UNIVERSITÄT BONN
Physikalisches Institut

Design and Realisation of Integrated Circuits for the Readout of Pixel Sensors in High-Energy Physics and Biomedical Imaging

von
Ivan Perić

**Abstract: Radiation tolerant pixel-readout chip for the ATLAS pixel detector has been designed, implemented in a deep-submicron CMOS technology and successfully tested. The chip contains readout-channels with complex analog and digital circuits.
Chip for steering of the DEPFET active-pixel matrix has been implemented in a high-voltage CMOS technology. The chip contains channels which generate fast sequences of high-voltage signals. Detector containing this chip has been successfully tested.
Pixel-readout test chip for an X-ray imaging pixel sensor has been designed, implemented in a CMOS technology and tested. Pixel-readout channels are able to simultaneously count the signals generated by passage of individual photons and to sum the total charge generated during exposure time.**

Post address:
Nussallee 12
53115 Bonn
Germany

BONN-IR-2004-13
Bonn University
August 2004
ISSN-0172-8733

UNIVERSITÄT BONN
Physikalisches Institut

Design and Realisation of Integrated Circuits for the Readout of Pixel Sensors in High-Energy Physics and Biomedical Imaging

von
Ivan Perić

Dieser Forschungsbericht wurde als Dissertation
von der Mathematisch - Naturwissenschaftlichen Fakultät
der Universität Bonn angenommen.

Angenommen am:	28.09.2004,
Referent:	Prof. Dr. Norbert Wermes
Korreferent:	Prof. Dr. Manfred Forker

Design and Realisation of Integrated Circuits for the Readout of Pixel Sensors in High-Energy Physics and Biomedical Imaging

Dissertation
zur
Erlangung der Doktorgrades (Dr. rer. nat.)
der
Mathematisch-Naturwissenschaftlichen Fakultät
der
Rheinischen Friedrich-Wilhelms-Universität Bonn

vorgelegt von
Ivan Perić
aus
Belgrad

Bonn 2004

Angefertigt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Universität Bonn

1. Referent:	Prof. Dr. Norbert Wermes
2. Referent:	Prof. Dr. Manfred Forker
Tag der Promotion:	23.12.2004

Contents

1 Introduction **1**

2 Front-End Chip for the ATLAS Pixel-Detector **5**

2.1 Large Hadron Collider and ATLAS Detector 5
2.1.1 Introduction 5
2.1.2 Large Hadron Collider 6
2.1.3 ATLAS Detector 6
2.2 Detection of Charged Particles 11
2.2.1 Amplification of Sensor Signal 14
2.2.2 Signal Response of Charge-Sensitive Amplifier 19
2.3 Requirements on the Front-End Chip for the ATLAS Pixel-Detector 25
2.3.1 Pixel-Detector Module 25
2.3.2 Requirements on the Front-End Chip 27
2.4 Radiation Effects in Silicon and Radiation-Tolerant Chip Design 31
2.4.1 Introduction 31
2.4.2 Displacement Damage in Silicon 31
2.4.3 Ionization Damage in $Si0_2$ 32
2.4.4 Radiation-Tolerant Design for the ATLAS Pixel-Detector 35
2.5 Architecture and Design of the Front-End Chip 41
2.5.1 Introduction 41
2.5.2 Architecture of the Front-End Chip 43
2.5.3 Analog Pixel-Electronic 45
2.5.4 Readout System 80
2.5.5 Small Circuits 88
2.5.6 Measurements 99

3 Fast Sequencer-Chip for a DEPFET Pixel-Sensor **105**

3.1 DEPFET Pixel-Sensor 105
3.2 Architecture and Design of the Switcher Chip 109
3.2.1 Main Blocks of the Switcher Chip 109
3.2.2 Synchronization of many Switcher Chips 110
3.2.3 High-Voltage Part 111
3.2.4 Sequencer 117
3.2.5 Measurements 118

4 X-Ray Imaging Pixel-Chip with Joint Counting and Integrating Capabilities 121
4.1 Counting and Integrating 121
4.2 Architecture and Design of the CIX Chip 125
4.2.1 Integrating Block 125
4.2.2 Counting Block 128
4.2.3 Interface between Counting and Integrating Block 130
4.2.4 Pixel Cell 131
4.2.5 Measurements 131

5 Summary and Outlook 135

A Noise in Charge-Sensitive Amplifier 137
A.1 Types of Noise 137
A.1.1 Mathematical Modelling of Noise 138
A.1.2 Thermal and 1/f Noise 140
A.1.3 Output Noise 141

B Modelling of MOS Transistors 145
B.1 DC-Current Equations 145
B.1.1 Gradual-Channel Approximation 146
B.1.2 Channel-Charge - Exact Solution 147
B.1.3 Channel Charge - Approximative Solutions 152
B.1.4 Drain-Current Equations 157
B.1.5 Drain Current in Strong Inversion 161
B.1.6 Short-Channel Effects 171
B.2 Transient-Current Equations 172

Chapter 1

Introduction

The goal of this work is to design various electronic microchips used to read out and process the signals generated by semiconductor radiation sensors. The semiconductor radiation sensors have application in many fields, such as particle physics, medical imaging *etc.* The main advantages of semiconductor sensors in comparison to the electronic sensors of older type, like gaseous ionization chambers, are their fast response time, low energy threshold and the possibility to integrate many small sensor units on the same piece of semiconductor. Very fast detectors with high spatial resolution can be produced.

The development of radiation sensors must be accompanied by the miniaturization and improvement of the electronic circuits used for sensor signal processing. High spatial resolution and fast signal response of a sensor does not mean a lot unless the processing circuits are small enough and able to cope with the fast sensor signals. The complexity of the electronic circuits for radiation sensors has continuously increased in recent years. This is correlated to the development of semiconductor technology, which gives the possibility to integrate many electronic elements in a small chip area. The first microchip in the chain of sensor signal processing, called "front-end" chip, has today much more functionality than before. It will not only amplify the charge delivered by a radiation sensor, but also measure and characterize the signal, produce appropriate binary information (for example the particle hit time in the case of a particle detector) and make the decision whether the information has to be disregarded or transmitted. The design of such microchips has become quite demanding task.

FE-I - Pixel-Readout Chip for the ATLAS Tracking Detector

A large part of this work is devoted to the design of the front-end microchip for the two-dimensionally segmented sensor that represents the innermost layer of the ATLAS particle detector. The ATLAS detector is currently under construction at CERN[1], Geneva. It is a multi-purpose detector on the new Large Hadron Collider (LHC). The requirements for the design of the ATLAS pixel front-end chip will be explained in detail later.

The ATLAS pixel front-end chip is connected with a sensor segmented into picture elements (pixels). The chip itself contains a pixel matrix with about 3000 elements. Every pixel comprises analog and digital circuits. The task of the analog pixel block is to amplify the sensor signal and to distinguish it from noise. The digital circuits measure the signal time, prepare the digital hit information and send it to the periphery elements. The

[1] Centre European pour la Recherche Nucleaire or European Laboratory for Particle Physics

periphery elements will decide whether the information is worth of being transmitted. This short description cannot illustrate the full complexity of the chip; instead we will only mention that a pixel cell in the latest version of the ATLAS pixel front-end chip contains almost 1000 transistors, while the whole chip includes about 3.5 million. It is important to note that the whole ATLAS pixel detector must be radiation tolerant. The total dose in the innermost detector layer is expected to reach 50 Mrad (500 kGy) after ten years of operation.

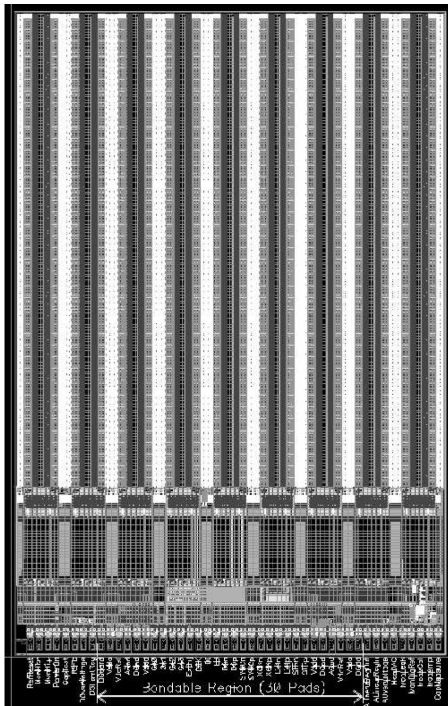


Figure 1.1: Layout of the FE-I chip.

SWITCHER - A Fast High-Voltage Sequencer-Chip for a DEPFET Detector

The second microchip that is designed within the scope of this work is the digital high-voltage chip used to generate the fast signals of variable amplitude needed for the steering of the DEPFET sensor matrix. This chip is called SWITCHER. The functional principle of the DEPFET sensor will be explained in detail later; to summarize briefly DEPFET is a novel conception of semiconductor pixel sensor, where sensor pixels and simple pixel amplifiers are integrated together on the same piece of semiconductor. This feature makes the DEPFET detectors suitable for very low noise applications. Recall that in the case of the ATLAS pixel detector, semiconductor sensors and pixel amplifiers are placed in separated integrated circuits. The pixel-segmented front-end chips are connected to the quite simple sensor module consisting of $p{-}n$ diodes. The readout concept of the DEPFET detector is completely different. We have two one-dimensionally segmented readout chips, whereas simple pixel amplifiers, consisting of one transistor, are placed on the sensor chip. A readout chip (SWITCHER) will generate the voltage sequence needed to select and clear DEPFET pixel rows. The other chip (current readout chip - CURO) will measure the currents generated by the DEPFET amplifiers selected by the SWITCHER.

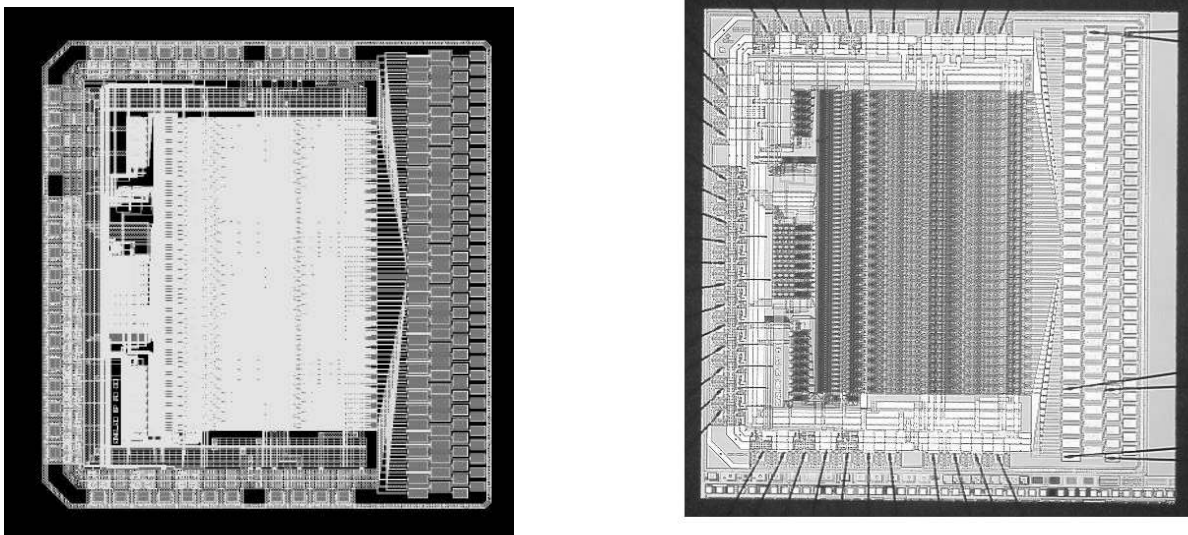


Figure 1.2: Layout and photomicrograph of the SWITCHER chip.

CIX - an X-Ray Imaging Pixel-Chip with Joint Counting and Integrating Capabilities

Both described detector concepts can have application not only as detectors in particle accelerators, they can be used for example as imaging detectors for medical diagnostic.

The study of different readout possibilities for imaging pixel detectors is also a goal of this work. The starting point is the application of a detector similar to the ATLAS pixel detector for medical X-ray imaging. The sensor and the readout chip are the separated entities. The readout chip contains the pixel cells that can amplify and process the sensor signals. In recent years, two complementary readout approaches for such imaging pixel detectors have been developed. One is to implement the pixel electronics that sums the charge generated by many photons that traverse a sensor pixel during the exposition time. The other is to implement a high gain amplifier followed by a counter and to count the charge signals, which means photons, individually. Each approach provides only one side of the full picture. The charge integration gives the information about the mean energy deposited in a pixel - the energy quantization remains, however, unknown. In the case of photon counting, the energy is not measured, however, it can be indirectly calculated if we know the energy spectrum of the incident radiation. Within the scope of this work, a small pixel-readout chip, called CIX, is designed. The chip is able to count the photons and to measure the deposited energy simultaneously. In this way, we hope to ease the identification of the imaged structures.

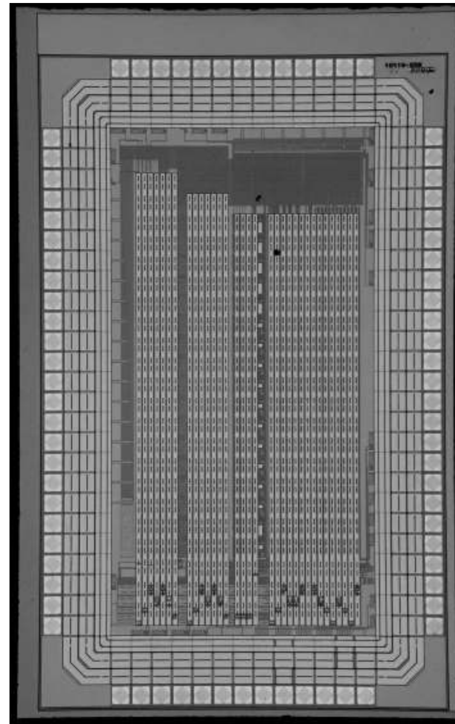


Figure 1.3: Layout and photomicrograph of the CIX chip.

Chapter 2

Front-End Chip for the ATLAS Pixel-Detector

In this chapter we will describe the design of the front-end chip for the ATLAS pixel detector. ATLAS is a multi-purpose particle detector on the large hadron collider, CERN, Geneva. In the first sections of this chapter we will shortly introduce the basic ideas of particle physics, describe the large hadron collider and the ATLAS detector. The ATLAS pixel detector will be introduced in the fourth section. The requirements on the front-end chip will be described as well. One of the biggest challenges in the design of the pixel detector is to produce the electronic components that are radiation tolerant. We devote the whole section (section five) to the topics of radiation damage and radiation tolerant chip design. The sixth section describes the design of the ATLAS pixel front-end chip.

2.1 Large Hadron Collider and ATLAS Detector

2.1.1 Introduction

The research of the structure of matter did immense progress in the last decades. The use of the large particle accelerators enabled the researchers to explore the structures with a size down to 10^{-18} m. One has succeeded to summarize all the experimental results, which led to the standard model of elementary particle physics. According to the standard model, all matter observed so far is built by particles with half-integer spin which obey the Fermi-Dirac statistics [63]. These particles are called fermions. The elementary fermions are classified into quarks and leptons. Quarks interact through the electroweak and strong interaction, while leptons interact only electroweakly. The dynamic interaction between fermions occurs through the exchange of particles with integer spin. These spin-one particles obey the Bose-Einstein statistics and they are called bosons.

