INAUGURAL-DISSERTATION

ZUR

ERLANGUNG DER DOKTORWÜRDE

DER

NATURWISSENSCHAFTLICH-MATHEMATISCHEN

GESAMTFAKULTÄT

DER

RUPRECHT-KARLS-UNIVERSITÄT

HEIDELBERG

vorgelegt von
Dipl.-Phys. Edgar Sexauer
aus Kenzingen

Tag der mündlichen Prüfung: 10.1.2001

DEVELOPMENT OF RADIATION HARD READOUT ELECTRONICS FOR LHCb

Gutachter:

Prof. Dr. Karl-Tasso Knöpfle

Prof. Dr. Volker Lindenstruth

Dissertation
submitted to the
Combined Faculties for the Natural Sciences and for Mathematics
of the Rupertus Carola University of
Heidelberg, Germany
for the degree of
Doctor of Natural Sciences

DEVELOPMENT OF RADIATION HARD READOUT ELECTRONICS FOR LHCb

presented by
Diplom-Physicist Edgar Sexauer
born in Kenzingen, Germany

Heidelberg, 10.1.2001

Referees:
Prof. Dr. Karl-Tasso Knöpfle
Prof. Dr. Volker Lindenstruth

Zusammenfassung:

Das Experiment LHCB am CERN ist zur Zeit in der Entwicklungsphase und hat zum Ziel, CP-Verletzung im System der B-Mesonen mit sehr hoher Präzision zu messen. In dem Experiment wird ein mit Silizium-Mikrostreifenzählern ausgestatteter Vertex-Detektor eingesetzt. Ein Mikrochip, der für die Auslese dieser Detektorkomponente geignet ist, wurde in einer Arbeitsgruppe im ASIC-Labor Heidelberg entwickelt.
Dieser Auslesechip Beetle-1.0 besteht aus 128 analogen Eingangsstufen mit einem ladungsempfindlichen Vorverstärker, einem Pulsformer und einer Treiberstufe. Das analoge Signal wird auf einen Komparator geführt, von dem ein schnelles Triggersignal abgeleitet werden kann. Der darauf folgende Ringspeicher (aufgebaut als Matrix von Gate-Kapazitäten) kann entweder zur Speicherung des analogen Signals der Eingangsstufe oder zur Speicherung des digitalen Ausgangs des Komparators benutzt werden. Ein externes Triggersignal markiert Ereignisse, die ausgelesen werden sollen, wobei die zugehörige Stelle in dem Ringspeicher in einem Derandomizer abgelegt wird. Markierte Ereignisse werden mittels einem rücksetzbaren ladungsempfindlichen Verstärker ausgelesen. Ein analoger Multiplexer, der die Signale abhängig vom Aulesemodus im Verhältniss 32:1, 64:1 oder 128:1 seriell weitergibt, schliest sich daran an. Durch einen Stromtreiber werden die Daten von dem dem Chip ausgegeben.
Beim Einsatz des Beetle im Vertexdetektor von LHCb muss er einer Strahlendosis von insgesamt 10 MRad standhalten. Dies wurde durch den Einsatz einer CMOS-Technology mit Strukturbreiten von 0,25μm sowie geschlossener Geometrie von nMOS-Transistoren erreicht.
Im Rahmen dieser Doktorarbeit wurden wesentliche Komponenten der analogen Auslesekette des Beetle-1.0 entworfen und getestet. Dazu wurde auf Ergebnisse von intensiven Messungen an dem Auslesechip HELIX128 zurückgegriffen. Ebenso wurden Teile des SCTA Auslesechips modifiziert, um diesen für das Experiment LHCb einsetzen zu können.

Abstract:

The experiment LHCb is under development at CERN and aims to measure CP-violation in the B-Meson system at very high precision. The experiment makes use of a vertex detector that is equipped with silicon microstrip detectors. A chip suitable for the readout of this detector has been developed in a working group at the ASIC-laboratory Heidelberg.
This readout chip 'Beetle-1.0' contains 128 analog input stages of a charge sensitive preamplifier, a pulse shaper and a buffer. The analog signal is fed into a comparator, from which a fast trigger signal can be derived. The following pipeline, realized as an array of gate capacitances, can be used to either store the analog output of the input amplifiers or to store the digital comparator output. External trigger signals mark events that have to be read out and the according pipeline location is stored in a derandomizing buffer. Pending events are read out from the pipeline via a charge-sensitive, resetable amplifier and an analog multiplexer, which serializes the signals with a factor of 32:1, 64:1 or 128:1, depending on the readout mode. The data are driven off-chip via a current buffer.
For the use of the Beetle in the vertex detector of LHCb, the chip has to withstand a radiation dose of about 10 MRad. This has been achieved by the use of a deep submicron CMOS with a feature size of 0.25μm and an enclosed layout geometry of nMOS transistors.
In the context of this thesis, essential components of the analog readout chain of the Beetle-1.0 have been developed. For that, results of intensive measurements with the HELIX128 readout chip have been used. Also, parts of the SCTA readout chip have been modified to make it usable for the LHCb experiment.

Contents

List of Figures

Introduction

This thesis describes the evolution of the Beetle-1.0 readout chip for LHCb. This VLSI chip has been developed at the ASIC-laboratory Heidelberg in collaboration with the Max-Planck Institute for Nuclear Physics Heidelberg, the University of Heidelberg, NIKHEF Amsterdam and the University of Oxford.

The specification phase of the readout chip started in late 1998, in parallel with the development of prototype components. The first submission of two test chips, containing different variations of analog input stages and bias circuits took place in May 1999. The successful test after their production in November 1999 triggered the assembly of the complete readout chip in one step, which has been submitted in April 2000. Together with the readout chip Beetle-1.0, three test chips containing different subcomponents have been designed and submitted. Table 1 gives an overview on the different chips with their submission dates and functionality.

Chapter 1 describes the basic physics goals of the experiment LHCb, as well as the apparatus and the detector technology. The principle of a silicon microstrip detector is described in more detail to give a picture of the requirements on a silicon microstrip readout chip. Chapter 2 details the important effects of radiation damage in microelectronic circuits, and an overview on existing architectures of microstrip readout chips is given in chapter 3. Chapter 4 describes the design and concepts of the readout chip Beetle-1.0 and in chapter 5 the measurements that have been done so far are reported.

name	submission date	size	functionality
Beetle-FE	May 1999	2×2 mm^2	3 different sets of analog input stages with different shaping times/noise behavior
Beetle-BG	May 1999	2×2 mm^2	10 bit current and voltage DACs, 3 different types of current sources
Beetle-CO	April 2000	2×2 mm^2	frontend comparators, current buffer
Beetle-PA	April 2000	2×2 mm^2	pipeline, pipeline readout amplifier I^2C interface
Beetle-MA	April 2000	2×2 mm^2	3 different sets of analog input stages for the readout of photomultiplier tubes
Beetle-1.0	April 2000	6.1×5.5 mm^2	complete readout chip with full functionality required by LHCb

Table 1: Summary of the chips forming the development steps of the Beetle-1.0

Chapter 1

LHCb: An Experiment to Study CP-Violation in the System of B-Mesons

Since the first observation of CP-violation in the $K^0 \leftrightarrow \overline{K^0}$-system in 1964 [1, 2], it has been the aim of physicists to search for this type of symmetry violation in other systems. The $B^0 \leftrightarrow \overline{B^0}$-system is very attractive for that purpose, since the standard model predicts a large CP-violation for that system. Therefore, the precise examination of CP-violation in the neutral B-mesons is a good test of the standard model and a promising way to find physics beyond the standard model. However, this task is challenging. Typical branching ratios for interesting decay channels of B-mesons are in the order of 10^{-5} and therefore demand a high precision in measuring the interesting effects. Several experiments aim to measure CP-violation in the B-system. The experiments BaBar and BELLE use colliding e^+e^--beams to produce about 10^8 pairs of B-mesons per year [3, 4]. Another possibility is to use the strong interaction to produce B-mesons. This is done in the experiments HERA-B and LHCb, where accelerated proton beams are used. In LHCb where the center-of-mass energy will be 14 TeV, B-mesons at a rate of 100 kHz can be easily produced and a large statistic of CP-violation data can be gained [5].

1.1 Overview of the LHCb detector [5]

LHCb is a forward spectrometer with an angular coverage from 10 mrad to 250 mrad in the non-bending plane and 300 mrad in the bending plane. A picture of the detector can be seen in fig.1.1. The LHCb features a vertex detector, which is built around the interaction point, a tracking system, two RICH counters, an electromagnetic calorimeter with pre-shower detector, a hadron calorimeter and a muon detector. The interaction point is located in the center of the vertex detector, which also comprises additional detectors to derive a pileup veto.

The vertex detector has to provide precise information on the location of the primary and secondary vertex. Its information is also used in the Level-1 trigger. In order to

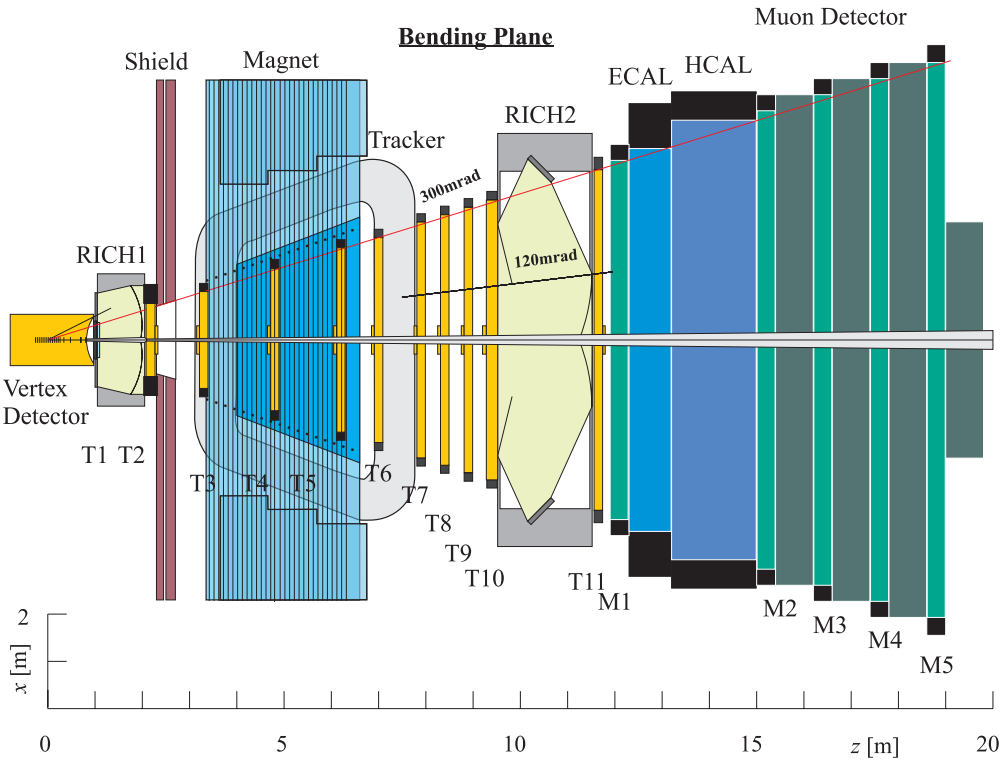

Figure 1.1: Schematic overview on the LHCb detector seen from above [5]

suppress events with more than one proton-proton interaction, a pile-up veto detector is integrated into the vertex detector system. A more detailed description of the vertex detector is given in 1.2.

The tracking system is built of two parts, the inner and the outer tracker. The type of detector technology is determined by the requirement of low occupancy to simplify tracking. For the outer tracker, a honeycomb-chamber technology has been chosen with a cell separation of 5 mm. For the inner tracking system, the higher track density requires a smaller granularity. Therefore, a combination of silicon microstrip detectors for the tracking stations close to the vertex detector and microstrip gaseous chambers for the stations far apart from the vertex detector has been chosen.

The RICH detector has to identify charged particles over the momentum range 1-150 GeV/c to provide an efficient kaon tagging facility. The system consists of two parts. RICH1 is located directly behind the vertex detector with a silica aerogel and a C_4F_{10} gas radiator. RICH2 uses CF_4 as active material. Hybrid photodiodes with pixel readout or multi-anode photomultipliers are foreseen to be used as Cherenkov photon detectors.

The calorimeter consist of three main parts. Its main task is to identify electrons and hadrons for triggering and offline analysis. A pre-shower detector is used to provide an accurate detection of pions. This detector is built from 14 mm thick lead plates followed by square scintillators that are read out by phototubes or photodiodes. The pre-shower detector is followed by an electromagnetic calorimeter (ECAL), which

provides an identification and energy measurement of electrons and photons and a track reconstruction of pions. The ECAL is built from 2 mm thick lead plates, also followed by scintillators. The hadronic calorimeter (HCAL) has to provide a single-particle transverse-energy measurement of high p_T hadrons and it improves the separation between electrons and hadrons at high energies. The HCAL is constructed from scintillator tiles, that are embedded in an iron structure. The scintillators are read out by photomultipliers.

The muon detector is used for muon identification and to provide Level-0 trigger information. As detecting material, Multigap Resistive Plate Chambers and Cathode Pad Chambers are embedded in an iron filter. The detector uses pad readout instead of strip readout to enable a fast trigger response.

It is planned to start building the detector in autumn 2003 and have it ready for operation together with the operation of LHC in the middle of 2005.

1.2 The Vertex Detector System of LHCb

The important tasks of the vertex detector are the location of the primary interaction point of the colliding protons and the detection of the so-called secondary vertex, which is the decay point of a produced B-Meson, with a resolution in the order of 6 μm to 10 μm. The efficiency in achieving this goal depends on the performance of the subsystems of the vertex detector as well as on their interaction.

1.2.1 Layout of the Vertex Detector

The geometrical layout of the detector arrangement is driven by the need of a precise measurement of the track coordinates close to the interaction region. For that purpose, silicon microstrip detectors have been chosen. In order to achieve an efficient reconstruction of secondary vertices originating from B-mesons in the Level-1 trigger, the layout of the detector strips has been optimized to that task. A strip pattern has been chosen, where detector discs with strips of constant radius alternate with detector discs with radial strips. The layout of the strips for a segment of a detector disc is sketched in fig.1.2 (left). Each disc is divided into six sectors, where the sensitive area of each sector covers 61° in ϕ-direction and 1 cm to 6 cm in radius. The arrangement of the detector stations can be seen in fig.1.2 (right). This configuration assures that practically all particles measured in the downstream spectrometer cross at least three stations. The detector modules are split into two halves, so that they can be retracted during beam injection. To allow an overlap of the two detector halves, the upper and lower modules are displaced by 2 cm along the z-direction.

1.2.2 Pile-Up Veto Counter

Two additional stations upstream of the LHCb detector act as a veto counter in the Level-0 trigger (see also sect. 1.3). Their geometrical layout is identical to the vertex

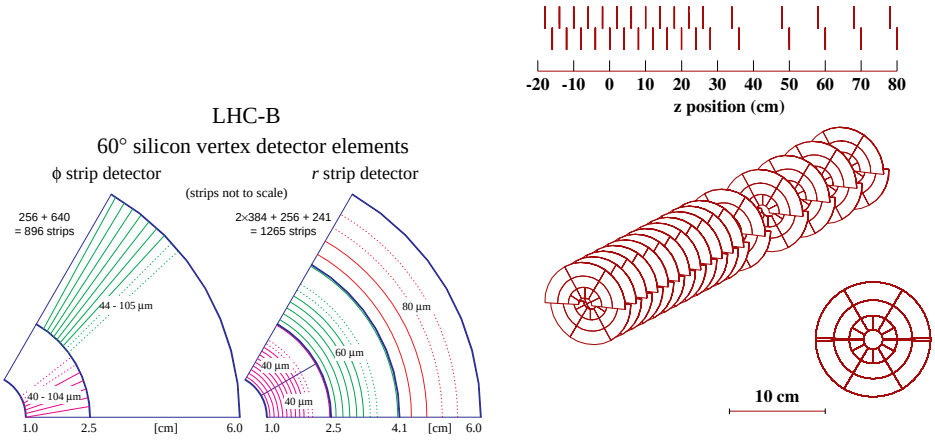

Figure 1.2: Schematic showing the strip layout of the ϕ- and r-measuring detectors (left) and the arrangement of the detector stations in the vertex detector (right)

detector stations, but the requirements on the frontend readout electronics are different. The signals from the veto counter detectors are amplified, discriminated and immediately read out. The discriminated signals are processed and in case of more than one interaction per bunch crossing, a veto for the Level-0 trigger is given. Simulations show that this method rejects 80% of the double interactions while retaining 95% of single interactions.

1.2.3 Silicon Detectors

The use of silicon microstrip detectors in vertex detectors has many advantages. Due to the small amount of energy that is needed to create an electron-hole pair in silicon and the high absorption (compared for example to gases), it is possible to detect ionizing radiation in a very efficient way. Semiconductor processing is a well developed technology for the production of integrated circuits and can be used to produce silicon detectors with high precision at relatively low prices.

Creation of the Charge

When a charged particle traverses silicon, the mean energy loss per path length can be described by the Bethe-Bloch formula [6]

$$\frac{dE}{dx} = 0.1535 \frac{MeVc^2}{g} \rho \frac{Z}{A} \frac{z^2}{\beta^2} \left[\ln \left(\frac{2m_e \gamma^2 v^2 W_{max}}{I^2} \right) - 2\beta^2 - \delta - 2\frac{C}{Z} \right] \qquad (1.1)$$

where

m_e is the electron mass

I is the effective ionization potential averaged over all electrons

Z is the atomic number of the medium

A is the atomic weight of the medium
ρ is the density of the medium
z is the charge of a traversing particle
$\beta = v/c$ with v=particle's speed, c=speed of light
$\gamma = \frac{1}{\sqrt{1-\beta^2}}$
δ is a density correction
C is a shell correction
W_{max} is the maximum energy transfer in a single collision

The charged particles scatter inelastically with electrons in the valence band and loose part of their energy, which in turn lifts the electrons into the conduction band. As can be seen from eq 1.1, for low energies, the energy loss rate is proportional to $\frac{1}{v^2}$. At a certain energy, a minimum ionization rate is reached and for relativistic energies, the energy loss rate rises logarithmically. In principle, the average energy loss for a sample of silicon of a defined thickness can be obtained by integrating the Bethe-Bloch formula. However, due to statistical fluctuations of the average energy loss, this is not possible in an analytical way. The distribution of the average value for thin targets follows a Landau-distribution, which is an asymmetrical distribution with a long tail for high energies. The high energetic tail is due to occasional energy transfers to single electrons (δ-electrons), which then travel several 10 μm through the sample, causing further ionization [6]. A phenomenological approach to obtain a value for the mean energy loss is the use of the Fano factor (F). The variance in the number of electrons N that are created by the charged particle passing through the semiconductor is then

$$\left\langle \Delta N^2 \right\rangle = F \cdot N \tag{1.2}$$

For silicon, the Fano factor is about 0.115 [6]. From eq. 1.1 and eq. 1.2, the most likely energy loss for a minimum ionizing particle can be determined, which is about 83 electron/hole pairs per μm or 295$\frac{eV}{\mu m}$. For a 150 μm thick detector material (as planned for the vertex detector of LHCb) the most likely charge signal amounts to 12,500 electrons.

Silicon Strip Detectors

Position-sensitive particle detectors need to provide information on the position where the particle has passed through the detector. In order to collect the charge created by the particle before electrons and holes recombine, a reverse bias voltage needs to be applied to the detection material. Silicon strip detectors provide an efficient method to obtain a position and energy-loss measurement. In silicon strip detectors, strip-like diodes are used to separate the produced electrons and holes. In principle, the diodes are realized as strip-like arranged p^+ layers (highly doped with donator atoms) on an n^- substrate (moderately to sparsely doped with acceptor atoms). The strip diodes are reverse biased and the width of the depletion region (which is the insulating region, where the intrinsic charge carriers are swept away due to the presence of an electric field) depends on the voltage applied to the strip diode. In normal operation of the strip diode as a particle detector, it is very desirable to have the complete substrate fully depleted so that a maximum of material is sensitive to ionizing radiation. Traversing

ionizing particles produce an ionization cloud which in turn produces a current pulse on the p^+ strip of typically 3ns width and 30 to 50pA height, depending on the thickness of the detector material [6]. If the amount of charge and the location of the strip is known, this technique provides a means of energy-loss and position measurement. The precision of the position measurement depends on the strip pitch and the method of readout. If only the digital information (strip is hit or not) is used to localize a particle track, the precision of the position measurement is given by

$$\left\langle \Delta x^2 \right\rangle = \frac{1}{p} \int_{-p/2}^{+p/2} x^2 dx = \frac{p^2}{12}$$

where
x is the measured position
p is the strip pitch

For a typical strip pitch of 50μm, this leads to a resolution of ~15μm. The resolution can be improved significantly, if the analog information of the strip signal is used. This allows to measure the height of the charge on more than one strip and therefore the center of gravity of the charge signals gives a more precise position information than in the case of digital readout. In the case of analog readout, the precision only depends on the noise performance of the readout electronics and on the readout pitch p as

$$\Delta x \approx \frac{p}{S/N}$$

where S/N is the signal-to-noise ratio of the readout electronics. In the case of the LHCb vertex detector, where a S/N value of about 12 is targeted, a resolution (neglecting multiple scattering and inclined tracks) of about 4μm can be reached. However, in a real detector the resolution is limited by the multiple scattering in the material between the vertex and the points of the track measurements.

1.2.4 Radiation Fluxes in the Vertex Detector [1]

During the operation of the LHCb experiment, the detector modules of the vertex detector are placed 1 cm away from the colliding beams. As a consequence, the detectors and the readout electronics will suffer from radiation damage. The particle fluxes decrease approximately with $1/r^2$ (qith r being the distance to the beam axis). Figure 1.3 shows the simulated flux of 1 MeV equivalent neutrons/cm^2 (normalized to 1 interaction in the interaction point) as a function of the distance to the beam axis. The equivalent neutron flux is obtained by using the non-ionizing energy loss (NIEL) of the different particles. It is expected that the NIEL is the major source of radiation damage of the silicon detectors [8].

