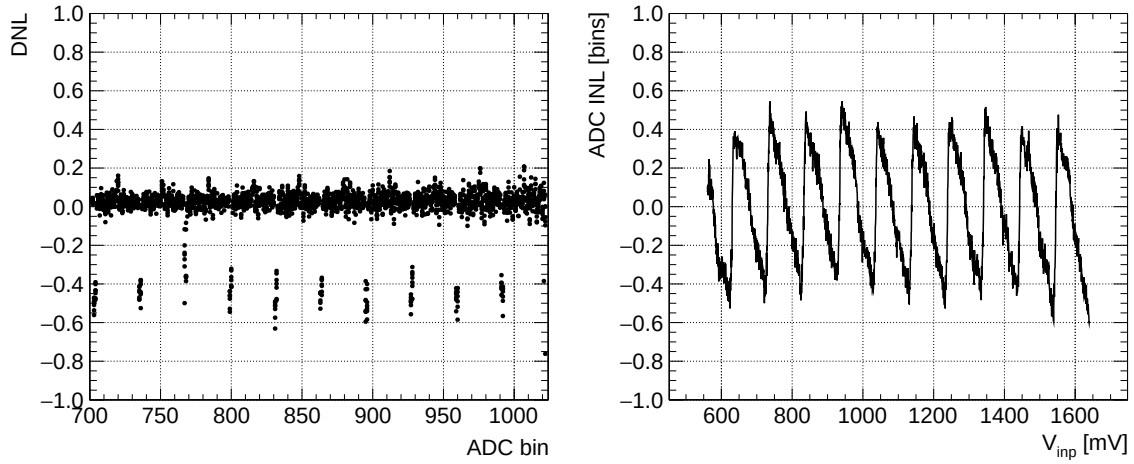
can be adjusted in a range of 100 ns around the expected peaking time of $\approx$ 93 ns with a resolution of 770 ps. This allows to keep the sampling voltage noise introduced by the imperfect sampling time configuration negligible. Due to the relatively long delay required to sample the peak voltage, the delay cell also introduces a jitter which is in fact dominating the time uncertainties of the hold time. Figure 7.7b shows the jitter generated by the delay circuit as a function of the average delay. The time delay is set by the crossing time of a voltage stored on a capacitor, discharged by a constant current. Thus, longer delays introduce higher jitter because the smaller derivative (Compare to Equation 5.16). In all cases, the total uncertainty of the hold delay is still generating a negligible noise contribution to the voltage sampled by the ADC.

7.3 ADC performance

The performance of the integrated analog to digital converter was characterized by connecting an external voltage to the inputs of the ADC channels instead of the analog front-end output signals. The differential nonlinearity of the ADC was measured by a code density test. Here, a fixed number of ADC conversions is performed for a constant voltage connected to the chip. The input voltage is swept over the dynamic range covered by the analog front-end. With a step size of 200 μV, i.e. much smaller than the intrinsic bin size, and the voltage noise generated by the external source, the converted input voltage has a uniform spectrum, and the number of conversions falling in to a specific bin is proportional to the bin size. The differential nonlinearity of bin i is then given by

$$\mathrm{DNL}_i = \frac{N_i}{< N >} - 1 \tag{7.2}$$

where N_i are the number of entries in bin i in the code density histogram and $< N >$ is the average number of entries in each bin covered by the voltage scan.



(a) **(b)**

Figure 7.8: ADC measurement nonlinearities
a) DNL as a function of the ADC bin (Seven channels in the same ASIC).
b) INL as a function of the voltage at the positive ADC input terminal (One channel).

The highest bin (1024) is excluded from the plot, since all input voltages above the ADC dynamic range are mapped to this code. The integrated nonlinearity is calculated from the integral of the differential nonlinearities (5.30). Figure 7.8 shows the measured differential and integral nonlinearity obtained for the ADC in 10-bit operation mode in the voltage range accessible by the front-end output[2].

A clear pattern of a pair of neighboring bins smaller by about 50% is observed repeating every 32 bins and very comparable between the different channels in one ASIC. The integrated nonlinearity obtained in this range is less than ±0.6 bins, dominated by the DNL nonlinearity pattern. This comes from the large parasitics to to ground on the MSB and LSB arrays, not precisely matched to the bridge capacitor. Due to some of the layout choices made (i.e. large parasitics to ground), the linearity is very sensitive to small variations of the parasitics and not reproducible by the layout parasitics extraction software. These parasitics also affect the average bin size of the ADC, and the dynamic range of the ADC is reduced from an envisioned range close to ±1.8 V, to about ±1.65 V. As an example on the high sensitivity to the parasitics, the linearity of the ADC is changed when applying a protective epoxy resin (glog-top) on the ASIC. The idividual parasitics (C_{pM}, C_{pB}, C_{pL}) are affected differently by the material surrounding the chip, increasing the nonlinearities as shown in Figure B.3 in Appendix B. To find the optimum layout strategy bejond the capabilities of the layout extraction tools, several versions of the ADC with differerent layout modifications were implemented in the 36 channel version of the ASIC.

Using the ADC in 12-bit operation mode the maximum and minimum differential nonlinearities are -0.44 and +0.36 LSB, respectively. No repetetive DNL pattern is visible in this case because the the bridge capacitor of the 10-bit SAR stage is not being used actively. The 8-bit pipelined stage also not introducing a systematic pattern, since the capacitor array

[2]The code transition 511 → 512 corresponds to a input voltage difference of 0V. Thus the pedestal voltage of the front-end at 600mV is converted to a code of ≈ 700.

is less affected by the matching requirements on the bridge capacitor due to it's different capacitor splitting structure. Here, the integrated nonlinearity is mainly affected by the amplifier and has a peak-to-peak magnitude of 2.83 bins. Both DNL and INL are shown in Figure B.4, Appendix B. The integrated nonlinearity is larger than the value expected from simulation, which is due to a mistake in the digital control circuit of the ADC. The amplifier is enabled just before the conversion in the pipeline stage starts, leaving too little time for the amplifier's common-feedback circuitry to settle. This issue was corrected in the 36 channel version of the chip.

ADC nonlinearity correction

In order to correct for the nonlinearities caused by the ADC's varying bin sizes, an algorithm remapping the digital informations to uniform bins was used. It is using the DNL informations obtained from code density tests to convert the measured spectra with non-uniform bin sizes to spectra with the DNL errors corrected. Figure 7.9 shows the procedure of the calibration in the case of a uniformly distributed input spectrum. Due to the differing bin sizes of the ADC as indicated in the left figure, the number of converted entries in a bin differs from the average value and is proportional to the input density integrated over the voltage range of this bin (in this specific case equal to the width of the bin). When the bin widths are known from characterization measurements, the number of entries in each bin can be remapped to a yield a uniform distribution as sketched in the right figure. In the case of the measurements presented here, the correction is performed on histograms and the original number of entries is distributed to the resulting histograms based on the overlap region of the bins in the two histograms. The same method can also be applied on an event-to-event bases using a weighted random distribution to the final histogram. Both methods give equivalent results for high statistics.

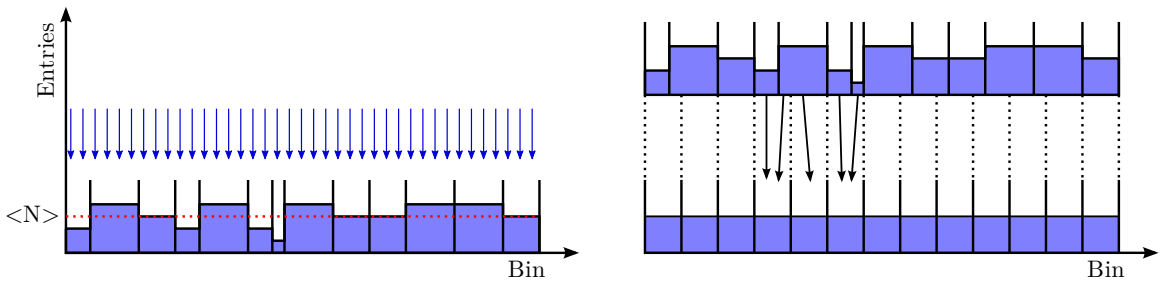


Figure 7.9: ADC nonlinearity correction using a look-up table. a) Nonlinearities generated by differing bin sizes from a uniform input spectrum. b) Redistribution of the entries to obtain the real spectrum, using the bin-size informations. Adopted from [32].

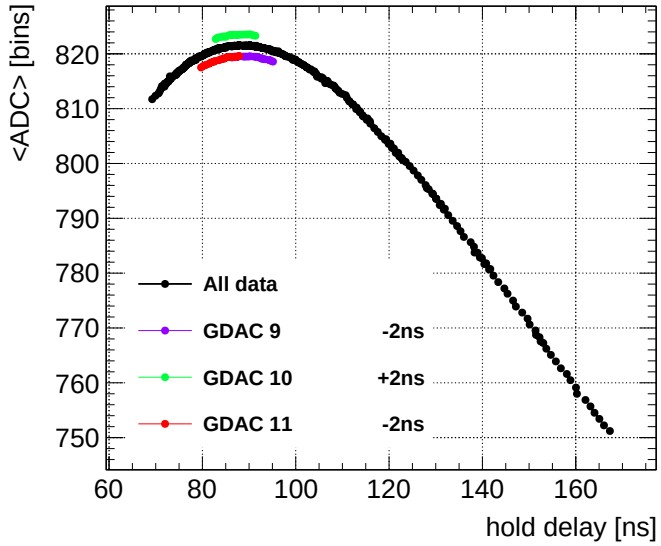


Figure 7.10: Pulse sampled by the internal ADC using the delay box to shift the the hold-time. The subranges tuned by the channel-wise DAC are shown for different global DAC settings around the peak and offset by ±2 ns.

7.4 Full-chain measurements

The performance of the full readout chain combining the front-end and the analog to digital converter in the KLauS channel was studied first by charge injection measurements to extract the full linearity including all effects of the individual building blocks in the processing chain. Tests with single silicon photomultipliers with different pixel sizes are conduced to ensure an adequate signal to noise ratio such that the single photon spectra can be obtained. These measurements will be presented in the following.

To ensure the signal is sampled and held correctly at the peak maximum, the hold delay needs to be adjusted by maximizing the ADC output response. In order to perform the maximization, the signal is kept constant for charge injection measurements. The maximization can also be performed using a SiPM signal, provided that the light input and SiPM gain does not change during the measurement. Figure 7.10 shows the average signal amplitude for a charge injection measurement with fixed amplitude. The x-axis shows the configured hold time set in the delay cell with an arbitrary offset. Thus, the plot represents the reconstructed signal shape seen by the ADC. By setting the hold time to the peak maximum, the obtimum signal and noise for the converted signal is obtained. The optimum hold time setting changes with the input signal,for example light pulse width from the LED or scintillator or other configuration settings influencing the peaking time. Also the comparator threshold setting has an influence on the optimum configuration due to the timewalk effect. In practice however, no significant changes of the peaking time with respect to the comparator decison time were observed which would motivate the requirement of frequent recalibrations of this parameter.

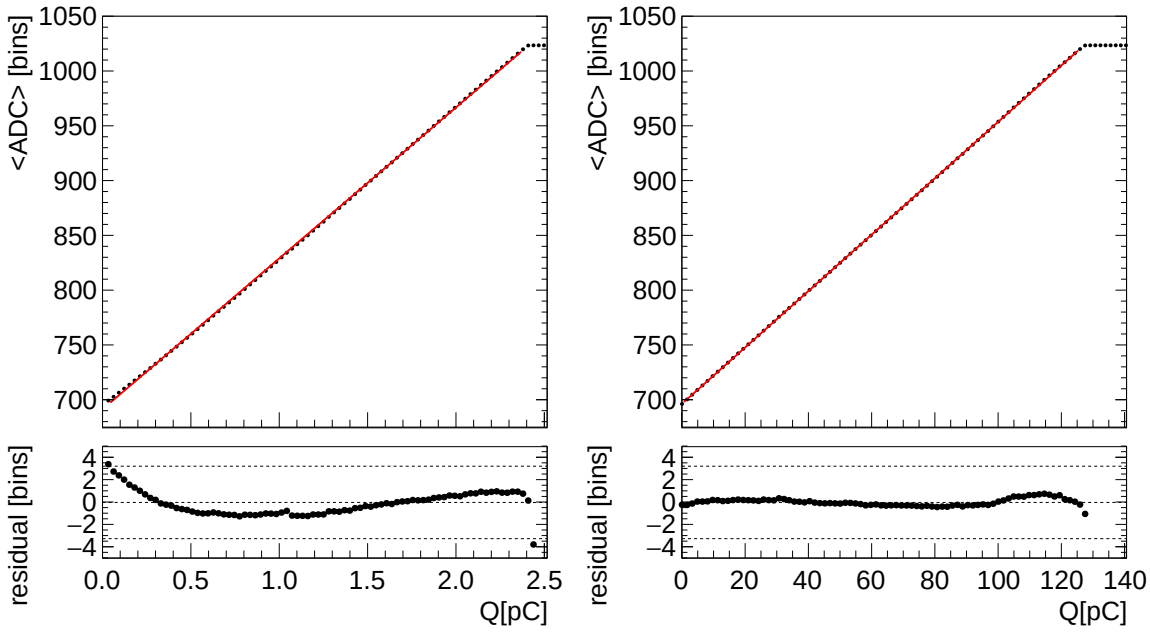


Figure 7.11: Combined linearity of the analog front-end and ADC in 10-bit mode. a) High gain branch. b) Low gain branch.