In the formulation of the standard model, the Lagrange density has been derived starting from the observed symmetries and the principle of local gauge invariance. Technically, the standard model is a local gauge theory with the $\mathrm{SU}(3)_\mathrm{C} \times \mathrm{SU}(2)_\mathrm{L} \times \mathrm{U}(1)_\mathrm{Y}$ gauge symmetry group. A theoretical problem arises if one tries to add the mass terms in the Lagrangian density. These terms should correspond to the masses of fermions and bosons. The gauge invariance of the theory becomes then destroyed. A solution is to postulate a scalar doublet with a potential which is called the Higgs-field. The vacuum expectation value of the neutral component of the Higgs-field does not vanish. The

existence of this non zero component causes the spontaneous symmetry braking of the electroweak interaction at the energies of about 100 GeV. The gauge bosons of weak interaction and the charged fermions acquire their effective masses through the interaction with the Higgs-field. Additionally a new particle is added to the particle spectrum of the standard model - the Higgs boson. The Higgs boson is the only unobserved particle of the standard model so far. The mass of the Higgs boson is a free parameter and it has to be determined by experiments.

Despite of the success of the standard model to explain experimental results, it is believed that it is only a low energy approximation of a more fundamental theory. The appearance of new physical phenomena is expected at energies above a few TeV. These new phenomena should be described by a new more fundamental theory. The main motivation for the construction of the large hadron collider is to explore the complete energy range for the standard model Higgs boson mass and to search for the new phenomena at the energy scale beyond a few TeV.

2.1.2 Large Hadron Collider

The large hadron collider [73] is currently under construction at CERN. Two proton beams will be accelerated in opposite directions and let collide in a few interaction points along the beam pipes. The accelerator will be placed in a 27 km circumference tunnel. The LHC beam energy will be limited only by the strength of magnets used to keep the beams in circular track. The magnetic field generated by special super-conducting dipole magnets allows a proton energy of 7 TeV. The center of mass energy will be 14 TeV. Since LHC works only with protons and not with protons and antiprotons, which was common for the accelerators of older design, it will be possible to achieve much higher beam intensities. LHC is expected to achieve a luminosity of 10^{34} cm^{-2}s^{-1}

The large hadron collider will be filled with 450 GeV protons, delivered from the Sp$\overline{\text{p}}$S[1] and its pre-accelerators. The protons will be accelerated to 7 to 7 TeV. After acceleration, the beams will counter-rotate for several hours colliding at interaction points until the beam degradation becomes such that the machine has to be emptied and refilled. LHC will have four interaction points. Large experimental halls located at these points will accommodate four experiments. The experiment called ALICE is dedicated to the study of the quark-gluon plasma.[2] The experiment called LHCb will study the CP violation in the B system. The two remaining interaction points serve to the multi purpose experiments CMS and ATLAS.

The LHC proton beams will be bunched in packets (bunches). The bunch separation expressed in time will be 25 ns.

2.1.3 ATLAS Detector

Introduction

The ATLAS detector [70, 71] has a multi-layer structure characteristic for such type of particle detectors, fig. 2.1. It is about 22 m high and has an overall weight of 7000 tons.

[1]Super Proton Antiproton Synchrotron

[2]In this experiment LHC will be filled with heavy Pb^{+} ions.

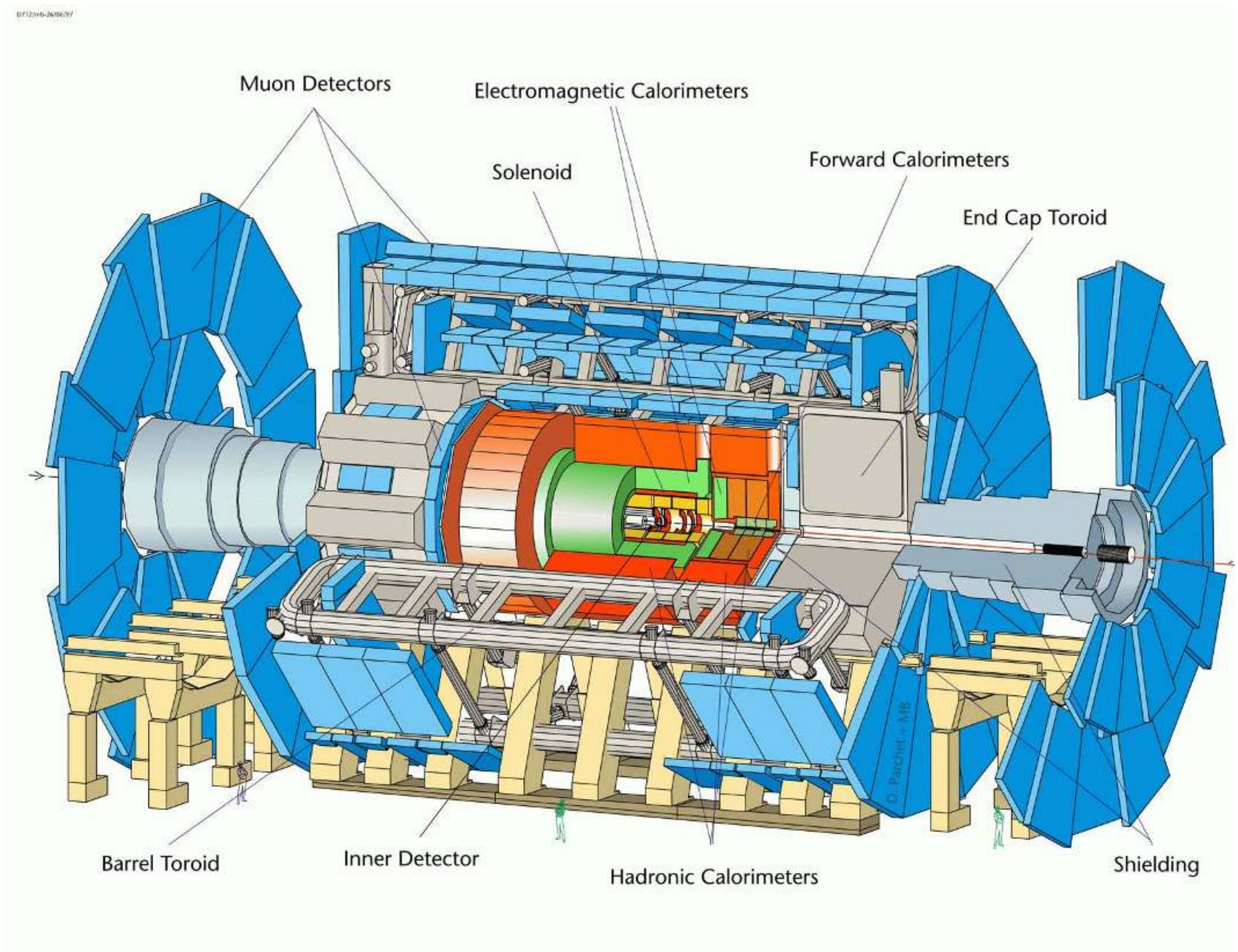


Figure 2.1: ATLAS detector.

In the next few sections we will shortly describe the detector, starting from its innermost sub-components.

Inner Detector

The inner tracking detector [69] is the innermost part of the ATLAS detector. The inner detector cavity (6.8 m long and 1.15 m in radius) is directly enclosed with a superconducting solenoid magnet generating an axial field of 2 T. The purpose of the inner detector is to measure the trajectories of charged particles. The trajectories will be bent in ϕ direction due to magnetic field, which allows momentum measurements. The inner detector consists of three components: silicon pixel detector, semiconductor tracker and transition radiation tracker.

Silicon Pixel-Detector

The silicon pixel detector [72] will be placed directly around the LHC interaction point. Since it is exposed to a large particle flux, a high granularity is needed. The pixel detector

is arranged in three cylindrical (barrel) layers and two times three disks in forward regions, fig. 2.2.

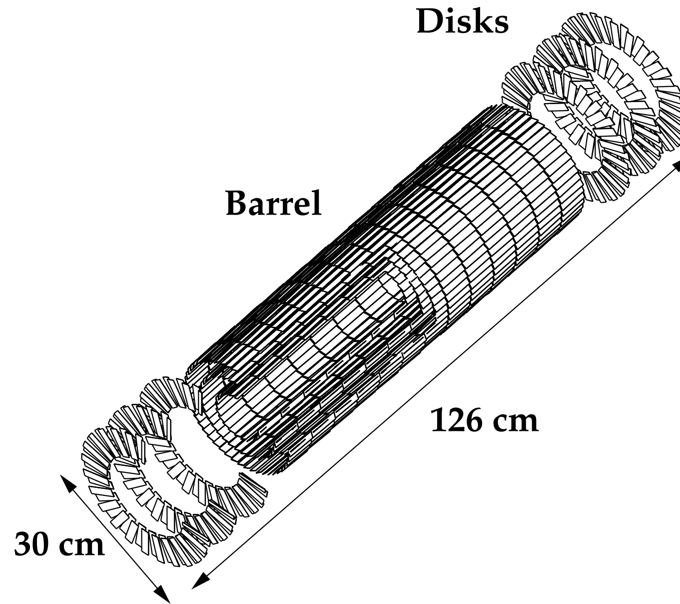


Figure 2.2: Pixel detector.

The barrel layers will have radii of 5 cm, 9 cm and 12 cm and a length of about 80 cm. Disk and barrel support structures are covered with identical sensing units - detector modules. A detector module has a sensitive area of 16.4×60.8 mm^2, segmented into 50×400 μm^2 pixels. The pixel detector measures three points on a particle trajectory. Because of the large LHC background radiation[3] it is not possible to reconstruct every trajectory by knowing only three points per track and additional tracking components are necessary for efficient track identification. However, the track coordinates measured by the pixel detector are of the greatest importance for the precise determining of the vertex position, due to the small value of the first ρ coordinate and the large ratio between the second (or third) ρ coordinate with respect to the first one. Fig. 2.3 shows an event display. The pixel detector has 1744 modules. This leads to a total sensitive area of about 1.74 m^2. The sensor signals are amplified with 80.4 million pixel amplifiers - more than in all other ATLAS components together.

Semiconductor Tracker

The next component of the inner detector is the semiconductor tracker (SCT). SCT consists of four barrels and two times nine disks in forward regions. The particle flux in the region covered by this detector is much lower than the flux through the pixel detector and the simpler silicon strip modules can be used.

Transition-Radiation Tracker

The outer component of the tracking system is the transition radiation tracker (TRT). The sensing elements are metal straws filled with a gas with a centrally placed high-

[3]We expect about 1000 tracks per bunch-crossing.

voltage wire. In the barrel region the straws are oriented along the z axis. They are placed one near the other, making 64 ρ layers. In the forward (wheel-shaped) detector part, the straws are oriented in the ρ direction. The passage of a charged particle ionizes the gas and produces electric signals which can be measured at the straw wires. In the barrel region 64 $\rho - \phi$ track coordinates are measured. This allows the identification of individual tracks even at highest luminosities, see fig. 2.3. The z track coordinate is not measured.

TRT is able to distinguish between electrons and the other charged particles. Very fast electrons emit transition radiation X-ray photons when pass through the radiator between straws. The photons additionally ionize the xenon gas in straws, which leads to higher electric signals.

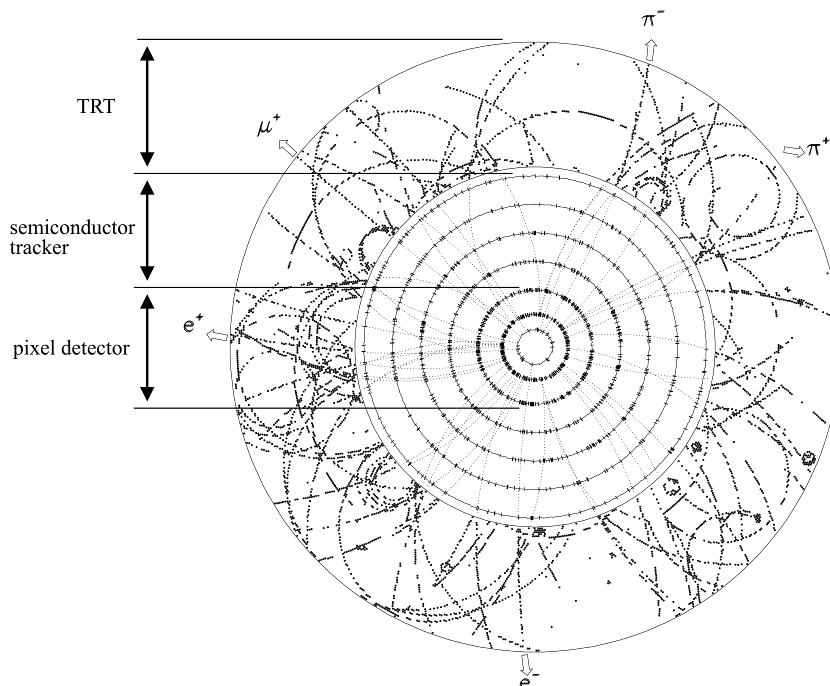


Figure 2.3: Inner detector and an event display.

Calorimeters

The calorimeters surrounds the the solenoid magnet, see fig. 2.1.

The calorimeter system of the ATLAS detector consists of an electromagnetic and a hadronic part.

Electromagnetic Calorimeter

The electromagnetic (EM) calorimeter encloses the solenoid magnet and the inner detector. Two end-caps are placed in the forward regions, a barrel part surrounds the solenoid, fig. 2.1. The purpose of the EM calorimeter is to absorb the electrons and photons coming from the interaction point and to measure their energies.[4] The calorimeter is of sampling

[4]The calorimeter is called "electromagnetic" since electrons and photons interact electromagnetically with the absorbing material. During the interaction with atom nuclei, electrons and positrons emit

type. Many layers of passive (absorbing) and active material are placed along the particle track (in the r direction). Lead is chosen as the passive material. The active material is liquid argon filled in the gaps between the lead plates. In the middle of each liquid argon gap there is a readout plate electrode. When an electron or a photon hits an absorber plate, it generates electromagnetic shower. The shower passes the next liquid argon gap and ionizes the argon atoms in it. The electrons produced by the ionization will be collected by electrodes. The electromagnetic shower will be multiplied in the next absorber layer. The signal measured in the following liquid argon gap will be therefore higher. If we measure the signals along the shower track, the energy deposited by the particle can be calculated from the signal increase.

Routing out the signal electrodes without making unsensitive regions between calorimeter segments is not a trivial task. In order to keep the hermeticity of the calorimeter, the absorbing plates of the electromagnetic calorimeter have a special accordion shape.

Hadronic Calorimeter

The hadrons are not absorbed in the electromagnetic calorimeter. Their energies are measured in the hadronic calorimeter which surrounds the electromagnetic one fig. 2.1.

The hadronic calorimeter is realized as a sampling type calorimeter. In the end-cap region, the radiation tolerant liquid argon technology is used. The absorbing layers are copper sheets separated by liquid argon gaps. The plate electrodes are placed in the middle of every gap.

Radiation is much weaker in the barrel region, which gives more freedom by the choice of technology. Iron absorbing plates are placed in the space between plastic scintillating tiles. The scintillation light, produced in tiles is read out by wavelength shifting fibres to the photomultipliers, outside of the calorimeter. The tiles are placed following the special pattern that enables the routing out of fibres by keeping of hermeticity.

Forward Calorimeters

The calorimeter components that directly enclose the beam pipe are exposed to very high radiation. They must be designed in a special way. The forward calorimeter has three modules. The first module is the electromagnetic calorimeter with copper as passive material. The second and third module are hadronic calorimeters with tungsten absorber. Mechanically, all modules are simple. They consists of single absorber bodies with arrays of cylindric electrodes, placed in the holes of the absorber body. The electrodes are oriented along the z axis. The gap between the electrodes and the absorber is filled with liquid argon.