1.2.5 Electronic Readout

The total number of detector strips of the complete vertex detector amounts to about 220,000. The strip pitch varies between $40\mu m$ and $60\mu m$ and the detectors are housed

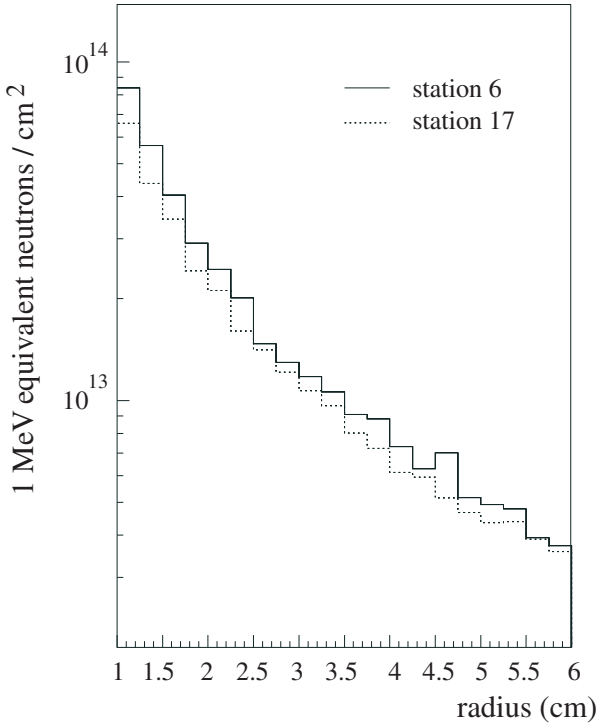

Figure 1.3: Total particle flux (normalized to 1 interaction in the interaction point) at two stations of the Vertex Detector as a function of the distance to the beam axis [1]

in a vacuum tank, which excludes individual feed-throughs of all channels . Custom made VLSI electronics is needed for an application with such requirements. A chip suitable for readout of the LHCb vertex detector has to be capable of amplifying the detector signals within a time window compatible to the bunch crossing frequency of the LHC of 40 MHz. In order to reduce the number of feedthroughs in the vacuum tank, a certain multiplexing ratio of the signals has to be fulfilled by the readout chip. Due to constrained power consumption (the readout chips are operated in vacuum and have to be cooled), the readout frequency is limited. As a consequence, not all events that are amplified can be read out and only events that are marked by the Level-0 trigger will be sent out of the detector. This requires a ring buffer, which stores the events. This architecture of the readout chip is similar in other experiments, like HERA-B, ATLAS and CMS. A more detailed treatment of several approaches follows in chapter 3. A comparison of analog and binary readout in the vertex detector can be found in [9].

The basic specifications of the LHCb vertex detector are summarized in table 1.1. They determine the requirements on the readout chip given in tab. 1.2. It should be noted that these numbers are a compromise between efficency in achieving the task of measuring decay vertices, technical limits, reliability of the system and financial limitations.

Total number of detector channels	220.000
Strip pitch at detector	40μm to 60μm
Sampling frequency	40 MHz
Level-0 trigger delay	$4\mu s$
Level-0 trigger rate	1 MHz
Maximum number of events pending to be read out	16
S/N at a given load capacitance	> 14 at 10 pF
irradiation dose at readout chip position	2 Mrad/year
charge created by 1 minimum ionizing particle	11,000 electrons
required linearity	better than 5% for $\pm$10 MIP
signal remainder at 25 ns after peaking time	< 30%
acceptance of consecutive triggers	yes
operation time	5 years

Table 1.1: Basic specifications of the LHCb vertex detector, determining the requirements on a readout chip for the LHCb vertex detector.

Input amplifier	preamplifier and pulse shaper peaking time 25 ns, signal remainder after 25 ns of <30%
Total noise	< 785 electrons at input load of 10 pF
dynamic range	-110,000 electrons to +110,000 electrons
Number of channels	128
Channel pitch	50 μm
Sampling frequency	40 MHz
Readout time per event	< 900 ns
Latency	160 clock cycles
Size of multievent buffer	16
Total ionizing dose	10 Mrad
Acceptance of consecutive triggers	yes
Generation of trigger for pile-up veto	yes
Deadtimeless readout	yes
Power consumption	< 4 mW / channel

Table 1.2: Requirements on a readout chip for LHCb

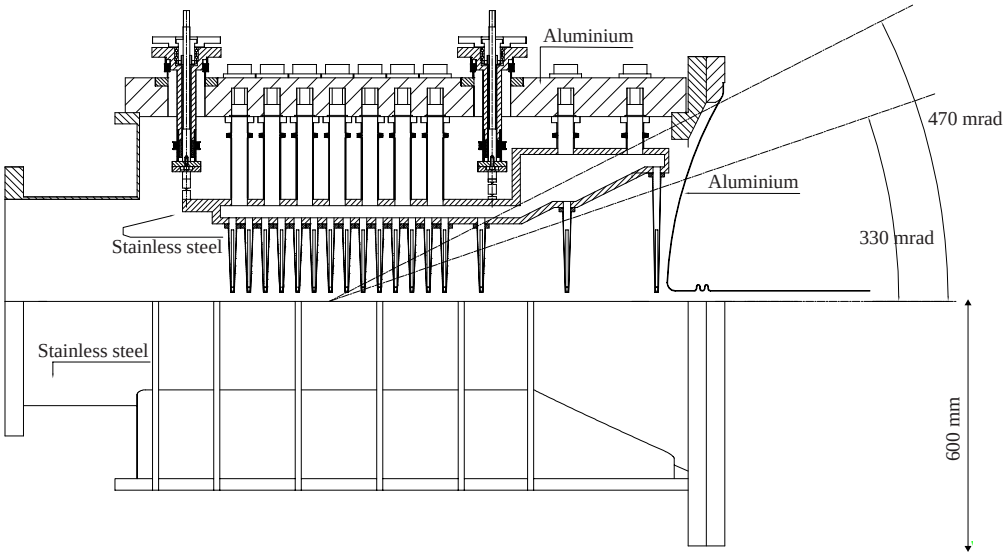

Figure 1.4: Side view of the vacuum tank and support structures for the Vertex Detector. The bottom half is omitted.

1.2.6 Mechanical Support Structure

The mechanical support of the detector modules is an important component in achieving the required performance of the vertex detector system. The detector modules have to be mounted so that they can be retracted from their nominal operating position during beam injection and be aligned to a precision of better than $50\mu m$. The detectors also have to be protected from RF-pickup from the proton beams, and in general the total mass in the acceptance region of the detector has to be kept to a minimum. These requirements can be satisfied by a roman pot system. The detector stations are mounted on aluminum support frames via linear actuators. The modules are housed in a secondary vacuum, which is separated from the vacuum of the beam pipe via a $100\mu m$ thick aluminum foil. The aluminum foil is folded around the detector modules in order to keep the effective thickness for the particles at a minimum. Figure 1.4 shows a preliminary sketch of the mechanical arrangement of the detector modules and their supporting structures [5].

1.3 The Trigger System of LHCb

The task of the trigger system is to reduce the amount of data during the operation of the detector and after the data have been acquired. Different trigger levels allow a very coarse filtering at an early stage of data acquisition with rather simple algorithms and a detailed decision with more complicated algorithms at later stages. In LHCb, the trigger decision is taken in four steps:

Level-0 trigger uses the information of the muon detector, the electromagnetic and the hadronic calorimeter. It searches for particles with a high transverse momentum

(high p_T-trigger), which are typical for events with B-Mesons. Events with more than one interaction are suppressed by the pile-up veto counter. The information of the Level-0 calorimeter trigger, the Level-0 muon trigger and the Level-0 pile-up veto trigger are combined by the Level-o decision unit. The complete Level-0 trigger will be implemented as hardware. The latency of the Level-0 trigger is fixed to $4\mu s$. The targeted average trigger rate of Level-0 is 1 MHz, which yields a suppression factor of 40.

Level-1 trigger uses the information of the vertex detector to select events with one or more secondary vertices (vertex trigger). In addition, the tracks of the high p_T-trigger are confirmed. The information of the Level-1 vertex trigger and the Level-1 track trigger are combined by the Level-1 decision unit. The complete Level-1 trigger will be implemented as a software solution. The date will be processed on a processor farm consisting of commercial CPUs. The suppression factor of Level-1 is 25 and the latency is variable up to a value of $256\mu s$.

Level-2 trigger eliminates events with a false secondary vertex. It uses the momentum information of the tracks and correlates the information of the tracking stations with the tracks from the vertex detector. The latency of the Level-2 trigger depends on the installed computing power and it is currently planned to achieve an average latency of $10ms$ with a suppression factor of 8.

Level-3 trigger partially and fully reconstructs the final states to select events with specific decay modes. It uses the complete detector information. The suppression factor at Level-3 is about 25 with a latency of about 200 ms. The data output rate is therefore $\approx$200 Hz.

Bibliography

[1] J. H. Christenson et.al, Phys. Rev. Lett. 13 (1964) 138

[2] B. P. Povh, Teilchen und Kerne, Springer Verlag 1994

[3] The BaBar Homepage, http://www.slac.stanford.edu/BFROOT

[4] The Homepage of the BELLE experiment, http://bsunsrv1.kek.jp/

[5] LHCb Technical Proposal, CERN/LHCC 98-4

[6] G. Lutz, Semiconductor Radiation Detectors, Springer Heidelberg 1999

[7] V. Talanov, Radiation Environment at the LHCb Vertex Detector Area, CERN, LHCb 98-019/TRAC

[8] G. Casse, et.al, Study of Radiation Induced Charge Trapping in Irradiated Silicon Detectors, internal LHCb note

[9] P. Koppenburg, Comparison of analog and binary readout in the silicon strips vertex detector of LHCb, LHCB/97-020

Chapter 2

Radiation Damage in Microelectronic Devices

The long-term performance of electronic devices in an environment like the LHCb vertex detector strongly depends on the amount of damage that is caused to the electronic devices by high energy particles. Since no technology or device yet exists that is completely immune to ionizing and non-ionizing radiation, the decision for a certain technology has to keep the radiation defects, failures and the consequences in mind, so as to judge their acceptability. This chapter summarizes most of the known defects due to ionizing and non-ionizing radiation in microelectronic integrated circuits. It shows up consequences and ways to overcome them.

2.1 Introduction

The composition of particle and electromagnetic radiation in a hadron-collider experiment like LHCb can be understood quite well, although the quantitative distributions of the particle types and their energies are difficult to calculate. Detailed distributions of radiative background as seen by electronic devices that are located inside the detector area or inside the detector itself are therefore usually obtained by means of simulations. The simulation takes into account the real distributions of material inside the detector and the distributions and spectra of incident particles. Simulations done for a proposed configuration of the LHCb vertex detector as from 1998 can be found in [1]. Since the configuration of the complete detector system and therefore the exact location of the front-end electronics was not frozen at that time, these calculations can not be regarded as final.

Section 2.2 explains the radiation defects of MOS field effect transistors and describes methods to enhance the radiation tolerance of such devices. In section 2.3, radiation defects of bipolar transistors are presented.

2.2 Radiation Damage in MOS Field Effect Transistors

The use of MOS (metal-oxide-semiconductor) microelectronics in radiation environments recently became popular, mainly due to the wide availability of MOS processes and their low price, compared to bipolar processes. Also, the improvement and development of MOS processes with reduced leakage currents and lower spread of process parameters, yields high performance devices that are available in standard processes. The availability of CMOS processes with structure sizes of less than 0.35 μm and the expected radiation tolerance of them has further driven the use of standard CMOS as a basis for the development of radiation hard microelectronics.

2.2.1 Total Ionizing Dose Effects

Trapped Charges in Silicon Oxide [12, 3]

Cumulative radiation effects occur during the complete lifetime of a transistor as long as it is exposed to radiation. As long as the dose applied to a transistor is known and the device is characterized completely, the failure or degradation in performance of the device can be predicted.

In a radiation environment like in the LHCb detector, the total ionizing dose effects are caused by electrons and by charged hadrons. γ-particles and neutrons do not ionize directly, but they can induce ionization by energy deposition. The most sensitive part of a MOS field effect transistor is the thin gate oxide, which separates the gate from the active channel. If an electron-hole pair is created in the silicon dioxide of the gate oxide, they do not recombine immediately due to the insulation characteristics of the gate oxide. If no electrical field is applied to the gate of the transistor, both the electron and the hole stay in the gate oxide and recombine eventually by diffusion. By application of an electrical field, the electrons can leave the gate oxide faster than the holes due to their higher mobility in silicon dioxide. The result is a growth of the hole concentration in the gate oxide. From this simple consideration, it is already obvious, that total ionizing dose effects on field effect transistors are smaller if there is no electrical field applied to the gate. As a consequence, transistors should always be set under operating condition during irradiation tests to ensure a realistic damage due to total ionizing dose. The effects on the performance of trapped charges in the gate oxide depends on the type of transistor. In an nMOS transistor, the field of the positive trapped charges (holes) adds towards the one, that is generated by supplying a positive voltage to the gate. In a pMOS transistor, the trapped holes 'shield' the electrical field that is generated by a usually, negative voltage applied to the gate.

The mechanism of the generation, distribution and trapping of the holes and electrons in the silicon oxide is visualized in fig.2.1. The amount of charge generated in the oxide depends on the energy that is deposited by the ionizing particle. The energy needed to create one electron/hole pair is about 3.6 eV [2]. The time scale for a recombination process of the electron/hole pair is given by the mobility of the electrons in SiO_2, which is much higher than the mobility of holes. At room temperature, the mobility of electrons is about $20 cm^2 V^{-1} s^{-1}$ and its velocity saturates at about $10^7 cm/s$ for high

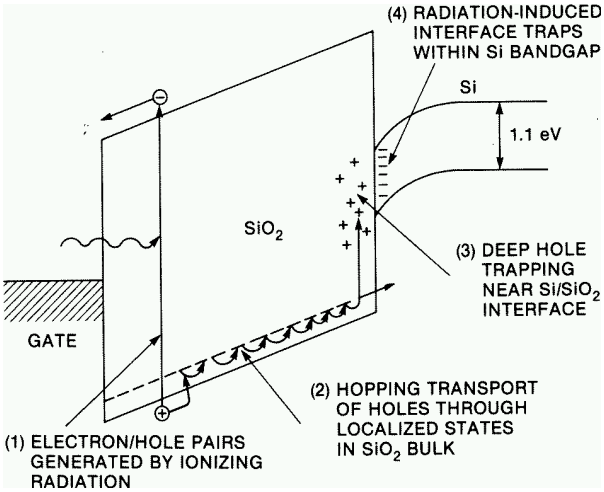

Figure 2.1: Charge generation, distribution and trapping due to ionizing radiation in silicon oxide (e.g. in the thin oxide of a MOS transistor) [3]

electrical fields [3]. The mobility of holes is always much lower than the electron mobility and typical values are $10^{-4} - 10^{-11} cm^2 V^{-1} s^{-1}$, depending on the temperature and the electrical field. Therefore, if an electrical field is applied to the silicon oxide, the electrons immediately leave the oxide in the order of 1ps. Any recombination process is limited by this time constant. Even without electrical field, any recombination takes places in several picoseconds due to the large difference in mobility of electrons and holes. The remaining holes, that do not initially recombine, follow any electrical field in the SiO_2-layer towards the negative electrode through a relatively slow transport mechanism, where they are collected or captured in deep trapping sites. The transport mechanism is highly dispersive and the time constants can vary over many decades. A satisfying model for the transport of holes in SiO_2 is given by the CTRW (continuous-time random walk) model [3]. A detailed understanding of this transport mechanism is especially important for the modeling of short-term (order of seconds at room temperature) post-irradiation response of a transistor. The long term radiation-induced behavior of a transistor is dominated by the holes trapped near the SiO_2/Si interface. Figure 2.2 shows a schematic of trapped charges in the gate oxide of a MOS transistor. The shift in the gate bias voltage ΔV_{th} originating from the trapped holes in the gate oxide can be written as

$$\Delta V_{th} = -\frac{q}{\varepsilon_{ox}} d_{ox} \Delta N_{ot} \tag{2.1}$$

where ΔN_{ot} is the charge density per unit area in the SiO_2/Si interface and d_{ox} the thickness of the oxide. ΔN_{ot} can be written as

$$\Delta N_{ot} = \frac{1}{d_{ox}} \int_0^{d_{ox}} n_{ht}(x) x dx$$

where n_{ht} is the local density of trapped holes. An analytical expression for 2.1 can be obtained by simplifying the assumptions on the distribution of the trapped holes. If the

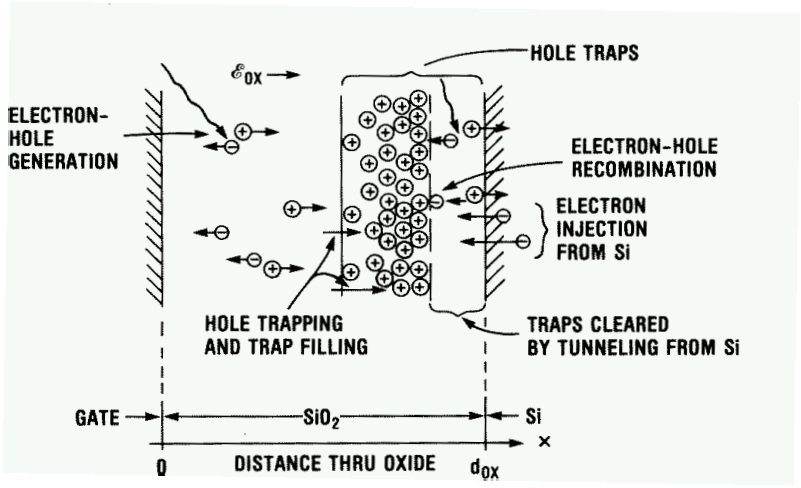

Figure 2.2: Schematic of hole trapping and removal in a MOS transistor at a positive gate voltage [3]

distribution of the holes is assumed to be a single layer with an average density $\bar{N}_{ht}$ and a thickness ΔX (small compared to d_{ox}) and the recombination of trapped holes with electrons is neglected, the fraction of radiation-generated holes that are trapped in the oxide is

$$f_T(E_{ox}) = \sigma_{ht}(E_{ox})\bar{N}_{ht}\Delta X$$

where E_{ox} is the effective electrical field in the oxide and σ_{ht} is the local cross-section of the hole traps for capturing holes. The total number of holes trapped in the oxide is then

$$F_h(E_{ox}, E) = d_{ox}[K_g(E)f_y(E_{ox}, E)D(E)]$$

where K_g denotes the charge generation coefficient, f_y the fractional free charge yield and D the total dose of ionizing radiation that has passed through the oxide. The energy of the ionizing radiation is E. Using $\Delta N_{ot}(E_{ox}, E) = F_h f_T$, one can deduce

$$\Delta V_{th}(E_{ox}, E) = -\frac{q}{\varepsilon_{ox}} K_g(E) f_y(E_{ox}, E) f_T(E_{ox}) d_{ox}^2 D(E) \qquad (2.2)$$

This expression [3] contains the important relationship $\Delta V_{th} \propto d_{ox}^2$ of the shift in the gate voltage due to trapped charges as a function of the gate oxide thickness for a fixed radiation dose.

The limit of this model can be deduced from fig.2.2 if the thickness is small enough so that the tunneling processes between the gate oxide and the silicon (or the gate, respectively) becomes dominant. Several measurements have shown a much faster decay of the threshold voltage shift with decreasing thickness when d_{ox} falls below 20 nm. Figure2.3 shows the shift in the flat-band voltage per unit dose as a function of the oxide thickness for various MOS structures. For thicker oxides, the curve follows the the d_{ox}^2 dependency as shown in 2.2. For a thickness below approximately 20 nm, the shift in the flat-band voltage drops much faster [4, 5]. This effect greatly favors the use of MOS technology with thin gate oxides to prevent a shift in the effective gate voltage.

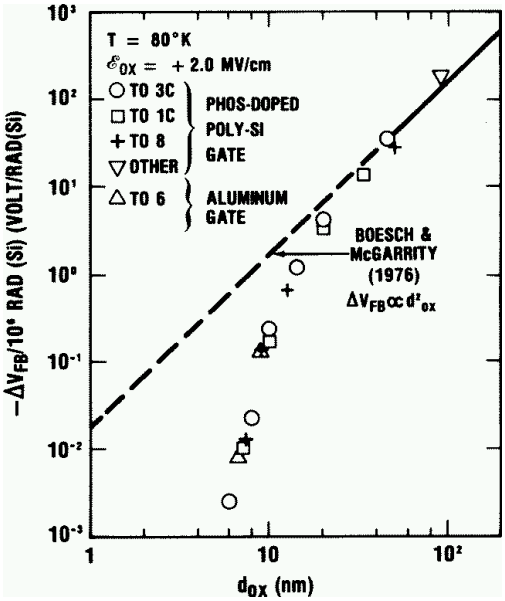

Figure 2.3: Shift of the flat-band voltage as function of oxide thickness for various MOS structures. The solid/dashed line shows the expected d_{ox}^2 dependency for thicker oxides [3].

Another part of transistors, namely nMOS transistors, that is sensitive to trapped charges due to ionizing particles is the so called lateral oxide. Also known as field oxide, this part separates the active channel region from the surrounding bulk material. In an nMOS transistor, trapped holes in the field oxide at the interface to the silicon can create a leakage path from the source to the drain of the transistor. The size of this parasitic leakage current rises with the total amount of holes trapped in the lateral oxide, and therefore with the total ionizing dose. The absolute size of the leakage current depends on the detailed geometrical structure of the lateral oxide and can hardly be predicted. Since the thickness of the field oxide is much larger than 20 nm, a tunneling of trapped holes from the field oxide into the silicon is excluded. Figure 2.4 shows a sketch of an nMOS transistor and the region where this parasitic leakage current occurs. Several methods to prevent the rise in leakage current in nMOS transistors due to ionizing radiation have been reported [6], but most of them use different processing steps and manufacturing techniques to achieve the goal. Only recently, when processes with structure sizes of 0.35μm and less became available, the approach of closed gate structures has become an alternative. By using a gate geometry as pointed out in fig. 2.5, any leakage path between source and drain under the field oxide of the transistor is avoided [3]. This method has the price of a larger area for transistors. Also, modeling of the effective behavior of enclosed transistors is not as well established as for standard transistor geometries. A description of models for enclosed transistors is given in chapter 4.