Charge measurement linearity

The charge measurement capabilities of the full KLauS readout chain was investigated by measuring the average ADC response response as a function of the injected charge for all charge scale settings. Figure 7.11 shows the ADC response as a function of the charge for the high gain (1:1 scaling) and low gain branch using an external signal for the hold time generation. The lower plots show the residual differences to a linear fit, limited to the range of 1% integral nonlinearity. This 1% limit on the nonlinearity is indicated by the dashed lines in the residual plots. Due to the reduced dynamic range of the ADC, also the charge measurement range is reduced accordingly. The nonlinearities introduced by the front-end are the same as for the measurements of the front-end alone, particularly visible for the high gain branch at small charges. The increase in the response visible here comes from a distinct feature of the input stage in combination with the current mirrors used in the front-end. To obtain an optimum dynamic range, the pmos-mirror copying the signal current to the integration RC-circuit is designed to withstand the large signal currents from the SiPM without introducing nonlinearities. At small detector capacitances, the input stage of the KLauS Front-end has a frequency response dominated by a pair of complex poles, and the output current copied from the input stage is ringing. Due to this ringing, the pmos mirror sees a "negative" current compared to the DC state, causing the cascodes to enter the triode regions. This introduces the nonlinearities seen for the high gain stage results.

To investigate the impact of the time walk from the internal comparator on the voltage sampling, the same measurement is repeated using the internal comparator. The hold delay is tuned to be optimal at 2pC, i.e. close to the maximum dynamic range of the high gain

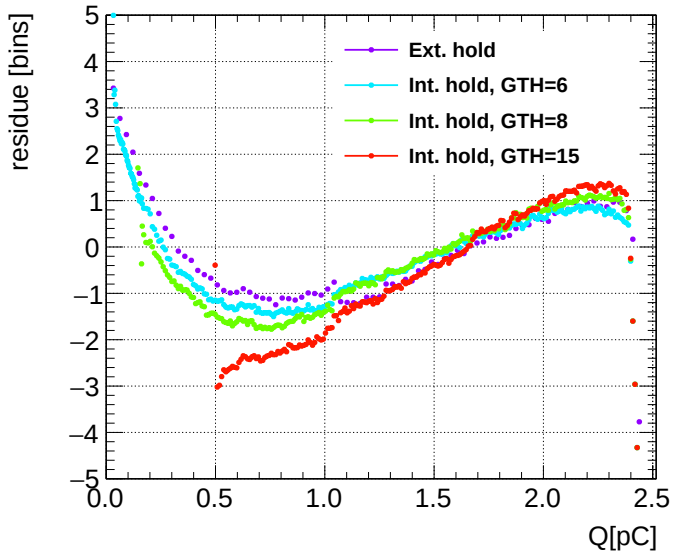


Figure 7.12: Residual errors for the high gain branch for different trigger configurations. The fit function has been fixed for all threshold configurations for comparability. The 1% nonlinearity limit corresponds to a residue of about 3 ADC bins.

branch which is used to study the linearity. At 2 pC, the time-walk is already small and not changing significantly towards higher charges, which would be covered by the low gain stage. Since the differences are small, only the residual graphs with respect to the fit result using the external trigger signal are shown in Figure 7.12. As expected, the response is reduced close to the trigger threshold and approaching the external trigger measurements with increasing signal amplitude. In fact, the time-walk effects partially compensates the nonlinearities from the front-end. The effect is more pronounced for higher thresholds, yielding an almost perfectly linear response for large thresholds up to the dynamic range of the high gain branch. For all configurations, the nonlinearities close to the trigger threshold are lower than 1%.

Automatic gain selection

Both of the two charge measurement branches devoted to the different signal ranges are active at the same time. The gain selection comparator is used to select the charge information to be digitized by the ADC. This decision bit is also added to the digital event data, which allows to combine these two charge ranges into a common charge measurement with high resolution for small signals and lower quantization resolution for the full dynamic range towards high charges. The automatic gain selection was tested in charge injection measurements using the 10-bit ADC. Figure 7.13a shows the ADC and gain selection input data before merging the informations from the high gain and low gain branches. For small charges, only the high gain branch is used since the gain selection comparator is not triggered. The efficiency curve for the gain selection comparator is shown in the lower part of the graph. The threshold was configured to be smaller smaller than the dynamic range of the high gain branch. For charges above the gain selection threshold, the charge informa-

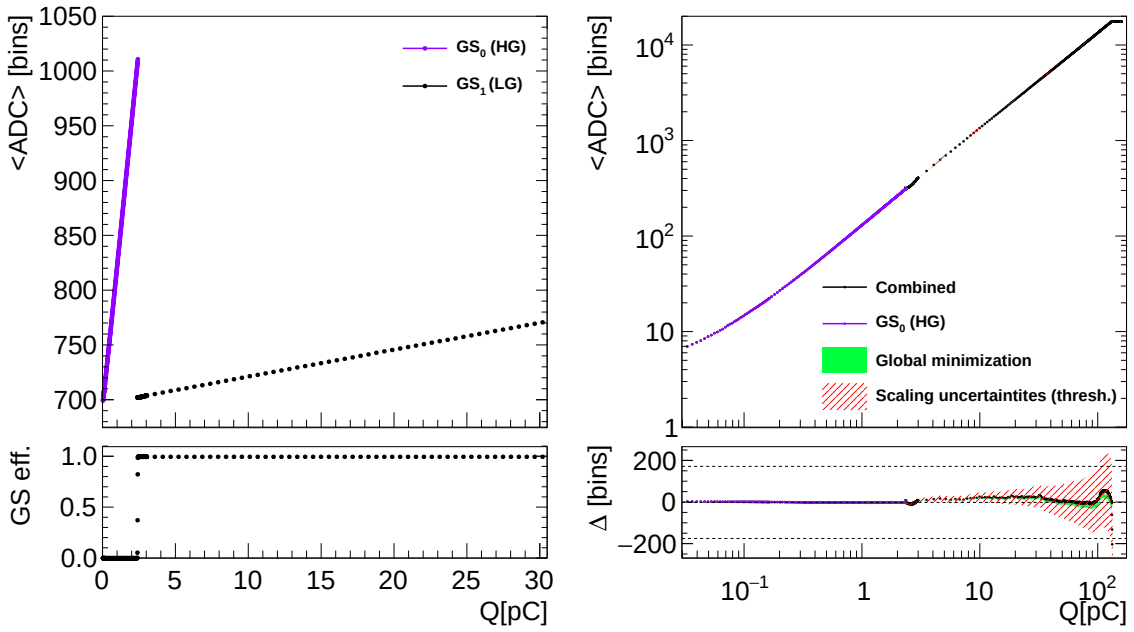
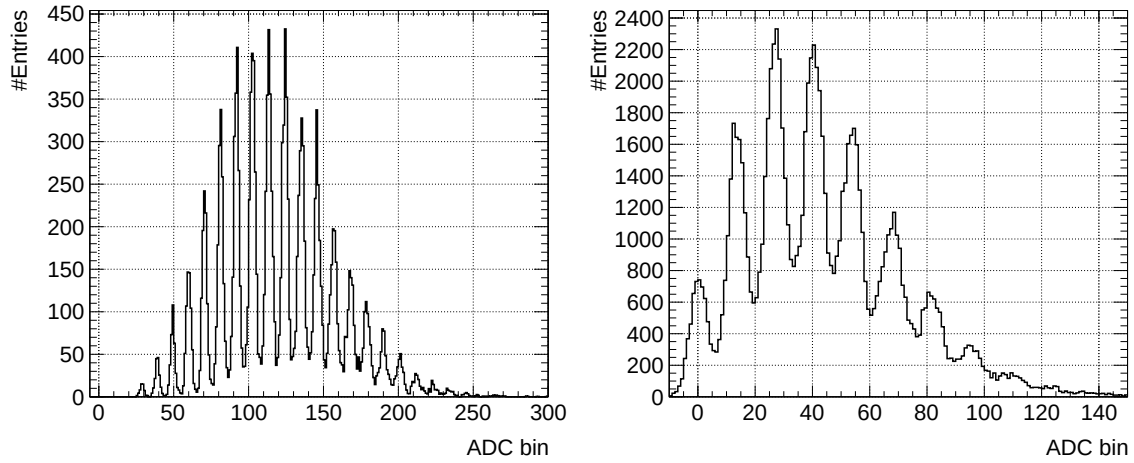


Figure 7.13: Automatic gain selection. a) Input data to be merged. b) Linearity of the merged data in double logarithmic axes.

tion is only coming from the low gain branch, with a smaller charge conversion factor (thus lower resolution) but high dynamic range.

In order to combine the two charge ranges, an inter-calibration factor needs to be determined and should given by the ratio of the charge conversion factors of the two branches. Since the use of these factors would infer linearity in the calibration, they can not be used to investigate the linearity of the merged charge information. Instead, the zero-suppressed response of the two branches at the gain selection threshold are used for calibration, i.e. the pedestal informations and the ADC response of the two branches at the gain selection threshold.

Figure 7.13b shows the obtained response after merging the informations of the two branches, with the residue to a linear fit on the lower plot. As indicated by the dashed lines, the 1% linearity is fulfilled over the full range. The 1% linearity limit corresponds to a residue of almost 200 bins due to the larger full scale range. Nonlinear behaviours are visible for small charges introduced by the high gain stage, and close to maximum dynamic range of the low gain stage limited by the ADC dynamic range. The switching process introduces a deviation of 0.07% close to the crossover point. The systematic uncertainty of the inter-calibration factor due to the limited information used yields an uncertainty of the response in the high charge region shown as the red shaded 1σ confidence limit area in the residual plot. A global χ^2 minimization of the residual errors gives almost the same result, but with much smaller uncertainties as shown by the green area in the residual plot.



(a) (b)

Figure 7.14: Sample single photon spectra obtained from different silicon photomultiplier models using a pulsed LED source. The DC pedestal voltage of about 0.6V is subtracted. a) $25\mu m$ pixel-pitch MPPC using the 10bit ADC and internal trigger. b) $10\mu m$ pixel-pitch MPPC using the 12bit ADC and external trigger.

Single pixel spectra

One of the design goals posed on the KLauS4 ASIC was to provide the possibility to measure single photon spectra using low gain Silicon Photomultipliers. Figure 7.14 shows the spectra for two types of sensors from Hamamatsu Photonics with different pixel sizes. To generate the light signal, pulsed LEDs were used. For the 25 μm SiPM spectra shown in the left figure, the gain is sufficient to use the 10-bit ADC, corrected for it's nonlinearities using the method described in 7.3. The internal comparator is triggering the ADC conversion, with a threshold set to about 3 fired pixels to minimize the triggers on dark count pulses. The right figure shows the obtained spectrum for a device with a pixel size of 10 μm. The sensor was biased at the nominal operating voltage given by the producer corresponding to a single pixel signal of $1.35 \cdot 10^5$. In this case, the low gain requires the 12-bit ADC, which can only work using an external trigger signal due to the bug in the digital part described in Section 7.3. For both spectra a clear separation of the single pixel signals is visible, proving the excellent noise performance of the full readout chain of KLauS.

7.5 Power pulsing

The power pulsing functionality of the chip is controlled by a single digital pad, which generates all necessary control signals to the front-end and the channel control logic with configurable delay.

As described in Chapter 6.7.1, the bias current of the input stage is reduced and the feedback branch is disabled during the low-power mode. In order to obtain a small offset voltage at the input terminal with respect to the enabled state, a subthreshold compensation current is added to the feedback branch. The subthreshold bias current of the input stage

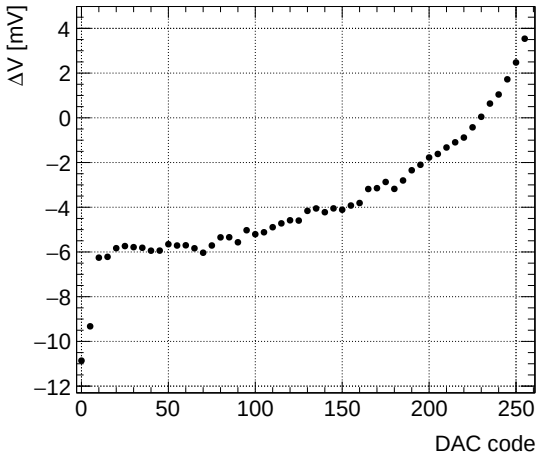
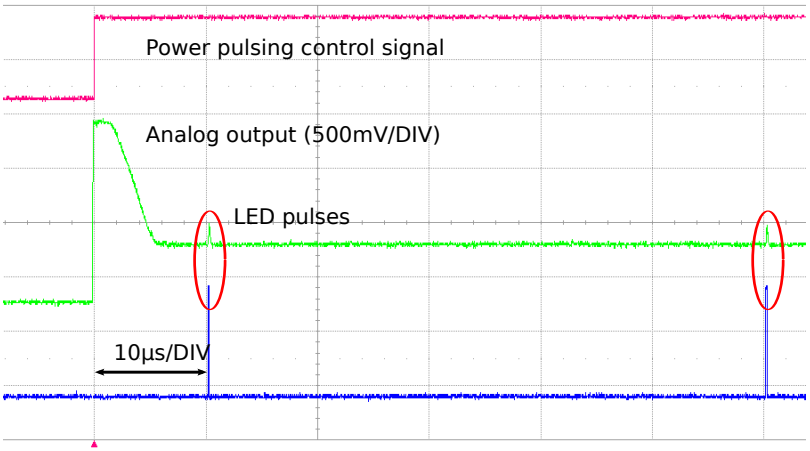


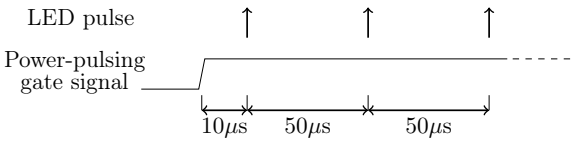
Figure 7.15: Voltage difference of the SiPM input terminal as function of the DAC code, comparing low power and active mode of the input stage.

as well as the compensation current can be adjusted using two subthreshold DACs with 3bit resolution. Figure 7.15 shows the difference of the input terminal voltage between the normal and low power mode of the input stage as a function of the configuration value of the voltage DAC controlling the DC voltage. With a maximum difference of about 6mV after optimizing the compensation current, the introduced offset is negligible compared to the SiPM overvoltage of a few volts, and the Silicon Photomultipliers are always held in the same operating conditions as required.