Muon Spectrometer

The muon system is the largest part of the ATLAS detector. It measures muon momenta completely independent of the inner detector. Muons usually follow the interesting events with large energy transfer and a fast detection of these particles is highly desirable. In the inner detector, a muon track is one between hundreds. It cannot be identified without

---

Bremsstrahlung photons. The photons are then converted into e - e^+ pairs. This leads to the creation of electromagnetic shower that contains electrons, positrons and photons.

a complex off-line analysis. The situation is completely different outside the calorimeters. Muons and neutrinos are the only particles that can escape from the hadronic calorimeter. A single signal measured in the outer areas can be immediately assigned to a muon. This signal can be used as a base for the trigger pulse. The trigger will automatically start the readout sequence, while the other detector components send their data concerning the marked event. [5]

The magnetic field needed for the measurement of muon momenta is produced by three big air core superconducting toroids. Track coordinates are measured by precision chambers. They are arranged in three barrels and four disks on each side. Sensing elements are the monitored drift tubes (MDT) everywhere except in the regions closest to the beam pipe, where the cathode strip chambers (CSC) are used. MDT are gas filled tubes with high-voltage wire running through the center. CSC are multiwire proportional chambers. The spatial resolution is here obtained by the cathode strips. Both types of detectors can measure only one coordinate (perpendicular to the tubes/strips). The z coordinate will be measured in the barrel region and the ρ coordinate on the disks.

In addition to the high-precision chambers, the muon spectrometer contains also several layers of the resistive plate chambers (RPC) and the thin gap chambers (TGC). These detectors generate fast signals and measure two coordinates, with a typical space time resolution of 1 cm $\times$ 1 ns. This fast signal is used for triggering. RPC are placed in the barrel region, close to the MDT middle barrel. TGC are located near the middle MDT disk. RPC is a gaseous strip detector with two layers of orthogonal strips. TGC is a multiwire proportional chamber with cathode strips orthogonal to the anode wires.

2.2 Detection of Charged Particles

Semiconductor detectors have many advantages over gaseous detectors. The average energy required to create an electron-hole pair is much smaller than the ionization energy in a gas. Due to greater density of semiconductors compared to gases, the stopping power is much larger. Semiconductor detectors are fast and can be very small.

In this section we will discuss the interaction of high energy charged particles and semiconductor material. We will also talk about the generation of electric signal in sensor and introduce the values like the charge collection distance and the charge collection time.

Interaction of Charged Particles and Matter

A heavy charged particle loses its energy mostly in Coulomb collisions with atomic electrons. The energy loss in a single collision is generally a very small fraction of particle's total kinetic energy. However, in normally dense materials the number of collisions per

[5] The calorimeters can be also used for triggering. The signature is here missing transversal energy. The transversal energy is defined as $E \sin\theta$. In the case of high-energy particles, the transversal energy is directly proportional to the transversal momentum. Due to momentum conservation, the total transversal momentum of all particles produced in a collision must be zero, since the colliding protons do not have the transversal momentum component. If the calorimeters measure the value different than zero, it means that a particle invisible for the calorimeter took out the missing momentum. It is most probably neutrino, but it can be also the identification of new physical phenomena. This triggering method requires hermeticity and a large spatial angle coverage of the calorimeter.

unit path length is so large that a large cumulative energy loss is observed even in relatively thin layers of material. The mean energy loss of a heavy charged particle per unit path length (the quantity called also the "stopping power") was calculated by Bethe and Bloch (see [3]):

$$-\left\langle \frac{dE}{dx} \right\rangle = D\frac{Z}{A}\rho z^2 \Phi(\beta) \tag{2.1}$$

with

$$\Phi(\beta) = \frac{1}{\beta^2}\left[\ln\left(\frac{2m_e c^2\beta^2\gamma^2 W_{max}}{I^2}\right) - 2\beta^2 - \delta - 2\frac{C}{Z}\right] ,$$

where

D:	$\mathrm{D} = 2\pi N_a {r_e}^2 m_e c^2 = 0.1535 MeV cm^2/mol$
r_e:	classical electron radius $= 2.817 \cdot 10^{-13}\ cm$
m_e:	electron mass
N_a:	Avogadro's number $= 6.022 \cdot 10^{23}\ mol^{-1}$
I:	mean excitation potential
Z:	atomic number of absorbing material
A:	atomic weight of absorbing material
ρ:	density of absorbing material
z:	charge of incident particle in units of e
β:	v/c of the incident particle
γ:	$1/\sqrt{1-\beta^2}$
δ:	density correction
C:	shell correction
W_{max}:	maximum energy transfer in a single collision $W_{max} \approx 2m_e c^2\beta^2\gamma^2$

The stopping power is to a good approximation proportional to electron density in absorbing material (given by $\rho N_a \frac{Z}{A}$) and to the square of the incident particle's charge (z^2). Otherwise it depends mainly on the velocity of the incident particle. It decreases with $1/\beta^2$ for increasing velocity until reaching a minimum around $\beta \approx 0.96$, *i.e.* when the particle's energy becomes about 3.6 times higher than its rest energy Mc^2. Beyond this point the mean energy loss reaches a plateau after a small relativistic rise. Another interesting property of the Bethe-Bloch formula is that the energy loss does not depend on the mass of the incident particle. This has the consequence that the minimal energy loss of different particles with the same charge has the same value. Generally, due to small increase of the energy loss beyond its minimal point, all particles with energies higher than $\approx 3.6Mc^2$ loose, approximately, the same energy per unit path length. Such particles ($E > 3.6Mc^2$) are called "minimum ionizing particles".

Mean and Actual Energy Loss

In order to calculate the thickness of absorbing material required for the detection of particles with certain energy, it is of great importance to know the total amount of energy

deposited by an incident particle passing through a detector of some given thickness. The mean deposited energy after the path length x can be calculated by integrating the equation

$$dE = \left\langle \frac{dE}{dx} \right\rangle dx$$

and calculating the inverse function of $x(\Delta E)$. $\left\langle \frac{dE}{dx} \right\rangle$ is the stopping power. The calculation of the mean deposited energy can be simplified if we suppose that the penetrating particle has so high initial energy E_0 that the energy loss ΔE can be neglected with respect to the initial energy E_0. The stopping power is than constant and the mean deposited energy after a path length x is simply

$$\Delta E = \left\langle \frac{dE}{dx} \right\rangle_{E_0} x \ . \tag{2.2}$$

Due to statistical nature of energy loss mechanism, the actual value of the lost energy has some distribution around the mean value given by eq. 2.2. In the case of thick absorbers the number of collisions is huge and actual energy loss can be described by a Gaussian distribution. For thin absorbers the distribution describing actual energy loss has a longer tail at higher energies. Such asymmetrical function is called the Landau distribution. The Landau distribution takes into account rare scattering events with high energy transfer. These events increase the mean value of energy loss with respect to its most probable value. Let us illustrate this with an example. A minimum ionizing particle ($E \gg 3.6Mc^2$) looses in a 250 μm thick silicon sensor about 97.5 KeV of its energy (mean value). However, the most probable energy loss is only 70 KeV. The energy is dissipated in the sensor material, a part is converted to thermal energy (phonon creation) and the other part goes to the creation of electron-hole pairs. In average, 3.62 eV of the energy deposited in silicon leads to the creation of one electron-hole pair. Therefore a minimum ionizing particle generates in average 97.5 $KeV/3.62\ eV \approx 26900$ electrons and most probably 70 $KeV/3.62\ eV \approx 19300$ electrons.

Signal Generation in a Semiconductor Radiation-Detector

Generated electrons and holes recombine unless they are kept separated by an electric field. The application of the electric field is therefore necessary for signal detection. The recombination rate can be described by a time-constant called the recombination life time (τ). In order to estimate the recombination in a sensor it is useful to define the parameter called the charge collection distance (λ):

$$\lambda = \mu \tau E \ . \tag{2.3}$$

μ is the carrier mobility, E is the electric field and τ is the carrier recombination life time. The charge collection distance describes the mean path length that the carriers, generated by radiation, drift before they get recombined.

In order to estimate the electric signal produced by a particle traversing a sensor, let us consider the simplest implementation of semiconductor sensor - solid state ionization chamber [3]. It is a piece of sensing material placed between two parallel plate electrodes,

fig. 2.4. Suppose that a high energy particle generate some amount of electrons and holes at the distance x from the positive sensor electrode. The electrons and holes start to drift in the direction of the electric force. The motion of the charge carriers induces current in the external circuit, fig. 2.4. The current can be easily calculated from the energy conservation law. If some charge (q) moves in a constant electric field (E) along a certain short path (dx), the work done by the electric force is $W_E = qEdx$. If we assume that the internal energy of the system does not change, the equal work has to be done by the generator connected to the sensor electrodes. The work done by the generator is $W_G = Udq_{sig}$, U is the voltage between the sensor electrodes, dq_{sig} is the signal charge that flows through the external circuit. In the case of one-dimensional geometry (the parallel plate capacitor), the electric field between electrodes, E, is constant and equal to U/l, where l is the distance between the sensor electrodes. Finally, the equation $W_G = W_E$ leads to $dq_{sig} = dx\ q/l$ or $i_{sig} = dq_{sig}/dt = (dx/dt)\ (q/l) = v\ q/l$; i_{sig} is the signal current and v is the drift velocity. The electrons generated at the distance x from the positive sensor electrode need the time $t_e = x/v_e$ to reach the electrode. The holes generated at the distance $l - x$ from the negative electrode will be collected after the time $t_h = (l - x)/v_h$. Due to lower mobility of holes in silicon, their drift velocity is smaller than the electron drift velocity. Fig. 2.5 shows the current signals induced by hole and electron motions as functions of time. If the electrons reach the positive and the holes the negative electrode, the total measured charge signal q_{sig} is equal to the total positive generated charge q. This can be seen in fig. 2.5: $q_{sig} = q\ (x/l) + q\ (l - x)/l = q$. However, if the electrons or holes get recombined on their way to the electrodes, the charge signal that is observed until the moment of recombination is followed by an equivalent charge signal of opposite sign. The measured current signal is bipolar and the total collected charge zero. In order to avoid such signal loss, the charge collection distance given by eq. 2.3 should be much larger than the distance between sensor electrodes.

Let us now suppose that the ionization in the sensor shown in fig. 2.4 occurs homogenously along the particle's track. In this case, the total current signal has the time dependence shown in fig. 2.6. This can easily be verified by the integration of the signals shown in fig. 2.5. It is assumed that electrons have three times higher mobility than holes with a consequence that the collection of electrons is three times faster. However, the areas below the electron and hole induced currents as functions of time are equal. Both areas correspond to one half of the total positive charge Q generated. Since holes are less mobile, the charge collection time is limited by the hole collection. The typical charge collection time, assuming the bias voltage of 100 V, the detector thickness of $l = 250\ \mu m$ and the hole mobility of $\mu_h = 500\ cm^2/Vs$, is $t = l/\mu_h E = 15\ ns$

2.2.1 Amplification of Sensor Signal

Introduction

The sensor, introduced in the previous section, can be electrically modelled as a parallel connection of a capacitor, constant current source[6] and a current pulse source, see fig. 2.7. Note the nomenclature used in fig. 2.7. Time dependent current or voltage is denoted by the lowercase letter i or v followed by an uppercase subscript (for example i_{SIG}). The

[6]Due to the finite resistance of sensor material a constant current always flows when a voltage is applied between the sensor electrodes.

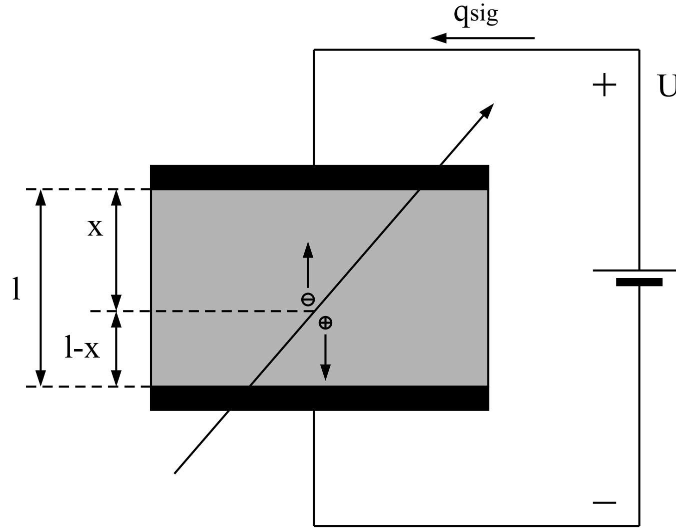


Figure 2.4: Solid state ionization chamber.

uppercase letter I or V is used to describe time independent current or voltage, or the mean value of the time dependent signal (for example I_{const}). The difference between a signal and its mean value (small signal) is denoted by the small letter i or v with a subscript in lowercase (for example i_{sig}). The typical values of electrical parameters describing a silicon pixel sensor are: $I_{const} = 10\,\text{nA}$ and $C_{det} = 400\ fF$. We will assume that the charge collection time is much smaller than the response time of the measuring system connected to the sensor. In this case, the current signal generated by the sensor can be modelled by a Dirac function, for example $i_{sig}(t) \approx 20000\ e \cdot \delta(t)$.

Let us calculate the voltage step at the sensor electrodes produced by the charge signal of 20000 e. We obtain $\Delta v = Q_{sig}/C_{det} = 20000\ e/400\ fF \approx 8\ mV$. This is a very small signal and if we want, for example, to measure the particle hit time by using a digital timing device, we must first amplify and shape the sensor signal in order to make it understandable for digital circuits.[7] In the following section we will introduce an amplifier that can be used to amplify a weak and fast sensor signal in a way that allows its further digital processing.

Field-Effect Transistor

Before we start talking about the structure of the sensor signal amplifier, we will briefly introduce the main building part used to implement any type of electronic circuit - transistor. Transistors can be used either as electronic switches, which allows the implementation of logic functions in digital devices, or as amplifying elements in analog circuits. The most frequently used transistor type is the Metal Oxide Semiconductor Field Effect Transistor (MOSFET) [1]. The field effect transistor is probably the most important innovation that caused the computer and information revolution. The schematic cross-section of an n-channel MOSFET is shown in fig. 2.8. We can recognize two islands of highly doped semiconductor (n^+ doped source and drain) placed in a semiconductor bulk of the oppo-

[7] The digital electronic devices are always implemented by using a certain type of logic gates. Such circuits operate only with well defined potential levels, for example the positive supply potential is logic 1 and the negative supply potential is logic 0.

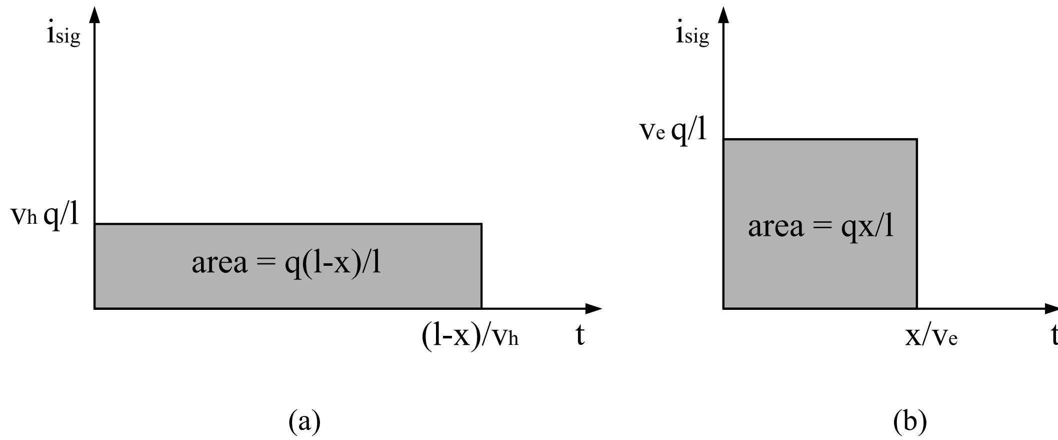


Figure 2.5: (a) - the current signal produced by the holes generated at the distance $l - x$ from the negative electrode of the solid state ionization chamber shown in fig. 2.4. (b) - the current signal produced by the electrons generated at the distance x from the positive electrode.

site doping type (p). The n^+ islands and the p bulk form two $p - n$ diodes which are connected serially, anode with anode. Such a diode connection prevents the current flow between drain and source. However, by applying a positive voltage between the third electrode, called gate and bulk, we can cause the "type inversion" of the surface bulk region and in this way establish an ohmic connection between source and drain contacts. The gate electrode is placed above the bulk area between the n^+ regions. The gate is isolated from the bulk by a thin insulating layer, which is in the case of a silicon based technology made of silicon-dioxide ($Si0_2$), see fig. 2.8.