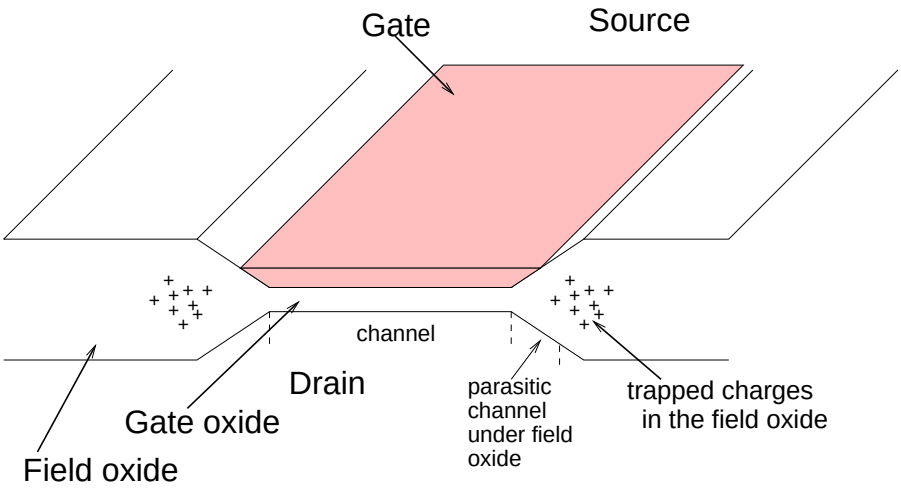

Figure 2.4: Schematic view of the parasitic channel due to trapped holes in the field oxide of an nMOS transistor which gives rise to a parasitic leakage current

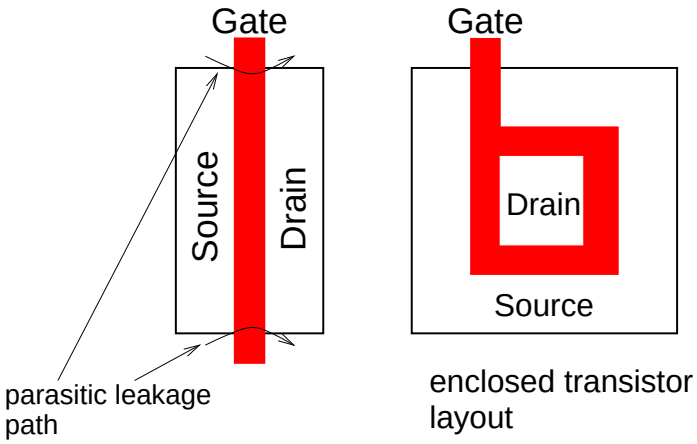

Figure 2.5: The parasitic leakage path between source and drain as in a regularly laid out transistor (left) is prevented in an enclosed transistor layout (right)

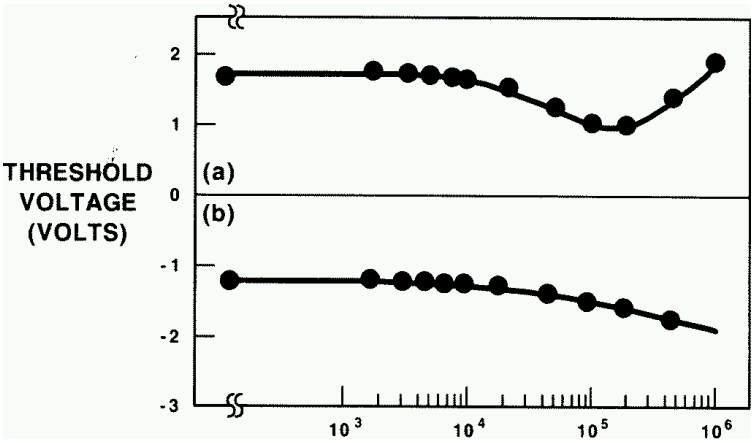

Figure 2.6: Threshold voltage of an nMOS transistor (a) and a pMOS transistor (b) as a function of the total ionizing dose [3]

Interface States at the Si/SiO_2 Interface

Defects in the crystal lattice that are formed at the interface of the gate oxide and the silicon are called interface states. These interface states trap charges from the channel of the transistor and therefore cause a shift of the threshold voltage and a degradation of the mobility of the charge carriers in the channel of the transistor. Interface states (also known as interface traps or surface traps) are electronic energy levels between the conduction and the valence band. They arise from lattice mismatch at the interface, disconnected bonds or impurities, and they can be induced by ionizing radiation. As for parasitic leakage currents under the field oxide induced by trapped charges, the total effect of interface states on the behavior of a MOS transistor cannot be predicted. It depends on processing details, gate material and on doping concentrations. The nature of the charge states in the silicon bandgap can be divided into acceptors and donors, where energy levels in the upper half of the bandgap are occupied by acceptors and energy levels in the lower half by donors [3]. The net charge that is captured in the interface states can be positive, negative or zero. However, by applying a voltage to the gate of a MOS transistor, the interface states move along with the valence and conduction band relative to the Fermi level. The effects of interface states on the behavior of transistors are manifest in three important effects:

Shift in the threshold voltage: Due to the additional charges captured in the interface states, the effective gate voltage is shifted which results in a shift of the threshold voltage. For pMOS and nMOS transistors, this effect is different: in pMOS transistors, positively charged donor interface traps reduce the threshold voltage, whereas in nMOS transistors, the threshold voltage is increased due to negatively charged acceptor interface traps. These effects, together with the threshold voltage shift due to trapped charges in the gate oxide, cause the typical shape of the threshold voltage as a function of the total ionizing dose shown in fig. 2.6. Since for pMOS transistors, the threshold voltage is decreased by the trapped charges and the interface states, the overall slope is negative. On the other hand, nMOS transistors show a decrease in th

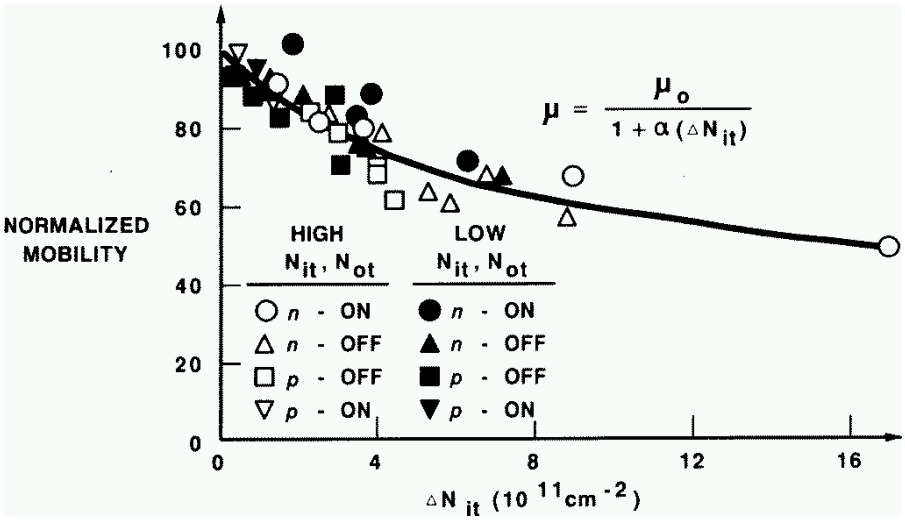

Figure 2.7: Normalized channel mobility as function of radiation induced interface trap density [8]

threshold voltage for a dose below 100 krad (typically). This decrease is caused by positive trapped charges in the gate oxide, which are dominating at low doses. Above 100krad, the increase of the threshold voltage is caused by negatively charged interface traps. In the literature, this effect is often called ”rebound”.

Degradation in the transconductance $g_m = \partial I_d / \partial V_g|_{V_d = const}$ **:** The transconductance of a transistor is reduced by radiation induced interface traps. This can be understood by considering a gate voltage sweep, which empties or fills the interface states at the gate surface. These additional states in turn modify the electrical field, so that more or less respectively, charge is needed on the gate to create the desired electrical field in the channel of the transistor. Also, the voltage needed to bring the transistor into strong inversion is higher. This affects (amongst others) the switching speed in digital circuits.

Change of the mobility: The mobility μ of charge carriers in the channel of a transistor is reduced by radiation induced interface traps. A model for the mobility degradation is given in [7, 8]. The mobility of the charge carriers is reduced by an increased amount of lattice and Coulomb scattering due to the interface states. The empirical relationship for the degradation of the channel mobility is

$$\mu = \frac{\mu_0}{1 + \alpha(\Delta N_{it})} \tag{2.3}$$

where μ_o is the mobility before irradiation, N_{it} is the concentration of trapped interface charges and the fitting parameter $\alpha = (8 \pm 2) \times 10^{-13} cm^2$. Figure 2.7 shows a comparison between measurement values and the empirical model 2.3.

Dynamics of Trapped Charges and Interface States

The dynamics of the two effects, trapped charges in the gate oxide and interface states, is very different. Trapped charges can easily be removed from the gate oxide by elevated temperatures. The holes are then thermally activated and de-trapped from their locations. In contrast, annealing of interface states has not been observed at normal operating temperatures. Annealing at temperatures above 100° C seems to be very sensitive to the processing of the oxide. An activation energy of 1.4 eV has been reported for the annealing of interface states [10].

2.2.2 Single Event Effects (SEE)

In contrast to the cumulative total ionizing dose effects, single event effects are usually triggered by single particles that cross sensitive regions of the device. The actual effect occurs very localized, but can propagate over the complete system. Single event effects can be classified depending on their impact on the systems behavior:

- Transient effects: yield in asynchronous signals, that propagate through the system. They can appear in analog as well as in digital circuits and can become static if they are latched.
- Static effects: change the content of a memory cell and can be overwritten.
- Permanent effects: are destructive events that cannot be recovered. In some cases, they can be prevented by a fast current limitation.

In the applications of high energy physics experiments, SEE have become a concern during recent years. This is due to the use of deep sub-micron processes, where the intrinsic sensitivity due the smaller gate capacitances is significantly higher than for processes with structure sizes of 0.8 μm or larger.

Single Event Upset (SEU)

If an ionizing particle passes through a semiconductor, it loses its energy by ionization by creating pairs of electrons and positive charged holes. In the semiconductor bulk of a MOS transistor, the electrons and holes recombine immediately. This is not the case in the active region of a transistor and in its surroundings. The electron-hole pairs are separated by the electrical field and therefore give rise to a current spike. This current spike has two components: a fast component (less than a ps) which yields from the electron-hole pairs created directly in the depleted region of the transistor channel, and a second component which is rather slow (in the order of ns) and arises from regions outside of the active area. Depending on the energy loss of the ionizing particle, the current spike can be rather large, and the impact on the circuit, in which the transistor is used, cannot be neglected.

The amount of energy that is deposited in the active region of a transistor by an incident particle is usually expressed by the linear energy transfer (LET). The energy needed to

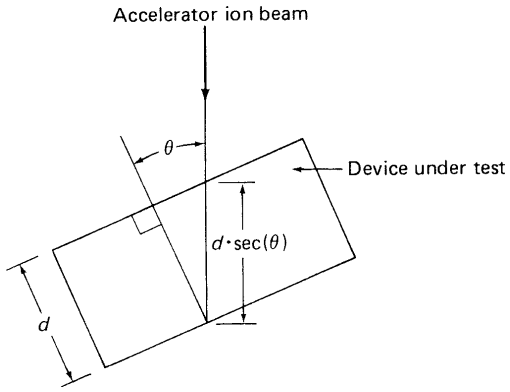

Figure 2.8: Schematic of the experimental setup for SEU cross section measurements [11]

create an electron/hole pair in silicon is 3.6 eV and that defines the amount of charge that is produced by the absorbed energy. Ions of the same energy have a higher linear energy transfer, the heavier they are. Particles, which do not have enough mass to create enough charge for a SEU to occur (like protons or α-particles) can initiate nuclear reactions, whose recoil products are heavy enough to produce direct ionization.

The minimum LET needed to change the state of a given device (e.g. a flip-flop) is called critical LET. The critical LET depends strongly on the fabrication process of the device that is considered, since it contains not only the amount of charge that is needed to upset a device, but also the detailed structure of the device, which in turn is responsible for the distribution of the charge.

The experimental determination of the critical LET is usually done by exposing the device to a beam of heavy ions (e.g. O, Li, Na). SEUs are then characterized using a beam threshold measure. The threshold is defined as the minimum perturbation that produces an error in the test device, and the upset cross-section is the maximum sensitivity to SEU caused by the ion beam. The upset cross section σ is then calculated as

$$\sigma = \frac{U}{\Phi t \cos\theta} \tag{2.4}$$

where U is the absolute number of upsets during the time interval t and Φ the particle fluence of incident ions per cm^2. The incident angle θ is visualized in fig. 2.8. During the measurement, the incident angle of the ion beam is varied across the device. As a result, the charge that is deposited into the sensitive region of the transistor is varied. Assuming the ion tracks inside the material are long compared to the active device geometry, the amount of charge deposited is proportional to $\frac{1}{\cos\theta}$. With that method, the LET can be varied over a limited range. By combining these results with results of different types of ions, a larger statistics can be gained. The data of different ions can be combined by assuming that two different ions with the same LET have the same impact on the device. This assumption is not exact, since the radial distribution of the charge that is created by an ion certainly depends on the type of ion. Also, the charge distribution along the track of the ion that penetrates the semiconductor

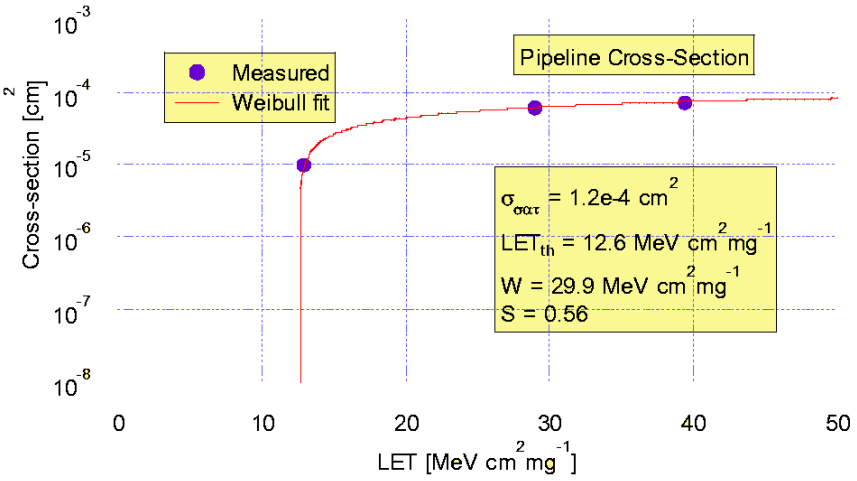

Figure 2.9: Typical result of an SEU cross-section measurement performed on the APV25 [9]

Ion Type	Si	Si	Cl	Ni	Br
LET [MeV cm^2 mg^{-1}]	9	10.4	12.9	30	39.4

Table 2.1: Ion types and according linear energy transfer for the measurement plotted in fig. 2.9

depends on the ion type. Nevertheless, combining results of different ion types is an accepted method to measure the critical LET. Figure 2.9 shows a typical result for a measurement of SEU cross-sections as a function of LET. This measurement was performed with different ion types of an energy of about 100 MeV as listed in tab. 2.1 The curve shows a steep rise in the SEU cross section at the critical LET value of 12.6 MeV cm^2 mg^{-1}. This measurement has been performed with the APV25 readout chip for CMS [9], which is manufactured in the same technology as the Beetle readout chip for LHCb (see chapt. 4).

Single Event Latchup (SEL)

SEL is a well studied phenomenon in CMOS devices since latchups can occur due to high temperatures, large transients on the power supply lines or due to wrong bias cycling of circuits. In a radiative environment, a latchup can also be induced by a heavily ionizing particle.

In general, an SEL occurs in a CMOS structure due to inherently present bipolar transistors, that can be completely turned on. The presence of these parasitic pnp- and npn structures can be seen in fig. 2.10. Of course, this model is only a simplification of the parasitic bipolar structures that are present. A more realistic model would need consideration of the three-dimensional distributions of the p-n junctions.

In this configuration, the collector of the npn-transistor is connected to the base of the pnp-transistor. In the case of an increasing collector current in the pnp-transistor

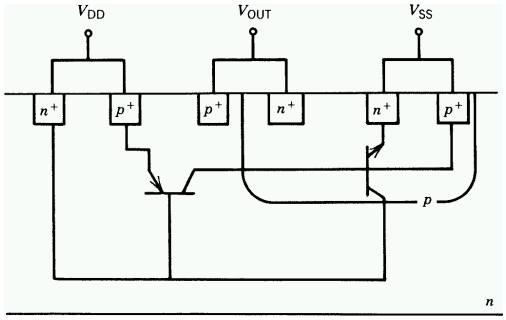

Figure 2.10: Parasitic pnp and npn bipolar structure in a CMOS inverter. The gates of the MOS-transistor are not shown for simplicity.

(which can happen due to a heavily ionizing particle crossing the active region of the device), the base current of the npn-transistor is also increasing, which in turn increases the collector current of it. The rising collector current of the npn-transistor then increases the base current of the pnp-transistor. This positive feedback leads to an almost shorted circuit between the power supplies if the gain of this parasitic pnpn-thyristor is high enough. This case is usually destructive for the adjacent MOS device, unless the current is limited e.g. by the resistance of the supply lines.

Effective methods to decrease the sensitivity of devices to latchup are possible by reducing the gain of the parasitic pnpn thyristor. This can be done by technological solutions, e.g. by using trench isolations between adjacent MOS structures, the gain of the lateral parasitic transistor is strongly reduced. Layout techniques can also reduce the resistance along the parasitic current path. The systematic use of guardrings around the wells of transistors ensures a low-ohmic connection of the lateral bipolar transistors and therefore effectively reduces the latchup sensitivity.

Single Event Gate Rupture (SEGR), Single Event Burnout (SEBO) Less often found in literature, SEGR and SEBO have been observed in power MOSFET applications. The threshold for these destructive events is higher than for SEU and therefore requires the presence of very heavily ionizing hadrons [12]. Due to their small likelihood, these effects are not considered to be a problem for the application of CMOS devices in a high energy physics experiment [12].

2.3 Radiation Damage in Bipolar Transistors [11, 16]

In bipolar transistors, radiation damage occurs exclusively in a cumulative way. The effects can be divided by the nature of interaction of the radiation with the devices. Damage caused due to ionizing energy loss is treated in 2.3.1, whereas the effects due to non-ionizing energy loss are described in 2.3.3.

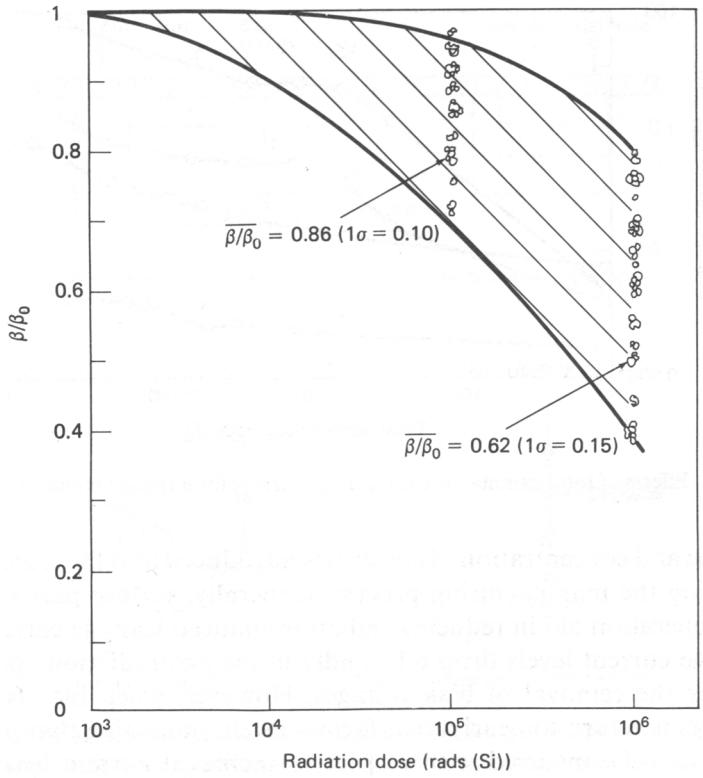

Figure 2.11: Measurement of the β-degradation with irradiation for different devices [11]

2.3.1 Total Ionizing Dose

The cumulative effects of ionizing particles in bipolar transistors are caused by trapped charges in oxide layers and by interface states. They are caused by the creation of a conductive channel under the thick oxide (field oxide) by inversion of the silicon due to trapped charges. There are three main locations where this may occur. First, a conductive channel can be opened between two buried layers. Second, a channel can be opened that creates a current path between collector and emitter at the sidewall oxide. Third, a current can flow under the thick oxide of the surface. These effects result in an increased base current, which in turn reduces the effective gain (β) of the bipolar transistor. The bipolar devices most sensitive to ionizing radiation are lateral PNP transistors, due to their relatively large interface to the field oxide. Figure 2.11 shows a measured β-degradation for different bipolar devices [11].

2.3.2 Dependency on the Dose Rate [11, 12]

For some bipolar processes, a variation of the damage with the dose rate has been reported, e.g. in [13]. In such cases, transistors (both NPN and PNP) show an excess base current by a factor of typically 10 to 20 more at a low dose rate (0.1 rad/s) compared to a high rate (1 krad/s). This effect is due to the formation of trapped charges

	Proton	Neutron	Electron
1 MeV		2	0.01
50 MeV	2		
1 GeV	1		0.1

Table 2.2: Relative displacement damage for various particles and energies [15]

in the oxide that is overlaying the emitter-base junction. The net trapped charge in that oxide is higher at lower dose rates.

A precise prediction on the dose rate behaviour of a bipolar device cannot be given. Therefore, an experimental characterization, usually performed at elevated temeperatures ($\approx 90^{\circ}$) to enhance the damage, is performed.

2.3.3 Displacement Damage

Particles or photons that cross silicon and transfer energy of about 20 eV to a silicon atom, can remove the atom from its lattice location. This displacement damage creates clusters of defects by changing the electronic structure around the displaced atom in the lattice. The total damage due to displacement is proportional to the non-ionizing energy loss (NIEL), which in turn depends on the particle type and its energy. The displacement damage is usually not measured by the absorbed energy, but measured by the effect on the device. For a particle of known energy, the NIEL in silicon can be calculated [14] and these distributions can be used to estimate the relative damage of radiation due to displacement. Table 2.2 shows a comparison for some particle types at different energies.

The damage in silicon due to displacement results in three different effects:

1. Mid-gap states are formed between the valence and the conduction band. Since the direct transition of electrons from the valence to the conduction band is rather unlikely, the formation of mid-gap states enables electrons from the valence band to jump to the conduction band via the mid-gap states. Theses electrons would then contribute to the current flow. On the other hand, mid-gap states can capture electrons from the conduction band which then capture holes from the valence band. This process would reduce the current flowing in the conduction band. Whether generation or recombination is dominating depends only on the relative concentration of carriers and empty defect states. Therefore, in a depletion region, where the conduction band is less populated, generation of charge carriers dominates and in a forward biased junction, where the conduction band is filled, recombination is the dominating effect.

2. The formation of states close to the conduction band allows trapping of charge carriers for a limited time. This effect reduces the charge carrier concentration in a forward biased region.

3. The effective doping characteristics is changed.

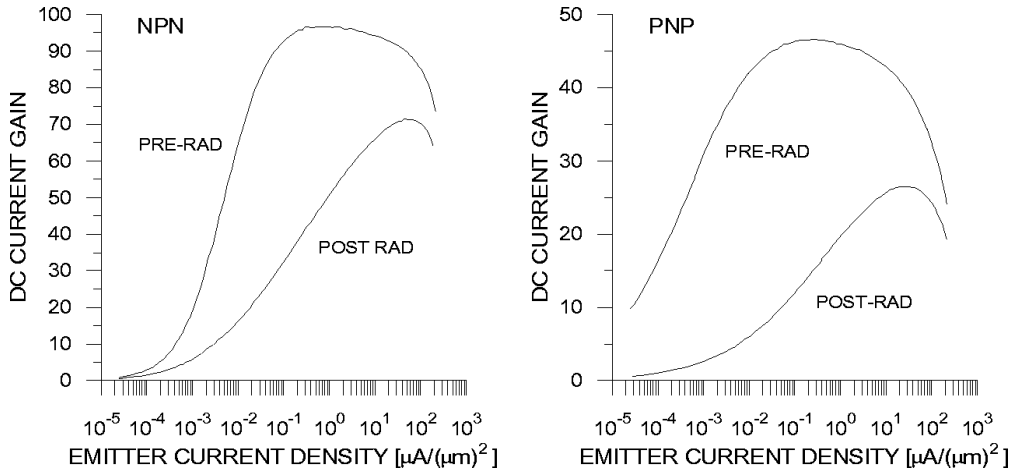

Figure 2.12: DC current gain of NPN and PNP transistors before and after irradiation with protons of 800 MeV to a fluence of $1.2 \cdot 10^{14} cm^{-2}$ [16].

As a result of the displacement damage, bipolar transistors suffer from a reduction in the DC current gain at low currents. As the base-emitter junction is usually forward biased, the base-emitter current density degrades. The relative degradation depends on the current density, since it is a function of the population density in the conduction band. Therefore, small devices suffer less from radiation damage due to displacement than large devices at a given collector current. Since the probability for a recombination depends on the transit time through the base, a reduced base width is favorable for radiation hardening. An example of a measured DC current gain reduction after irradiation with protons of 800 MeV is shown in fig. 2.12 [16].