Figure 7.16a shows the settling behaviour of the analog front-end after correct configuration of the relative delays between the power pulsing control signals. The voltage at the output of the analog part is shown as a green line. $6\,\mu$s after enabling the front-end, the pedestal voltage reaches the foreseen voltage level and the additional current to the pedestal feedback circuits is disabled. After an additional time of $10\,\mu$s, the pedestal voltage is fully settled to the final value within an accuracy of 1 mV. This is not visible in this measurement due to the scale.
The response to charge signals using the powergating functionality was tested using a Silicon Photomultiplier with $25\,\mu$m pixel size. As sketched Figure 7.16b, the SiPM is lit by light pulses from an LED at fixed delays after taking the chip out of the idle state. Figure 7.17 shows the obtained single pixel spectra for the power-pulsing disabled (black) and the spectra obtained with power pulsing $10\,\mu$s and $60\,\mu$s after enabling the front-end. A clear single pixel separation is found for all spectra. From the analysis of the single pixel spectra, the SiPM gain and pedestal position is obtained. Within the 0.1% accuracy of the measurement, the gain is constant for all spectra. In the spectrum at $\Delta t = 10\,\mu$s, a systematic shift of the pedestal of about 0.7bins ($\approx 2.25\,$mV) is observed because the front-end is not fully settled. In all following spectra, the pedestal position is not changed within the 100μV accuracy of the measurement.



(a)



(b)

Figure 7.16: Settling of the front-end in power pulsed operation. a) Scope measurement of the analog output signal after enabling the front-end. b) Sketch of the sipm-response measurement.

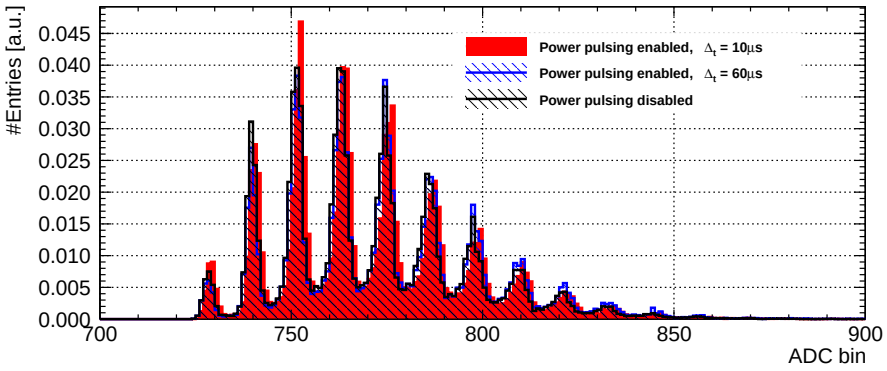


Figure 7.17: Single photon spectra in power-pulsed operation: 10 μs after enabling the front-end (red), 60 μs after, and with the power pulsing disabled (black).

7.6 Power consumption

For the analog front-end, the power consumption in the enabled state was measured to be about 3 mW per channel excluding the common bias block. This result is in line with the circuit simulations obtained for the seven channel prototype, and allows to extrapolate to the final 36 channel version of the ASIC where the common bias block is shared between all channels and therefore does not contribute significantly to the total power consumed. The power consumption of the analog blocks in the low power state could not be measured with the test board, since the power consumed by external circuits required for the characterization measurements is much larger than power consumed by the ASIC in this configuration. The subthreshold bias current used by the different blocks active during the low power state is generated by a single cell. This allows to measure the bias current in one point and extrapolate to the overall power consumption. The bias current of the input stage in the low power mode was extracted from a measurement of the transient voltage of the input terminal under resistive and capacitive load during power pulsing. The results are comparable to the expectations from the circuit simulation, allowing the extrapolate to a real power-pulsed operation of the ASIC. For the 36 channel version, simulations suggests a power consumption of 26μW per channel for a power pulsing duty cycle of 0.5%.

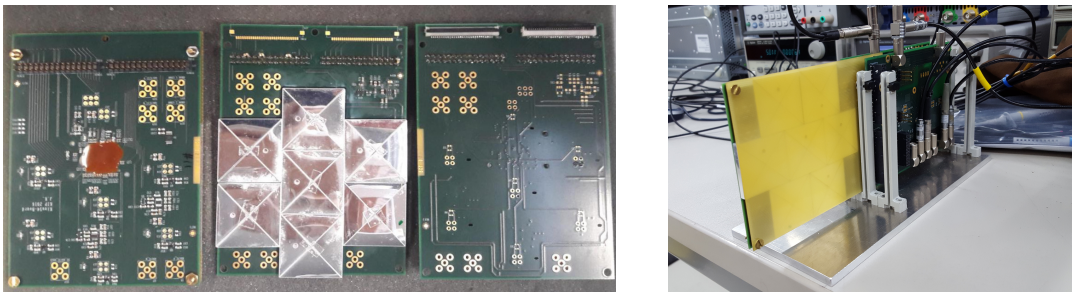
7.7 Test-beam

In order to test the chip with multiple channels active at the same time in a realistic environment, measurements at the DESY electron test beam facility were conducted. The system is shown in Figure 7.18 and is comprised of 3 ASICs with 15 equipped channels read out by a common data aquisition. Surface-mountable SiPMs with an active area of $1.3 \times 1.3\,\mathrm{mm}^2$ and a pixel size of $25\,\mu$m were soldered on the test-boards. Scintillating tiles produced at Hamburg University with a size of $30x30x3\,\mathrm{mm}^3$ are air coupled to the sensors, using a spherical dimple to obtain a good uniformity and individually wrapped with reflecting foil to reduce crosstalk and acheive a high light yield [97]. For two of the ASICs, all seven channels were equipped with sensors and scintillating tiles. One ASIC was equipped with one channel to allow measurements of the trigger efficiency of electrons passing all three boards. For the SiPM gain calibration, the integrated LED system is used. During the week of data-taking, all key characteristics of the chip could be tested successfully, some results outlining the most important aspects of the KLauS ASIC will be presented in the following paragraphs.

System calibration

Before the beam was used, the system was calibrated to obtain a uniform behaviour of all channels. For most of the steps during calibration, the LED system was used. As a starting point, a configuration common for all channels was used, selected to give response to the light pulses from the LED using the internal trigger. To obtain a uniform response for all channels, the following steps were followed.

- The comparator thresholds were adjusted to yield a noise rate of approximately 1Hz for all channels.



(a) (b)

Figure 7.18: Setup used at the DESY test-beam facility.
a) Test board front side with assembled ASIC and LED system, back side of the board with assembled tiles and backside without tiles and SiPMs equipped.
b) Setup used in the test beam with one of the tree boards inserted in the mechanical holder.

- The hold delay was swept for all channels under incident LED light pulses. A global DAC settings was selected for each ASIC, and the fine-tuning setting was selected for each channel to maximize the response.
- Three sets of single photon spectra with different settings for the input terminal voltage DAC were recorded. Using the gain extracted from the spectra, a configuration setting was selected for every channel to obtain the same single pixel gain.

A single pixel spectrum obtained after the calibration procedure was already shown in Figure 7.14a. Qualitatively, no changes in the measured single photon spectra were observed in the test-beam setup compared to the single-channel measurements under laboratory conditions.

Automatic gain selection

The automatic gain selection was tested for the response to electrons passing the scintillator in order to investigate the possibility to find an inter-calibration factor without prior knowledge from characterization measurements. In contrast to the procedure finding an inter-calibration factor in Chapter 7.4, no direct information of the injected charge is available in this case. The medium gain configuration was used in order to have sufficient overlap with the low gain branch, required to evaluate the linearity after combining the two charge informations. When the automatic gain selection feature is disabled, the input of the ADC is fixed to one of the analog outputs of the charge measurement branches. However, the information of the gain selection comparison is also available. This allows to measure the response of both branches at the threshold of the gain selection comparator. Figure 7.19 shows the measured spectra for the two charge measurement branches, separating the all measured events from the part where the gain selection comparator is not triggered.

From the ratio of these spectra, the gain selection efficiency is extracted and shown below the spectra. By fitting a Gaussian error function to the efficiencies, the ADC response S_{MG} and S_{LG} at the threshold position is obtained directly. With the pedestal positions P_{MG} and P_{LG} obtained from a dedicated noise measurement run, the inter-calibration factor is

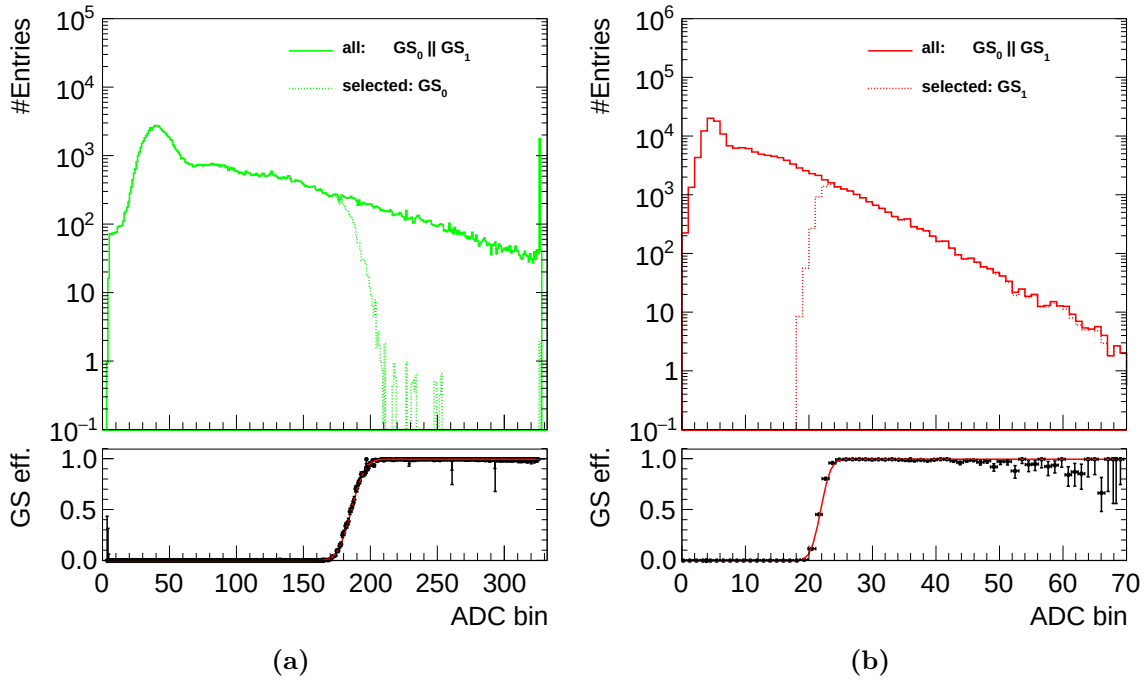


Figure 7.19: Spectra to estimate the gain intercalibration factor from testbeam data in fixed gain mode. a) MG spectrum (solid line) and part selected by the gain selection comparator (dashed line). b) LG spectrum and part selected by the gain selection comparator. The lower plots show the gain selection efficiency curve as a function of the converted ADC value, fitted by a Gaussian error function.