The functional principle of a MOSFET is simple. If we bias gate positively with respect to source and drain, the electrons from the bulk will gather at the interface between the semiconductor and the insulator, see fig. 2.8. In this way, a negatively charged channel will be formed. The electron density in the channel and therefore the channel resistance, depends on the applied gate-source (and gate-drain) voltage. One defines here the threshold voltage. When the gate-source voltage exceeds the threshold, an ohmic connection is established between drain and source and the transistor is considered to be "on".

Let us assume that the gate-source voltage is above the threshold. If we increase the drain-source voltage, by keeping the other electrode potentials constant, the transistor (drain-source) current will increase until the gate-drain voltage becomes lower than, typically, 0.5 V (threshold voltage). For higher drain potentials the drain end of the channel vanishes. The drain-source current saturates and becomes virtually independent of the drain potential. This behavior can be seen in fig. 2.9 (a). Fig. 2.9 shows the input and output current-voltage characteristics of an n-channel MOSFET. The source potential is taken as the reference and the bulk-source voltage is assumed to be zero. The value of the saturated current depends on the gate-source voltage. The current saturation effect is of great importance in analog electronics, it allows the implementation of high gain voltage amplifiers by using MOSFET transistors.

Beside n - channel MOSFETs, it is also possible to produce p - channel devices. A p - channel transistor has p^+ electrodes in n-doped bulk. The channel of such transistor is made of holes. In the following text we will refer to an n - channel MOS transistor as

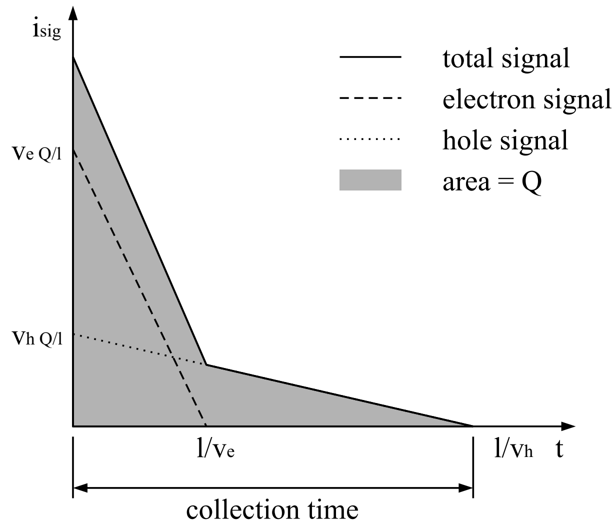


Figure 2.6: The current signal produced by a particle traversing the solid state ionization chamber. Q is the total generated charge.

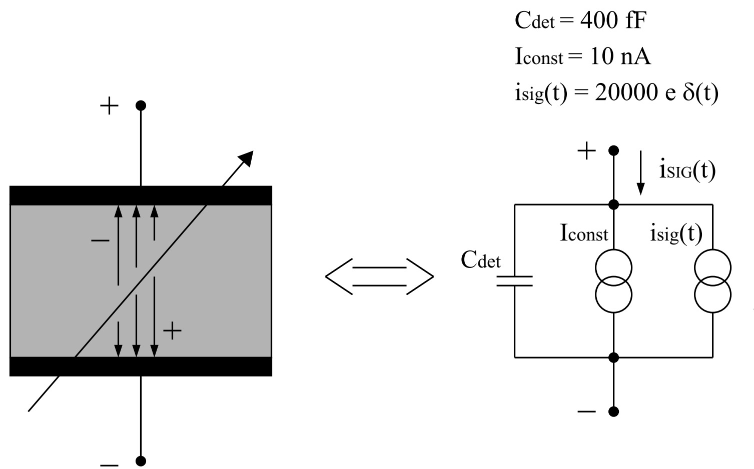


Figure 2.7: Electrical model of a radiation sensor.

NMOS and to a p - channel transistor as PMOS. The current-voltage characteristic of a PMOS transistor is equivalent to the characteristic of an NMOS with the exception that the signs of all voltages have to be changed. Fig. 2.10 shows the transistors of both types.

Charge-Sensitive Amplifier

Fig. 2.11 shows a simple implementation of a MOSFET based amplifier. The quasi-static input-output characteristic of the amplifier is shown in fig. 2.11 (c). This transfer characteristic can be derived by combining the output characteristics of both transistors T1 and T2, see fig. 2.11 (b). The amplification of the amplifier is large only in a very small region, where both transistors operate in saturation. Outside this region, the amplifier saturates and the amplification decreases, fig. 2.11 (c). Such a nonuniform characteristic is not suitable for sensor signal amplification. In order to make a linear amplifier starting from the nonlinear one, a negative feedback must be applied. The purpose of the negative feedback is to keep the input and output potentials of an amplifier in the region with

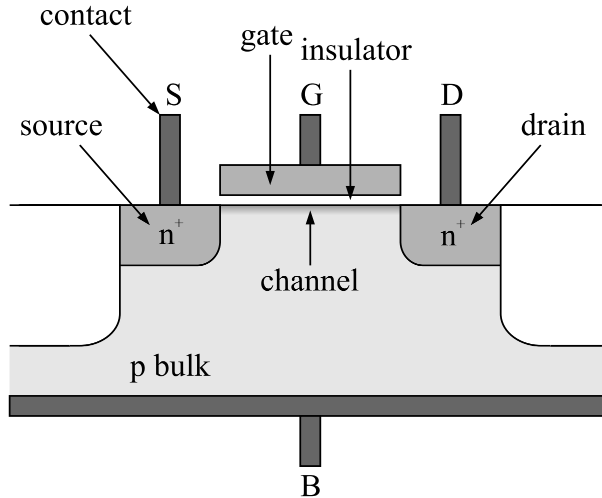


Figure 2.8: N-channel MOSFET.

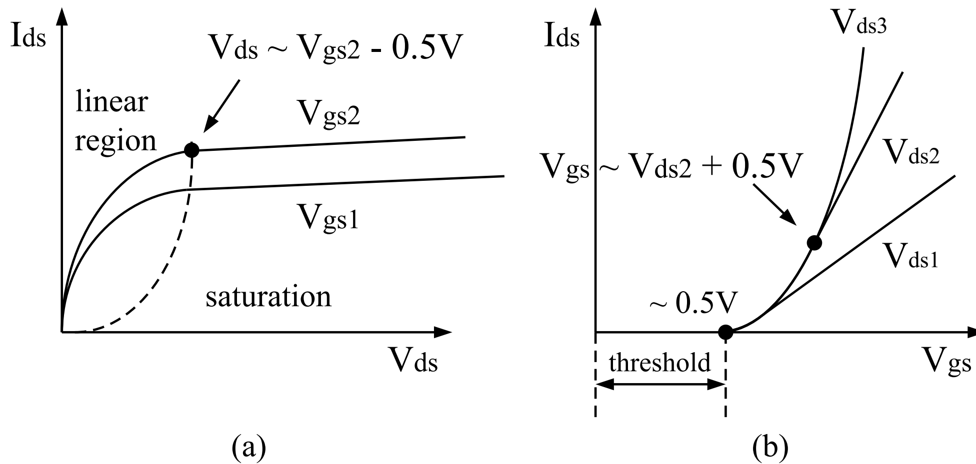


Figure 2.9: Current-voltage characteristic of an n-channel MOSFET. (a) - output characteristics, (b) - input characteristic.

high amplification, *i.e.* to prevent the saturation of the amplifier. A negative feedback circuit is very often implemented by using only passive components such as resistors and capacitors. The advantage of such type of feedback is its linear characteristic. It can be shown that the amplification of an amplifier with negative feedback depends mostly on the parameters of the feedback elements and not on the intrinsic amplification of the amplifier. If the feedback components are linear, the amplification is linear as well. In the case of the amplifier shown in fig. 2.11 the simplest implementation of the negative feedback is to connect a resistor between the nodes *in* and *out*, fig. 2.12. Usually an capacitor is connected parallel to the resistor. The capacitor influences only the dynamic behavior of the circuit. We will investigate this in the next subsection. The amplifier with the resistive and capacitive feedback (fig. 2.12) is called "charge-sensitive" amplifier.

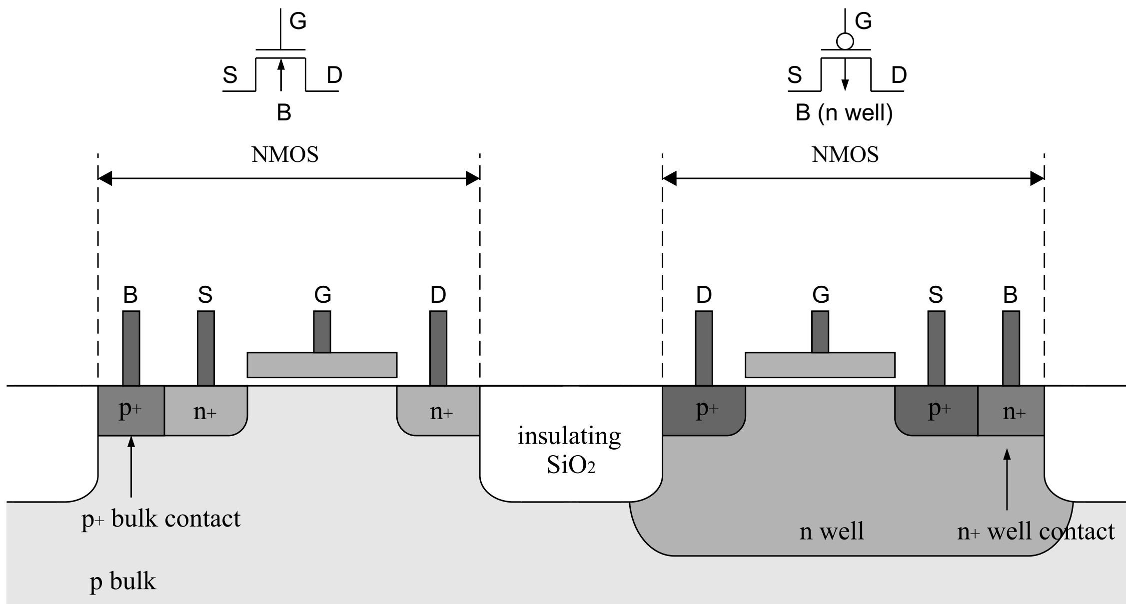


Figure 2.10: NMOS and PMOS transistors, structure and electrical symbols.

2.2.2 Signal Response of Charge-Sensitive Amplifier

In this section we will calculate the pulse form at the output of the charge-sensitive amplifier shown in fig. 2.12 that is connected with the sensor shown in fig. 2.7. The calculation will be done in a few steps, following the common procedure used in analog circuits design, see for example [4].

The first step is to calculate the currents and voltages in the stationary state that assumes the absence of sensor signal. The stationary values of electrical parameters are called DC[8] values or DC operation point. The DC values can be found by solving the system of nonlinear equations characterizing all electrical components in a circuit. This is usually done by using a Computer Aided Design (CAD) software. In order to estimate the DC operation point in the case of the charge-sensitive amplifier, we can simply plot the quasi-static transfer characteristics of the amplifier and the feedback on the same diagram, fig. 2.12 (b).[9] The intersection of the two lines defines the DC operating point. As we see, the operating point is in the high gain region of the amplifier characteristic.

The second step in the analysis of the circuit is to linearize the characteristics of all transistors close to the DC current/voltage values. If exact quasi-static transistor characteristic has the form

$$I_{GS} = 0;\ I_{DS} = f(V_{GS}, V_{DS}) \ ,$$

and its linearized form is

$$i_{ds}(t) = \frac{\partial I_{DS}}{\partial V_{GS}} v_{gs}(t) + \frac{\partial I_{DS}}{\partial V_{DS}} v_{ds}(t) \equiv g_m v_{gs}(t) + g_{ds} v_{ds}(t) \ . \tag{2.4}$$

[8] Direct Current.

[9] Assuming the stationary state (no signal at the input) no current flows through the resistor R (fig. 2.12). The voltage drop across the resistor is therefore zero, it leads to the quasi-static transfer characteristic: $V_{in} = V_{out}$.

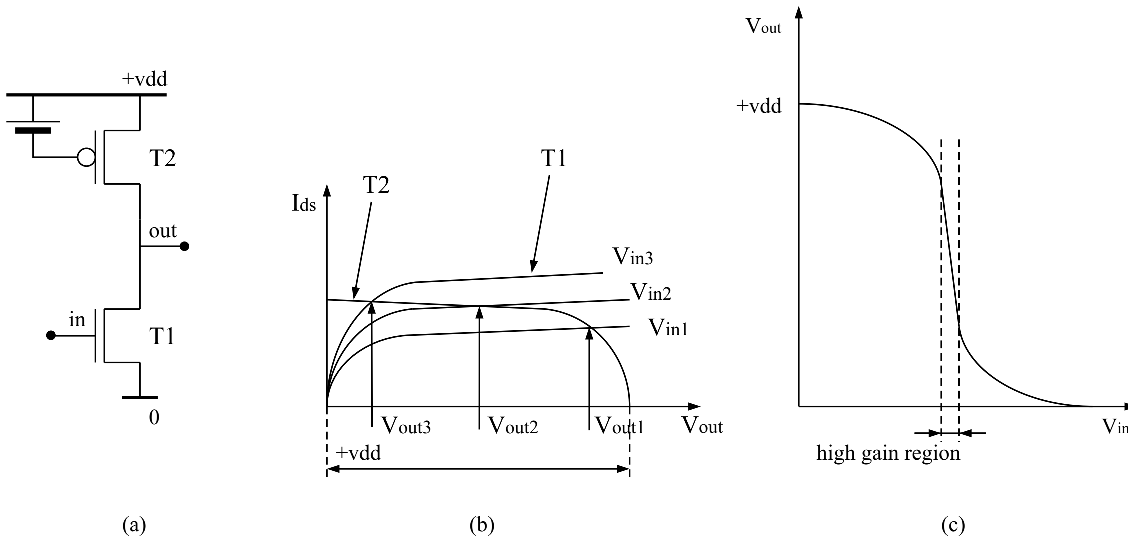


Figure 2.11: (a) - amplifier with an NMOS input-transistor (T1) and a PMOS active load transistor (T2). (b) - drain-source currents of both transistors versus V_{out} for different V_{in} values. Since the drain-source currents are equal, the intersection of two lines determines the V_{out} value. (c) - quasi-static transfer characteristic of the amplifier is derived from the plot (b).

$i_{ds}(t)$, $v_{ds}(t)$ and $v_{gs}(t)$ are the deviations from DC values (small signals). The parameter g_m is called transconductance, it describes the current gain of a transistor. Starting from the linearized transistor equations (eq. 2.4), we can construct a small-signal equivalent circuit. In this circuit the behavior of each transistor is described by using the linear components like voltage controlled current sources, resistors and capacitors. Small-signal circuit does not contain the generators of constant voltage or current. The generators of constant voltage are substituted by short connections and the generators of constant current become open circuits. Since a small-signal circuit describes the dynamical behavior of an electronic device, transistor internal capacitances must be taken into account. A correct model that describes transistor capacitances will be derived in appendix B. For the moment, we take into account only the gate-source capacitance C_{gs} and the drain-bulk capacitance C_{db}. The small-signal circuit of the charge-sensitive amplifier (fig. 2.12) is shown in fig. 2.13. The gate of the load transistor T2 (fig. 2.13) is connected to a constant bias potential, the small signal v_{gs2} is therefore zero. For small signals, the transistor T2 behaves as a parallel connection of the resistor $1/g_{ds2}$ and the capacitor C_{ds2}, fig. 2.13. By combining the equivalent circuits of the sensor, shown in fig. 2.7 and the charge-sensitive amplifier, shown in fig. 2.13, we obtain the small-signal equivalent circuit which is shown in fig. 2.14. We have here omitted the controlled current source that belongs to the load transistor T2 (fig. 2.13) since it generates no current. We have also omitted the shorted capacitor C_{gs2}.