Bibliography

[1] V. Talanov, Radiation Environment at the LHCb Vertex Detector Area, CERN, LHCb 98-019/TRAC

[2] G. A. Ausman, F. B. McLean, Electron-Hole Pair Creation Energy in SiO_2, Appl. Phys. Lett. 26, 173 (1975)

[3] T. P. Ma, P. V. Dressendorfer (Editors), Ionizing Radiation Effects in MOS Devices and Circuits, John Wiley & Sons 1989

[4] N. S. Saks, M. G. Ancona, J. A. Modolo, Radiation Effects in MOS Capacitors with very Thin Oxides at 80K, IEEE Trans. Nucl. Sci. NS-31, 1249 (1984)

[5] J. M. Benedetto, H. E. Boesch, F. B. McLean, J. P. Mize, Hole Removal in Thin-Gate MOSFETs by Tunneling, IEEE Trans. Nucl. Sci. NS-32, 3916 (1985)

[6] T. J. Sanders, CMOS Hardness Assurance Through Process Controls and Optimized Design Procedures, IEEE Trans. Nucl. Sci. NS-24(6), 2051 (1977)

[7] A. G. Stanley, Effect of Electron Irradiation on Carrier Mobilities in Inversion Layers of Insulated Gate Field Effect Transistors, IEEE Trans. Nucl. Sci. NS-14(6), 266-275 (1967)

[8] F. W. Sexton, J. R. Schwank, Correlation of Radiation Effects in Transistors and Integrated Circuits, IEEE Trans. Nucl. Sci. NS-32(6), 3975-3981 (1985)

[9] G. Hall et.al, Status of APV25, Front End Electronics for Future High Luminosity Trackers, Conference Contribution, Perugia 2000

[10] A. G. Sabnis, Characterization of Annealing of Co^{60} Gamma-ray Damage at the Si/SiO_2 Interface, IEEE Trans. Nucl. Sci. NS-30(6), 4094-4099 (1983)

[11] G. C. Messenger, M. S. Ash, The Effects of Radiation on Electronic Systems, Van Nostrand Reinhold, New York, 1986

[12] F. Faccio, Radiation Effects in the Electronics for CMS, Tutorial Script, CERN, 1999

[13] A. H. Johnston et.al, Total Dose Effects in Conventional Bipolar Transistors and Linear Integrated Circuits, IEEE Trans. Nucl. Science NS-41/6, 2427 (1994)

[14] A. Van Ginneken, Non Ionizing Energy Deposition in Silicon for Radiation Damage Studies, FNAL FN 522 (October, 1989)

[15] J. R. Srour et.al, Radiation Damage Coefficients for Silicon Depletion Regions, IEEE Trans. Nucl. Sci. NS26/6, 4784 (1979)

[16] I. Kipnis et.al, An Analog Front-End Bipolar Transistor IC for the SDC Silicon Tracker, IEEE Trans. Nucl. Sci. NS-41/4, 1095-1103 (1994)

Chapter 3

Different Approaches towards Microstrip Readout Chips

3.1 Overview

The development of high energy physics experiments at particle accelerators has lead to more advanced requirements on microelectronics over the years. High performance, radiation tolerance, reliability, availability of the technology and low cost are cornerstones in the development of frontend electronics for tracking detectors. Several approaches have been undertaken to develop readout chips for high energy physics experiments . This chapter gives an overview on important outcomes of this development and describes those readout chips in more detail, which have been dealt with in the context of this dissertation.

For experiments like ATLAS, CMS, HERA-B and LHCb, the general requirements on the analog readout electronics for microstrip tracking detectors are similar, they differ in parameters and technical details. The cornerstones of these requirements are listed in tab. 3.1. A concept that covers the requirements to a large extent has been developed by the RD-20 collaboration [1, 3]. The so-called RD-20 architecture of a readout chip is described in detail in chapter 4. It has been realized in several variations, the most important outcomes are listed in tab. 3.2.

The APV series of chips has been developed for the CMS experiment. Up to the version APV-6, these chips were manufactured in the radiation tolerant Harris AVLSIRA 0.8μm CMOS process, which lead to a radiation resistance level of 3 Mrad. The most recent version, the APV-25 has been ported to and manufactured in a 0.25 μm standard CMOS process and has been irradiated to a total ionizing dose of 10 Mrad without significant loss of performance.

The SCTA analog readout chip, manufactured in the radiation hard DMILL process, has been developed for the semiconductor tracker of the ATLAS experiment. It serves as an analog fallback for the case that the binary readout might not be feasible. The SCTA will be described in more detail in section 3.3.

The readout chip used for the silicon vertex detector and the microstrip gaseous chambers of the inner tracking system of HERA-B is the HELIX128. It is also used in the

Requirement on Readout Chip	Reason
Amplification and pulse-shaping of the detector signal	Detector signal of typically 10,000 to 30,000 electrons
Low-noise amplification	High efficiency required for detection of small signals
Intermediate storage of the detector signal	Cover the latency of an external trigger processor
Serial readout of some of the stored data	Reduction of number of signal transmission lines
Deadtimeless readout	Data taking should not be interrupted by readout of data
Integration of typically 128 channels on one readout chip	Readout pitch of typically 50 μm
Minimum power consumption	Cooling of the chips is usually difficult
Radiation tolerance	Employment of the chips in a radiative environment
Integration of bias generators and error detection	Simplification of the use in large systems

Table 3.1: General requirements on a microstrip detector readout chip

Readout Chip	Targeted Experiment / accelerator		Process Technology	References
APV 25	CMS	LHC	0.25 μm CMOS	[4]
SCTA	ATLAS	LHC	DMILL 0.8 BiCMOS	[5, 6]
HELIX 128	HERA-B, ZEUS, HERMES	HERA	0.8 μm CMOS	[7, 8]
Beetle	LHCb	LHC	0.25 μm CMOS	[9, 10]

Table 3.2: Important outcomes of the RD-20 concept: The readout chips are listed with the experiment, they have been developed for.

ZEUS and HERMES experiments and is manufactured in the AMS 0.8 μm CMOS process. It is described in more detail in section 3.2.

For the LHCb experiment, the analog readout chip Beetle is under development. This chip is also produced in a 0.25 μm standard CMOS process and its architecture is based on the HELIX128 readout chip. The Beetle is described in detail in chapter 4 and chapter 5.

A fundamental difference in the outcomes of the so-called RD-20 concept is the realization of the readout chips in different processes, which, amongst other impacts, leads to a different radiation tolerance level.

3.2 The HELIX128 Analog Readout Chip for HERA-B

3.2.1 Overview

The HELIX128 [7] readout chip has been developed for the HERA-B experiment [11] in the AMS 0.8 μm CMOS process. The use of the HELIX128 in the silicon vertex detector of HERA-B requires the readout chip to withstand a total ionizing dose of about 100 krad per year, if the readout chips are mounted $\approx$10 cm away from the beam axis.

The AMS 0.8 μm process features a gate oxide thickness of 16 nm and no special measures are taken by the manufacturer to achieve immunity to ionizing or non-ionizing radiation. As an estimate for the expected threshold voltage shift of this process, fig. 2.3 gives about $\frac{200mV}{100krad}$. Measurements reported in [12] show a change of the threshold voltage of about 150 mV after 130 krad total ionizing dose for pMOS transistors. Due to the relatively large minimum feature size of 0.8 μm and the rather high gate capacitance of minimum size devices, SEU effects are of no concern.

3.2.2 The Architecture of the HELIX128

The HELIX128 consists of 128 input channels with charge sensitive preamplifiers followed by active RC-CR shapers. The output of the shapers are sampled into an analog memory at the nominal operating frequency of 10 MHz, where the pulse height is stored for about 12 μs. Trigger signals can be derived from a comparator stage, which is located directly behind the pulse shaper. The charge that is stored in the analog memory is readout out via a charge sensitive pipeline readout amplifier upon an external trigger signal and the data of all 128 channels are multiplexed onto one transmission line at a frequency of 40 MHz. The control of all digital signals is provided by a completely synthesized digital block. The standard cells of this digital block use standard minimum size transistors. Other than the input transistors of the preamplifier, the shaper and the pipeline readout amplifier, all transistors are laid out in a regular way. The input transistors of the analog stages mentioned use a so-called waffle transistor layout, which has the advantage of a maximum W/L ratio per area. Figure 3.1 shows a schematic overview on the architecture of the HELIX128.

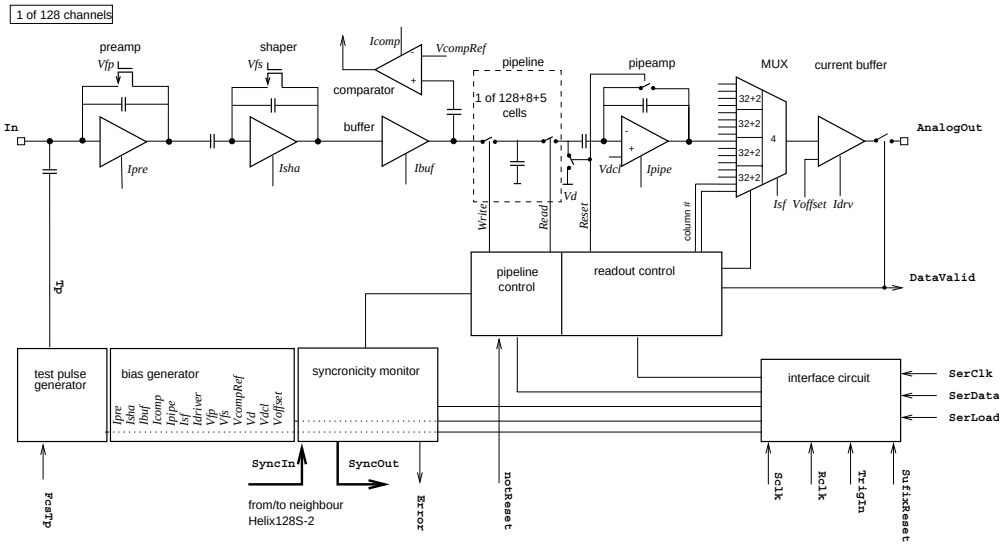

Figure 3.1: Schematic overview on the architecture of the HELIX128 analog readout chip

3.2.3 Radiation Tolerance of the HELIX128

The employment of the HELIX128 in the HERA-B vertex detector is possible, since the detector modules are replaced after one year of operation due to the radiation damage of the silicon detectors. From a technical point of view, it is not desirable to reuse the chips that once have been connected to the silicon detectors. This limits the total ionizing dose to 100 krad, the HELIX128 has to withstand.

To maintain the performance of the HELIX128 during its employment in the detector, all analog stages are biased by means of current mirrors. This technique prevents a change of the bias parameters of the amplifiers and buffers due to a shift in the threshold voltage. The increase in leakage current of the nMOS transistors leads to a higher power consumption of the chip. As mentioned above, the input transistors of some analog stages are layed out as waffle transistors. This layout prevents, like the enclosed transistor layout, a rise in leakage current of these nMOS transistors. The degradation of the transconductance of the transistors increase the noise and reduce the speed and gain of the analog stages. The digital control circuits suffer from an increased power consumption. They remain functional up to a total dose of about 500 krad. Detailed measurements on the radiation tolerance of the HELIX128 can be found in [12].

3.3 The SCTA Analog Readout Chip for ATLAS

The SCTA is one member of a family of readout chips that have been developed for the semiconductor tracker of the ATLAS experiment. Despite the readout architecture of the ATLAS experiment is based on binary readout, the SCTA is the fallback solution for the case that the binary readout will not be feasible.

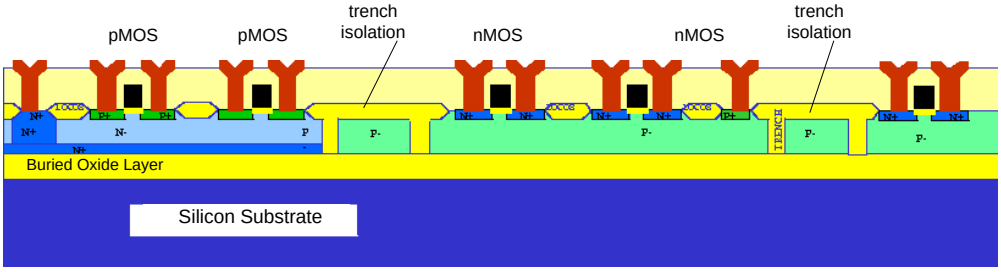

Figure 3.2: Cross section of DMILL MOS devices and trench isolations

The SCTA is developed and manufactured in the TEMIC 0.8 μm BiCMOS silicon-on-insulator (SOI) process (DMILL). The process is described in section 3.3.1 and special features of the SCTA are described in section 3.3.2.

3.3.1 The BiCMOS SOI DMILL Process [13]

The DMILL process is a dedicated process for applications, where radiation hardness is required. It combines bipolar and CMOS components in a silicon-on-insulator technology. This means, that the active substrate is kept very thin ($\approx$70 nm) and separated from the silicon carrier substrate by a thin oxide layer. The minimum feature size is 0.8 μm and the thickness of the gate oxide is 14 nm. Several processing steps and features are added, so that radiation damage effects are kept at a minimum.

- The combination of SOI and trench isolation reduces the latchup likelihood by cutting possible parasitic paths, which could be build up by strongly ionizing particles (cf. chapter 2).
- Due to the SOI technology, the active bulk volume underneath the MOS transistors is much smaller than in a standard bulk CMOS process. Therefore, the likelihood for SEU is reduced strongly, since only a small amount of the charge created can reach the transistor.
- Special processing steps are introduced to clean the surface between oxides and diffusion layers. These cleaning steps reduce the amount of trapped interface states.
- The use of high frequency plasma etching is reduced to avoid the generation of trapped interface states during the processing.

Figure 3.2 shows a cross section of DMILL nMOS and pMOS transistors. The trench isolations separate the transistors at different potentials from each other.

The specified radiation hardness of the DMILL process for total dose effects is 10 Mrad. This includes a threshold voltage shift of less than 100 mV for nMOS and pMOS transistors after 10 Mrad of total ionizing dose. A degradation of 20% in the current gain factor β of bipolar transistors after irradiation with 10^{14} neutrons/cm^2 is also specified by the manufacturer. Due to the technological precautions in this process, the SEU rate is very small.

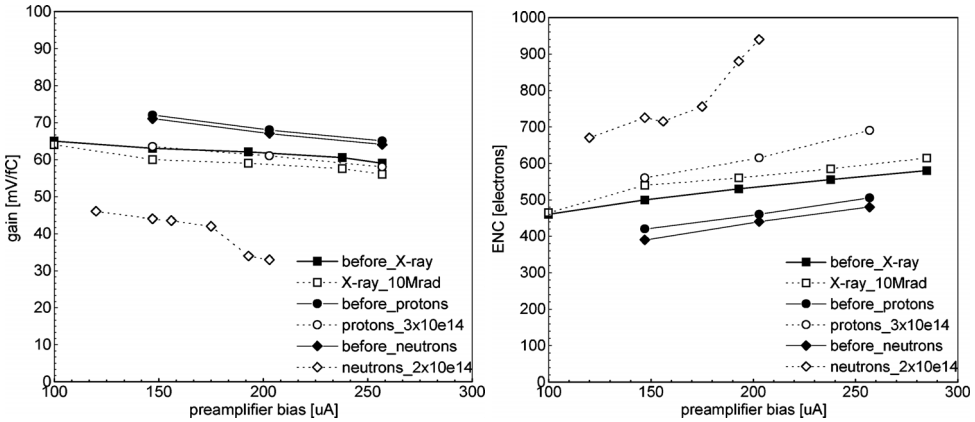

Figure 3.3: Gain (left) and noise (right) as a function of the preamplifier current of the ABCD frontend, which is identical to the SCTA frontend [14]

3.3.2 Architecture and Radiation Hardness of the SCTA

As already mentioned, the SCTA is also based on the RD-20 concept of readout chips. For simplification, only the differences to architecture of the HELIX128 are described here.

The input stage of the SCTA is completely realized in bipolar technology. Due to the smaller intrinsic capacitances and the large current gain of bipolar transistors, this allows to build fast amplifier at a relatively low power consumption. The latest version of the SCTA (SCTA-VG) uses a transresistance (current-)amplifier as preamplifier followed by an CR^2-RC^5 shaping stage. A concern in the input stage of the SCTA is the degradation of the β in the preamplifiers input transistor after irradiation. Figure 3.3 shows the gain (left) and the noise (right) as a function of the preamplifier bias current of the ABCD frontend (which is identical to the SCTA frontend) [14]. A relatively large drop of the gain and a rise in the noise figures is observed after neutron irradiation. Both, the decrease in gain and the increase in noise can be related to a drop of the input transistors β by about 50% after neutron irradiation. This is typical for bipolar transistors, since they are very sensitive to displacement damage.

All nMOS transistors are laid out in an edgeless design to prevent an increased leakage current after irradiation. The resulting space consumption is compensated in the pipeline control block by a completely hand-made layout of the pipeline control.

Currently, a modified version of the SCTA is under development (SCTA-VELO) which suits the requirements of the LHCb experiment. Mainly, the readout-time per event of the SCTA-VELO is faster than in the original SCTA. Parts of the SCTA-VELO have been developed within this thesis.

Bibliography

[1] R. Brenner, et.al, Performance of a LHC front-end running at 67 MHz, Nuclear Instruments and Methods A339 (1994) 477-484

[2] R. Brenner, et.al, Design and Performance of an analog delay and buffer chip for use with silicon strip detectors at LHC, Nuclear Instruments and Methods A339 (1994) 564-569

[3] R. Horisberger, et.al, A novel readout chip for silicon strip detectors with analog pipeline and digitally controlled analog signal processing, Nuclear Instruments and Methods A326 (1993) 92-99

[4] L. Jones, et.al, The APV25 Deep Submicron Readout Chip for CMS Detectors, Proceedings of the Fifth Workshop on Electronics for LHC Experiments, Snowmass, Colorado, CERN/LHCC/99-33

[5] J. Kaplon, et.al, DMILL Implementation of the Analogue Readout Architecture for Position Sensitive Detectors at LHC Experiments. Proceedings of the Fourth Workshop on Electronics for LHC Experiments, Rome, CERN/LHCC/98-36

[6] J. Kaplon, SCTA128 & SCTA128HC, Internal Note, Feb. 1998, CERN

[7] M. Feuerstack-Raible, et.al, Helix128 Users Manual, HD-ASIC

[8] W. Fallot-Burghardt, A CMOS Mixed-Signal Readout Chip for the Microstrip Detectors of HERA-B, Dissertation, Universität Heidelberg, 1998

[9] E. Sexauer, et.al, Design of a Prototype Frontend and Bias Generator for a new Readout Chip for LHCb, Proceedings of the Fifth Workshop on Electronics for LHC Experiments, Snowmass, CERN/LHCC/99-27

[10] D. Baumeister, et.al, Design of a Readout Chip for LHCb, Proceedings of the Sixth Workshop on Electronics for LHC Experiments, Cracow, CERN/LHCC/2000-??

[11] The HERA-B collaboration, HERA-B Technical Design Report, DESY-PRC 95/01 (1995)

[12] U. Trunk, Radiation Tolerance of the HELIX128-2.2 Readout Chip for the HERA-B Microstrip Detectors, Ph.D. Thesis in preparation, Heidelberg 2000

[13] T. Corbiere, DMILL for HEP, Space and Automotive, Frontend Electronics Meeting, Perugia 2000

[14] W. Dabrowski, et.al, Radiation Hardness of the ABCD Chip for Binary Readout of Silicon Strip Detectors in the ATLAS Semiconductor Tracker, Proceedings of the Fifth Workshop on Electronics for LHC Experiments, Snowmass, CERN/LHCC/99-27

Chapter 4

Beetle - A Readout Chip suitable for LHCb

4.1 Overview

The Beetle readout chip has been developed to fulfill the requirements of the LHCb vertex detector. Some features have been integrated so that it could also be used to read out multi-anode photomultiplier tubes, which are one technology option for for the RICH detector of LHCb. The first version of the chip (Beetle 1.0) has been manufactured in a 0.25 μm CMOS process. The chosen process features one polysilicon layer, 3 metal layers and a gate oxide thickness of 6.2 nm. The development of the Beetle profited immensely from the experience with the HELIX128 readout chip for HERA-B [1]. The architecture of both chips is based on the RD-20 concept [2, 3], with adaption to specific constraints set by the application in the LHCb and HERA-B experiments.

The Beetle integrates 128 channels of an analog input stage, which consists of a charge sensitive preamplifier and a pulse shaping stage. The output of each analog input stage is sampled into an analog ring buffer (pipeline), realized by gate capacitances. The charge from the pipeline can be retrieved by a charge sensitive amplifier (pipeline readout amplifier), triggered by an external signal. An analog multiplexer serializes the amplified pipeline values and they are brought off-chip via a current driver. Behind each input stage, a comparator with adjustable threshold is located. The (binary) output of the comparator can be brought off-chip in two ways. One way is to build the logical OR of four neighboring channels. This value is multiplexed by a factor of two and brought off-chip via LVDS drivers. The other way is to sample the digital output into the pipeline and read it out via the pipeline/multiplexer data path.