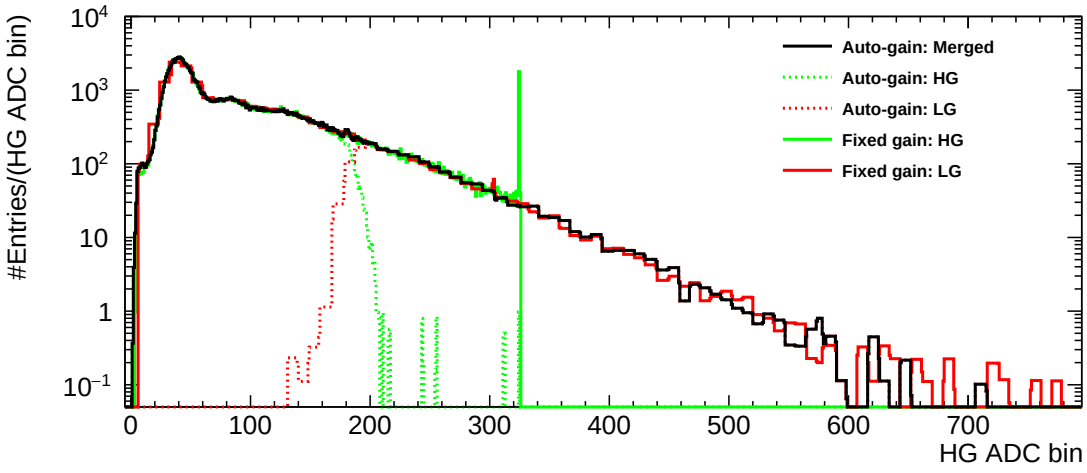


Figure 7.20: Merged spectrum (black line) in automatic gain selection mode after intercalibration. The input data from the two branches are plotted with dashed lines. For comparison, spectra from the two stages in fixed gain configuration are also shown with solid lines.

given by

$$G = \frac{S_{\mathrm{MG}} - P_{\mathrm{MG}}}{S_{\mathrm{LG}} - P_{\mathrm{LG}}} \tag{7.3}$$

and the merged and zero-suppressed output code of an event is given by

$$M = \overline{g} \cdot S + g \cdot G \cdot S - P_{\mathrm{MG}} \tag{7.4}$$

where g denotes the binary gain selection comparator flag and S is the ADC value of the event. The black line in Figure 7.20 shows the resulting spectrum with the automatic gain selection enabled, with dashed lines being the two parts of the spectrum. Using the intercalibration factor, the measured response of events from the low gain branch are scaled to match the charge conversion factor of the MG branch. Correspondingly, the gain-selection threshold in the low gain spectrum is scaled from ≈ 22 ADC bins to a response of 180 ADC bins after merging the spectra. Measurements obtained with the automatic gain selection function disabled are shown as solid lines. The limited dynamic range of the high gain branch is clearly visible at about 320 ADC bins, as well as the reduced resolution of the low gain information used above the gain selection threshold. As in the case of the measurements performed in the laboratory setup, the deviation at the takeover point is quite small, and the shape of the merged spectrum fits the fixed-gain counterparts well within their dynamic range.

Trigger efficiency to electrons

Because all of the chips in the system are running with a common clock signal, also the timestamps recorded for every event are synchronized. This allows to build clusters in time between the individual layers. By requiring a hit in the central tiles of the first and last ASIC, the trigger efficiency to passing electrons is determined from the rate a hit of the central tile of the middle ASIC is recorded. To minimize random coincidences, the triplet of events is only accepted if the time stamps are different by a maximum of 1 bin, resulting in a window of about 51 ns Figure 7.21 shows the obtained efficiency as a function of the selected electron energy. For the full measured range electron beam energies, the efficiency is found to be above 99.2%. This proves the stable operation of the readout chain also at high rates and low threshold settings, in a potentially noisy environment with multiple channels being active or triggering.

Power pulsing

The power pulsing functionalities have also been tested in the electron beam. In order to achieve a good data-taking efficiency, the gate signal was synchronized to the DESY repetition period of 80 ms. This ensures data is only taken while electrons with the right energy are generated. Figure 7.22 shows the response to electrons with the power pulsing function enabled (red) and disabled (black). The changed number of single, two and three electron events most likely comes from different conditions of the delivered electron beam, as the power pulsing gate is synchronized to the DESY extraction cycle in the case of power pulsing. The single electron response of the ASIC due to the power pulsing is not affected.

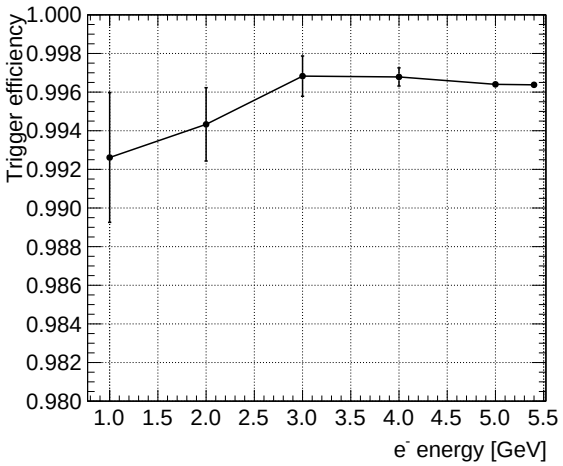


Figure 7.21: Trigger efficiency as a function of the selected electron energy.

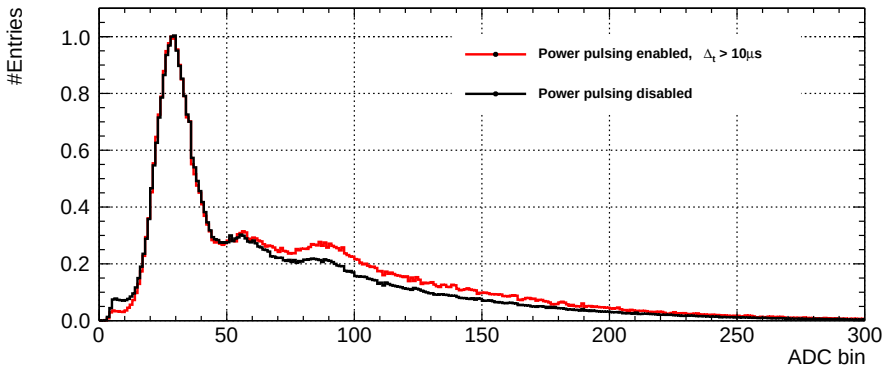


Figure 7.22: Electron spectrum obtained at the DESY test-beam during power pulsed operation.

8 Summary

The detectors being developed for future linear colliders are optimized for the use of particle flow algorithms in order to achieve a jet energy resolution of 3-4%, sufficient to distinguish the hadronic decay of W and Z bosons based on the jet energy measurements. This allows to use these final states also for precision measurement, significantly increasing the sensitivity in measurements of standard model processes and searches beyond the latter. The particle flow approach requires all particles in a jet to be measured separately in the tracking and calorimetry system. This allows to use the sub-detector providing the best energy resolution for the given particle. Apart from an excellent energy resolution, the calorimeters are required to have a high spacial resolution and compactness, needed to make full use of the particle flow idea which relies on the separation of individual showers from a jet. Different readout concepts for the electromagnetic and hadronic calorimeters are studied within the CALICE collaboration. For both scinillator-based calorimeters, the Analog Hadronic Calorimeter (AHCAL) and the Scintillator-strip ECAL concepts, sandwich calorimeters using segmented plastic scintillator cells read out by Silicon Photomultipliers will be used.

In this work, the development of integrated readout ASICs for the scintillator-based calorimetry systems at the ILC detectors is discussed. The front-end ASICs are required to deliver a self-triggered, fully integrated readout solution for charge measurements of Silicon Photomultiplier signals, spanning a dynamic range from single photons (or fired pixels) up to the fully dynamic range of the sensors consisting of up to $\gtrsim$ 10000 pixels. A key requirement for the ASICs to be developed for the calorimeters is the limited power consumption. The high channel density and requirements on the compact active layers requires the readout electronics placed inside the active layers. To be able to operate the calorimeters without an active cooling system for the readout ASICs, the power consumption is limited to an average power consumption of 25μW per channel, which is only accomplished by power pulsing the front-end ASICs with the bunch train scheme of the ILC.

In order to facilitate the calibration of a detector consisting of millions of readout channels, the gain calibration of the Silicon Photomultipliers using single photon spectra is required. Due to the large detector capacitance of the sensors, a small input impedance is required to achieve a good charge conversion efficiency. Different noise terms contribute to the smearing of the single pixel signals. In this thesis, dedicated measurements and simulations were presented and used to investigate the effect of dark count rate avalanches from the Silicon Photomultipliers. An analysis of the different noise terms contributing to the single pixel resolution shows that the effective noise from dark-count avalanches pile up and add significantly to the total noise smearing the single photo-electron peaks. This also affects the optimum shaping time constants used in the signal shaping circuits of the ASIC. The peak voltage sampling strategy used in the developed ASIC does not add significantly to the noise. Also the timewalk effect introduced by the leading edge discrimination does

not lead to significant nonlinearities in the charge measurement.

The KLauS ASIC is designed provide a readout solution for high dynamic range Silicon Photomultipliers which have a small single pixel gain due to the reduced pixel size. Charge injection measurements using the analog front-end of KLauS show a dynamic range of 140 pC, sufficient to cover the dynamic range of Silicon Photomultipliers with an active area of $1\,mm^2$. With an equivalent noise charge of 4.5 fC for small area sensors, the ASIC shows an excellent charge measurement capability in the view of the power budget, allowing to use high dynamic range SiPMs required in particular by the scintillator-based electromagnetic calorimeters at ILD. In order to mitigate differences in the sensor gain, the input stage of KLauS allows to adjust the SiPM bias voltage in a range of 1.8 V, sufficient in view of the device to device spread obtained for modern Silicon Photomultipliers.

The analog front-end has a power consumption of approximately 3 mW per channel, and implements power pulsing functionalities to reduce the power consumption to the requirements of the ILD detector. The SiPM bias voltage is stabilized in the power pulsed operation by the specific choice of the input stage topology and additional compensation currents. To minimize the required duty cycle of the power pulsing and thus the average power consumption, the settling time of the front-end is reduced to about $10\,\mu$s after an appropriate configuration of the power pulsing sequence. For the 36 channel version of the chip, a power consumption of $26\,\mu$W is expected, based on the characterization measurements of the 7 channel ASIC.

For reasons of power-efficiency as well as simple and stable calibration, the SAR topology was chosen for the integrated analog to digital converter which has a nominal quantization resolution of 10 bit. For the gain calibration of small gain sensors, an additional pipelined stage allows to increase the resolution to 12 bit. The digitization of the SiPM response to small light pulses using the integrated ADC show good single pixel separation for commonly used as well as small gain sensors.

To cope with the dynamic range and the requirements calibrating the sensor gain using single photon spectra, two charge measurement branches with different scaling factors are implemented in the chip. An automatic gain selection function is used to cover the full charge measurement range with the required resolution. The linearity is preserved also after merging the charge measurement data from the two branches in the automatic gain selection mode of the chip.

Measurements using a setup of multiple ASICs were carried out at the DESY test beam facility, validating the performance and stability of the KLauS ASIC also in a realistic environment.

Only minor changes were implemented in the 36 channel version of the ASIC. After characterizing this chip is expected to be available shortly after submitting this thesis, the integration of the ASIC into the AHCAL technological prototype is foreseen.

A Description of additional sub-circuits

Bootstrapped sampling switch

Figure A.1 shows the schematic diagram of the switch used for sampling the input voltage at the 10 bit ADC and the pipelined stage. The gate voltage of the switch transistor M6 is designed to follow the input voltage plus a constant offset in order to keep the V_{gs} difference at M6 and thus the impedance of the switch in conduction mode constant. In the open state, the "clock" signal is at a logic low state forcing the gate voltage of M6 to ground level. The capacitor included in the switch is charged to a constant voltage close to the 1.8V supply voltage. With the input signal going to a logic high level, both capacitor potentials are offset to the input terminal voltage, causing M6 to see the constant voltage difference of the capacitor plus the input terminal potential. The switch is based on the design by Abo and Gray, [79] and [80] In this original design, a charge pump was implemented to generate the higher voltage charging the switch. This solution is not applicable here since the sampling is supposed to start immediately after a trigger from the front end. A separate supply voltage of 3.3V is used instead.

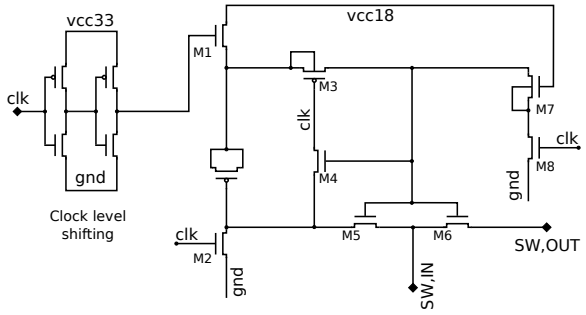


Figure A.1: Schematic diagram of the hold switch used in the KLauS ADC

Dynamic comparator

Figure A.2 shows a schematic diagram of the dynamic comparator used in the KLauS4 ADC. Consisting of a bias current generation circuit, a differential preamplifier, the comparator core and digital logic, it provides the decision signal for the SAR search algorithm. For the clock signal at logic zero, the comparator core is at a reset state the bias current to the comparator core is disabled. At the falling edge of the clock signal, the voltage difference is amplified in the positive feedback loop of the comparator core, yielding digital signals which are combined to a single digital output. The comparison procedure introduces a digital switching noise at the input of the comparator core. A single stage preamplifier is added to reduce the voltage change at the comparator input. The bias current is generated

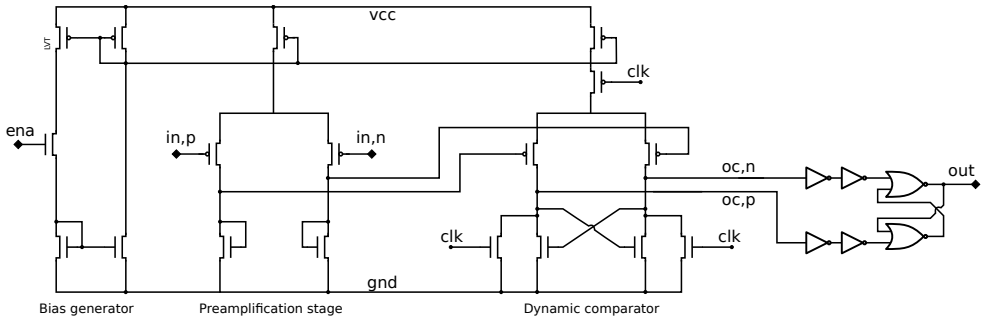


Figure A.2: Schematic diagram of the voltage comparator used in the KLauS ADC

in each of the comparator cells. It is used for the preamplifier and the comparator core, switched off while there is no conversion in progress.