The small-signal circuit shown in fig. 2.14 can be used for the calculation of the pulse at the output of the charge-sensitive amplifier. The use of the small-signal circuit gives reliable results in the case of amplifiers with passive feedback. The feedback keeps the potential of the input node of a high gain amplifier nearly constant and the linear expansion of the input transistor characteristic (fig. 2.9 (b)) is quite accurate. The signal amplitude at the output of the amplifier is usually not small. However, this does not

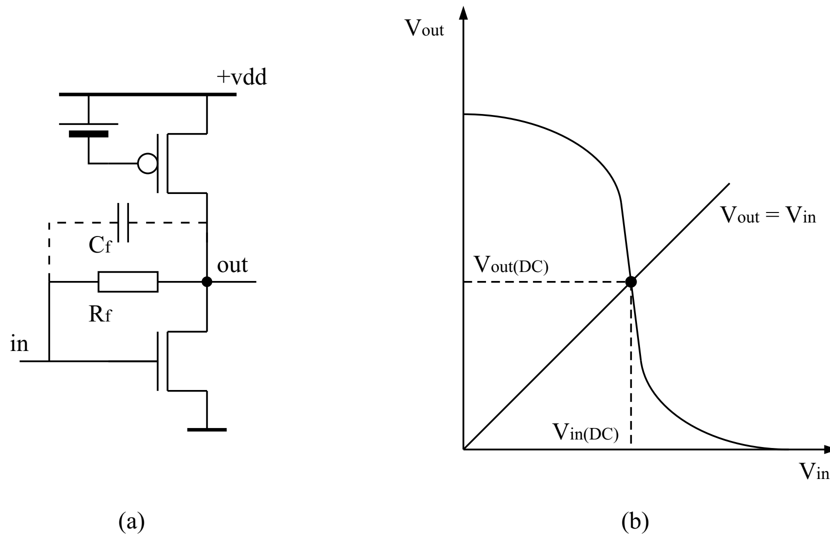


Figure 2.12: (a) - amplifier with feedback. (b) - quasi-static transfer characteristics of the amplifier and the feedback circuit (resistor).

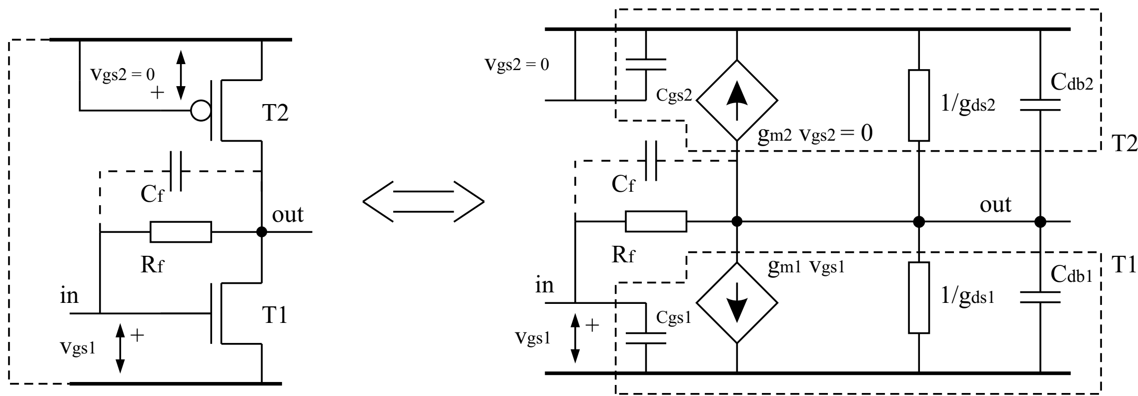


Figure 2.13: Small-signal circuit of the charge-sensitive amplifier.

represent a problem, since the output characteristic of a transistor in saturation (fig. 2.9 (a)) behaves, approximately, linear in a large voltage range.

In order to calculate the pulse form at the output of the charge-sensitive amplifier we have to derive the differential equation for the output voltage. This can be done in the complex frequency domain. The time derivative is substituted either by the imaginary frequency $i\omega$, or by the complex number s. In the frequency domain, each capacitor can be characterized by its imaginary impedance, as shown in fig. 2.14.

The differential equation for the output voltage in the frequency domain (transfer function) can be directly derived by using the Kirchoff laws for currents and voltages. We will derive the transfer function in a slightly different way in order to illustrate the common procedure followed in the analysis of feedback systems. Every circuit with a single feedback path is equivalent to the generalized circuit shown in fig. 2.15. It can be shown that

$$\frac{x_{out}}{x_{in}} = \frac{\alpha A}{1 + \beta A} \ . \tag{2.5}$$

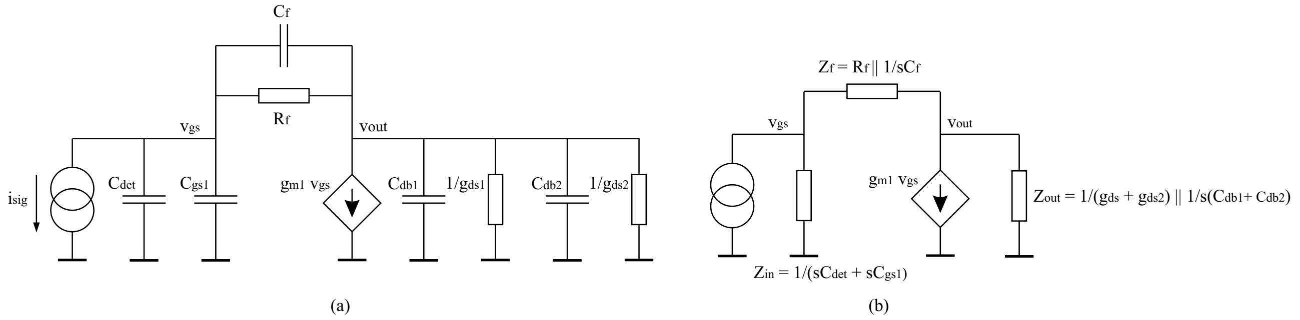


Figure 2.14: (a) - Small-signal circuit of the sensor and the charge-sensitive amplifier. (b) - Small-signal circuit. Parallel connections of resistors and capacitors are described with equivalent impedances.

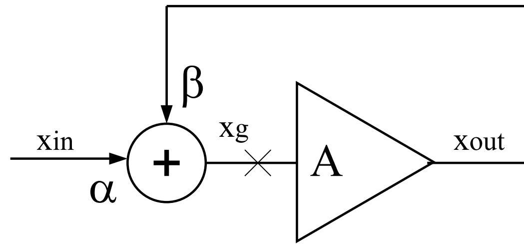


Figure 2.15: Generalized feedback scheme.

The result given by eq. 2.5 can be also applied in the case of the small-signal circuit shown in fig. 2.14. The only non trivial task is to determine the values of the parameters α, β and A. These parameters can be calculated in the following way. Let look at the circuit shown in fig. 2.15. α can be found when we "switch off" the amplifier A.[10] in order to disable the feedback, connect a test generator (which generates some known signal x_{test}) to the node x_{in} and calculate the signal x_g, fig. 2.16 (a). In order to calculate the forward amplification A we can break the line x_g, as shown in fig. 2.15, connect the test generator to the input of the amplifier and calculate the signal x_{out}, fig. 2.16 (b). If we leave the test generator at the input of the amplifier and calculate the signal x_g left from the cut, the

[10]To switch off a voltage source means to replace it with a short connection. To switch off a current source means to replace it with an open connection.

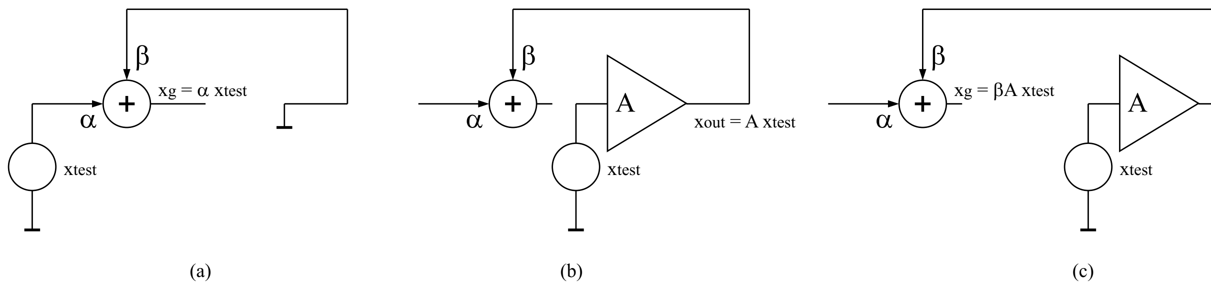


Figure 2.16: (a) - test circuit for the calculation of the parameter α. (a) - test circuit for the calculation of the parameter β. (c) - test circuit for the calculation of the open loop gain βA.

open loop gain βA can be found from the ratio between the signals x_g and x_{test}, fig. 2.16 (c). Such procedure can be used in the case of the circuit in fig. 2.14. The corresponding "test" circuits are shown in fig. 2.17. The results are simple to obtain:

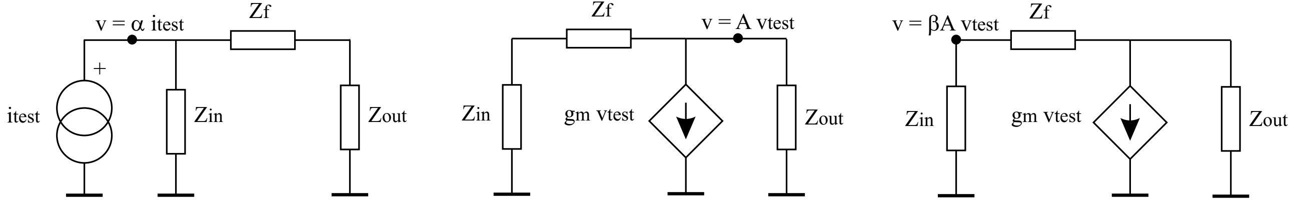


Figure 2.17: (a) - test circuit used for the calculation of the α parameter, see fig. 2.15. (b) - test circuit for the calculation of the parameter A (forward amplification). (c) - test circuit used for the calculation of the open loop gain βA.

$$\alpha = Z_{in}||(Z_f + Z_{out});\ A = -g_m[Z_{out}||(Z_{in} + Z_f)];\ \beta A = \frac{Z_{in}}{Z_{in} + Z_f}A. \tag{2.6}$$

The symbol $||$ denotes the parallel connection of two impedances: $a||b \equiv ab/(a+b)$. We can now substitute the last results into eq. 2.5. One obtains[11]

$$\frac{V_{out}(s)}{I_{in}(s)} = \frac{R_o\left(\frac{g_m R_f - 1}{g_m R_o + 1}\right)\left(1 - \frac{sR_f C_f}{g_m R_f - 1}\right)}{s^2\left(\frac{R_f R_o(C_o C_f + C_o C_i + C_i C_f)}{g_m R_o + 1}\right) + s\left(\frac{R_o C_o + (R_f + R_o)C_i + R_f(g_m R_o + 1)C_f}{g_m R_o + 1}\right) + 1}\ , \tag{2.7}$$

where[12] $C_i \equiv C_{det} + C_{gs1};\ C_o \equiv C_{db1} + C_{db2};\ R_o \equiv 1/g_{ds1} + 1/g_{ds2};\ g_m \equiv g_{m1}$. The last equation can be simplified if $g_m R_o \gg 1;\ g_m R_f \gg 1$; and $g_m R_o C_f \gg C_i, C_o$:[13]

$$\frac{V_{out}(s)}{I_{in}(s)} \approx \frac{R_f\left(1 - \frac{sC_f}{g_m}\right)}{s^2\left(\frac{R_f R_o(C_o C_f + C_o C_i + C_i C_f)}{g_m R_o}\right) + sR_f C_f + 1}\ . \tag{2.8}$$

If we factorize the denominator in the last equation, we obtain

$$\frac{V_{out}(s)}{I_{in}(s)} = \frac{R_f(1 - s\tau_z)}{(s\tau_r + 1)(s\tau_f + 1)}\ . \tag{2.9}$$

The constants in the last equation have the following approximative values:[14]

the feedback time	$\tau_f \approx R_f C_f$;
the rise time	$\tau_r \approx (C_o C_f + C_o C_i + C_i C_f)/(g_m C_f)$;
the zero time-constant	$\tau_z \approx C_f/g_m$.

[11] In our case the output signal is a voltage, therefore $x_{out}(t) = v_{out}(t)$. The input signal is a current $x_{in}(t) = i_{in}(t)$.

[12] $V_{out}(s)$ and $I_{in}(s)$ are the Laplace transforms of $v_{out}(t)$ and $i_{in}(t)$. The Laplace and Fourier transforms will be always denoted by using uppercase.

[13] It is highly desirable to fulfill these conditions in order to maximize the charge amplification.

[14] The result of the assumptions that led to the simplified form of the transfer function is $\tau_f \gg \tau_r$.

Using eq. 2.9 we can calculate the pulse at the output of the amplifier. The Laplace transform of the input signal, which has a delta pulse form $i_{in}(t) = Q\delta(t)$, is the constant function $I_{in}(s) = Q$. Therefore

$$v_{out}(t) = \mathcal{L}^{-1}\{Q\frac{V_{out}(s)}{I_{in}(s)}\} \approx \frac{Q}{C_f}\frac{\tau_f}{\tau_f - \tau_r}\left(\exp\left(\frac{t}{\tau_f}\right) - \exp\left(\frac{t}{\tau_r}\right)\right) \quad . \tag{2.10}$$

In order to simplify the solution, we have neglected the zero time-constant τ_z. The function in eq. 2.10 has the form shown in fig. 2.18. The peaking time is:

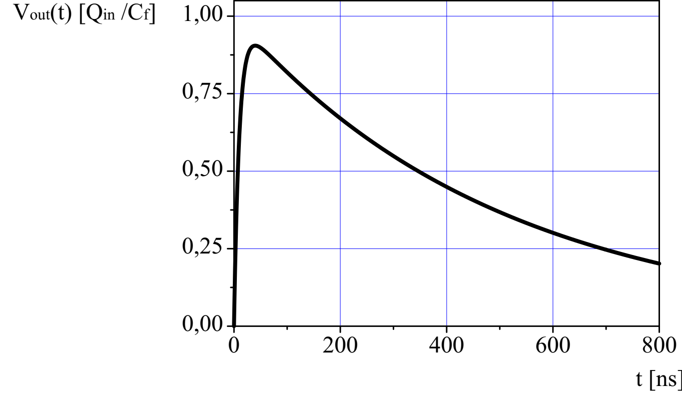


Figure 2.18: Output pulse of the charge-sensitive amplifier with the rise time-constant $\tau_r = 10\ ns$ and the feedback time-constant $\tau_f = 500\ ns$.

$$t_p = \frac{\tau_r \tau_f}{\tau_f - \tau_r} \ln\left(\frac{\tau_f}{\tau_r}\right) \tag{2.11}$$

and the maximal amplitude:

$$v_{out}(t_p) = \frac{Q}{C_f}\left(\frac{\tau_r}{\tau_f}\right)^{\frac{\tau_r}{\tau_f - \tau_r}} \quad . \tag{2.12}$$

The maximal amplitude of the output pulse will be larger for the larger difference between time constants. For $\tau_f \gg \tau_r$, the amplitude approaches its asymptotic value Q/C_f.[15] A typical value for C_f is 5 fF. In this case, assuming $v_{out}(t_p) \approx Q/C_f$, we have the voltage gain of 80 with respect to the voltage pulse measured at the electrode of a detector without

[15]This holds only in the case when $g_m R_o C_f \gg C_i$. If this condition is not fulfilled, the simplified eq. 2.9 is no longer valid and the original eq. 2.7 must be solved in order to find the output signal amplitude. If we suppose that the feedback resistance R_f has a very large value, eq. 2.7 can be approximated as:

$$\frac{V_{out}(s)}{I_{in}(s)} \approx \left(\frac{g_m R_o + 1}{C_i + (g_m R_o + 1)C_f}\right)\frac{(1 - s\tau_z)}{s(s\tau_r + 1)} \ ,$$

where $\tau_z \equiv C_f/g_m$ and $\tau_r \equiv \frac{R_o(C_oC_f + C_oC_i + C_iC_f)}{C_i + (g_mR_o + 1)C_f}$. The maximal signal amplitude at the output of the charge-sensitive amplifier is

$$v_{out}(t_p) = \left(\frac{Q}{C_f}\right)\left(\frac{(g_m R_o + 1)C_f}{C_i + (g_m R_o + 1)C_f}\right) \ .$$

amplifier ($Q/C_{det} = Q/400\ fF$). The maximal amplitude is $v_{out}(t_p) = 20000\ e/5\ fF = 0.64\ V$. The amplitude of the output signal depends only on the collected charge, it does not depend on the input pulse form, unless the pulse is slow with a characteristic length comparable to the feedback time-constant. This is an important feature of the charge-sensitive amplifier. It can also be shown that the voltage pulse seen at the input node has an amplitude which is only $1/(g_m R_o)$ of the voltage amplitude seen at the output. The potential at the input is nearly constant since the voltage gain $g_m R_o$ has usually very large values (> 100). The input node is said to be on virtual ground.