The switches for sampling the analog frontend output into the pipeline (write-switch) and to read out the sampled value from the pipeline (read switch) are controlled by a completely synthesized standard cell logic block. The write-switches are closed one after the other at the nominal LHC bunch crossing frequency of 40 MHz. Incoming trigger signals that mark events to be read out, store the corresponding pipeline location in a FIFO and the corresponding columns of the pipeline are protected from being

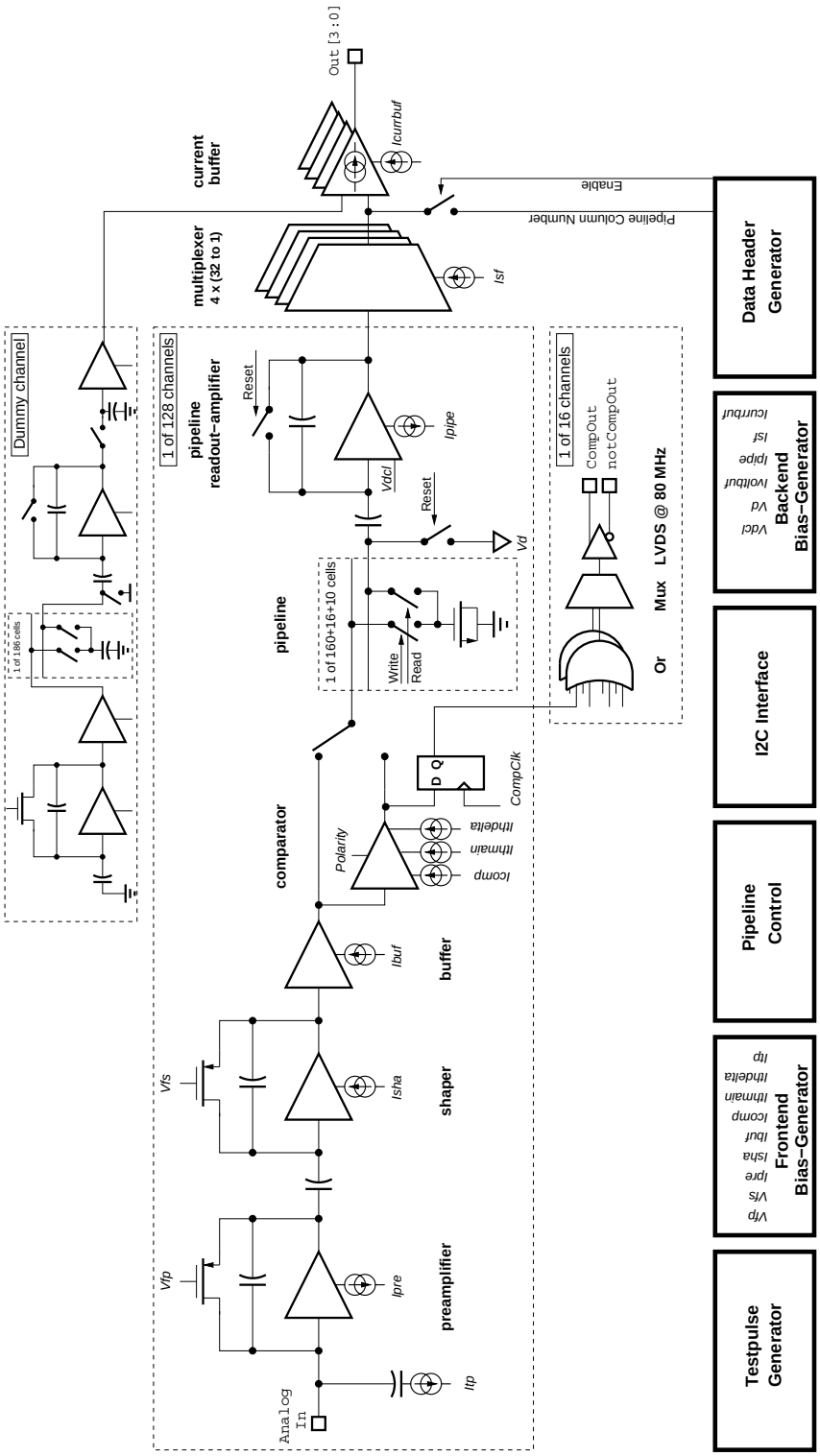

Figure 4.1: Block diagram of the Beetle-1.0 readout chip

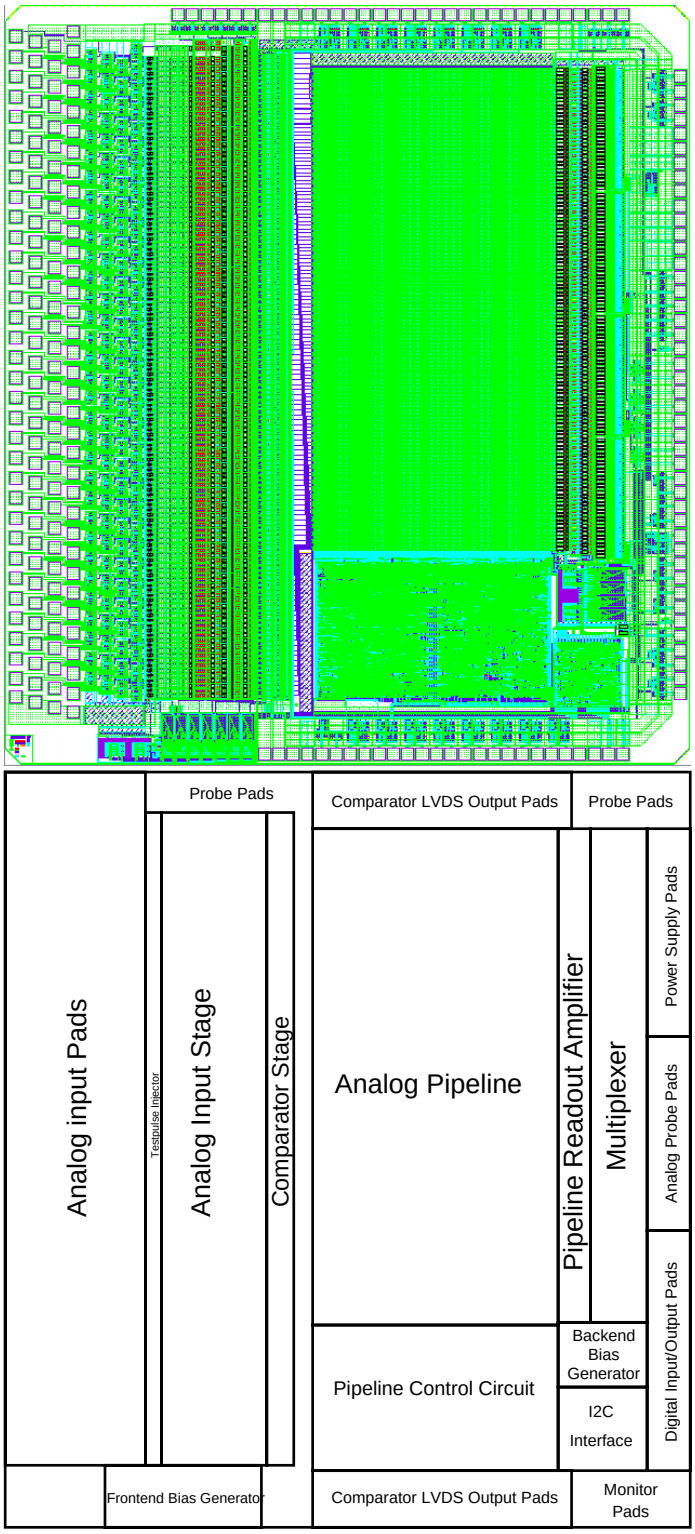

Figure 4.2: Layout (top) and floorplan (bottom) of the Beetle-1.0 readout chip.

overwritten. Pending events in the FIFO are read out by asserting the control signals to the pipeline readout amplifier and the multiplexer. The serialized analog signal is brought off-chip by a current buffer. To operate, the control logic needs the sampling clock, the trigger signal to start the readout of a stored event, and a reset signal for startup. The sampling clock is synchronous to the LHC bunch crossing frequency and the trigger signal is received from the L0-trigger. All bias currents and voltages that are needed to operate the analog stages of the readout chip are generated on the chip. All of them are programmable by a standard I^2C interface. The block diagram of all components is shown in fig. 4.1.

The floorplan of the Beetle readout chip as well as the layout is depicted in fig. 4.2. The layout is described from left to right. The first block consists of the 128 analog input pads. The fourfold staggered layout of the pads leads to an effective pitch on the chip of 40.42 μm and enables to read out strip detectors of a pitch of $50\mu m$ without using a fanout. The size of each pad is $105 \times 105\mu m^2$. ESD protection diodes are integrated into the block of the input pads. The testpulse injection block for each channel is located behind the input pads. The next block is the analog input stage, consisting of the preamplifier (with the clearly visible, large input transistor), the shaper and the buffering source follower. The comparator circuit is located directly behind the analog frontend. The digital comparator outputs are fanned to the top and bottom side of the chip, where the LVDS output driver pads are located. The analog output of the comparator (which is actually the analog frontend output fed through the comparator) is routed to the pipeline. The pipeline consists out of 129×186 cells, each cell including a gate capacitor for the charge storage and a write- and read switch to connect the pipeline cell with the frontend or the pipeline readout amplifier, respectively. Below the pipeline, the pipeline control logic block is located. The pipeline is followed by the pipeline readout amplifier, the multiplexer and the backside pads, which also contain the analog current buffer. The bias generators for the frontend- and the backend components are located below the frontend and below the pipeline readout amplifier/multiplexer, respectively. The I^2C interface is located in the bottom right of the chip. The overall size of the chip is $6.1 \times 5.5mm^2$. A list and description of all bond pads is given in Appendix A.

4.2 Comparison with other Readout Chips

To show the need for a special readout chip for the experiment LHCb, tab. 4.1 shows a comparison between the Beetle-1.0 and other available readout chips. For several reasons, none of the existing readout chips can be used for the LHCb vertex detector. The HELIX128 is excluded due to the slow shaping time of the frontend and the sampling frequency of 10 MHz. The existing SCTA cannot be used due to the latency, which is not sufficient. However, a modified version of the SCTA is under development which will feature a higher latency and a shorter readout time. The APV-25 cannot be used due to the shaping time of the input stage. Finally, a readout chip for the LHCb vertex detector needs to be able to generate fast trigger signals, derived from hit channels.

	Beetle-1.0	HELIX128	SCTA	APV-25
radiation tolerance	> 10 Mrad	≈400 krad	10 Mrad	> 10 Mrad
Latency	160	128	112	156
Multievent buffer	16	8	8	10
Deadtimeless readout	yes	yes	yes	yes
Sampling clock	40 MHz	10 MHz	40 MHz	40 MHz
Readout clock	40 MHz	40 MHz	40 MHz	40 MHz
Peaking time of frontend	25 ns	100 ns	25 ns	50 ns
Fall time	≈40 ns	≈100ns	≈60 ns	≈100ns
Noise	303e+ 33.6e/pF	400e+39.4e/pF	720e+33e/pF	246e+36e/pF
Power consumption [mW/ch.]	2.5	2.2	2	1.8
Trigger signal output	yes	yes	no	no

Table 4.1: Comparison between the Beetle-1.0 and the readout chips HELIX128, SCTA and APV-25 [4]

4.3 Radiation Hardening Methods

Measurements in the frame of the RD49 collaboration have shown very good radiation hardness of standard CMOS processes, if certain design rules are applied [10, 13]. Different test structures and complete designs have been irradiated up to a total ionizing dose of 30 MRad and no significant loss in performance of transistors and circuits has been observed. Single event effects on test structures and complete circuits can not be neglected and the impact of these effects on the performance of the Beetle readout chip has to be measured. The following list summarizes the considerations that are implemented in the design of the Beetle 1.0.

- Choice of technology: the use of a standard CMOS process with a feature size of 0.25 μm and a gate oxide thickness of 6.2 nm ensures enough resistivity against a threshold voltage shift due to ionizing radiation (see. chapter 2). Irradiation of test transistors have shown a threshold shift of 35 mV for nMOS transistors and 30 mV for pMOS transistors after a total dose of 10 Mrad [12]. However, since the chosen process is designated for digital designs by the vendor, analog stages have to be designed carefully with respect to process parameter spread.

- Enclosed layout of nMOS transistors: all nMOS transistors are laid out with an enclosed geometry. The extraction of effective geometrical parameters of the transistors follow rules, described in [10]. As well as the larger area consumption of enclosed transistors, extra parasitic capacitances appear. Measurements of leakage currents with enclosed transistors show no change in the leakage current even after a total dose of 30 Mrad [12]. It should be mentioned that the waffle transistor layout, which is often used if a large W/L ratio is needed, is also an enclosed transistor layout, since source and drain are completely separated by the gate.

- Systematic use of guardrings: all nMOS transistors are surrounded by guardrings to prevent inter-device leakage currents. Substrate contacts ensure a low-ohmic

connection of the substrate to the ground level to reduce the latchup sensitivity.

- Biasing of analog stages using current mirrors compensates even for minimal shifts in the threshold voltage

4.4 Layout and Device Parameter Extraction of Enclosed Transistors

The use of an enclosed transistor layout to prevent a rise in the leakage current due to irradiation of the device has three main drawbacks:

- The extraction of the device parameters of enclosed transistors is usually not supported by the manufacturer of the process and therefore, a special means of modeling of the devices has to be established.
- Due to the different perimeter of source and drain of an enclosed transistor layout, the output conductance and the drain capacitance depends on whether the inner or outer diffusion is chosen as drain of the transistor.
- The area consumption of an enclosed transistor is larger than of the regularly laid out transistor, if the aspect ratio (W/L) is comparable.

4.4.1 Extraction of the Aspect Ratio

To extract the aspect ratio of an enclosed transistor, a method has been developed in [11]. The aspect ratio adds up from three contributions [13]:

$$\left(\frac{W}{L}\right)_{eff} = \underbrace{4\frac{2\alpha}{\ln\frac{d'}{d'-2\alpha L_{eff}}}}_{1} + \underbrace{2K\frac{1-\alpha}{1.13\cdot\ln\frac{1}{\alpha}}}_{2} + \underbrace{3\frac{\frac{d-d'}{2}}{L_{eff}}}_{3} \tag{4.1}$$

where
$(W/L)_{eff}$ is the effective aspect ratio of the enclosed transistor
L is the drawn length of the transistor
d, d‘, c are the geometrical values of the transistor as depicted in fig.
K=3.5 for short channels (L$\leq$0.5μm), K=4 for L>0.5μm
α=0.05 (fitting parameter)

The first part of expression 4.1 takes into account the contribution by the linear edges in between the corners. The second part describes the contribution of the triangles between the linear parts and the rectangular corner parts (as depicted in fig. 4.3), and the third part describes the the rectangular corner part. This formula describes the aspect ratio of enclosed transistors to a precision of 6% for short transistors and to an even better precision for long transistors.

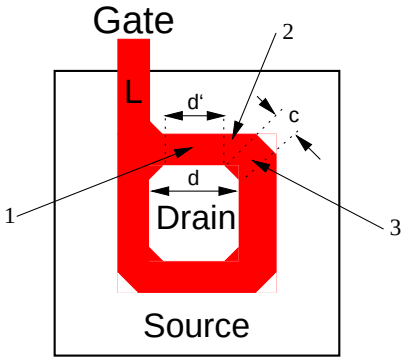

Figure 4.3: Geometry of an enclosed transistor. The numbers point to the corresponding parts in eq. 4.1.

4.4.2 Output Conductance

Due to the non-symmetric length of drain and source in an enclosed transistor layout, the output conductance depends on the location of the source/drain. Also, the drain and source overlap capacitance\s depend on whether the inner or outer diffusion is chosen as drain (or source, respectively). Measurements reported in [12] show an output conductance of 25% if the drain is chosen as the outer diffusion with respect to the inner diffusion for long transistors (L=5 μm). For short transistors (L=0.28 μm), this value is about 80%. In the application of enclosed transistors, the output conductance and the overlap capacitance have to be traded off against each other.

4.5 Analog Stages

Within the context of this dissertation, the analog frontend, the multiplexer and several smaller analog components have been designed. However, in order to give a complete description of the analog data path, all analog stages are described.

4.5.1 Frontend

Preamplifier

The preamplifier integrates the current signal delivered by the silicon strip detector. Its design is crucial for the performance of the analog data path, since the noise is dominated by the input device. Apart from the low noise requirement, the charge sensitive preamplifier must have a sufficiently fast rise time. This is especially important, since the following shaping stage would lose significant amount of gain if the time constant of the preamplifier was larger than the time constant of the shaping stage. The fast rise time requires an amplifier core with a high gain-bandwidth. Also, a means of reset needs to be provided so that the integrator does not charge up by subsequent integration events. General requirements like stability and power consumption have to be kept in mind.

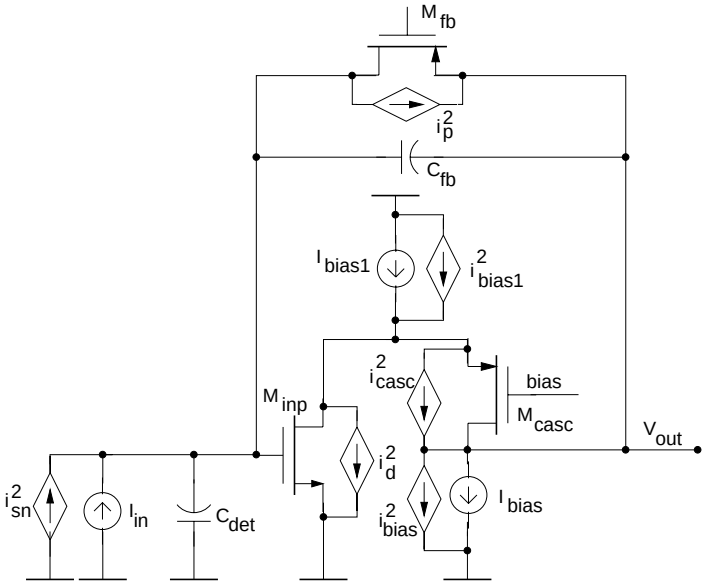

Figure 4.4: Noise sources in the charge sensitive preamplifier

The noise optimization of the charge-sensitive preamplifier can be reduced to the input transistor of the amplifier. This is due to the quadratic contribution of the noise currents of the amplifiers transistors. Since the drain noise current of a MOS-transistor is given by $i_{d,n}^2 = \frac{8}{3}kT\frac{g_m}{(1+\eta)}$, the transconductance g_m of the (usually very large) input transistor dominates (with $\eta = \frac{g_{mb}}{g_m}$). To achieve a minimum equivalent noise charge (ENC), the design parameters of the input transistor have to be chosen for given detector capacitance. These design parameters are in general the geometry of the device (width W and length L) and drain current through the device. The noise sources that are relevant in the application treated in this context are the serial contributions from the channel thermal noise of the input transistor, the $1/f$-noise, the parallel contributions of the detector leakage current and the feedback current due to the resistive feedback in the preamplifier. In general, the shaping stage that follows the preamplifier influences the overall noise of the analog frontend due to its bandpass characteristics. However, since the requirements on the shaping stage are given by system aspects (the pulse width has to be within a window of the order of $25ns$), the stress in this section is put on the optimization of the preamplifiers input transistor.

The noise performance of the input transistor can be fully characterized by an equivalent input voltage noise v_i^2 and an input current noise i_i^2. A schematic of the noise sources in the circuit is shown in fig. 4.4. The equivalent input noise voltage and current are [6, 7]

$$v_{ia}^2 = \frac{K_f}{C_{ox}^2 WLf} + \frac{8}{3}kT\frac{(1+\eta)}{g_m} + \frac{i_{sn}^2}{g_m^2} + \frac{i_p^2}{g_m^2} \quad (4.2)$$

$$i_{ia}^2 = |j\omega C_{in}|^2 v_{ia}^2 \quad (4.3)$$

where

K_f is the flicker noise coefficient

C_{ox} is the specific oxide capacitance
W, L is the width/length of the transistor
k is Boltzmann's constant
T is the absolute temperature
g_m is the transconductance of the transistor
η is $\frac{g_{mb}}{g_m}$with g_{mb}=bulk-source transconductance

The different terms in eq. 4.2 can be explained as follows: the first term represents the $1/f$-noise of the input MOS transistor, the second term represents the contribution due to the channel thermal noise. The last two terms are the noise currents due to the detector shot noise and the parallel noise contribution due to the resistive feedback of the amplifier. It should be noted that both the noise voltage and the noise current are given in units per frequency $[V^2/Hz]$ and $[A^2/Hz]$, respectively. In order to get an equivalent noise voltage referred to the input signal, the absolute input noise voltage is transfered to

$$v_{equi}^2 = \left[\frac{(C_{gate} + C_{in} + C_{fb})}{C_{fb}}\right]^2 v_{ia}^2 \tag{4.4}$$

where C_{gate} is the total gate capacitance of the input transistor, C_{in} the total input load capacitance (e.g. due to the detector) and C_{fb} the feedback capacitance of the charge sensitive amplifier. Considering only the channel thermal noise contribution for the time being, it can be concluded, that a minimum noise can be achieved. Calculating the minimum of eq. 4.4 with respect to W (where $C_{gate} = WLC_{ox}$) gives an optimum at a width of [5]

$$W_{opt} = \frac{C_{in} + C_{fb}}{2C_{ox}L} \tag{4.5}$$

The relationship 4.5 can be rewritten as

$$C_{gate} = \frac{C_{in} + C_f}{3} \approx \frac{C_{in}}{3} \tag{4.6}$$

As will be explained later, this rule for the optimization of the input transistor geometry is independent of the design of the following shaper stage. To find an optimum geometry that minimizes the $1/f$-noise contribution, eq. 4.4 is now examined while ignoring the noise contributions other than the $1/f$-noise. Minimizing the noise voltage v_{ia}^2 yields an optimum gate area [5]

$$(WL)_{opt} = \frac{3(C_{in} + C_{fb})}{2C_{ox}} \tag{4.7}$$

For a fixed transistor length L, the optimum width W of the input transistor for an optimum $1/f$-noise is three times larger than for an optimum channel thermal noise. As in the case of eq. 4.5, the optimum geometry is independent of any frequency and therefore independent of the time constants of a following shaper stage.

To study the transient characteristics of the charge sensitive preamplifier in the time domain, the response of the amplifier on a δ-like current pulse must be calculated. The small signal model used for this considerations is depicted in fig. 4.5. The feedback of the amplifier (generally a transconductance g_m with an output impedance of

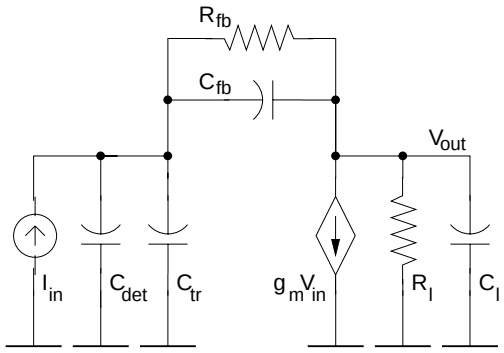

Figure 4.5: Small signal model of the preamplifier

$R_{load}||C_{load}$) is provided by the capacitor C_{fb} and the large resistor R_{fb}. The transfer function in the frequency domain is then given by:

$$\frac{V_{out}(s)}{I_{in}(s)} = \frac{g_m}{\frac{g_m}{R_{fb}} + sg_mC_{fb} + s^2C_{input}(C_{fb} + C_{load})} \tag{4.8}$$

where $C_{input} = C_{det} + C_{tr} + C_{fb}$ and the assumptions of a high output impedance ($g_mR_l \gg 1$) and a high feedback resistance ($g_mR_{fb} \gg 1$) are made. This second order transfer function (with respect to s) contains two poles, which are spread widely in this application. The two poles are:

$$p_1 = \frac{1}{2\pi\tau_1} = \frac{1}{2\pi R_{fb}C_{fb}} \tag{4.9}$$

$$p_2 = \frac{1}{2\pi\tau_2} = \frac{g_mC_{fb}}{2\pi C_{input}(C_{fb} + C_{load})} = GBW\frac{C_{fb}}{C_{input}} \tag{4.10}$$

where GBW denotes the gain-bandwidth product of the amplifier cell. The first pole is produced by the large feedback time constant $R_{fb}C_{fb}$ and determines the continuous reset time of the integrator. The second pole is the result of the capacitive feedback. Its position is very important for the stability of the amplifier. The response of the charge sensitive amplifier in the time domain $V_{out}(t)$ can be obtained by taking the inverse Laplace transformation of the transfer function in the frequency domain. The input current signal is therefore approximated by a δ-like current pulse with a total charge of Q. The output signal in the time domain then becomes:

$$V_{out}(t) = \frac{Q\tau_1}{C_f(\tau_1 - \tau_2)}(e^{-\frac{t}{\tau_1}} - e^{-\frac{t}{\tau_2}}) \tag{4.11}$$

As already assumed earlier, τ_2 is much smaller than τ_1. Therefore $V_{out}(t)$ represents an exponentially rising step signal with a slowly decaying tail. A simulation result using the chosen design values is plotted in fig. 4.5. It is obvious that the reset method with a high feedback resistor does not affect the transient response significantly. The time constant τ_1, however, needs to be determined by the expected signal rate in the targeted application. For practical reasons, the feedback resistance is realized as a long MOS-transistor, which also gives the benefit of adjustability of the feedback resistance.