Linearity analysis of a 4 bit SAR ADC

As discussed in chapter 6.4, the static linearity of the ADC is degraded by the nonlinearities of the DAC due to parasitics. The bin width of the ADC (6.26) of a specific code is given by the upper and lower boundaries of the bin. In Figures A.3 and A.4, the upper and lower boundaries are visualized as the voltage level where the relevant comparator decision marked with the red line passes the bin boundary. The upper boundary of code X is the same as the lower boundary of X+1, and the upper and lower boundaries of the highest and lowest code are not defined (and thus not shown). The boundaries can also be calculated from the given code using the common bits of the codes X,$X+1$.

In table A.1, the upper and lower bin boundaries and the bin width are listed for all codes of the 4 bit ADC. The mismatch of the capacitor pairs for the differential structure of the DAC are ignored here.

Delay cell

Figure A.5 shows a schematic diagram of the delay cell used to generate the sampling signal for the ADC. It is based on the discharge of a capacitance with a constant current. With the input signal D rising and $\overline{D}$ falling, the charge stored on the capacitance connected to the node $\overline{Q}$ is slowly discharged by the current source M_{16}. Once the voltage falls below the threshold voltage of transistor M_{12}, the Q node is charged quickly, starting a positive feedback loop which causes $\overline{Q}$ to quickly discharge to a logic zero state. To reduce the jitter on the generated delay due to noise on the power supply, the cell is designed differentially. The delay can be adjusted by changing the bias current discharging the capacitors. This is used for the global configuration of the delay in the KLauS ASIC. In addition, the capacitance value can be adjusted in each cell in order to fine tune the delay value.

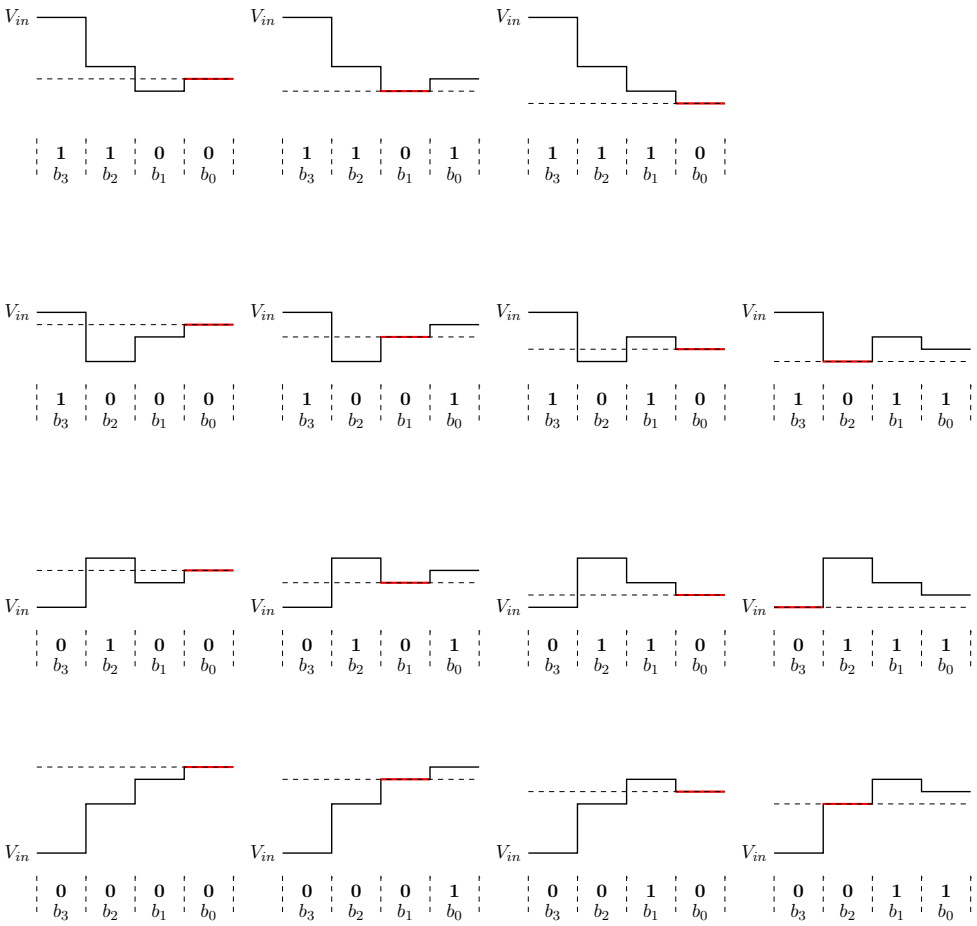


Figure A.3: 4-bit SAR ADC example: Upper bin boundaries

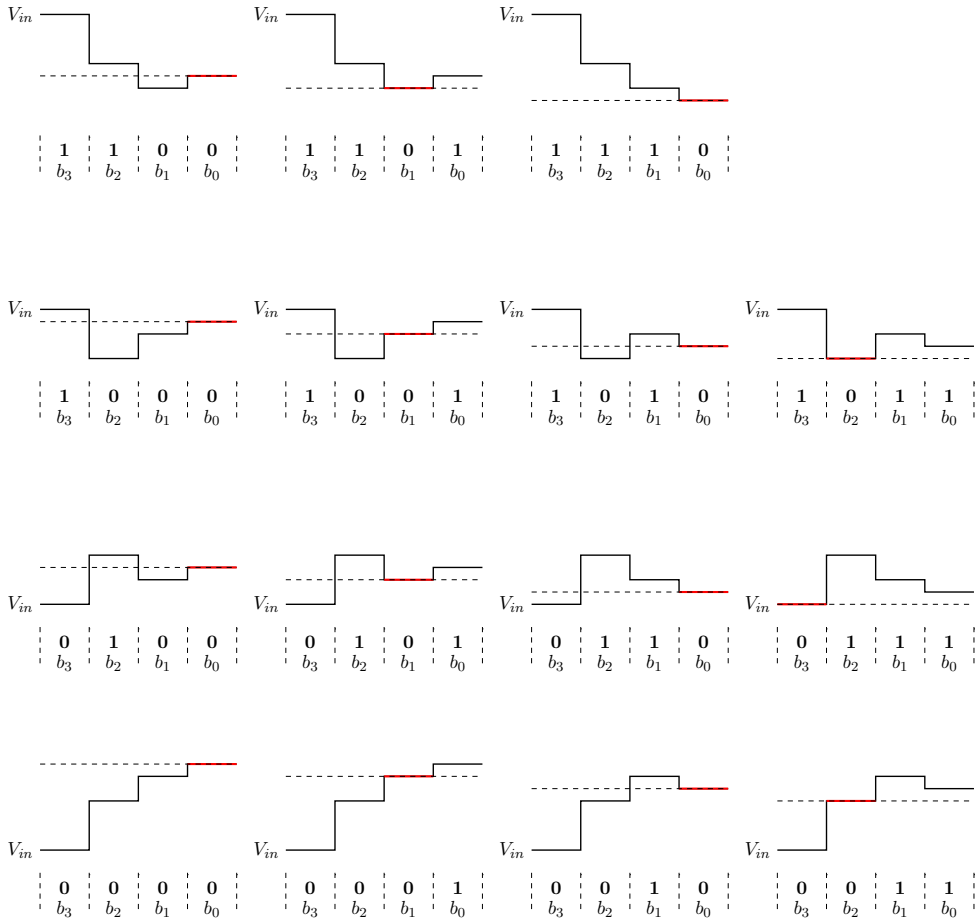


Figure A.4: 4-bit SAR ADC example: Lower bin boundaries

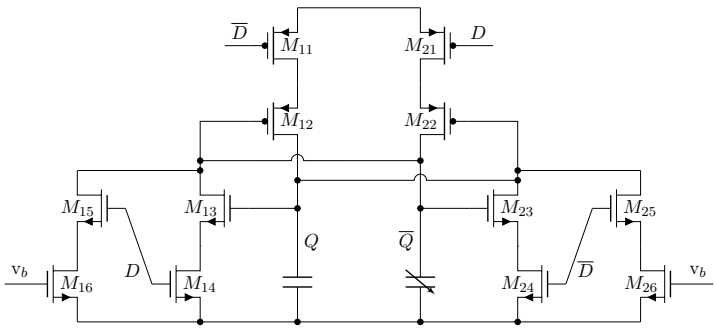


Figure A.5: Schematic sketch of the delay cell implemented in the hitlogic block to generate the sampling signal of the ADC.

Code	Lower boundary	Upper boundary	Bin width
0000	–	$0 = V_{in} + \Delta_3 + \Delta_2 + \Delta_1$	–
0001	$0 = V_{in} + \Delta_3 + \Delta_2 + \Delta_1$	$0 = V_{in} + \Delta_3 + \Delta_2$	Δ_1
0010	$0 = V_{in} + \Delta_3 + \Delta_2$	$0 = V_{in} + \Delta_3 + \Delta_2 - \Delta_1$	Δ_1
0011	$0 = V_{in} + \Delta_3 + \Delta_2 - \Delta_1$	$0 = V_{in} + \Delta_3$	$\Delta_2 - \Delta_1$
0100	$0 = V_{in} + \Delta_3$	$0 = V_{in} + \Delta_3 - \Delta_2 + \Delta_1$	$\Delta_2 - \Delta_1$
0101	$0 = V_{in} + \Delta_3 - \Delta_2 + \Delta_1$	$0 = V_{in} + \Delta_3 - \Delta_2$	Δ_1
0110	$0 = V_{in} + \Delta_3 - \Delta_2$	$0 = V_{in} + \Delta_3 - \Delta_2 - \Delta_1$	Δ_1
0111	$0 = V_{in} + \Delta_3 - \Delta_2 - \Delta_1$	$0 = V_{in}$	$\Delta_3 - \Delta_2 - \Delta_1$
1000	$0 = V_{in}$	$0 = V_{in} + \Delta_3 + \Delta_2 + \Delta_1$	$\Delta_3 - \Delta_2 - \Delta_1$
1001	$0 = V_{in} - \Delta_3 + \Delta_2 + \Delta_1$	$0 = V_{in} + \Delta_3 + \Delta_2$	Δ_1
1010	$0 = V_{in} - \Delta_3 + \Delta_2$	$0 = V_{in} + \Delta_3 + \Delta_2 - \Delta_1$	Δ_1
1011	$0 = V_{in} - \Delta_3 + \Delta_2 - \Delta_1$	$0 = V_{in} + \Delta_3$	$\Delta_2 - \Delta_1$
1100	$0 = V_{in} - \Delta_3$	$0 = V_{in} + \Delta_3 - \Delta_2 + \Delta_1$	$\Delta_2 - \Delta_1$
1101	$0 = V_{in} - \Delta_3 - \Delta_2 + \Delta_1$	$0 = V_{in} + \Delta_3 - \Delta_2$	Δ_1
1110	$0 = V_{in} - \Delta_3 - \Delta_2$	$0 = V_{in} + \Delta_3 - \Delta_2 - \Delta_1$	Δ_1
1111	$0 = V_{in} - \Delta_3 - \Delta_2 - \Delta_1$	–	–