2.3 Requirements on the Front-End Chip for the ATLAS Pixel-Detector

In the first part of this section we will describe the ATLAS pixel detector module. The requirements on the ATLAS pixel front-end chip will be described in the second subsection.

2.3.1 Pixel-Detector Module

The support structures of the pixel detector will be covered by identical detector modules. The central part of the module is a rectangular silicon sensor element. The total area of the sensor element is 63.0 mm $\times$ 18.6 mm and the thickness is 250 μm (the active sensor area is 60.8 mm $\times$ 16.4 mm). The sensor element is divided into 8×2 fields, fig. 2.19. Each of the 16 fields has a pixel matrix with 18 columns and 164 rows, fig. 2.19.

The charge deposited in a pixel is collected by electrodes which are implemented as highly doped n^+ islands in a low doped n bulk. The back side of the sensor is p^+ doped (see the cross section in fig. 2.19). The p^+ rear electrode and the n bulk form a $p-n$ junction. In order to deplete the n bulk, a negative voltage is applied between the p^+ electrode and the bulk.[16] The pixels are 50 μm $\times$ 400 μm large.

The detector module is a hybrid structure. The front-end chip has a size similar to that of the fields on the sensor, fig. 2.20. The surface of the chip is segmented into 50 μm $\times$ 400 μm large pixels. Every front-end pixel contains a charge-sensitive amplifier used to amplify the charge signal generated by the pixel sensor and logic circuitry used to measure the hit time and to generate the digital hit information. The input of every amplifier is connected to a bonding pad implemented by using the last metal layer on the chip, fig. 2.20. Similarly, every n^+ doped electrode in the sensor is connected to its bonding pad. The bonding pads on the chip and the sensor side are arranged in identical way, so that direct "one-by-one" connection can be established between them when the chip is flipped upside-down and mounted on the sensor, fig. 2.20. It is done by using the flip - chip and bump bonding technique. The bonding pads are connected by solder or indium bumps. On the front-end chip, beside the pixel matrix there is an area which is used to implement the memory for hit data, control circuits and signal/power pads. The chips are placed in the way that their insensitive areas extend beyond the longer edge of the sensor, fig. 2.20. The pixel matrices of neighboring chips are placed as close as possible to each other. However, small gaps between the chips cannot be avoided.

[16] The depletion of the bulk material is necessary for the sensor operation - the n^+ pixel electrodes are electrically disconnected only if the bulk is emptied from free carriers (electrons).

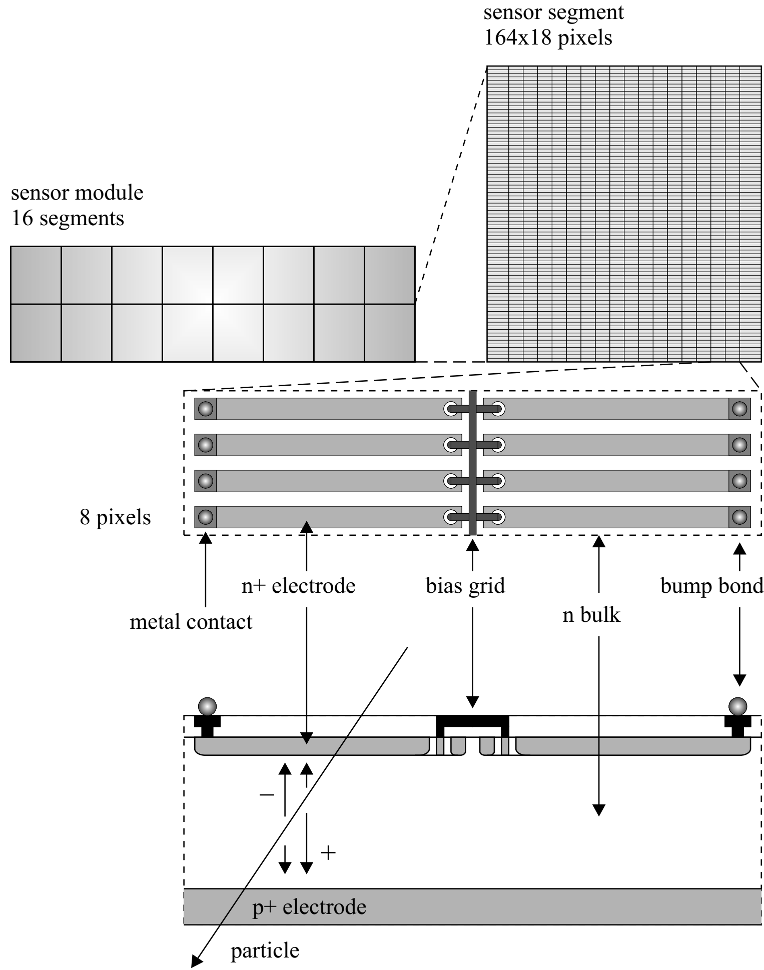


Figure 2.19: Sensor element, top view and cross-section.

The sensor pixels above the long middle gap (fig. 2.20) cannot be directly bonded to the amplifiers of a front-end chip. They are connected on the sensor side to the pixels which have direct connection to the amplifiers. The pixels in the first and 18th column of a sensor field are longer to cover the gaps perpendicular to the longer module side.

Beside the sensor element and the 16 front-end chips, a detector module contains a module controller chip (MCC), passive components, power distribution busses, local signal interconnections and an connector (see fig. 2.21 and fig. 2.22). These electronic devices are placed on a flexible hybrid with wire bond connection from the front-end chips to the power and signal busses of the hybrid. The front-end chips extend beyond the long sides of the sensor. This gives access to their output pads so that wire bond connections can be used to make the contact to the flexible hybrid glued to the back of the sensor, fig. 2.21. Aluminium and copper bus lines route the signals to the MCC mounted on the flexible hybrid.

The modules are finally mounted on mechanical support structures in a chip-down geometry. The task of the support structure is to provide mechanical stability and efficient cooling of modules to the temperatures below 0° C. The support structures, which make the barrel detector layers, have the form of staves. The staves are made of carbon - carbon material. This material offers a low radiation length and a high rigidity. An aluminium

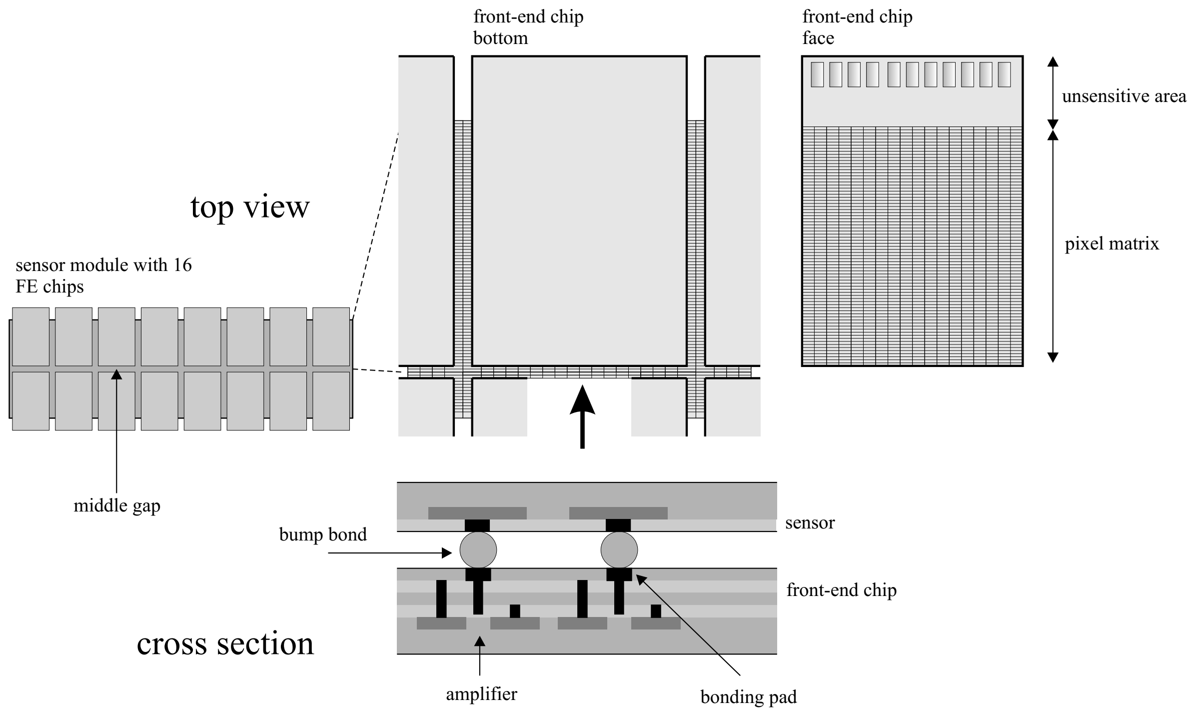


Figure 2.20: A sensor element and front-end chips bonded with bumps. Top view and cross section.

cooling pipe is integrated in the stave in an Ω shaped channel, fig. 2.21. The modules on a stave overlap in longitudinal direction to avoid insensitive areas. The neighboring staves overlap in ϕ direction. In the case of the disk detector layers, the basic support structures have the form of disk segments.

2.3.2 Requirements on the Front-End Chip

Let us start from three basic facts [6, 13, 14]. First, the front-end chip must be able to cope with a particle flux of $4.5 \cdot 10^{-4}$ hits per pixel and bunch-crossing. Second, the total amount of material in a pixel detector layer should be smaller than 1.5 % of the radiation length. Third, the pixel sensor must operate at -7^0 C. The working principle of the ATLAS pixel detector should be kept in mind. The front-end circuits must have a sufficient time resolution (25 ns) to be able to distinguish between hits that originate from different bunch crossings. Every hit data, which comprises pixel address and amplitude information, must be stored on the chip until the trigger signal delay is off. The trigger signal delay is not longer than 3.2 μs, which determines the size of internal buffers.

Requirement 1 - Power

The detector modules must be cooled to -7° C in order to avoid the radiation induced degradation of the sensor material. The heat generation of the electronic circuits on the module must be therefore minimized. The power consumption per front-end pixel should be less than 40 μW.

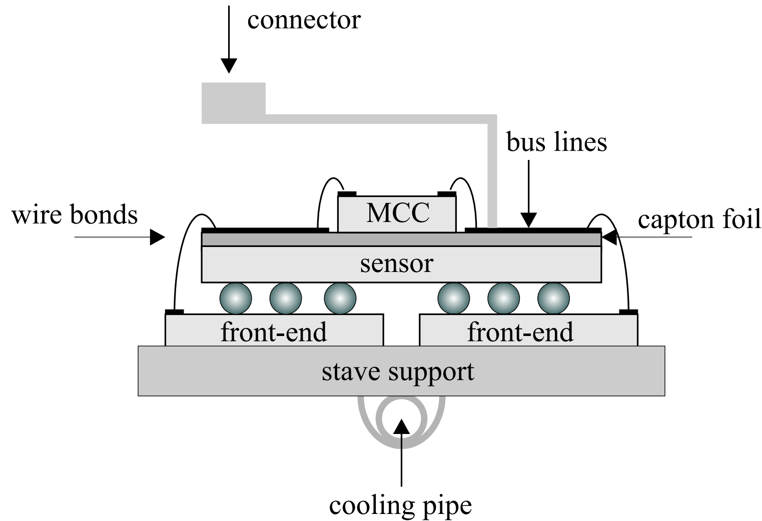


Figure 2.21: Module mounted on the stave support structure, cross section.

Requirement 2 - Signal Amplitude

The amount of material in the pixel detector should be minimized in order to avoid the multiple coulomb scatterings and the generation of electromagnetic showers on the pixel detector - a tracking detector must not disturb the detector components that surround it. The total amount of material per detector layer should be less than 1.5 % of the radiation length. The mechanical support, sensor and chips should be therefore as thin as possible. The sensor thickness is chosen to be 250 μm. Such a sensor generates a signal of about 19000 e. The signal can be shared between two pixels, which gives, approximately, 9500 e. If we take into account that charge collection loss increases with total radiation flux and that sensor depletion becomes more difficult due to radiation damage, we could expect that the minimum signal per pixel drops to about 5000 e after a few years of operation.

Requirement 3 - Memory Space

The simulations of proton-proton collisions lead to the following important result: the probability that a ionizing particle traverses a sensor pixel per bunch-crossing is about $4.5 \cdot 10^{-4}$. A front-end chip has 2880 pixels and every hit must be recorded for the maximal delay time of 3.2 μs.

Requirements on the Analog Pixel-Electronics

Before we introduce the requirements on the analog pixel electronics, let us repeat a few facts. We introduced the charge-sensitive amplifier in section 2.2. Every pixel on the front-end chip contains such an amplifier. The output signal of the amplifier will be electronically compared with a threshold. In the moment when the threshold is exceeded, the comparator responds and the chip accepts the hit. The task of the front-end chip is to assign a hit to the correct bunch-crossing. The correct timing of the comparator signal is therefore of great importance.

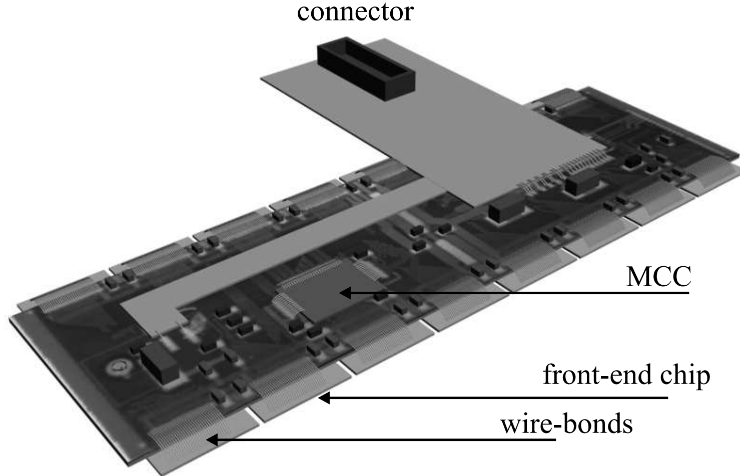


Figure 2.22: ATLAS pixel-detector module.