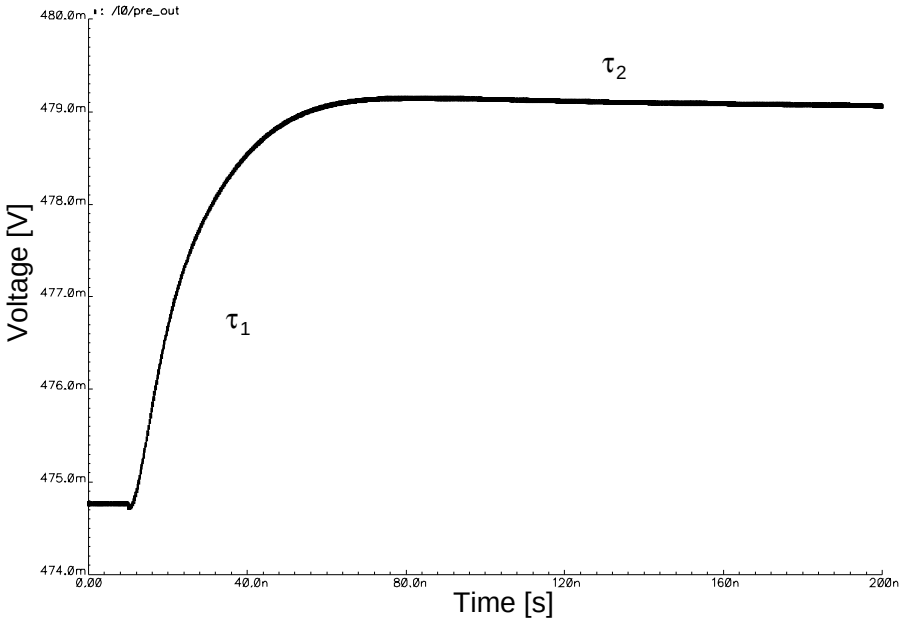

Figure 4.6: Simulated output of the preamplifier. The two time constants are represented by τ_1 and τ_2

The rise t_{rise} time of the output signal is usually defined as the transition time between 10% and 90% of the asymptotic amplitude. From 4.11 it can be calculated:

$$t_{rise} = 2.2 \cdot \tau_2 = 2.2 \frac{C_{input}}{2\pi \cdot GBW \cdot C_{fb}} \tag{4.12}$$

From 4.12 it is clear that minimizing the input capacitance and increasing the feedback capacitance and the GBW of the amplifier reduces the rise time.

The minimal rise time, which is equivalent to the maximum GBW is limited by the stability constraint. Since the second pole p_2 is equal to the unity loop gain frequency of the feedback loop, it is required, that all non-dominant poles of the amplifier cell must be higher than the second pole from eq. 4.10 [5]. For the chosen folded cascode configuration of the amplifier cell [6], this criterion is easily fulfilled. The reason is that the first non-dominant pole in this configuration is determined by the cutoff frequency of the cascode transistor. The position of the non-dominant pole is given by

$$p_{nd} = \frac{g_{mcasc}}{C_2} \tag{4.13}$$

where g_{mcasc} is the transconductance of the cascoded transistor and C_2 the total capacitance connected to the node between the cascode transistor and the current source transistor (bias 1). Since the contribution of the gate capacitance of the input transistor is usually quite large (for minimizing the noise of the charge sensitive amplifier), a relatively high W/L-ratio for the cascoded transistor needs to be chosen.

Figure 4.7 shows the detailed schematic of the preamplifier with the according design parameters and tab. 4.2 depicts the small signal values of the preamplifier of the Beetle 1.0.

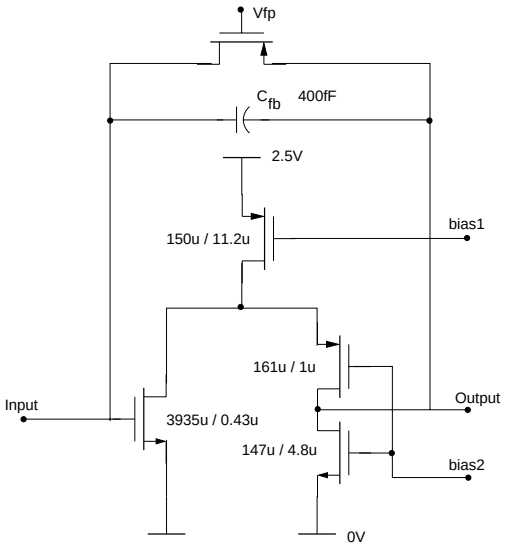

Figure 4.7: Detailed schematic of the preamplifier with the according design parameters

g_m	9.8$\frac{mA}{V}$	τ_1	10.5 ns
C_{tr}	4 pF	τ_2	8 μs
C_{fb}	400 fF	t_r	23.2 ns
R_{fb}	20 MΩ	GBW	711 MHz
R_l	1.8 MΩ	C_{out}	1.1 pF

Table 4.2: Small signal parameters for the preamplifier in the Beetle 1.0

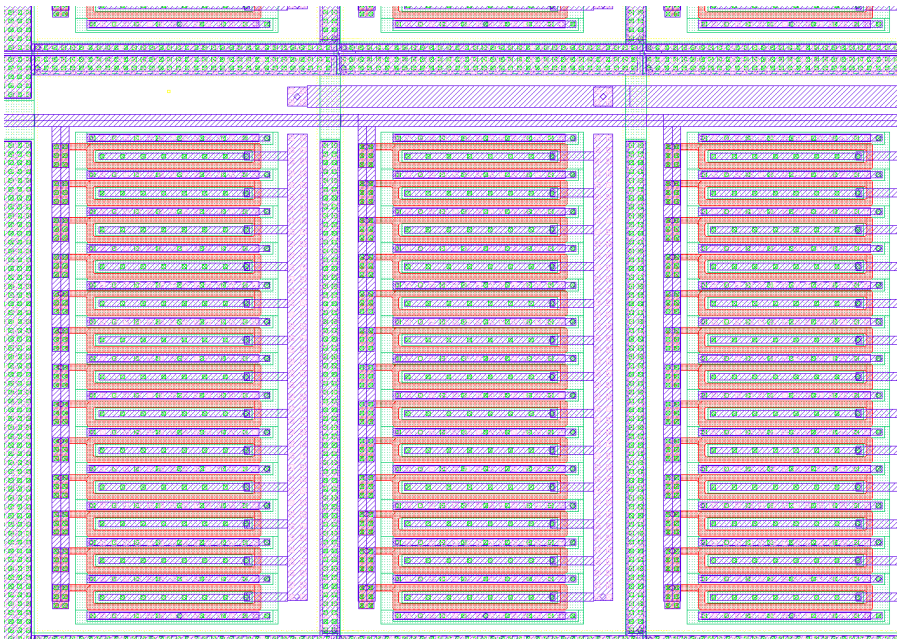

Figure 4.8: Layout of the input transistor: The transistor adds up from 12 blocks each consisting of 13 smaller transistors with enclosed gate structure. Three out of 12 blocks are shown for simplicity.

Layout of the Input Transistor

The layout of the input transistor is crucial for the performance of the preamplifier. Especially a low-ohmic connection to the ground node is important to prevent a susceptibility to common mode variations. A part of the layout is shown in fig. 4.8. The enclosed gates of the transistors are interleaved with rows of source and drain contacts. To achieve the large aspect ratio of W/L=4000μm/0.43μm, the transistor has been split into 156 transistors with an aspect ratio of W/L=25.6μm/0.43μm. This leads to very low-ohmic connection of the source node to ground and allows to add columns of contacts between the individual transistors, since a low-ohmic bulk connection is important to exclude possible noise contributions due to the bulk resistance.

Pulse Shaper

The task of the pulse shaper is to convert the voltage step at the output of the preamplifier into a time-limited pulse, with a peak height that is proportional to the charge delivered by the detector channel. The design boundaries on a pulse shaping stage are manyfold. Some of the requirements are even contradictory in their physical implementation and certain compromises have to be made. The most demanding requirements on the shaping stage are given by S/N performance, time limitation of the shaped pulse, dynamic range, power consumption and physical space requirements.

The decision to use an active CR-RC bandpass filter as shaping stage and a single sampling method was driven by several considerations. The use of an CR-RC filter is

quite common, since it is simple to realize in a CMOS-process and the obtainable S/N-ratio is reasonable at a limited power consumption. The use of a higher order shaping filter usually consumes more power and increases the circuit complexity. The double correlated sampling method, which would be favorable in terms of noise performance, has not been used since the method would introduce additional dead time (if higher sampling frequencies are excluded) and the need of more than one sampling clock. For a detailed comparison of the different shaping and sampling methods, see [1, ch. 6].

The semi-Gaussian shaper implemented is a first order shaper, which means that one differentiator is followed by one integrator. The time constants of the differentiator and the integrator should be identical in order to optimize the band pass characteristics. The time constant of the shaper influences the total noise behavior of the complete frontend. For a given shaper configuration, the total ENC varies with the shaping time. This is due to the channel thermal noise which rises in inverse proportion to the shaping time and the ENC due to the detector leakage current, which rises proportional to the shaping time. The contribution due to the 1/f-noise is independent from the shaping time constant. Therefore, an optimum shaping time (with respect to the total noise of the frontend) exists. However, in the case of the Beetle, this optimization cannot be applied, since the pulse width has to lie within a window of 25 ns with a maximum remainder of 30% in the following bunch cycle, as required by the Level-1 trigger algorithm (see also chapter 1).

To obtain the transient behavior of the pulse shaper, the Laplace transform of the output signal in the frequency domain needs to be calculated. The output signal in the frequency domain is given by the product of the transfer function $H(s)$ and the Laplace transform of the output signal of the preamplifier $V_{in}(s)$. For simplicity, the output of the preamplifier is approximated by an ideal step function. The output voltage is then given by

$$V_{out}(s) = \underbrace{\left[\frac{s\tau_0}{1+s\tau_0}\right]\left[\frac{A}{1+s\tau_0}\right]}_{H(s)}\underbrace{\frac{Q}{sC_f}}_{V_{in}(s)} \tag{4.14}$$

Taking the inverse Laplace transformation of eq. 4.14 gives the the output signal in the time domain:

$$V_{out}(t) = \frac{QA}{C_f}\left(\frac{t}{\tau_0}\right)e^{-\frac{t}{\tau_0}} \tag{4.15}$$

One characteristic of eq. 4.15 is that the peak voltage, reached after τ_0, is proportional to the input voltage and therefore proportional to the charge that is coupled into to the preamplifier. Figure 4.9 shows the simulated response of the pulse shaper as realized on the Beetle 1.0. Curve A represents the response to an ideal voltage step, curve B shows the response to the preamplifier output.

The detailed schematic of the pulse shaper is depicted in fig. 4.10 and the small signal parameters of the shaper can be found in tab 4.3.

As can be seen from the simulation results (which are in good agreement with test results, chapter 5), the remaining 30 % of the peak voltage after 25 ns leaves no margin for a further slowing down of the signal due to for example a degradation of the detector signal. A safety margin would therefore be very desirable, also because process

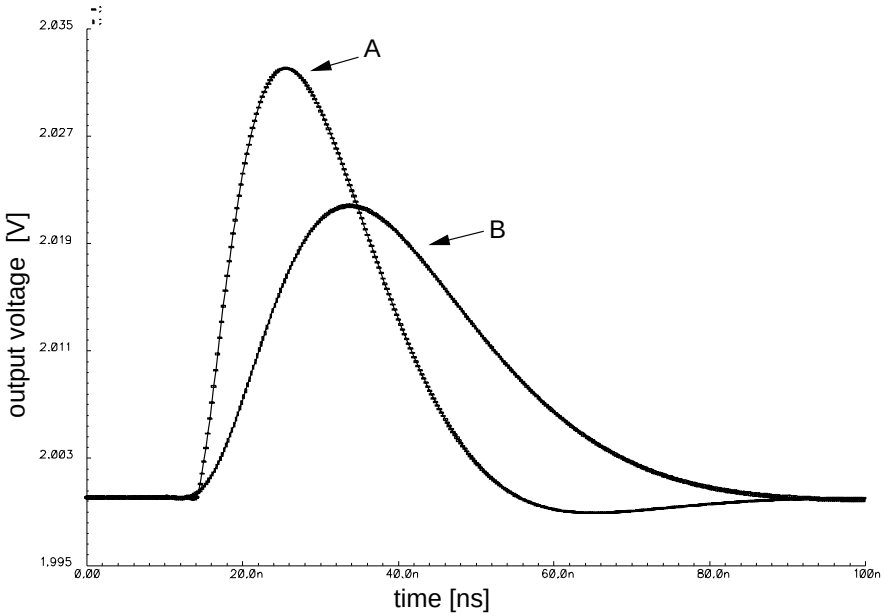

Figure 4.9: Simulated output voltage of the pulse shaping stage. Curve A shows the response to an ideal voltage step, curve B the response on the preamplifier output voltage.

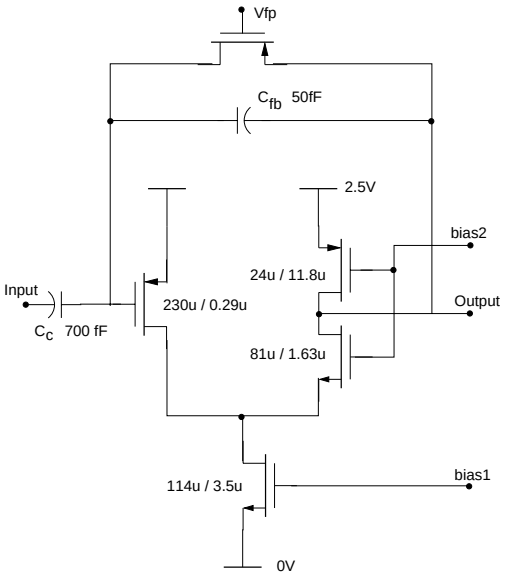

Figure 4.10: Detailed schematic of the pulse shaper with the according design parameters

g_m	$1.1\frac{mA}{V}$	t_{peak}	12.9 ns
C_c	700 fF	GBW	112 MHz
C_{fb}	75 fF	C_l	89 fF
C_{tr}	420 fF	R_l	8.7 MΩ

Table 4.3: Small signal parameters of the shaper

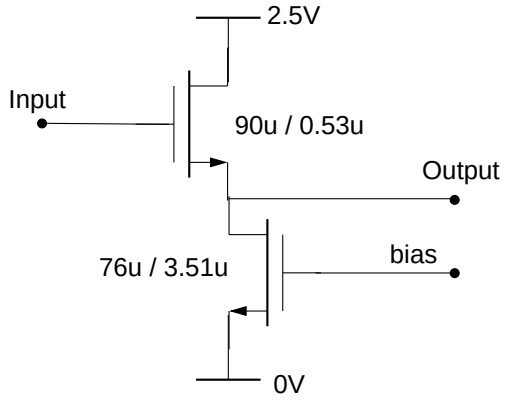

Figure 4.11: Schematic of the frontend buffer

g_m	1.22 $\frac{mA}{V}$
gain	0.79
C_{out}	1.49 pF
t_r	2.2 ns
slew rate	67.6 $\frac{mV}{ns}$

Table 4.4: Small signal parameters of the frontend buffer

parameter variations are not taken into account in the simulation. One way to achieve this would be to speed up the preamplifier. As can be seen in fig. 4.9, a fast preamplifier output reduces the remainder significantly. However, the preamplifier already operates at its limits given by the power consumption. An increase in the bias current of the preamplifier (and with that an increase in the power consumption) to a value of 1 mA would result in a peaking time of the shaper output of about 22 ns and a remainder of 27%. Compared to a peaking time of 25 ns and a remainder of 30% in the case of a bias current of 600 μA, this higher power consumption does not pay back. A more promising way is the introduction of a second shaping stage, that would differentiate the signal once more. A detailed examination and comparison of the trade-offs should be made.

Buffer

The buffer of the analog input stage has been designed as a standard source follower [5]. It must be able to drive the pipeline capacitor, the parasitic capacitance of the write line and the input capacitance of the comparator, which amounts to 1.84 pF.

The schematic together with the design parameters are given in fig. 4.11 and the small signal parameters are listed in tab. 4.4.

4.5.2 Comparator

A comparator, designed by Hans Verkooijen [8, 9], is located directly behind each analog frontend channel in the signal path. It compares the output of the frontend

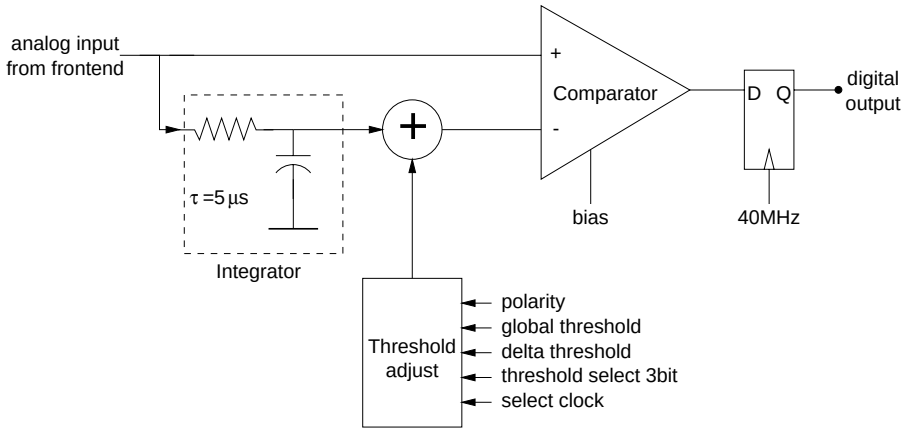

Figure 4.12: Schematic blocks of the comparator circuit [8]

with an adjustable threshold value and derives a digital signal if a frontend output signal exceeds the threshold. The digital output of the comparator can be handled in two two different ways. For a prompt output of the digital signals (which will be made use of in the pile-up veto counter of LHCb), the logic sum of four neighboring channels is derived and the resulting digital signals are multiplexed by a factor of two. These signals are brought off-chip via 16 LVDS drivers, that are located at the top and bottom side of the readout chip (see fig. 4.2). For a pipelined readout of the digital information, the output signal of the comparator is stored in the pipeline. To prevent an overdrive of the analog stages after the pipeline, the levels of the digital comparator output are shifted so that they correspond to a signal height of about 0 MIP and 10 MIP, respectively. The pipelined readout of the comparator output has been implemented for the application of the Beetle in the RICH detector of LHCb, where the analog information of the events is not needed.

The requirements on the comparator are driven by the need to detect the signals of minimum ionizing particles with a high efficiency. The efficiency of the detection of signals should only be limited by the noise level of input the amplifier of the Beetle, and not by the comparator. Therefore, the noise contribution by the comparator should be well below 800 electrons to be able to detect all minimum ionizing particles. The response of the comparator should be fast enough to localize the time of the hit with respect to the LHC bunch crossing clock. Within a time window of 25 ns, a signal at the shaper output should be detected and the time walk of the digital output should not influence the precision of the detection. The channel homogeneity of the comparators should be high enough so that a common threshold can be applied to all channels. Although a trim-DAC for each channel is implemented in the Beetle-1.0 to adjust channel to channel variations in the offset voltage, a common threshold would ease a lot the application of the chip in a large system like the LHCb vertex detector.

A configuration as shown in fig.4.12 has been chosen for this task. The analog output of the frontend is integrated with a time constant of about $5\mu s$ and the resulting DC level is added to an adjustable threshold voltage. The DC level and the analog signal is then fed to the actual comparator. The output of the comparator is then latched with the 40 MHz bunch crossing clock.

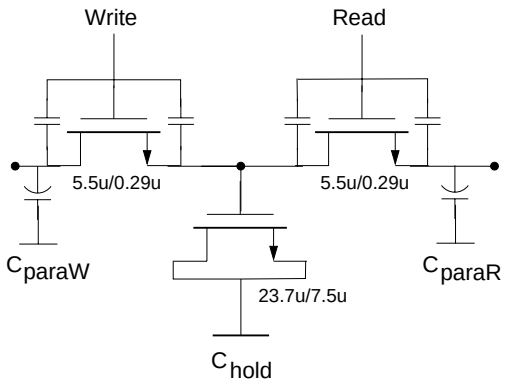

Figure 4.13: Schematic of one pipeline cell, including the parasitic capacitances

The comparator core amplifier consists of a simple differential pair, that has been matched carefully to keep offset variations at a minimum. The differential output of the first amplifier is fed into a second differential amplifier. The total gain of the comparator amounts to about 1300 [8].

4.5.3 Pipeline

The pipeline consists of an array of 129×186 storage cells, each cell containing an nMOS write switch, an nMOS read switch and an nMOS gate capacitance to store the output of the frontend as a charge for later readout by the pipeline readout amplifier. nMOS transistors instead of pMOS transistors have been chosen as switches due to the lower on-resistivity of the nMOS transistors. As will be shown, the larger charge injection of the nMOS transistors (they have to be laid out in an edgeless geometry and therefore exhibit a larger gate capacitance) do not distort the signal significantly. The storage cell capacitor has been realized as nMOS transistor with source and drain tied to ground potential, since the use of metal-metal capacitors is limited due to a manufacturers design rule. This also has the advantage of a compact layout due to the very large specific gate capacitance of 5.56 fF/μm^2.

Figure 4.13 shows the schematic of one pipeline cell including the parasitic capacitances of the write- and read-line and the gate capacitance of the switches. A falsification of the signal takes place during the opening of the switches. Two different mechanisms are responsible for this clock feed-through. The first contribution is due to the charge remaining in the channel of the switch transistor, after the transistor has been turned off. This charge is added to the sampled signal and causes a falsification. The second contribution occurs due to the coupling of the digital switch signal into the transistor channel via the gate capacitance. The first effect gives a constant offset contribution, which was simulated to be about 2 mV. The second effect is absolutely negligible due to the small gate capacitance of the switch of 8.9 fF.

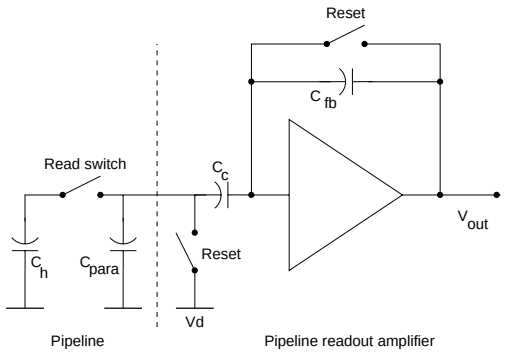

Figure 4.14: Schematic of the pipeline readout amplifier

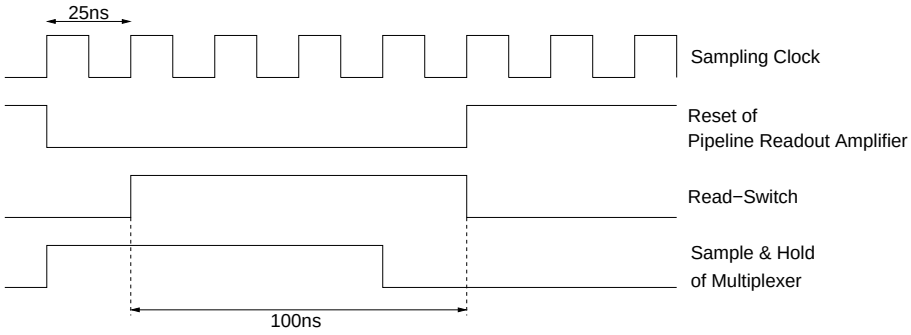

Figure 4.15: Control signals of the pipeline readout amplifier during readout of one event

4.5.4 Pipeline Readout Amplifier

The pipeline readout amplifier (designed by D. Baumeister) reads the charge, that is stored in a pipeline cell. It is designed as a resetable, AC-coupled charge-sensitive amplifier with a folded cascode configuration as an amplifier cell. The reset of the amplifier and of the read-line is done by two transmission gates to reduce clock feed-through. Figure 4.14 shows a schematic of the pipeline readout amplifier. The sequence of the control signals is depicted in fig. 4.15. If a trigger signal arrives at the readout chip, the reset signal of the pipeline readout amplifier is released. After a pause of one clock cycle (nominally 25ns), the read-switch of the pipeline cell, that will be read out, is closed. The charge that is stored in the pipeline cell is now amplified by the pipeline readout amplifier. The rise time of the amplifier therefore defines the time until the hold switch of the sample-and-hold stage of the multiplexer is opened. Three clock cycles after the read switch is closed, the sample-and-hold switch of the multiplexer is closed and the output of the pipeline readout amplifier is sampled.