Table A.1: Upper and lower boundary and the width of the bins in a 4 bit SAR ADC

B Additional characterization measurements

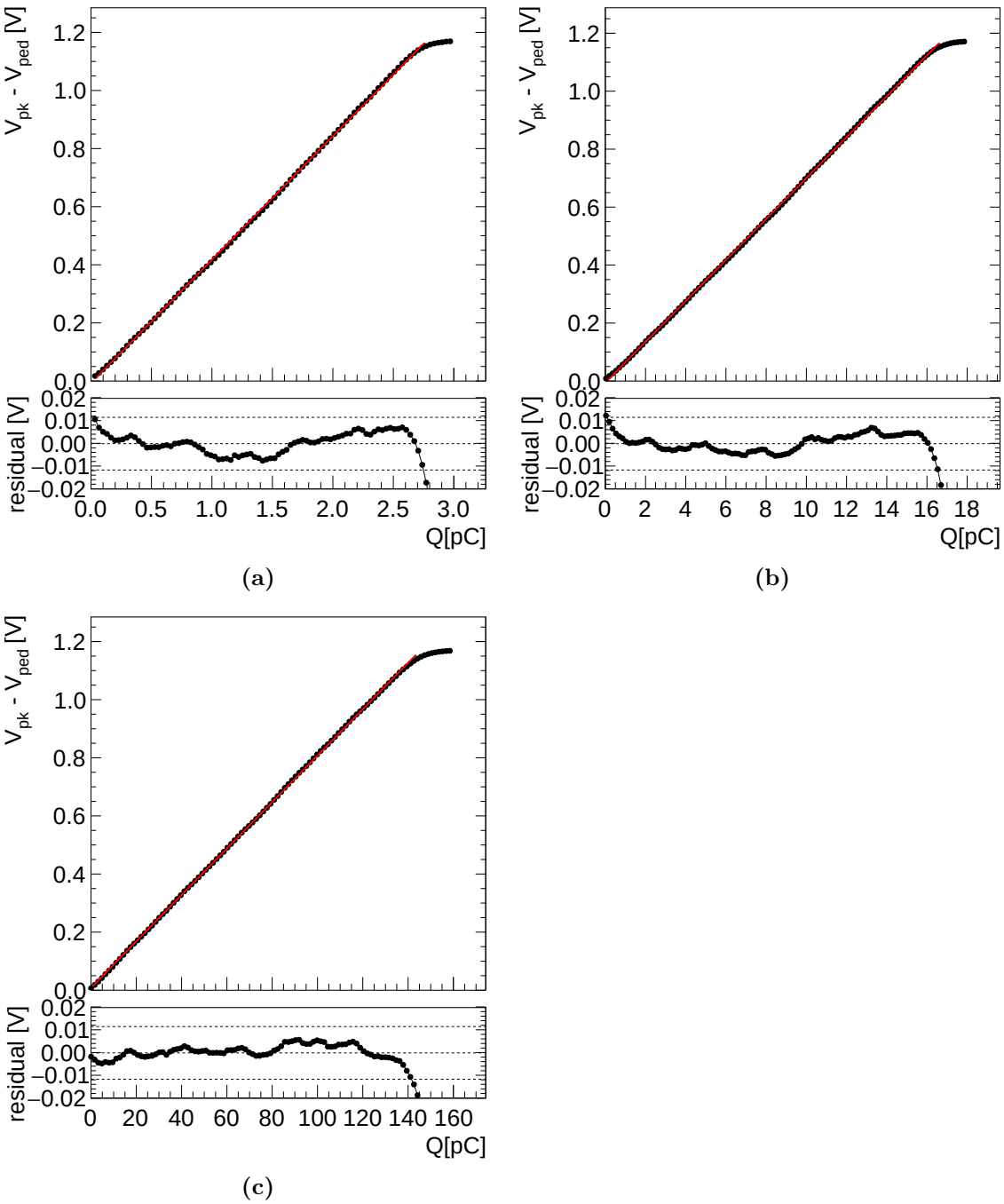


Figure B.1: Charge measurement response and residuals from a linear fit. The fit range is limited to an integrated nonlinearity 2% FSR INL, and the dashed lines correspond to the 1% FSR limits. a) High gain branch, HG setting. b) High gain branch, MG setting. c) Low gain branch

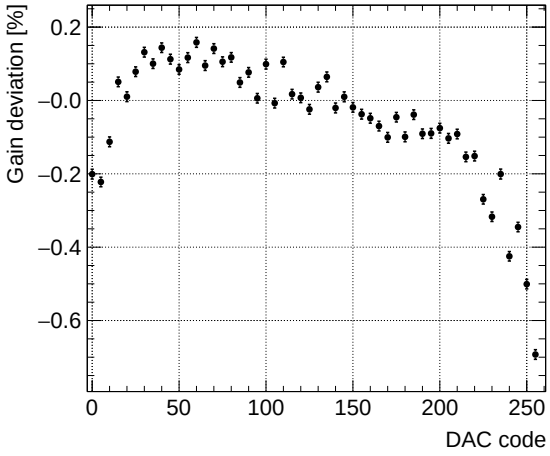
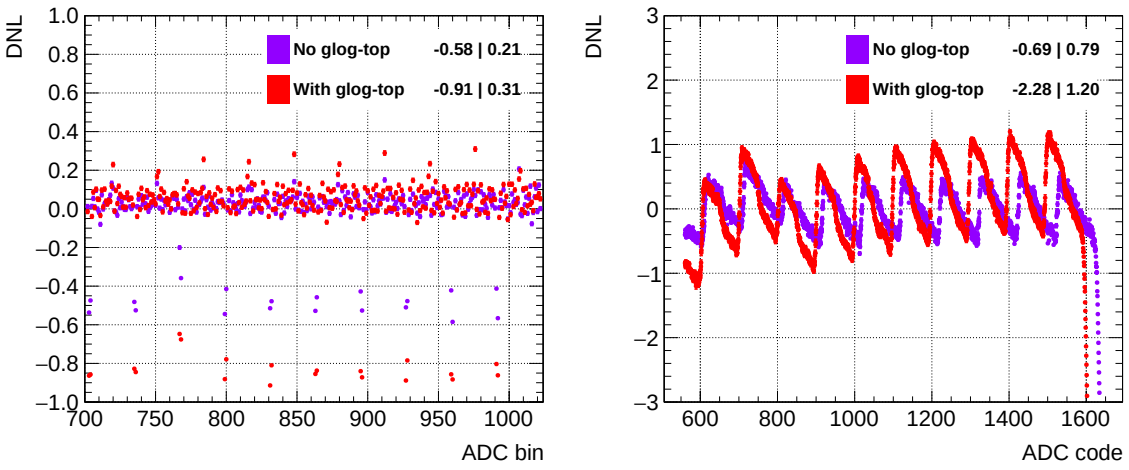
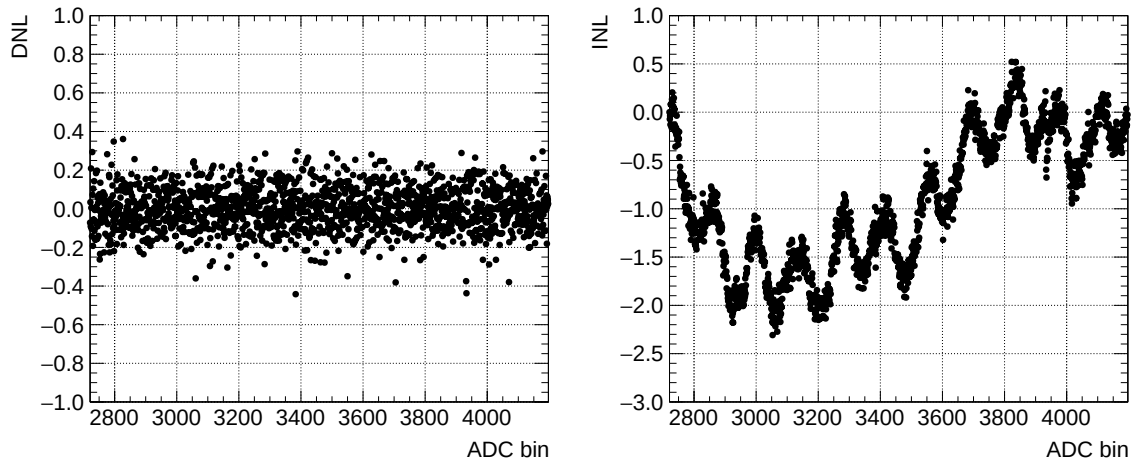


Figure B.2: Relative change of the charge conversion factor as a function of the SiPM bias voltage DAC configuration



(a) **(b)**

Figure B.3: Impact on ADC measurement nonlinearities from the application of glog-top to the ASIC.
a) DNL as a function of the ADC bin.
b) INL as a function of the voltage at the positive ADC input terminal.



(a) **(b)**

Figure B.4: Differential nonlinearity (a) and Integrated nonlinearity (b) of the ADC in 12-bit resolution mode.

C Lists

C.1 List of Figures

2.1 Formation of showers in the calorimeter . . . 6
2.2 Structure of a sampling calorimeter . . . 7
2.3 Particle flow and sources of confusion . . . 9
2.4 Event reconstruction and energy resolution in Particle flow calorimetry . . . 10
2.5 The international linear collider . . . 12
2.6 View of the acceleration tunnel at the European XFEL facility . . . 13
2.7 The ILD detector . . . 14
2.8 Calorimeters at ILD, AHCAL module . . . 16
2.9 Scintillator strip ECAL . . . 17

3.1 Sketch of the APD doping profile and electric field . . . 19
3.2 Analog Silicon Photomultipliers . . . 21
3.3 SiPM single photon spectrum . . . 22
3.4 Correlated avalanche processes induced by photons from hot carriers [43] . . . 23
3.5 Photon counting resolution of Silicon Photomultipliers . . . 24
3.6 High density Silicon Photomultipliers . . . 25
3.7 Electrical model of a SiPM . . . 26
3.8 SiPM response simulation . . . 27

4.1 Schematic symbol and sketch of an n-MOS transistor . . . 32
4.2 Operation regions of the n-MOS transistor in strong inversion . . . 33
4.3 Small signal model of a MOS transistor . . . 35
4.4 Other (MOS) transistor types in CMOS technologies . . . 36
4.5 Illustration of commonly-used pulse shaping systems (filters) . . . 41
4.6 Comparison of shaper topologies . . . 42

5.1 Electronic noise sources . . . 45
5.2 Illustration of signal pileup from dark-count pulses. . . . 47
5.3 Simulated pedestal spectra of single-pulse DCR pileup for different impulse response functions . . . 48
5.4 Effective DCR pileup noise measurement and simultaion . . . 50
5.5 Pulse contributions of the effective DCR pileup noise . . . 51
5.6 Simulation of pileup noise as a function of the shaping time for different filter topologies . . . 52
5.7 Illustration of the peak voltage sampling and the impact of comparison time uncertainties . . . 53
5.8 Noise introduced by sampling time jitter . . . 55
5.9 Comparator decision time (time walk) as a function of the input charge. . . . 56

5.10 Voltage sampling errors due to the time-walk effect 57

6.1 Layout of the KLauS4 ASIC 60
6.2 Block diagram of the analog front-end of a KLauS channel 62
6.3 Input stage schematic diagram 63
6.4 SiPM bias tuning DAC schematic 66
6.5 Bias current generation of the low power DAC 67
6.6 Block diagram of the charge measurement branches 68
6.7 Schematic of the HG integration block 69
6.8 Schematic sketch of the active filter used for pulse shaping 70
6.9 Schematic diagram of the input stage including the noise sources 73
6.10 Noise components of the KLauS4 front-end 76
6.11 Ballistic deficit and optimum pSNR calculation 77
6.12 Schematic sketch of the KLauS comparator branches 79
6.13 SAR ADC topology 81
6.14 Schematic of the 10-bit SAR-ADC 82
6.15 SAR ADC control logic core 84
6.16 Simulation of a 10 bit ADC conversion cycle 85
6.17 Illustration of capacitor DAC parasitics in the SAR ADC 86
6.18 Schematic of the 12-bit SAR-ADC 88
6.19 Simulation of a 12 bit ADC conversion cycle 89
6.20 Timing diagram 90
6.21 Front-end hit-logic 91
6.22 Block diagram of the channel control digital part 92
6.23 Block digagram 93
6.24 I^2C transaction types used to communicate with the ASIC. 95
6.25 On chip event validation features 97
6.26 Illustration of the switching procedure in power-pulsed operation 99
6.27 Input stage schematic including circuits for powergated operation 99
6.28 Pedestal feedback circuit 101
6.29 Simulation of the pedestal voltage settling behaviour in power pulsing mode 101

7.1 Test setup for characterization 106
7.2 SiPM input DAC range and linearity 107
7.3 Charge measurement response of the analog front-end 108
7.4 Charge measurement response and equivalent noise as function of the detector capacitance 109
7.5 Trigger threshold adjustment 110
7.6 Trigger jitter and timewalk 111
7.7 Delay box characterization 112
7.8 10-bit ADC DNL and INL 113
7.9 ADC nonlinearity correction using a look-up table 114
7.10 Pulse sampled by the internal ADC by means of hold-delay shifting 115
7.11 Combined front-end and ADC linearity 116
7.12 Effect of timewalk on the front-end+ADC linearity 117
7.13 Automatic gain selection 118

7.14 Sample single photon spectra . . . 119
7.15 SiPM input DAC in power-pulsed operation . . . 120
7.16 Settling of the front-end in power pulsed operation . . . 121
7.17 Single photon spectra in power-pulsed operation . . . 121
7.18 Setup used at the DESY test-beam facility . . . 123
7.19 Automatic gain selection input spectra . . . 124
7.20 Merged spectrum after gain intercalibration . . . 124
7.21 Trigger efficiency to electrons . . . 126
7.22 Electron spectrum in the DESY testbeam: Power pulsing . . . 126

A.1 Schematic diagram of the hold switch used in the KLauS ADC . . . 131
A.2 Schematic diagram of the voltage comparator used in the KLauS ADC . . . 132
A.3 4-bit SAR ADC example: Upper bin boundaries . . . 133
A.4 4-bit SAR ADC example: Lower bin boundaries . . . 134
A.5 Delay cell schematic . . . 134

B.1 Charge measurement response and fit residuals . . . 138
B.2 Charge conversion factor change due to the SiPM bias voltage configuration 139
B.3 10-bit ADC DNL and INL after applying glog-top . . . 139
B.4 12-bit ADC DNL and INL . . . 140

C.2 List of Tables

3.1 Parameters in the SiPM electrical model . . . 29

4.1 Small signal model parameters for the n-MOS Transistor . . . 35

6.1 Potential of the capacitor back-planes connected to the charge steering DAC. 85

7.1 Charge conversion factors, linear range and equivalent noise charge at small detector capacitances for all charge ranges of the front-end . . . 109

A.1 Upper and lower boundary and the width of the bins in a 4 bit SAR ADC . 135

D Bibliography

[1] ATLAS - Georges Aad et al. „Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC“. In: *Phys. Lett.* B716 (2012), pp. 1–29. DOI: 10.1016/j.physletb.2012.08.020. arXiv: 1207.7214 [hep-ex].