Time Walk and the Maximal Threshold

Fig. 2.23 shows the response of the charge-sensitive amplifier to two different input charge signals. The comparator will respond faster in the case of the higher input signal. The range of possible sensor signals is large ($\approx 5000\, e - 19000\, e$). Since the chip has to assign a hit to the correct bunch-crossing, the time skew of the comparator response over the whole input signal range, called the time walk, must be smaller than the required time resolution (25 ns), fig. 2.23. There are two ways to diminish the time walk. The first way is to make the amplifier faster *i.e.* to decrease the rise time-constant τ_r. In section 2.2 we derived the formula for the rise time:

$$\tau_r \approx (C_o C_f + C_o C_i + C_i C_f)/(g_m C_f) \ .$$

g_m is the transconductance of the input transistor, C_i and C_o are the capacitances of the input and output amplifier nodes respectively and C_f is the feedback capacitance. Provided that we optimized the capacitances, which means that we decreased the output capacitance as much as possible, the rise time becomes: $\tau_r \approx C_i/g_m$. The transconductance g_m is proportional to the DC (bias) current that flows through the input transistor of the charge-sensitive amplifier in the stationary state. However, a higher bias current leads to a higher power dissipation and power dissipation is limited. The second way to decrease the time walk is to decrease the threshold, see fig. 2.23. The measurements performed on the front-end chips showed that the threshold value should be at least 2000 e lower than the minimal signal (5000 e)[17] in order to keep the time walk small enough. This leads to the maximal threshold of about 3000 e.

Noise and the Minimal Threshold

Let us estimate the minimal threshold value that can be used. Here we are theoretically limited by a phenomenon that has the origin in the statistical nature of the charge motion

[17] The threshold is here expressed as the equivalent input signal.

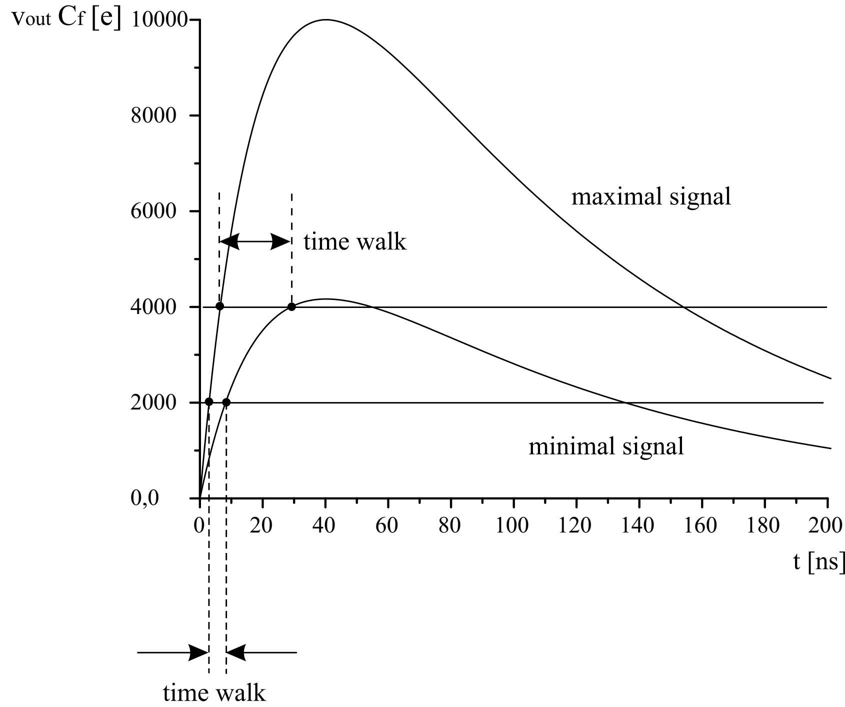


Figure 2.23: Response of the charge-sensitive amplifier to two different input signals. Time walk.

- noise.[18] If the threshold value is low, the probability that the voltage noise at the output of the amplifier exceeds the threshold becomes considerable. The rate of the comparator pulses generated by the noise can be calculated by using the following formula [51]

$$f_n = \frac{1}{4\sqrt{3}\tau_f} \exp\left(\frac{-Q_{Thr}^2}{2ENC^2}\right) \ . \tag{2.13}$$

$\tau_f \approx 200\, ns$ is the feedback time-constant (introduced in section 2.2), Q_{Thr} is the threshold expressed as the equivalent input charge signal, ENC is the "equivalent noise charge" - the root-mean-square (RMS) value of the output noise expressed as the equivalent input signal. The probability that the noise activates the comparator (noise occupancy) should be much smaller than the particle hit probability. Assuming the noise occupancy of less than 10^{-6} per pixel and bunch-crossing, eq. 2.13 gives the condition:

$$ENC < Q_{Thr}/4.5 \approx 600\ e \ . \tag{2.14}$$

Here, we must also take into account that the threshold have a certain pixel - pixel variation. If we set a threshold value for the whole chip, the thresholds in individual pixels will deviate from the desired value. The threshold dispersion from pixel to pixel has a similar effect as noise, it can be treated as a time independent noise component. The ENC value in eq. 2.14 should be therefore interpreted as the sum of the squares of the real (time dependent) noise ENC and the equivalent charge that corresponds to the standard deviation of the threshold dispersion.

[18]For the moment, we will not consider the crosstalk.

Radiation Tolerance

The last requirement we mention is radiation tolerance. The ATLAS pixel detector is located only a few cm away from the interaction point, where the high energy protons collide. It will be exposed to a life time flux of 10^{15} n_{eq}/cm^2 and ionizing radiation dose of 50 $Mrad$. All electronic component must survive this hostile environment. The radiation tolerance will be discussed in the next section.

2.4 Radiation Effects in Silicon and Radiation-Tolerant Chip Design

The front-end chip for the ATLAS pixel detector is one of the first complex microchips implemented in a deep-submicron technology by using the radiation tolerant layout. The chip has been successfully tested up to the radiation dose of 100 Mrad. In the first part of the section we will talk about the mechanisms of radiation damage. We will distinguish between the non-ionizing displacement damage of semiconductor bulk and the surface damage caused by the ionization in $Si0_2$ insulation layer. In the second subsection we will concentrate on the specific LHC environment. We will first introduce the design of a radiation tolerant pixel sensor. After that we will talk about the radiation effects in complementary MOS (CMOS) microchips and the design techniques used to harden the electronic devices.

2.4.1 Introduction

Radiation induced damage takes place in many different ways. We can distinguish between permanent and transient effects. The permanent effects are related to structural modifications of the material. The transient effects are related to the instantaneous generation of charge carriers in semiconductor material, which can lead to failure in signal processing.

We will now focus our attention to the permanent effects. A common way to classify the permanent effects is based on the distinction between the displacement damage and the damage caused by ionization. The displacement of the atoms is caused mainly by neutrons and high-energy charged particles, while the ionization is caused by charged particles and photons.

2.4.2 Displacement Damage in Silicon

When a neutron or a high-energy charged particle knocks a silicon nucleus, the latter can move from its original position, which leads to a vacancy formation. Very often, the incoming particle has so high energy that it generates many vacancies and the recoiled silicon atoms itself impinge other atoms in the crystal lattice. A cluster of vacancies can be produced in this way. The recoiled atoms usually find space between the lattice atoms and become interstitials. The described distortion of the crystal structure leads to the formation of additional energy levels. The new energy levels can be classified as shallow and deep trap levels. The trap levels influence the electric properties of the material. Radiation-induced shallow traps usually compensate donors and act as acceptors. This

can lead to the change of effective dopant density - originally low doped n-type silicon becomes p-type after irradiation.[19] This effect is called the type inversion. The deep traps open new recombination-generation channels, which has as a consequence a decrease of the carrier life time and an increase of the thermal generation rate. The increase of carrier-trap scattering rate lowers the carrier mobility.

The displacement damage is linearly proportional to the particle flux and to the non-ionizing energy loss of an incident particle per collision. Different particles with the same energy will have different non-ionizing energy loss (NIEL). For example, a 200 MeV proton has roughly the same NIEL as an 1 MeV neutron. If we want to characterize the displacement damage caused by different particles with different energies, we calculate the equivalent flux of 1 MeV neutrons that have the same NIEL and cause the equivalent displacement damage.

Since neutrons and high energetic charged particles are very penetrating, the displacement damage is distributed homogeneously in the entire semiconductor bulk. All devices having the electrical properties that rely on the minority carrier transport in large bulk areas are sensitive to the displacement damage. Especially sensitive are the high-voltage devices with low doped active region. Semiconductor radiation detector is an example of such device.

On the other hand, MOS transistors are not sensitive to the displacement damage. The region relevant for the electrical behavior of a MOSFET is confined to a very small surface volume and the damage of deeper bulk regions does not influence the electrical properties of the device. Also, the doping concentrations of source and drain are so high that the radiation-induced change of dopant density cannot cause the type inversion.

2.4.3 Ionization Damage in $Si0_2$

The electrical characteristics of MOS transistors depend mainly on the properties of the $Si0_2$ insulating layer and the quality of the Si-$Si0_2$ interface, fig. 2.24 (a). Since it is amorphous[20], the oxide layer is not sensitive to the displacement damage. However, the ionization in the insulating oxide is an irreversible process. Negative long term effects can occur. The ionization induced damage will be described in this section.

Oxide Charge

The ionization damage depends on the total energy of ionizing radiation absorbed in $Si0_2$. The common unit used to quantify the ionization damage is the *rad*. The MKS unit is the *Gray*, equal to 1 J/kg or 100 rad. Fig. 2.24 shows the MOS structure (a) and the enlarged cross section of the insulating $Si0_2$ layer (b). The energy deposited by ionizing radiation in $Si0_2$ leads to the creation of electron-hole pairs.[21] Let us assume that there is an electric field in the oxide produced by a positively charged electrode placed above it. The electrons that escape recombination will be quickly swept out by the electric field. The electron mobility in the $Si0_2$ is about 20 cm^2/Vs. The remaining

[19] The formation of the phosphor-vacancy centers deactivates the phosphor as donor. On the other hand, the vacancy-vacancy centers act as acceptors.

[20] The silicon tetrahedron is the fundamental unit of silicates. Silicon atom is surrounded by four oxygen atoms. In the amorphous $Si0_2$ layer the tetrahedrons are linked by sharing an oxygen atom. The structure does not display a longer range order as that in the case of quartz.

[21] About 18 eV is needed for the generation of one electron-hole pair in $Si0_2$.

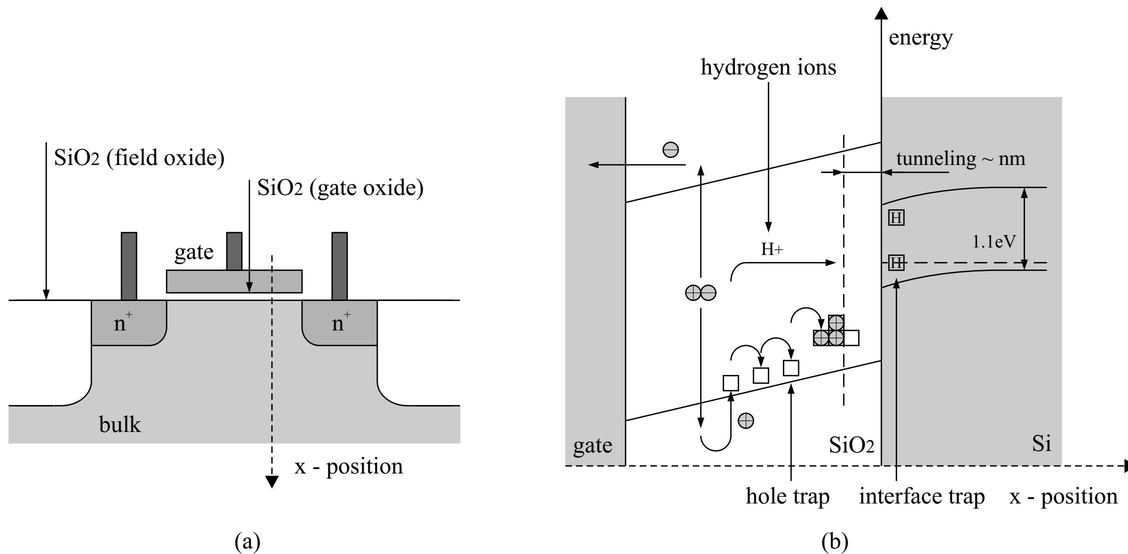


Figure 2.24: (a) - MOS structure, (b) - enlarged cross section of the insulating $Si0_2$ layer.

holes will begin a slow transport toward the Si-$Si0_2$ interface. Their mobility is only about $2 \cdot 10^{-5}$ cm^2/Vs. The reason for so small mobility is that the holes get captured in shallow traps. Their walk through the oxide can be described as hopping between neighboring trap sites. In the vicinity of the Si-$Si0_2$ interface the oxide structure is more disturbed and the hole traps become deeper. The holes captured in these deep traps cannot move further. The deep trap centers will collect the holes generated by radiation, which leads to the formation of the positive oxide trapped charge, see fig. 2.24. The deep trap centers that are responsible for the formation of the positive trapped charge are identified to be the point defects in the $Si0_2$ structure with a missing oxygen atom (oxygen vacancy V_0). The V_0 defects in the $Si0_2$, which are filled with holes, are referred to as the E' centers [65]. The positive charge in the vicinity of the Si-$Si0_2$ interface attracts the electrons from the bulk, which can lead to the generation of parasitic negatively charged channels. Such channels can influence electrical properties of a transistor. For example, in the case of an NMOS transistor one needs a lower gate-source voltage to switch on the transistor - the threshold voltage decreases. The parasitic channels can also be formed between neighboring NMOS transistors making shorts between their n^+ electrodes. We will explain this in detail in the following section.

The holes trapped in the deep oxide-traps can be removed by electron trapping which acts compensatory. This can be done either by the thermal excitation of electrons from the valence band (thermal annealing) or by the electron tunnelling from the silicon surface. The thermal annealing requires the temperatures up to 300^0 C. Since the tunnelling probability increases with decreasing of the distance to be tunnelled, all deep trap centers closer than ≈ 5 nm from the Si-$Si0_2$ interface will be free from holes and therefore neutral [28, 27], see fig. 2.24. The annealing caused by the tunnel effect is of great importance for the application of deep submicron semiconductor technologies in the radiation environment. In such technologies the gate-oxide thickness is only about 5 nm. The charge in the gate-oxide will be efficiently removed through the tunnel-annealing since the relevant dimensions to be tunnelled are very small.

Interface Traps

The second effect caused by the ionization in the $Si0_2$ is the activation of the trap states at the interface between the oxide and the silicon bulk, see fig. 2.24. The interface trap-states are related to the lack of homogenous oxidation of the silicon atoms at the bulk surface. The unoxidized silicon atoms at the Si-$Si0_2$ interface generally contain unsaturated valence, which leads to the formation of dangling bonds [65]. Such defects are also referred to as the *Pb* centers.[22] The *Pb* centers are generally considered to belong to the substrate, while the E' centers are located in the silicon-dioxide layer. The *Pb* centers are amphoteric, they produce two trap levels. One level is donor-like (positive when electron state is empty) and placed about 0.3 eV above the valence band. The other acceptor-like level (negative when the electron state is full) is about 0.3 eV below the conduction band. The *Pb* centers can freely exchange the electrons with substrate, which makes distinction between them and the E' centers, embedded deeper in the oxide.

The *Pb* trap centers cause the increase of the voltage needed to switch on a MOS transistor. Let us consider the case of an NMOS transistor. When we apply a positive gate-source voltage, the electrons from the bulk are attracted to the Si-$Si0_2$ interface. The potential energy of the electrons at the interface will be lower than the potential energy of the electrons deeper in the bulk. Due to the lowering of electron energy levels, the acceptor-like traps are pulled below the Fermi energy, they become filled with electrons and thus negatively charged. Assuming a constant gate potential, which means that we have a constant amount of the positive charge on the gate electrode, the collection of negative charge in traps leads to the decrease of negative charge in channel. There are less free carriers available for current conduction between drain and source. We need therefore a higher gate potential to switch on the transistor than in the case when the traps are deactivated. In the case of a PMOS, in order to switch on the transistor we have to apply a negative gate-source voltage. The electron energy levels are than bent up and the donor-like traps are emptied from electrons, so that they become positive. Recall that the channel of a PMOS transistor consists of holes. More positive charge in the traps leads to less positive charge in the channel and a higher channel resistance. In order to switch on the transistor we need a higher gate-source voltage than needed in the absence of traps.