The time constant of the charge transfer from the pipeline capacitor to the couple capacitor of the amplifier and onto the parasitic capacitance of the read line follows a RC-law. The time constant is therefore given by

$$\tau_{RC} = R_r \cdot C_{tot} \tag{4.16}$$

where R_r is the resistance of the closed read-switch and $C_{tot} = \left(\frac{1}{C_h} + \frac{1}{C_c + C_{para}}\right)^{-1}$ the

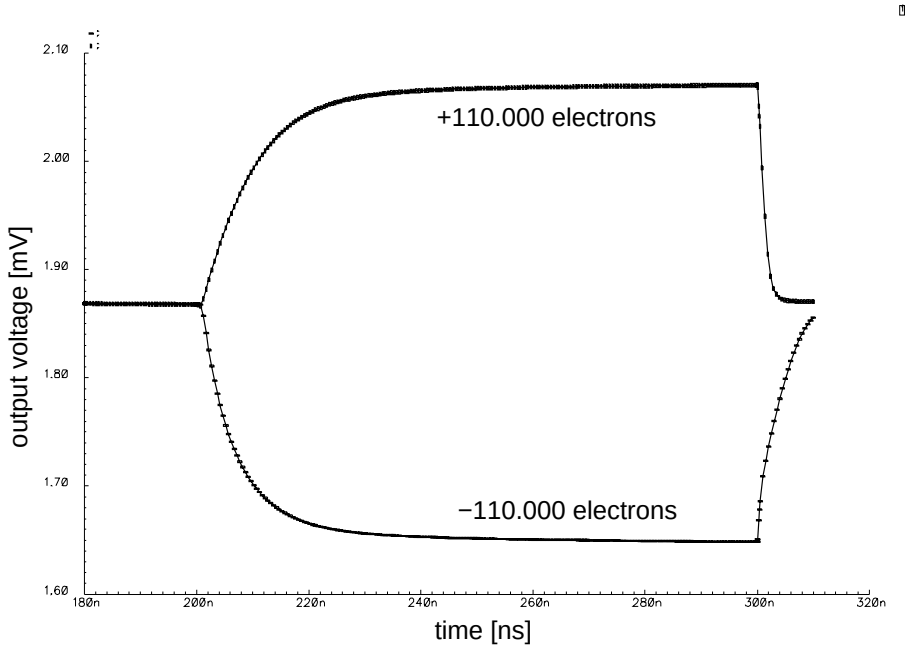

Figure 4.16: Simulated output of the pipeline readout amplifier for signals that are equivalent to -110.000e^- to +110.000e^-.

capacitance, that needs to be charged. This amounts to a rise time of $t_{RC} = 2.2 \cdot \tau_{RC} \approx 1ns$, and can be neglected compared to the rise time of the pipeline readout amplifier.

The determination of the small signal model and the transient behavior of the pipeline readout amplifier is therefore similar to the preamplifier case. The rise time of the amplifier can be calculated as done in eq. 4.12

$$t_r = 2.2 \cdot \tau_r = 2.2 \cdot \frac{C_{input}}{2\pi C_{fb} \cdot GBW} \tag{4.17}$$

where $C_{input} = C_{fb} + C_{gate}$ and $GBW = \frac{g_m}{C_{out}}$ the gain-bandwidth product, as in the case of the preamplifier. The gain of the pipeline readout amplifier has been maximized, so that the whole dynamic range of the following stages is used. Figure 4.16 shows the transient response of the amplifier for signals that are equivalent to -110.000e^- (-10 MIP) to +110.000e^- (+10 MIP). The size of the input transistor has been chosen, so that a rise time can be achieved, that is sufficient to cope with the given timing constraint of 100 ns. The low-ohmic voltage sources Vd and Vdcl are used to discharge C_c and C_{fb}. The output of the amplifier needs to drive the input capacitance of the following stage, which is mainly the the hold capacitor of the sample-and-hold stage of the multiplexer and the input capacitance of the according source follower. The reset switches are realized as fully compensated transmission gates to eliminate clock feed-through by charge injection.

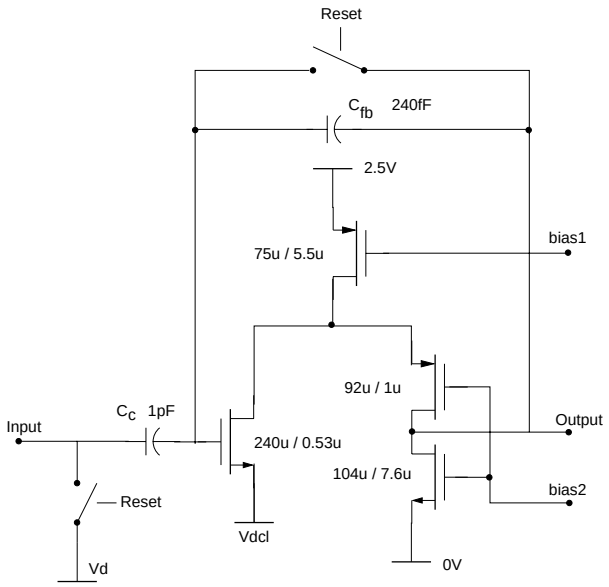

Figure 4.17: Detailed schematic of the pipeline readout amplifier

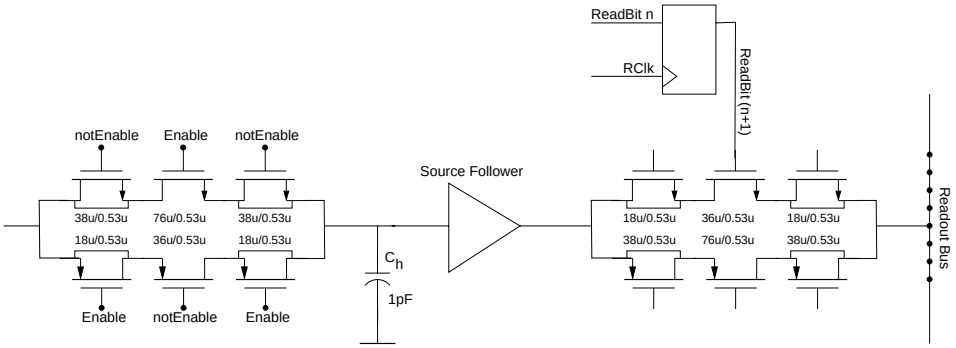

Figure 4.18: Schematic of one channel of the multiplexer

4.5.5 Multiplexer

The task of the multiplexer is to switch the data of a number of channels onto one common line sequentially. Depending on the mode the chip is operated in, this can be one, two or four lines. Therefore, the multiplexer consists of a sample-and-hold stage, a source follower to drive the bus capacitance, and switches including a shift register. By switching the so-called Read-bit onto different sections of the shift register, the different readout modes are controlled.

Table 4.5 lists the three different modes and their foreseen application. Mode 1 is the default operation mode for the Beetle chip used in the LHCb vertex detector. The 128 channels are multiplexed 32-block wise onto 4 output ports in parallel. At a nominal readout frequency of 40 MHz, the multiplexing of the analog channels takes $32 \times 25ns = 800ns$. The readout time for one complete event is the sum of the rise time of the pipeline readout amplifier (100ns) and the time for the multiplexing of 800 ns. The readout mode 2 is mainly implemented for the application of the Beetle chip in

Mode		Readout time per event	application
1	128 channels multiplexed on 4 ports @ 40 MHz	900ns	LHCb vertex detector
2	128 channels multiplexed on 2 ports @ 80 MHz	900ns	LHCb RICH detector
3	128 channels multiplexed on 1 port @ 40 MHz	3.2 μs	test setup

Table 4.5: The three different readout modes of the multiplexer

g_m	1.84 $\frac{mA}{V}$
gain	0.78
C_{out}	1.40 pF
t_r	2.1 ns
slew rate	14.7 $\frac{mV}{ns}$
C_{load}	6.8 pF

Table 4.6: Parameters of the source follower in the multiplexer.

the RICH detector of LHCb. In this mode, 2 blocks of 64 channels are multiplexed on 2 output ports. The multiplexing clock of 40 MHz is inverted for half of the channels with respect to the rest of the channels. This leads to an effective readout frequency of 80 MHz. The multiplexing time per event amounts to $64 \times 12.5ns = 800ns$. Including the setup time of the pipeline readout amplifier, this also leads to a total readout time of 900ns per event. In principle, mode 2 can be used to read out analog or digital pipeline samples, the default application is to read out the digital pipeline samples. The third readout mode multiplexes all 128 channels on one output port. The multiplexing time at a readout frequency of 40 MHz for one event amounts to $128 \times 25ns = 3.2\mu s$. This mode is intended to be used in applications, where a fast readout time per event is not required, for example in test setups.

In all three readout modes, the time interval between the multiplexing of the analog/digital event date, which is actually the settling time of the pipeline readout amplifier, is used to transfer some digital information onto the output ports. In the first version of the Beetle readout chip, this is the encoded location of the pipeline column from which the event has been read out.

Although the the multiplexer of the Beetle 1.0 uses a standard method of sample-and-hold stage with a following buffer, some points have required special care. The switches are realized as fully compensated transmission gates using pMOS and nMOS transistors switching simultaneously and dummy transistors of half the size connected to source and drain of the switching transistors [14]. The source follower of each channel has to be able to drive the capacitance given by all switches that are connected to the same output bus and the capacitance of the bus itself. An nMOS source follower has been chosen for this task. The detailed parameters of the source follower are shown in tab. 4.6. The clock signals for the shift register of the multiplexer have to be routed in parallel to all 128 channels (which is a total length of about 5mm). These long clock lines have a distributed capacitance of about 5pF to the substrate. Therefore a layer of

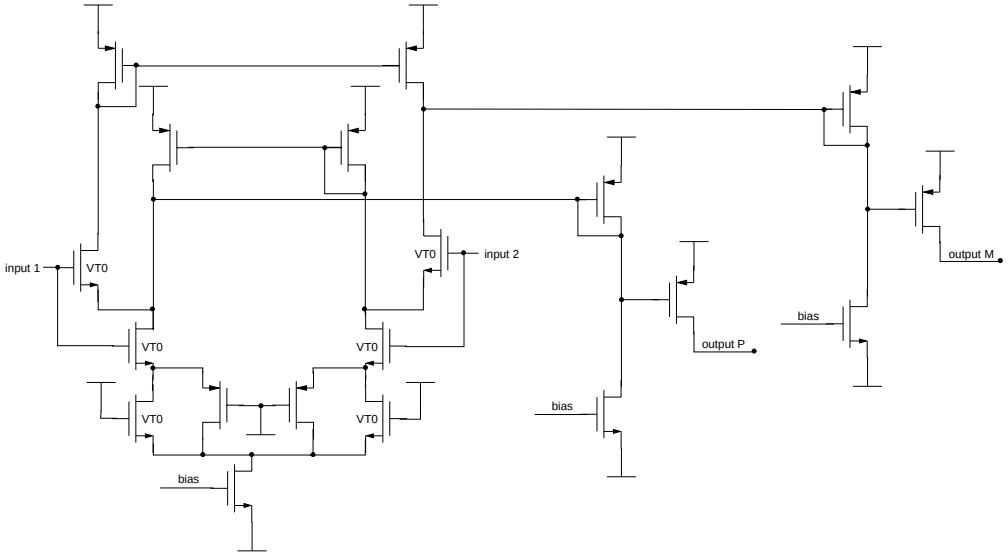

Figure 4.19: Schematic of the analog output buffer (dimensions of the transistors are not shown for simplicity). Transistors with the extremely small threshold voltage are labeled 'VT0'

polysilicon has been introduced as a means of shielding of the substrate.

4.5.6 Analog Output Buffer

The task of the analog output buffer is to drive the analog output signal to the stage following the readout chip. In the case of the LHCb vertex detector, this is an ADC located about 10 m apart from the detector module.

The buffer (designed by Hans Verkooijen) is realized as a current buffer with a differential output. A current buffer instead of a voltage buffer has been chosen to minimize the power consumption of the buffer. In addition, a current buffer can be designed without the use of a feedback amplifier, which is prone to instability if the capacitive load at the output of the buffer exceeds the given limits.

In the case of the binary readout mode, the multiplexer runs at an effective frequency of 80 MHz, which defines the maximum operating frequency of the output buffer. The dynamic range should be large enough to be insensitive to crosstalk from neighboring signal lines and the differential output has to be able to drive a 50 Ω impedance.

The chosen schematic is depicted in fig. 4.19. The differential input amplifier is completely symmetric and uses transistors with an extremely low threshold voltage (some 10 mV), which is offered as an option in the chosen technology. The advantage of using these so-called zero-Vt transistors is a simple biasing in combination with a large dynamic range, which would otherwise be limited by the biasing transistors. The output stages use a very large pMOS-transistor (W/L=200u/0.51u) to drive the current off-chip.

4.6 Control Circuits

4.6.1 Pipeline Control Logic

The pipeline control logic (PCL) has several tasks to accomplish:

- Controlling the read and write switches of the pipeline
- Storing pipeline locations that are marked by a trigger signal until they are read out
- Asserting the reset signals to the pipeline readout amplifier
- Starting the multiplexer readout

The PCL operates synchronously with the sampling clock in phase with the bunch crossing clock from the LHC accelerator. The write switches of one complete column are closed if the write pointer contains the column's address. The write pointer counts upwards until the physical size of the pipeline is reached and starts again at zero. The trigger pointer follows the write pointer at a fixed distance, which represents the latency of the pipeline. Upon an external trigger signal, the content of the trigger pointer is stored in a FIFO and the corresponding column is protected from being overwritten. Up to 16 pipeline locations can be stored in the FIFO. Stored events are read out by loading the charge from the according pipeline column into the multiplexer via the pipeline readout amplifier. If the event is completely read out, the according pipeline location is again set free and can be overwritten by new data. The write and read of the pipeline cells operate independently and thus allows a deadtimeless readout within the limits of the FIFO size.

The complete logic is described in the hardware description language Verilog and synthesized using Synopsis. Besides different parameters, the code is identical to the PCL of the HELIX128-3.1[1].

4.6.2 I^2C Interface

For the programming of the slow-control of the Beetle, an I^2C [15] protocol has been chosen. It has been implemented by D. Baumeister. Being a standardized protocol, I^2C has the advantage of easy access to commercial components that allows a simple setup of an operating facility for the readout chip. Using a read and write decoder allows the programming of all registers on the chip and a readout of their contents. The possibility of reading the register contents is very important in an environment, where SEUs can change the contents of the registers at any time. The interface on the Beetle acts as an I^2C-slave and the slave addresses of chips sharing one bus line are generated in an automatic way via a token chain.

4.6.3 Bias Generators

The operation of a readout chip is simplified to a large extent, if the amount of supply signals is reduced to an absolute minimum. Therefore, all bias currents and bias voltages that are needed to operate the analog stages of the Beetle are generated on-chip. This concept relies of course on the immunity of the bias circuits against parameter changes due to irradiation, temperature variation and process parameter spread.

Bias Voltage Generators

The R-2R resistor ladder configuration is used for the bias voltage generation. This configuration is well established and provides a simple means of a digital-to-analog converter that can operate between the power supply rails [14]. The resistor ladder has a resolution of 10 bit and uses resistors that are realized with polysilicon resistors of 6.8 kΩ. The (simulated) differential non-linearity amounts to about 4 mV, which corresponds to 1.5 LSB [16].

Bias Current Generators

The bias currents are generated by $2^{(n-1)}$ minimum size transistors in parallel, where n (10 bit) denotes the bit-value that is set. The reference voltage of the transistors is derived by means of current mirrors from a current source, that consists of a regular cascode of transistors and a 20 kΩ reference resistor. The nominal current of the current source is 100 μA and the total range of the current digital-to-analog converter is 1 mA.

Bibliography

[1] W. Fallot-Burghardt, A CMOS Mixed-Signal Readout Chip for the Microstrip Detectors of HERA-B, Dissertation, Universität Heidelberg, 1998

[2] R. Brenner et. al, Nucl. Instr. and Meth., A339 (1994) 477-484

[3] R. Brenner et. al, Nucl. Instr. and Meth., A339 (1994) 564-569

[4] M. Feuerstack-Raible, Overview of Microstrip Read-out Chips, Nucl. Instr. and Meth., A451 (2000)

[5] Z. Y. Chang, W. Sansen, Low-Noise Wide-Band Amplifiers in Bipolar and CMOS Technologies, Kluwer Academic Papers, 1991

[6] Y. Hu, E. Nygard, Nucl. Instr. and Meth., A365 (1995) 193-197

[7] W. Sansen, Nucl. Instr. and Meth., A253 (1987) 427-433

[8] H. Verkooijen, Technical Report about the Status of the Comparator on the Beetle Chip, November 1999, NIKHEF

[9] R. Kluit, NIKHEF Beetle Page, http://www.nikhef.nl/pub/departments/et/vlsi/BEETLE/beetle.html

[10] F. Faccio et. al, Total Dose and Single Event Effects (SEE) in a 0.25μm CMOS Technology, CERN/LHCC/98-36

[11] A. Giraldo, Ph.D thesis, University of Padua, 1998

[12] F. Faccio, et.al, Total Dose and Single Event Effects (SEE) in a 0.25 μm CMOS Technology, Proceedings of the Fourth Workshop on Electronics for LHC Experiments, Rome, CERN/LHCC/98-36

[13] P. Jarron et. al, RD49 Status Reports, http://rd49.web.cern.ch/RD49/Welcome.html

[14] L. Geiger et. al., VLSI Design Techniques for Analog and Digital Circuits, McGraw-Hill Publishing Company, 1990

[15] Philips, The I^2C-bus and how to use it, I^2C Interface specification, April 1995

[16] N. Smale, personal communication

Chapter 5

Measurement Results

5.1 Introduction

The design of the Beetle 1.0 has been submitted to the manufacturer in April 2000 and due to an unfortunate error in the design checking software, the chip is not functional. Therefore, measurements on the Beetle 1.0 could not be performed in time for this thesis. A patch of the malfunctioning part of the chip is being undertaken at the time of writing up this thesis.

The measurements reported here refer to the test chips that have been manufactured in October 1999. The components that have been characterized on the test chips are identical to the corresponding components on the Beetle 1.0.

5.2 Frontend Test Chip

5.2.1 Introduction

The frontend test chip contains three sets of analog input stages (that is preamplifier, shaper, buffer), each set consisting of four channels. The three different sets differ in certain parameters that mainly influence the pulse shape like shaper feedback resistance and couple capacitance. Only the measurements of set three are reported here, since they match closest the requirements of the LHCb vertex detector.

bias current/voltage	nominal value
Preamplifier bias current (Ipre)	600 μA
Shaper bias current (Isha)	80 μA
Frontend buffer bias current (Ibuf)	100 μA
Preamplifier feedback voltage (Vfp)	1 V
Shaper feedback voltage (Vfs)	500 mV

Table 5.1: Nominal bias settings for the BeetleFE

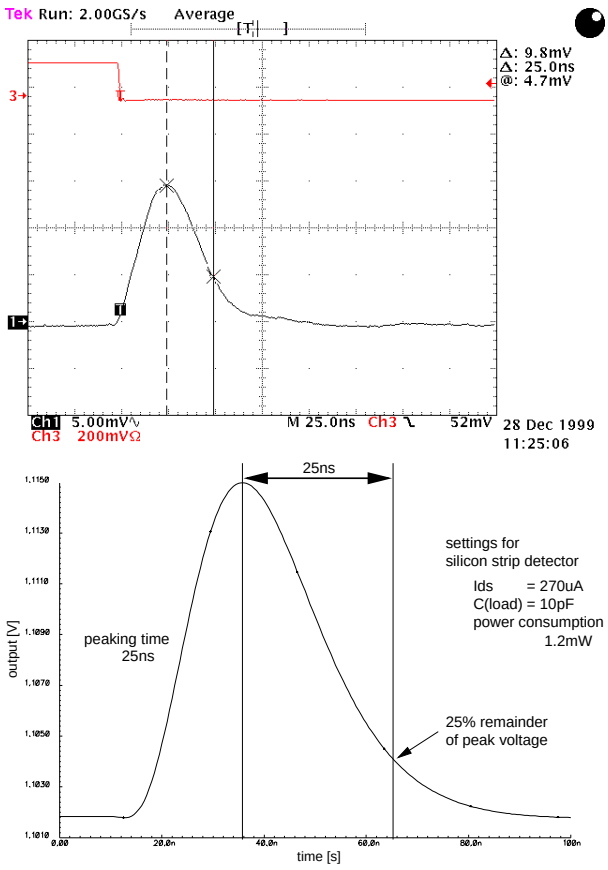

Figure 5.1: Transient response of the frontend to a signal of 11.000 electrons: measurement (top) and simulation (bottom)

The measurements of the frontend test chip BeetleFE has been performed with bias settings as listed in tab. 5.1. They will be referred to as nominal settings.

5.2.2 Transient Behavior

The transient pulse response has been tested by injecting a defined charge into the input of the amplifier by means of a couple capacitor of 1.5 pF. The biasing parameters of the preamplifier and the shaper have been swept over a range to compare the measured behavior of the amplifier with simulated results. Exemplarily, a pulse shape optimized for the LHCb vertex detector is shown in fig. 5.1. The measured rise time of the amplifier of 26 ns agrees well with the simulated value of 25 ns. The remainder of the signal maximum after 25 ns (which is an important number for the performance of the Level-0 trigger) is larger (30%) than in the simulation (25%) and leads to the conclusion that the internal resistance of the preamplifier is not small enough. The remainder of 30% fits just the requirements given by the Level-0 trigger system (see chapt. 1) and points to a possible place for improvement of the design.