[2] CMS - S. Chatrchyan et al. „Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC“. In: *Physics Letters B* 716.1 (2012), pp. 30–61. ISSN: 0370-2693. DOI: https://doi.org/10.1016/j.physletb.2012.08.021. URL: http://www.sciencedirect.com/science/article/pii/S0370269312008581.

[3] Halina Abramowicz et al. *The International Linear Collider Technical Design Report - Volume 4: Detectors*. Ed. by Ties Behnke et al. 2013. arXiv: 1306.6329 [physics.ins-det].

[4] Mark Thomson. „Model-independent measurement of the $e^+ e^- \to HZ$ cross section at a future $e^+ e^-$ linear collider using hadronic Z decays“. In: *Eur. Phys. J.* C76.2 (2016), p. 72. DOI: 10.1140/epjc/s10052-016-3911-5. arXiv: 1509.02853 [hep-ex].

[5] Wei Shen. „Development of high performance readout ASICs for silicon photomultipliers (SiPMs)“. PhD dissertation. Ruprecht-Karls Universität Heidelberg, Germany, 2012.

[6] K. Briggl et al. „KLauS: a low power Silicon Photomultiplier charge readout ASIC in 0.18 UMC CMOS“. In: *Journal of Instrumentation* 11.03 (2016), p. C03045. URL: http://stacks.iop.org/1748-0221/11/i=03/a=C03045.

[7] K. Briggl, H. Chen, W. Shen, and H.C. Schultz-Coulon. „Low power Analog Digital Converter for a silicon photomultiplier readout ASIC“. In: *Journal of Instrumentation* 10.04 (2015), p. C04041. URL: http://stacks.iop.org/1748-0221/10/i=04/a=C04041.

[8] W. Shen, K. Briggl, H. Chen, and H. C. Schultz-Coulon. „A dedicated analog digital converter for silicon photomultiplier readout“. In: *2014 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC)*. Nov. 2014, pp. 1–7. DOI: 10.1109/NSSMIC.2014.7431044.

[9] Claus Grupen and Boris A. Shwartz. *Particle detectors*. 2. ed. Cambridge monographs on particle physics, nuclear physics, and cosmology. Cambridge University Press, 2008, XXIII, 651 S. ISBN: 0-521-84006-6 and 978-0-521-84006-4.

[10] Alexander Kaplan. „Hadronic Imaging Calorimetry“. PhD thesis. Universität Heidelberg, 2011.

[11] Mark A. Thomson. „Particle Flow Calorimetry“. In: *Journal of Physics: Conference Series* 293 (2011), p. 012021. ISSN: 1742-6596 and 1742-6596.

[12] J. S. Marshall and M. A. Thomson. „The Pandora software development kit for pattern recognition“. In: *The European Physical Journal C* 75.9 (2015), pp. 1–16. ISSN: 1434-6044 and 1434-6052.

[13] Mark A. Thomson. „Particle flow calorimetry and the PandoraPFA algorithm“. In: *Nuclear Inst. and Methods in Physics Research, A* 611.1 (2009), pp. 25–40. ISSN: 0168-9002 and 1872-9576.

[14] The CALICE collaboration - C Adloff et al. „Hadronic energy resolution of a highly granular scintillator-steel hadron calorimeter using software compensation techniques“. In: *Journal of Instrumentation* 7 (2012), P09017. ISSN: 1748-0221 and 1748-0221.

[15] The CALICE collaboration - C Adloff et al. „The time structure of hadronic showers in highly granular calorimeters with tungsten and steel absorbers“. In: *Journal of Instrumentation* 9.07 (2014), P07022. URL: http://stacks.iop.org/1748-0221/9/i=07/a=P07022.

[16] The CALICE Collaboration. „The Time Structure of Hadronic Showers in Tungsten with FastRPC“. In: *CAN-043* (2013). ISSN: 1748-0221. URL: https://twiki.cern.ch/twiki/pub/CALICE/CaliceAnalysisNotes/CAN-043.pdf.

[17] A. Benaglia et al. „Space-Time Development of Electromagnetic and Hadronic Showers and Perspectives for Novel Calorimetric Techniques“. In: *IEEE Transactions on Nuclear Science* 63.2 (Apr. 2016), pp. 574–579. ISSN: 0018-9499. DOI: 10.1109/TNS.2016.2527758.

[18] The ILC Strategy Council. *Press release - "Announcement of the results of the ILC candidate site evaluation in Japan"*. 2013. URL: http://ilc-str.jp/topics/2013/08281826/.

[19] Chris Adolphsen et al. *The International Linear Collider Technical Design Report - Volume 3.II: Accelerator Baseline Design*. 2013. arXiv: 1306.6328 [physics.acc-ph].

[20] Ties Behnke et al. *The International Linear Collider Technical Design Report - Volume 1: Executive Summary*. 2013. arXiv: 1306.6327 [physics.acc-ph].

[21] European XFEL. *Official webpage*. September 2017. URL: https://www.xfel.eu.

[22] Linear Collider ILD Concept Group - Toshinori Abe et al. *The International Large Detector: Letter of Intent*. 2010. DOI: 10.2172/975166. arXiv: 1006.3396 [hep-ex].

[23] The CALICE collaboration - C Adloff et al. „Construction and performance of a silicon photomultiplier/extruded scintillator tail-catcher and muon-tracker“. In: *Journal of Instrumentation* 7.04 (2012), P04015. URL: http://stacks.iop.org/1748-0221/7/i=04/a=P04015.

[24] The CALICE collaboration - C Adloff et al. „Construction and commissioning of the CALICE analog hadron calorimeter prototype“. In: *Journal of Instrumentation* 5.05 (2010), P05004. URL: http://stacks.iop.org/1748-0221/5/i=05/a=P05004.

[25] The CALICE collaboration - C Adloff et al. „Electromagnetic response of a highly granular hadronic calorimeter“. In: *Journal of Instrumentation* 6.04 (2011), P04003. URL: http://stacks.iop.org/1748-0221/6/i=04/a=P04003.

[26] The CALICE collaboration - C Adloff et al. „Tests of a Particle Flow Algorithm with CALICE test beam data“. In: *Journal of Instrumentation* 6.07 (2011), P07005. URL: http://stacks.iop.org/1748-0221/6/i=07/a=P07005.

[27] The CALICE collaboration - K. Francis et al. „Performance of the first prototype of the CALICE scintillator strip electromagnetic calorimeter“. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 763.Supplement C (2014), pp. 278–289. ISSN: 0168-9002. DOI: 10.1016/j.nima.2014.06.039.

[28] Katsushige Kotera, Daniel Jeans, Akiya Miyamoto, and Tohru Takeshita. „A novel strip energy splitting algorithm for the fine granular readout of a scintillator strip electromagnetic calorimeter“. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 789.Supplement C (2015), pp. 158–164. ISSN: 0168-9002. DOI: 10.1016/j.nima.2015.04.001.

[29] The CALICE collaboration - Katsushige Kotera. „Scintillator Strip ECAL Optimization“. In: *International Workshop on Future Linear Colliders (LCWS13) Tokyo, Japan, November 11-15, 2013*. 2014. arXiv: 1404.1672 [physics.ins-det].

[30] K. Kotera. „$5{\cdot}5\,\mathrm{mm}^2$ granular electromagnetic calorimeter using $45{\cdot}5\,\mathrm{mm}^2$ scintillator strips and SiPMs“. In: *2014 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC)*. Nov. 2014. DOI: 10.1109/NSSMIC.2014.7431152.

[31] The CALICE Collaboration and ILD concept group. „Calibration of the Scintillator Hadron Calorimeter of ILD“. In: *CAN-018* (2009). URL: https://twiki.cern.ch/twiki/pub/CALICE/CaliceAnalysisNotes/CAN-018.pdf.

[32] Tobias Harion. „The STiC ASIC - High Precision Timing with Silicon Photomultipliers“. PhD thesis. Universität Heidelberg, 2015.

[33] S. Cova et al. „Avalanche photodiodes and quenching circuits for single-photon detection“. In: *Applied Optics* 35.12 (Apr. 1996), pp. 1956–1976. DOI: 10.1364/AO.35.001956.

[34] Patrick Eckert. „The Mu3e Tile Detector“. PhD thesis. Universität Heidelberg, 2015.

[35] S. Vinogradov et al. „Probability distribution and noise factor of solid state photomultiplier signals with cross-talk and afterpulsing“. In: *2009 IEEE Nuclear Science Symposium Conference Record (NSS/MIC)*. Oct. 2009, pp. 1496–1500. DOI: 10.1109/NSSMIC.2009.5402300.

[36] Patrick Eckert et al. „Characterisation studies of silicon photomultipliers“. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 620.2 (2010), pp. 217–226. ISSN: 0168-9002. DOI: 10.1016/j.nima.2010.03.169.

[37] D. Renker. „Geiger-mode avalanche photodiodes, history, properties and problems“. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 567.1 (2006). Proceedings of the 4th International Conference on New Developments in Photodetection, pp. 48–56. ISSN: 0168-9002. DOI: 10.1016/j.nima.2006.05.060.

[38] SensL. *C-Series Blue-Sensitive Silicon Photomultipliers (Rev 2.8)*. 2017. URL: http://sensl.com/products/c-series/.

[39] A. G. Chynoweth and K. G. McKay. „Photon Emission from Avalanche Breakdown in Silicon“. In: *Phys. Rev.* 102 (2 Apr. 1956), pp. 369–376. DOI: 10.1103/PhysRev.102.369.

[40] A. L. Lacaita, F. Zappa, S. Bigliardi, and M. Manfredi. „On the bremsstrahlung origin of hot-carrier-induced photons in silicon devices“. In: *IEEE Transactions on Electron Devices* 40.3 (Mar. 1993), pp. 577–582. ISSN: 0018-9383. DOI: 10.1109/16.199363.

[41] F. Retière and K. Boone. „Delayed avalanches in Multi-Pixel Photon Counters“. In: *2012 IEEE Nuclear Science Symposium and Medical Imaging Conference Record (NSS/MIC)*. Oct. 2012, pp. 1585–1588. DOI: 10.1109/NSSMIC.2012.6551378.

[42] C. Piemonte et al. „Development of an automatic procedure for the characterization of silicon photomultipliers“. In: *2012 IEEE Nuclear Science Symposium and Medical Imaging Conference Record (NSS/MIC)*. Oct. 2012, pp. 428–432. DOI: 10.1109/NSSMIC.2012.6551141.

[43] F. Acerbi et al. „High Efficiency, Ultra high-density Silicon Photomultipliers“. In: *IEEE Journal of Selected Topics in Quantum Electronics* PP.99 (2017). ISSN: 1077-260X. DOI: 10.1109/JSTQE.2017.2748927.

[44] S. Vinogradov et al. „Efficiency of Solid State Photomultipliers in Photon Number Resolution“. In: *IEEE Transactions on Nuclear Science* 58.1 (Feb. 2011), pp. 9–16. ISSN: 0018-9499. DOI: 10.1109/TNS.2010.2096474.

[45] E. B. Johnson et al. „New developments for CMOS SSPMs“. In: *2008 IEEE Nuclear Science Symposium Conference Record*. Oct. 2008, pp. 1516–1522. DOI: 10.1109/NSSMIC.2008.4774701.

[46] K. Sato et al. „The reliability and reproducibility of MPPC“. In: *2013 IEEE Nuclear Science Symposium and Medical Imaging Conference (2013 NSS/MIC)*. Oct. 2013. DOI: 10.1109/NSSMIC.2013.6829584.

[47] F. Corsi et al. „Modelling a silicon photomultiplier (SiPM) as a signal source for optimum front-end design“. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 572.1 (2007). Frontier Detectors for Frontier Physics, pp. 416–418. ISSN: 0168-9002. DOI: https://doi.org/10.1016/j.nima.2006.10.219.

[48] Florian Scheuch et al. „Electrical characterization and simulation of SiPMs“. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 787.Supplement C (2015). New Developments in Photodetection NDIP14, pp. 340–343. ISSN: 0168-9002. DOI: https://doi.org/10.1016/j.nima.2015.01.066.

[49] E. Popova et al. „Amplitude and timing properties of a Geiger discharge in a SiPM cell“. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 787 (2015). New Developments in Photodetection NDIP14, pp. 270–274. ISSN: 0168-9002. DOI: http://dx.doi.org/10.1016/j.nima.2014.12.050.

[50] Adam Nepomuk Otte, Distefano Garcia, Thanh Nguyen, and Dhruv Purushotham. „Characterization of three high efficiency and blue sensitive silicon photomultipliers“. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 846 (2017), pp. 106–125. ISSN: 0168-9002. DOI: 10.1016/j.nima.2016.09.053.

[51] D. Marano et al. „Silicon Photomultipliers Electrical Model Extensive Analytical Analysis“. In: *IEEE Transactions on Nuclear Science* 61.1 (Feb. 2014), pp. 23–34. ISSN: 0018-9499. DOI: 10.1109/TNS.2013.2283231.

[52] Hamamatsu Photonics. *S12571-010C MPPC Datasheet*. 2015. URL: http://www.hamamatsu.com/resources/pdf/ssd/s12571-010_etc_kapd1044e.pdf.

[53] Andreas Grübl. „VLSI implementation of a spiking neural network“. PhD thesis. Universität Heidelberg, 2007.