The maximum height of the output signal of the analog frontend as function of the input signal is plotted in fig. 5.2. The required linear range of $\pm$110,000 electrons is

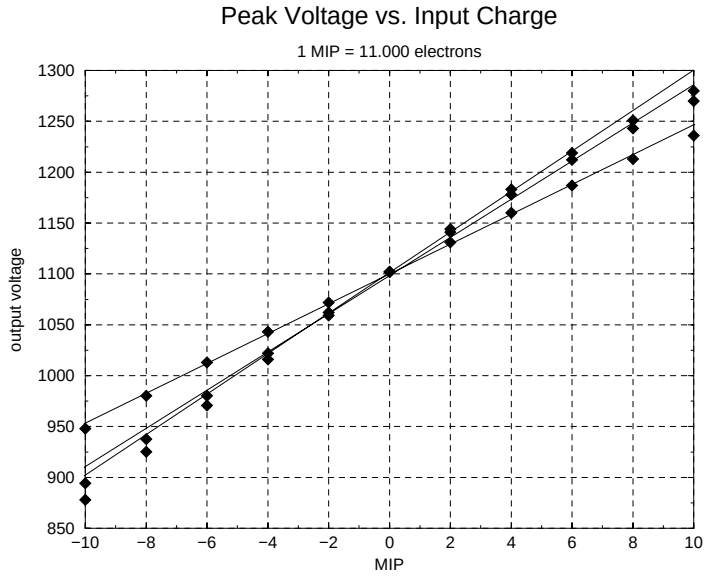

Figure 5.2: Maximum height of the frontend's output versus input charges ranging from -110,000 electrons to +110,000 electrons for three different load capacitances of 5.6 pF, 10 pF and 15.6 pF (highest slope to smallest slope)

Preamplifier bias current	Measurement	Calculation
I_{pre}=350 μA	409 e^-+38.0 $\frac{e^-}{pF}$	470 e^-+39.5 $\frac{e^-}{pF}$
I_{pre}=600 μA	303 e^-+33.6 $\frac{e^-}{pF}$	340 e^-+33.6 $\frac{e^-}{pF}$

Table 5.2: Comparison of calculation ad measurement of the noise function

given.

5.2.3 Frequency Behavior

The output amplitude as function of the input signals frequency is shown in fig. 5.3. The measurement has been performed by means of a network analyzer. The curve shows the typical behavior of an integrator with a following RC-CR bandpass filter. The rising part of the curve is proportional to the frequency f, the falling tail is proportional to $\frac{1}{f^2}$. The measured maximum of the frequency is located at 12.5 MHz, whereas the simulation yields 13 MHz.

5.2.4 Noise Performance

The noise performance of the analog input stage has been measured for different values of the load capacitance at the input of the preamplifier. For each noise value, 10,000 samples were taken with an oscilloscope by means of a Labview program. Figure 5.4 shows the noise values as a function of the load capacitance for two different preamplifier bias currents. A comparison of the measured values with the values obtained

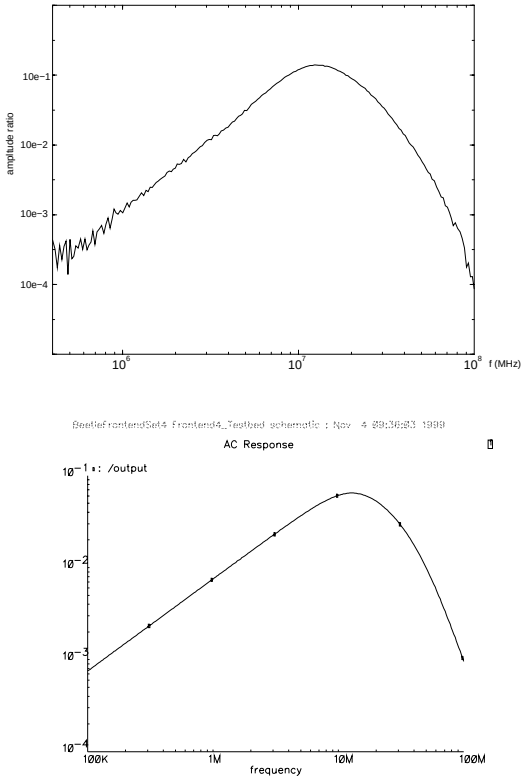

Figure 5.3: Frequency response of the input stage: measurement (top) and simulation (bottom)

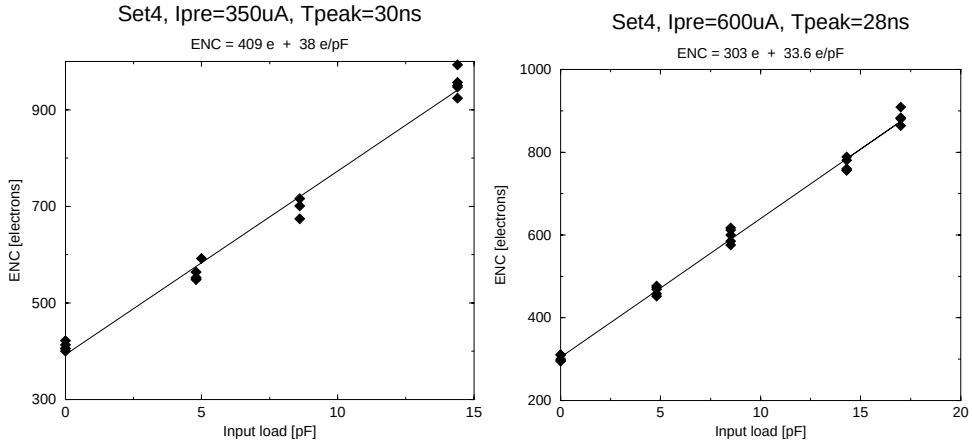

Figure 5.4: Noise functions for a preamplifier bias current of 350 μA (left) and 600 μA (right).

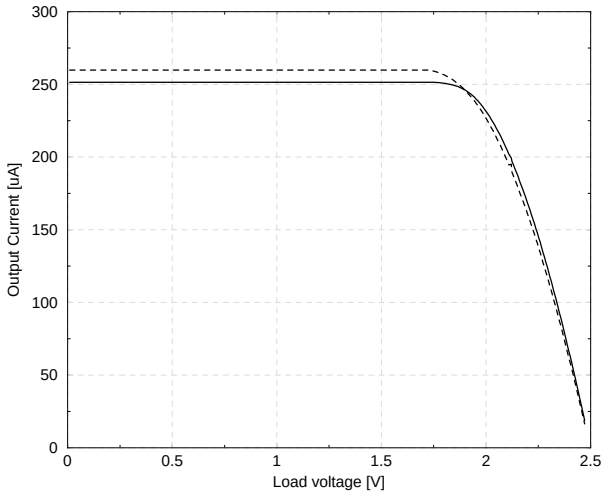

Figure 5.5: Measured output current of the current source (dashed line) compared with simulation (solid line) versus load voltage

from calculation (using the Hspice parameters given by the manufacturer) is shown in tab. 5.2. The deviation of the noise offset value from the expectation is not clear, it may be due to a un-precise extraction of the parasitic capacitances of the input protection diodes. Using the measured noise value for the nominal preamplifier bias current of 600 μA, a S/N value of 17 can be obtained (for an input signal of 11,000 electrons, as expected in the silicon vertex detector).

5.3 Bias Generator Test Chip

5.3.1 Introduction

The bias generator test chip BeetleBG has been developed for two purposes:

- to have a set of programmable current and voltage sources on hand, especially for test setups of amplifier test chips
- to study the accuracy of simulation models and the matching of devices over several chips

The Beetle BG test chip contains a set of voltage and current DACs and current sources, the same versions as implemented on the Beetle 1.0 readout chip.

5.3.2 Current Source

Figure 5.5 shows the output current of the current source compared with simulation. The nominal value of 250 μA at a load of 1 V is located safely in the plateau region.

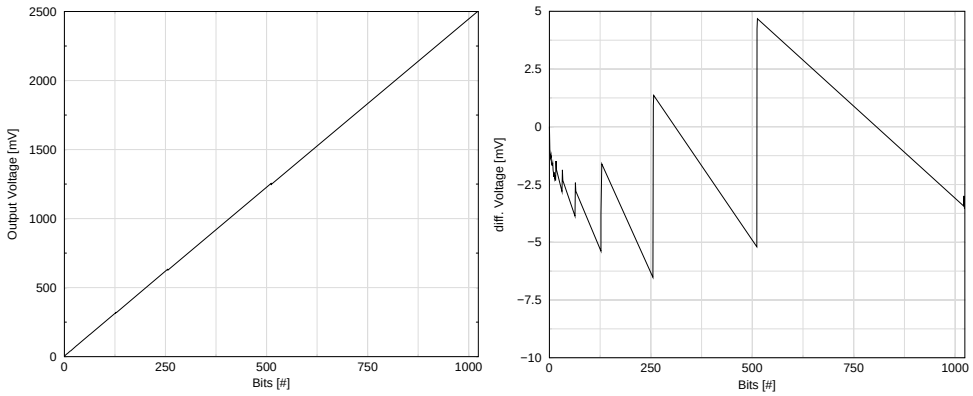

Figure 5.6: Output voltage (left) and differential non-linearity (right) of the 10bit voltage DAC as function of the bit number

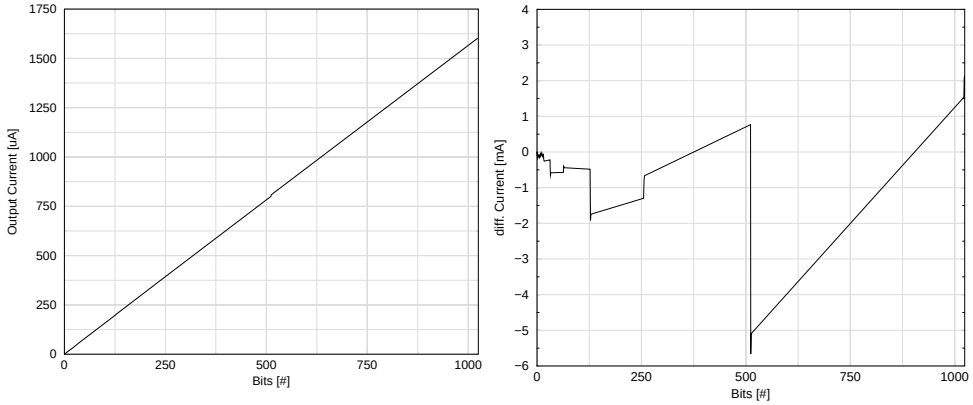

Figure 5.7: Output current (left) and differential non-linearity of the 10bit current DAC as function of the bit number

5.3.3 Voltage DAC

The output of the 10bit voltage DAC is plotted in fig. 5.6. The DAC operates between the power supply rails with an offset of 10 mV. The overall differential non-linearity is $\approx$11 mV, which corresponds to 4.5 LSB. The source of the differential non-linearity is the on-resistance of the switches that connect the resistors to eachother. The linearity is more than sufficient for the application in the Beetle 1.0.

5.3.4 Current DAC

The output of the current DAC is plotted in fig. 5.7. The maximum current is determined by the input current, and in this case the on-chip current source has been used, yielding in 1.6 mA. The differential non-linearity is 7 μA, corresponding to 4.5 LSB. As in the case of the voltage DAC, the linearity is more than sufficient for the application in the Beetle readout chip.

Conclusions

Within this thesis, a readout chip suitable for the LHCb vertex detector (Beetle 1.0) has been developed. The choice of the architecture, the use of radiation hard layout techniques and the standard 0.25μm CMOS process, the Beetle is manufactured, are expected to lead to a sufficient radiation hardness. From measurements of test structures and comparable chips, it is expected that the Beetle will withstand a total ionizing dose of 10 MRad. However, the impacts of single event effects during the operation of the Beetle in the LHCb vertex detector are not yet evident. Measurements of SEU rates will show the sensitivity of the Beetle to heavy ionizing particles, but due to the lack of knowledge on the distribution of heavy ionizing particles in the LHCb vertex detector, an operating scenario can hardly be predicted.

Therefore, an improved version of the Beetle 1.0 is under development, which makes use of redundancy checks in the control circuits and majority voting for important register contents. A detailed characterization and employment of the readout chip in a close-to-final system should be the prerequisite to an improved version of the readout chip.

Within the near future, when CMOS technologies with structure sizes below 0.25 μm become available and replace the todays technologies, the customers in the high energy physics domain will encounter two effects:

- The availability of the mainstream standard CMOS processes gets more and more difficult due to the increasing costs of the fabrication and the growing fabrication batches, that are necessary for an efficient throughput.
- The shrinking feature size of standard CMOS processes and with that the smaller gate oxide thickness increases the tolerance of devices against ionizing radiation. The increase in SEU sensitivity is the price to pay. Only if silicon-on-insulator CMOS processes become available to the high energy physics community at reasonable costs, this spiral of rising SEU sensitivity can be stopped.

Appendix A

Pad Description

Figure A.1 shows the arrangement of the bond pads on the Beetle 1.0. The size of the bond pads is 105×105 μm^2, and all pads other than the analog input and the frontend power pads are placed on a grid of 115 μm. The name, the function and the type of all pads are listed in tab. A.1 to tab. A.4.

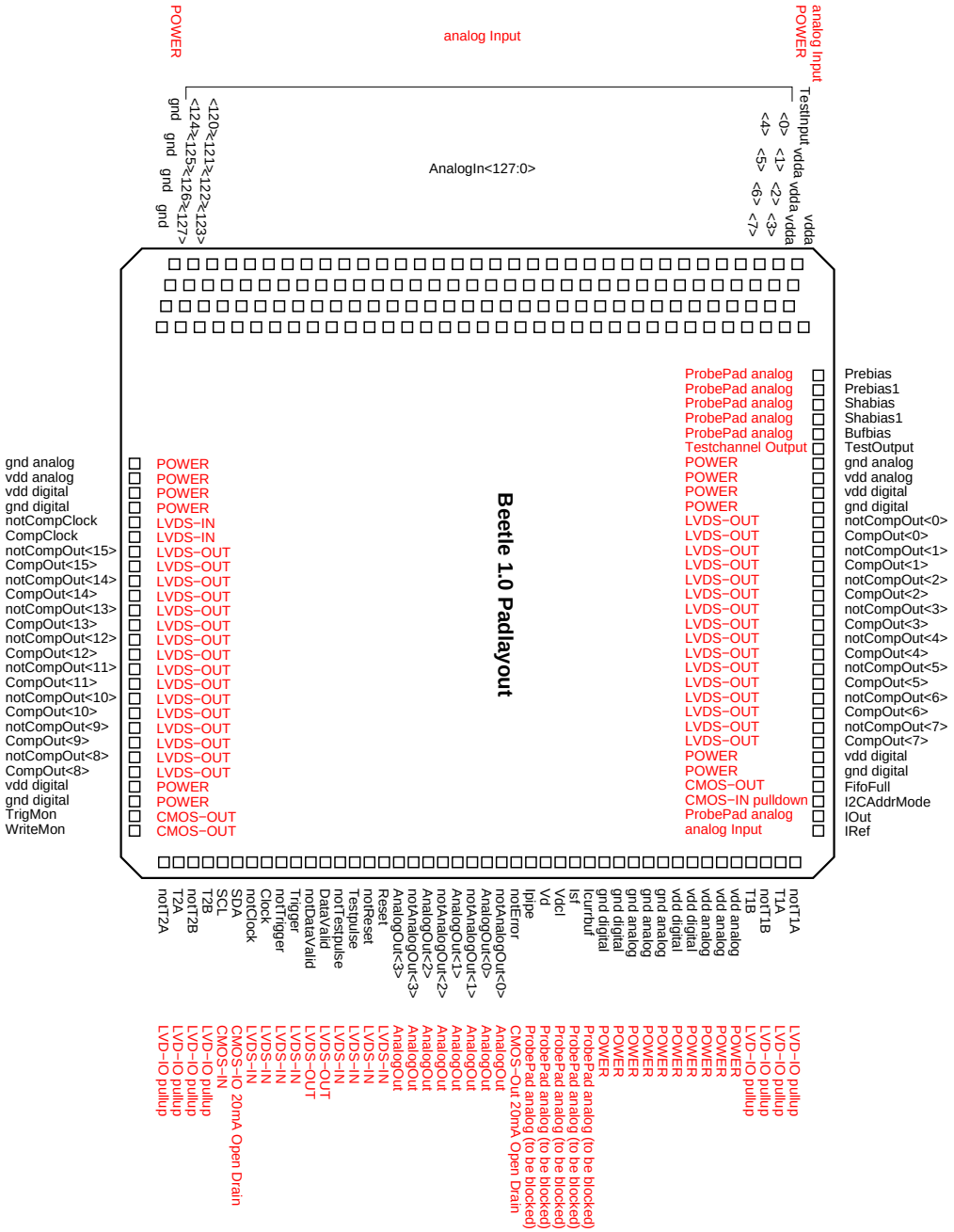

Figure A.1: Arrangement of the bond pads on the Beetle 1.0

Pad name	Pad type	Description
Prebias	Probe Pad	Preamplifer bias node
Prebias1	Probe Pad	Preamplifer bias node
Shabias	Probe Pad	Shaper bias node
Shabias1	Probe Pad	Shaper bias node
Bufbias	Probe Pad	Frontend buffer bias node
TestOutput	Analog Output	Testchannel output
gnd analog	Power Pad	Comparator power supply (0V)
vdd analog	Power Pad	Comparator power supply (2.5V)
vdd digital	Power Pad	Comparator pads power supply (2.5V)
gnd digital	Power Pad	Comparator pads power supply (0V)
notCompOut<0>...<7>	LVDS Output	Comparator output pads
CompOut<0>...<7>	LVDS Output	Comparator output pads
vdd digital	Power Pad	Comparator pads power supply (2.5V)
gnd digital	Power Pad	Comparator pads power supply (0V)
FifoFull	CMOS Output	Pipeline's fifo overflow
I2CAddrMode	CMOS Input pulldown	Adresse mode of I^2C interface (8bit/10bit)
IOut	Probe Pad	Current source bias node
IRef	Probe Pad	Current source bias node

Table A.1: Pads at the top side of the Beetle 1.0 from left to right

Pad name	Pad type	Description
notT1A	LVDS-IO pullup	Address/Daisy chain token
T1A	LVDS-IO pullup	Address/Daisy chain token
notT1B	LVDS-IO pullup	Address/Daisy chain token
T1B	LVDS-IO pullup	Address/Daisy chain token
vdd analog	Power Pad	Analog power supply (2.5V)
vdd digital	Power Pad	Digital power supply (2.5V)
gnd analog	Power Pad	Analog power supply (0V)
gnd digital	Power Pad	Digital power supply (0V)
Icurrbuf	Probe Pad	Current buffer bias node
Isf	Probe Pad	Multiplexer source follower bias node
Vdcl	Probe Pad	Readline reset level
Vd	Probe Pad	Pipeline readout amplifier bias node
Ipipe	Probe Pad	Pipeline readout amplifier bias node
notError	CMOS Open Drain	Synchronicity of mux readbit
AnalogOut<0>...<3>	Analog Output	Current buffer analog output
notAnalogOut<0>...<3>	Analog Output	Current buffer analog output
Reset	LVDS Input	Reset for pipeline control logic
notReset	LVDS Input	Reset for pipeline control logic
Testpulse	LVDS Input	Testcharge injection signal
notTestpulse	LVDS Input	Testcharge injection signal
DataValid	LVDS Output	Valid data on analog output
notDatavalid	LVDS Output	Valid data on analog output
Trigger	LVDS Input	Trigger signal input
notTrigger	LVDS Input	Trigger signal input
Clock	LVDS Input	Sampling/operation clock of PCL
notClock	LVDS Input	Sampling/operation clock of PCL
SDA	CMOS I/O Open Drain	Data input/output of I^2C interface
SCL	CMOS Input	Clock input of I^2C interface
T2B	LVDS-IO pullup	Address/Daisy chain token
notT2B	LVDS-IO pullup	Address/Daisy chain token
T2A	LVDS-IO pullup	Address/Daisy chain token
notT2A	LVDS-IO pullup	Address/Daisy chain token

Table A.2: Pads at the back side of the Beetle 1.0 from top to bottom

Pad name	Pad type	Description
gnd analog	Power Pad	Comparator power supply (0V)
vdd analog	Power Pad	Comparator power supply (2.5V)
vdd digital	Power Pad	Comparator pads power supply (2.5V)
gnd digital	Power Pad	Comparator pads power supply (0V)
notCompClock	LVDS Input	Comparator clock
CompClock	LVDS Input	Comparator clock
notCompOut<15>...<8>	LVDS Output	Comparator output pads
CompOut<15>...<8>	LVDS Output	Comparator output pads
vdd digital	Power Pad	Comparator pads power supply (2.5V)
gnd digital	Power Pad	Comparator pads power supply (0V)
WriteMon	CMOS Output	Write pointer passes pipeline column zero
TrigMon	CMOS Output	Trigger pointer passes pipeline column zero

Table A.3: Pads at the bottom side of the Beetle 1.0 from left to right

Pad name	Pad type	Description
TestInput	Amplifier Input	Test channel input
vdda	Power Pad	Frontend power supply (2.5V)
AnalogIn<0>...<127>	Amplifier Input	128 channel amplifier input
gnd	Power Pad	Frontend power supply (0V)

Table A.4: Pads at the front side of the Beetle 1.0 from top to bottom

Danke

Herrn Prof. Knöpfle danke ich für die sehr angenehme Betreuung meiner Doktorarbeit. Die mir gewährten Freiheiten und die ermutigenden Diskussionen mit ihm haben sehr dazu beigetragen, daß ich während meiner Doktorarbeit viel gelernt habe und mir auch das nötige Durchhaltevermögen angeeignet habe.

Ich danke Herrn Prof. Lindenstruth für die freundliche Übernahme des Zweitgutachtens.

Stellvertretend danke ich Herrn Prof. Hofmann und Herrn Prof. Eisele für die finanzielle Unterstützung durch das Max-Planck-Institut für Kernphysik und das "Graduiertenkolleg für experimentelle Methoden in der Kern- und Teilchenphysik".

Den Mitgliedern der HERA-B Vertexdetektorgruppe danke ich sehr für die Unterstützung meiner Arbeit und für alle interessanten Gespräche: Christian Bauer, Martin Bräuer, Matthias Eberle, Thorsten Glebe, Valery Pugatch, Michael Schmelling, Bernhard Schwingenheuer, Lothar Seybold, Frau Suppanz

Meinen Kollegen in der HERA-B/LHCb-Gruppe des ASIC-Labors danke ich für Hilfe, Kritik und die sehr angenehme Arbeitsatmosphäre im ASIC-Labor: Nils van Bakel, Daniel Baumeister, Harald Deppe, Wolfgang Fallot -Burghardt, Martin Feuerstack-Raible, Boris Glass, Sven Löchner, Josef Schweda, Nigel Smale, Uwe Stange, Ulrich Trunk

Allen anderen Mitgliedern des ASIC-Labors sei für die schönen Stunden vor und nach Feierabend gedankt: Ralf Achenbach, Joachim Boelsems, Dirk Droste, Dan Husmann, Michael Keller, Jörg Langeheine, Markus Loose, Ullrich Pfeiffer, Johannes Schemmel, Cornelius Schumacher, Achim Stellberger

Beatrice Bär, Frau Dannenberg, Herrn Damascheck, Frau Grimm, Herrn Hanke, Dagmar Hufnagel, Herrn Knebel, Klaus Schmidt, Frau Schmidt, Herr Ulses und Herrn Vörg danke ich für die Hilfe und die nette Zusammenarbeit bei meiner Arbeit im IHEP.

Rolf Lindner und Herrn H.-J. Hilke danke ich für die angenehme Zeit am CERN.

Meiner Familie und meinen Freunden sei für alles gedankt, was sich in Worten nicht ausdrücken lässt.