[54] Paul R. Gray. *Analysis and design of analog integrated circuits*. 5. ed., international student version. Hoboken, NJ: Wiley, 2010. ISBN: 978-0-470-39877-7.

[55] S. M. Sze and Kwok K. Ng. *Physics of semiconductor devices*. 3. ed. Hoboken, NJ: Wiley-Interscience, 2007, X, 815 S. ISBN: 0-471-14323-5 and 978-0-471-14323-9.

[56] Swapnadip De, Manash Chanda, Chandan Kumar Sarkar, and Angsuman Sarkar. *Low Power VLSI Design fundamentals*. Berlin ; Boston: De Gruyter Oldenbourg, 2016, XIV, 310 S. ISBN: 978-3-11-045526-7 and 978-3-11-045555-7.

[57] Frédéric Cohen Tenoudji. *Analog and Digital Signal Analysis. From Basics to Applications*. eng. SpringerLink : Bücher. Cham: Springer, 2016, Online–Ressource (XXIII, 608 p. 256 illus., 9 illus. in color, online resource). ISBN: 978-3-319-42382-1. DOI: 10.1007/978-3-319-42382-1.

[58] Shoab Ahmed Khan. *Digital design of signal processing systems. a practical approach*. eng. Chichester, West Sussex, U.K.: J. Wiley & Sons, 2011, Online–Ressource (1 v. p.) ISBN: 978-0-470-74183-2.

[59] Behzad Razavi. *Design of analog CMOS integrated circuits*. International ed. McGraw-Hill series in electrical and computer engineering. McGraw-Hill, 2001. ISBN: 0-07-118839-8 and 0-07-118815-0 and 978-0-07-118839-5 and 978-0-07-118815-9.

[60] R. P. Sallen and E. L. Key. „A practical method of designing RC active filters“. In: *IRE Transactions on Circuit Theory* 2.1 (Mar. 1955), pp. 74–85. ISSN: 0096-2007. DOI: 10.1109/TCT.1955.6500159.

[61] Helmuth Spieler. *Semiconductor detector systems*. Repr. Series on semiconductor science and technology. Oxford [u.a.]: Oxford Univ. Press, 2008. ISBN: 978-0-19-852784-8.

[62] M. Centis Vignali et al. „Neutron irradiation effect on SiPMs up to $\Phi_{neq} = 5 \times 10^{14}$ cm^{-2}“. In: (2017). arXiv: 1709.04648 [physics.ins-det].

[63] N. Campbell. „The study of discontinuous phenomena“. In: *Proc Cambr. Phil. Soc* vol. 15 (1909), pp. 117–136.

[64] M Dorn et al. „KLauS – A charge readout and fast discrimination chip for silicon photomultipliers“. In: *Journal of Instrumentation* 7.01 (2012), p. C01008. URL: http://stacks.iop.org/1748-0221/7/i=01/a=C01008.

[65] P Eckert, R Stamen, and H -C Schultz-Coulon. „Study of the response and photon-counting resolution of silicon photomultipliers using a generic simulation framework“. In: *Journal of Instrumentation* 7.08 (2012), P08011. URL: http://stacks.iop.org/1748-0221/7/i=08/a=P08011.

[66] S. Vinogradov. „Analytical model of SiPM time resolution and order statistics with crosstalk“. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 787 (2015). New Developments in Photodetection NDIP14, pp. 229–233. ISSN: 0168-9002. DOI: http://dx.doi.org/10.1016/j.nima.2014.12.010. URL: http://www.sciencedirect.com/science/article/pii/S016890021401451X.

[67] S. Seifert et al. „A Comprehensive Model to Predict the Timing Resolution of SiPM-Based Scintillation Detectors: Theory and Experimental Validation“. In: *IEEE Transactions on Nuclear Science* 59.1 (Feb. 2012), pp. 190–204. ISSN: 0018-9499. DOI: 10.1109/TNS.2011.2179314.

[68] J. Blair. „Histogram measurement of ADC nonlinearities using sine waves“. In: *IEEE Transactions on Instrumentation and Measurement* 43.3 (June 1994), pp. 373–383. ISSN: 0018-9456. DOI: 10.1109/19.293454.

[69] Marco Corrado, Sergio Rapuano, and Jan Šaliga. „An overview of different signal sources for histogram based testing of ADCs“. In: *Measurement* 43.7 (2010), pp. 878–886. ISSN: 0263-2241. DOI: http://dx.doi.org/10.1016/j.measurement.2010.04.002. URL: http://www.sciencedirect.com/science/article/pii/S0263224110000977.

[70] Hendrik F. Lundin. „Characterization and Correction of Analog-to-Digital Converters“. PhD dissertation. KTH, Stockholm, Sweden, 2005.

[71] Oskar Hartbrich. „Scintillator Calorimeters for a Future Linear Collider Experiment“. PhD thesis. Bergische Universität Wuppertal, 2016. DOI: 10.3204/PUBDB-2016-02800.

[72] E. M. Camacho-Galeano, C. Galup-Montoro, and M. C. Schneider. „A 2-nW 1.1-V self-biased current reference in CMOS technology“. In: *IEEE Transactions on Circuits and Systems II: Express Briefs* 52.2 (Feb. 2005), pp. 61–65. ISSN: 1549-7747. DOI: 10.1109/TCSII.2004.842059.

[73] A. de la Plaza. „Nanoampere supply independent low-voltage current reference“. In: *2011 Faible Tension Faible Consommation (FTFC)*. May 2011, pp. 9–11. DOI: 10.1109/FTFC.2011.5948930.

[74] C. H. Lee and H. J. Park. „All-CMOS temperature independent current reference“. In: *Electronics Letters* 32.14 (July 1996), pp. 1280–1281. ISSN: 0013-5194. DOI: 10.1049/el:19960827.

[75] H. Träff. „Novel approach to high speed CMOS current comparators“. In: *Electronics Letters* 28 (3 Jan. 1992), 310–311(1). ISSN: 0013-5194. URL: http://digital-library.theiet.org/content/journals/10.1049/el_19920192.

[76] Behzad Razavi. „Digital to Analog Converter Architectures“. In: *Principles of Data Conversion System Design*. Wiley-IEEE Press, 1995, pp. 79–95. ISBN: 9780470545638. DOI: 10.1109/9780470545638.ch5.

[77] Imran Ahmed. *Pipelined ADC Design and Enhancement Techniques*. Springer Netherlands, 2010. ISBN: 9048186528 and 9789048186525.

[78] B. Murmann. *ADC Performance Survey 1997-2017*. August 2017. URL: http://web.stanford.edu/~murmann/adcsurvey.html.

[79] A. M. Abo. „Design for reliability of low-voltage, switched-capacitor circuits“. PhD dissertation. Univ. California, Berkeley, CA, 1999.

[80] A. M. Abo and P. R. Gray. „A 1.5-V, 10-bit, 14.3-MS/s CMOS pipeline analog-to-digital converter“. In: *IEEE Journal of Solid-State Circuits* 34.5 (May 1999), pp. 599–606. ISSN: 0018-9200. DOI: 10.1109/4.760369.

[81] C. C. Liu, S. J. Chang, G. Y. Huang, and Y. Z. Lin. „A 10-bit 50-MS/s SAR ADC With a Monotonic Capacitor Switching Procedure“. In: *IEEE Journal of Solid-State Circuits* 45.4 (Apr. 2010), pp. 731–740. ISSN: 0018-9200. DOI: 10.1109/JSSC.2010.2042254.

[82] Alan Hastings. *The Art of Analog Layout (2nd Edition)*. Prentice Hall, 2005.

[83] Yuan Zhou and Yun Chiu. „Digital calibration of inter-stage nonlinear errors in pipelined SAR ADCs“. In: *ANALOG INTEGRATED CIRCUITS AND SIGNAL PROCESSING* 82.3, SI (Mar. 2015). IEEE 56th International Midwest Symposium on Circuits and Systems (MWSCAS), Ohio State Univ, Ohio Union, Columbus, OH, AUG 04-07, 2013, 533–542. ISSN: 0925-1030. DOI: {10.1007/s10470-015-0493-3}.

[84] Gyudong Kim, Min-Kyu Kim, Byoung-Soo Chang, and Wonchan Kim. „A low-voltage, low-power CMOS delay element“. In: *IEEE Journal of Solid-State Circuits* 31.7 (July 1996), pp. 966–971. ISSN: 0018-9200. DOI: 10.1109/4.508210.

[85] P. V. Manjunath, H. R. Baghyalakshmi, and M. K. Venkatesha. „A Low-Power Low-Voltage CMOS Thyristor Based Delay Element“. In: *2009 Second International Conference on Emerging Trends in Engineering Technology*. IEEE, Dec. 2009, pp. 135–140. DOI: 10.1109/ICETET.2009.11.

[86] Faraday Technology Corporation. *UMC 0.18 µm Generic Library*. URL: http://freelibrary.faraday.com.tw.

[87] F. Gray. *Pulse code communication*. US Patent 2,632,058. Mar. 1953. URL: https://www.google.com/patents/US2632058.

[88] NXP Semiconductors. *I2C-bus specification and user manual*. UM10204. Rev. 6 (2014). URL: http://www.nxp.com/docs/en/user-guide/UM10204.pdf.

[89] System Management Interface Forum, Inc. *System Management Bus (SMBus) Specification*. Version 3.0 2014. URL: http://smbus.org/specs/SMBus_3_0_20141220.pdf.

[90] H. Chen et al. „MuTRiG: a mixed signal Silicon Photomultiplier readout ASIC with high timing resolution and gigabit data link“. In: *Journal of Instrumentation* 12.01 (2017), p. C01043. URL: http://stacks.iop.org/1748-0221/12/i=01/a=C01043.

[91] David Doman. *Engineering the CMOS library - enhancing digital design kits for competitive silicon*. Hoboken, N.J.: J. Wiley & Sons, 2012. ISBN: 978-1-118-24304-6.

[92] Hubert Kaeslin. *Top-down digital VLSI design - from architectures to gate-level circuits and FPGAs*. Waltham, MA: Morgan Kaufmann, 2015. ISBN: 0-12-800772-9 and 978-012-800-730-3 and 978-012-800-772-3 and 978-0-12-800772-3.

[93] „IEEE Standard for the Functional Verification Language e“. In: *IEEE Std 1647-2016 (Revision of IEEE Std 1647-2011)* (Jan. 2017), pp. 1–558. DOI: 10.1109/IEEESTD.2017.7805158.

[94] R. Brun and F. Rademakers. „ROOT: An object oriented data analysis framework“. In: *Nucl. Instrum. Meth.* A389 (1997), pp. 81–86. DOI: 10.1016/S0168-9002(97)00048-X.

[95] Mathias Götze. „Systematische Untersuchungen zum LED Kalibrationssystem des analogen hadronischen Kalorimeters der CALICE Kollaboration“. German. Diploma Thesis. Bergische Universität Wuppertal, 2011.

[96] Konrad Briggl. „KLauS und STiC - Zwei ASICs zur präzisen Energie- und Zeitmessung mit Silizium-Photomultipliern“. German. Diploma Thesis. Universität Heidelberg, 2012.

[97] Yong Liu et al. „A design of scintillator tiles read out by surface-mounted SiPMs for a future hadron calorimeter“. In: *2014 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC)*. Nov. 2014. DOI: 10.1109/NSSMIC.2014.7431118.

Danksagungen

Zuletzt bleibt mir, mich bei allen zu bedanken die Anteil an dieser Arbeit hatten oder mich während der Promotionszeit begleitet haben.

Zuallererst gilt mein Dank meinem Doktorvater Prof. Hans-Christian Schultz-Coulon, für die Unterstützung über die Jahre, das Vertrauen und die Möglichkeit die Entwicklung dieses Chips voranzutreiben.
Prof. Norbert Herrmann danke ich für die Bereitschaft, für diese Arbeit als Zweitgutachter zur Verfügung zu stehen.

Allen Korrektoren danke ich für die hilfreichen Tipps und die Unterstützung während dem Schreiben. Die Mitglieder der Detektorentwicklung die mich über all die Jahre hinweg begleitet haben, allen voran Wei, Yonathan, Huangshan und Vera, und natürlich auch die Ehemaligen, Tobias und Patrick, hatten durch die vielen Diskussionen einen Wesentlichen Anteil an dieser Arbeit.
Dem ASIC-Team danke ich fürs durchhalten vor und nach stressigen Submissonsdeadlines. Ohne euch wäre nichts fertig geworden!
Der ganzen Elektronikabteilung und dem ASIC-Labor möchte ich für die Hilfe bei den PCBs und die vielen Nützlichen Tipps danken. Ein besonderer Dank gilt auch Zhenxiong und Yonathan für die Hilfe bei den Characterisierungsmessungen und der Woche am DESY (lange Nächte...).

Allen Mitgliedern von F8 & F11 danke ich für die schöne Arbeitsatmosphäre und all die Aktivitäten im und jenseits des Instituts.

Zum Schluss möchte ich mich bei meiner Familie bedanken, die mich über die ganze Zeit der Ausbildung in jeder denkbaren Weise unterstützt hat. Ganz besonders bei Mareile, für die Unterstützung über die Jahre und die Geduld während stressigeren Zeiten. Gerade in der Letzen Zeit der Promotion, wo andere Dinge von einem Tag auf den anderen viel Wichtiger werden.