In order to reduce the unwanted effects caused by the interface traps, one tries to passivate the dangling bonds. The post-oxidation hydrogen treatment has become the standard industry practice to improve the quality of the interface between gate-oxide and bulk. A hydrogen atom satisfies a dangling bond and make it electrically inactive. The ionizing radiation can cause a new activation of the passivated interface traps. This is believed to happen in the following way [65]. The holes generated by the ionizing radiation start the hopping transport towards the interface. They interact also with the hydrogen containing defects and release the protons. The protons are mobile, they move from an

[22] Silicon dangling bonds contain an unpaired electron and because of that they are detectable by means of the electron paramagnetic resonance spectroscopy (EPR). The first observation of silicon dangling bonds was reported from the EPR measurement by Nishi. In his experiments on a standard sample of thermally grown oxide on silicon substrate, Nishi observed three distinct signals and assigned them to three different paramagnetic species, a, b and c, with unknown identity. It was many years later that Poindexter et al. and Brower independently identified the source of Nishi's paramagnetic species b (*Pb*) as a dangling silicon bond at the Si-$Si0_2$ interface. The sources of the other two signals (a and c) were identified as impurities [65].

oxygen to an oxygen atom in the SiO_2 and finally reach the interface, fig. 2.24. At the Si-SiO_2 interface the protons combine with the hydrogen atoms populating the dangling bonds. The formed H_2 leaves behind the activated *Pb* center.

2.4.4 Radiation-Tolerant Design for the ATLAS Pixel-Detector

In this section we will focus our attention to a specific radiation environment - the ATLAS pixel detector. The pixel detector is placed in the closest vicinity of the LHC interaction point, where the bunches of 7 TeV protons collide. The silicon sensor and the readout electronic chips are the most sensitive elements of the pixel detector. They must survive ten years operating in very hostile environment. The total flux for that time is expected to reach 10^{15} n_{eq}/cm^2 and the total dose calculated for SiO_2 will be 50 Mrad.

Pixel Sensor and Displacement Damage

As we discussed in the previous section, a radiation sensor is sensitive to the displacement damage. The displacement damage will cause the increase of sensor leakage current, decrease of charge collection efficiency and degradation of sensor structure due to inversion of effective dopant concentration.

Leakage Current

The ATLAS pixel sensor is introduced in section 2.3. The sensor is a $p-n$ structure. The active (low doped) side can be totally depleted by proper biasing in order to achieve the full charge collection efficiency. In the case of a $p-n$ diode the inverse saturation current is given by the following formula [2]

$$I = \frac{en_i}{\tau_p + \tau_n} \cdot A \cdot W \ . \tag{2.15}$$

n_i is the intrinsic carrier density, τ_p and τ_n are the minority carrier life times (ambipolar life times), $A{\cdot}W$ is the volume of the active area, e is the elementary charge. As we already discussed, the radiation-induced displacement of silicon atoms leads to the formation of new generation-recombination centers. The carrier life time decreases, which leads to the increase of the leakage current, see eq. 2.15. The radiation-induced leakage current increase can be estimated by the following empirical formula [61]

$$\Delta I = \alpha \Phi_{eq} V \ .$$

Φ_{eq} is the neutron equivalent flux, V is the active volume, the constant α is equal to $8 \cdot 10^{-17}$ A/cm. The growth of the leakage current can be reduced by annealing. In the case of the ATLAS pixel detector there will be a few annealing cycles in a year, in each cycle the detector will be kept, typically, two weeks at room temperature. In this way, the leakage current increase ΔI can be reduced to $\Delta I = \alpha_\infty \Phi_{eq} V$, $\alpha_\infty = 3 \cdot 10^{-17} A/cm$. The charge-sensitive amplifier used to amplify the sensor signal must be able to cope with the remaining leakage current.

Type Inversion

The most serious degradation of sensor properties is caused by the radiation induced change of the effective dopant concentration. Suppose that we have a low doped n-type sensor material with the donor concentration of $N_d \approx 10^{12}\ 1/cm^3$. Typically, when the total radiation flux achieves $2 \cdot 10^{12}\ n_{eq}/cm^2$, almost all donor centers will be deactivated [61]. The number of point defects that act as acceptors will further increase linearly with the radiation flux and the material will convert to p-type. The material can be recovered to some extent by thermal annealing. However, after about two weeks of beneficial annealing at room temperature, the effective acceptor concentration starts again to increase. This negative effect is called the reverse annealing. In order to avoid the reverse annealing, the ATLAS pixel sensor will be kept at low temperature (-7^0 C) during the whole year except for the two weeks of room temperature annealing. The effective acceptor concentration after beneficial annealing, assuming the standard silicon detector, can be calculated by using the following formulas [61]: $N_{eff} \approx 1.5 \cdot 10^{-2}\ cm^{-1}\Phi_{eq}$ for neutrons and $N_{eff} \approx 1.9 \cdot 10^{-2}\ cm^{-1}\Phi_{eq}$ for protons. As we can see from these formulas, after ten years of operation and a radiation flux of $10^{15}\ n_{eq}/cm^2$ a low doped n-type material will convert to a moderately doped p-type material. The sensor must be therefore designed in the way to be unsensitive to such type inversion.

Let us recall the structure of the pixel sensor (section 2.3.1). The bulk of the sensor is low doped n-type silicon. The positive collection electrodes are implemented as n^+ doped islands, fig. 2.25. On the back side of the sensor, there is a p^+ doped high-voltage electrode (anode). Negative voltage is applied to the anode in order to deplete the n-type bulk. Such structure might look strange - in order to isolate the positive electrodes one from another, the n-type bulk must be completely depleted, fig. 2.25 (a). However, after only a few months of operation in the LHC environment and a total flux of $2 \cdot 10^{12}\ n_{eq}/cm^2$ the bulk converts to p-type material. The sensor obtains the "classical" structure, where the n^+ electrodes are separated by p silicon. The depleted zone grows then from the segmented side, fig. 2.25 (b) and the sensor can be operated without a full depletion of its bulk.

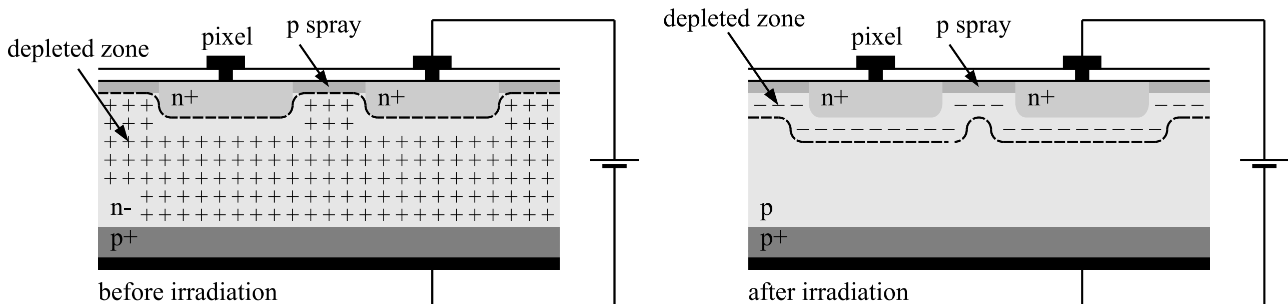


Figure 2.25: (a) - ATLAS pixel sensor before irradiation. (b) - The sensor after irradiation.

The depletion of the sensor bulk is of great importance for the efficient charge collection. In depleted material charge recombination is negligible and the electric field high. Because of that, virtually all electron-hole pairs, which are generated in depleted area, escape recombination. If the bulk is totally depleted, the charge collection efficiency is nearly 100 %. On the other hand, the carriers generated in quasineutral field-free bulk areas recombine before they can be collected.

The thickness of depleted area in an asymmetrically doped $p-n$ junction is given by the following formula [2]:

$$W = \sqrt{\frac{2\epsilon_{Si}V}{eN_{eff}}} \ . \tag{2.16}$$

ϵ_{Si} is the dielectric constant in silicon ($\epsilon_{Si} \approx 11 \cdot \epsilon_0$), N_{eff} is the effective dopant density in the low doped side of the $p-n$ junction, V is the inverse bias voltage. Applying the last formula in the case of the pixel sensor we can conclude that the bias voltage needed for full depletion of the sensor bulk increases with increase of the total equivalent flux due to the radiation induced growth of the effective dopant density N_{eff}. Because of technical reasons, the bias voltage should not be larger than 600 V. It could therefore happen that the sensor cannot be fully depleted when the total flux exceeds some value. This can lead to the decreased collection efficiency. Fortunately, it has been recently discovered [61] that the buildup of acceptor defects decreases by the factor of 3 - 4 if the silicon is "oxygenated" to the oxygen density of about 10^{17} cm^{-3} (indeed only in the case of charged hadron radiation). The use of an oxygenated sensor will probably allow to operate the fully depleted sensors during the whole ATLAS experimental period (10 years and the flux of 10^{15} n_{eq}/cm^2).

Radiation-Tolerant CMOS Circuits

The electronic chips used to amplify and process the signals generated by the ATLAS pixel sensor have been implemented by using a CMOS technology. The basic electronic device of a CMOS integrated circuit is MOS field-effect transistor. As we already mentioned, MOS transistors are sensitive to the radiation induced ionization in $Si0_2$ layer. To recapitulate, we distinguish between two negative effects. One is the creation of the positive trapped charge near the Si-$Si0_2$ interface. The positive trapped charge can induce the parasitic channels between source and drain of a transistor and between contacts of neighboring transistors. The second negative effect is the activation of interface traps. The interface traps cause the increase of the voltage necessary to switch on a transistor.

Fig. 2.26 shows two NMOS transistors. The transistors are implemented in the same p-type bulk. The neighboring transistors are electrically isolated by a trench filled with amorphous $Si0_2$, fig. 2.26. This insulation oxide between transistors is called the field oxide. The field oxide is much thicker than the gate-oxide; in the deep-submicron technology used to implement the ATLAS pixel front-end chip its thickness is about $1\,\mu$m, while the gate-oxide thickness is only 5.5 nm. Let us explain the purpose of the field oxide. A positively biased interconnection line routed between two transistors can act as a gate electrode, fig. 2.26. The structure consisting of two n^+ doped layers, field oxide and an interconnection line above it behaves as a parasitic MOS transistor, see fig. 2.26. Such transistor can produce short connection between the n^+ doped layers of two neighboring regular MOS devices, as shown in fig. 2.26 (a). A parasitic transistor can also be formed in parallel with a regular transistor connecting its drain and source, fig. 2.26 (b). In order to assure that all parasitic transistors are switched off, the oxide between transistors has to be thick enough. Due to the large thickness of the field oxide, the gate potential needed to switch on a parasitic transistor (threshold voltage) is much higher than the threshold voltage of a regular MOSFET. This is easy to understand. The capacitance of the gate electrode above the field oxide is much lower than the capacitance of the gate of a regular

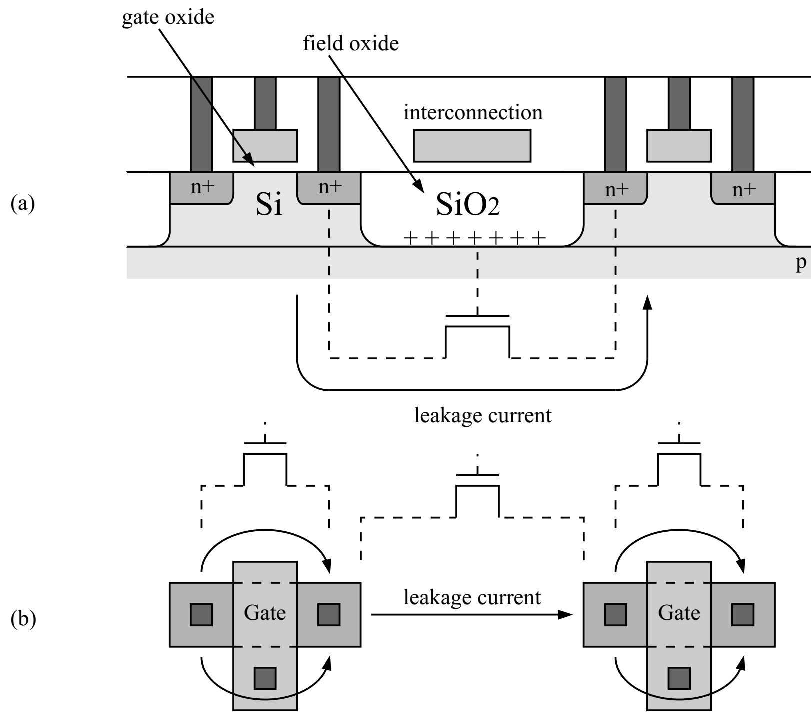


Figure 2.26: Two NMOS transistors and parasitic field-oxide transistors. (a) - vertical cross section. (b) - top view.

transistor. Therefore in the case of a field oxide transistor one needs much higher voltage to induce the same channel charge than in the case of a regular transistor.

Before we precede with the discussion of ionizing damage in CMOS circuits, let us say a few words about the technology used in the design of the ATLAS pixel front-end chip. The development of the semiconductor technology, meaning the decrease of transistor size, improvement of process reliability *etc.* have influenced the number of electrical components in a typical integrated circuit. Since nineteen sixties the optimal number of transistor in a standard digital IC (for example processor) doubles every couple of years.[23] This is illustrated with the graph in fig. 2.27 which shows the number of transistors in a few generation of Intel processors that are fabricated between 1970 and 2000. The prototypes of the ATLAS front-end chip is shown in the same graph. The front-end chip is a mixed-mode analog-digital chip. The transistors in analog circuits are usually much larger than the transistors in logic gates. This explains the smaller number of transistors in a front-end chip then in a processor of the same generation. Assuming a CMOS technology, the scaling down of planar transistor dimensions must be accompanied by the decrease of operating voltages. One tries in this way to reduce the power consumption of a chip and to keep the electric fields constant in order to avoid the negative small-channel effects. The gate-oxide thickness is also scaled down in order to increase the transistor speed.

The decrease of the gate-oxide thickness in newer generations of integrated circuits has a positive effect on the intrinsic radiation tolerance of transistors. One could expect to a

[23]Such exponential technology development was predicted by Gordon. E. Moore in 1965. The optimal number of components per IC is the number which leads to the minimal cost per component. In 1965 it was only 50 components per IC.

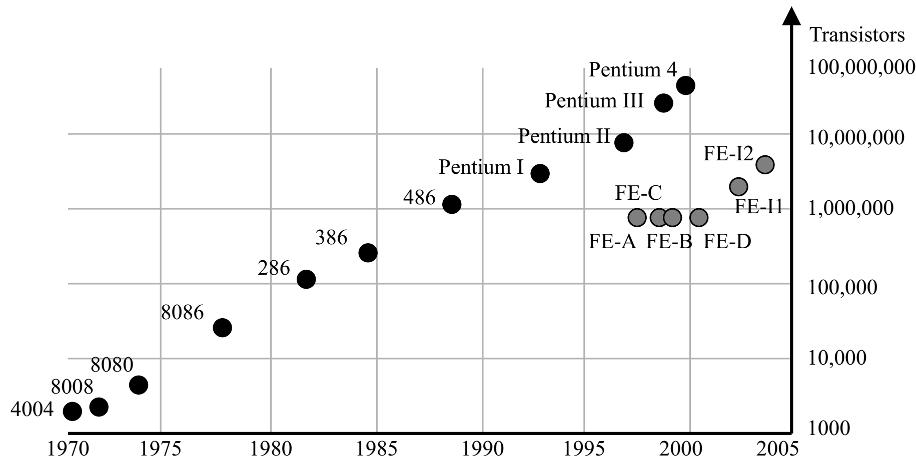


Figure 2.27: Different generations of Intel Processors and different prototypes of the ATLAS pixel front-end chip. The front-end chip is a mixed-mode analog-digital IC. Transistors in analog electronic circuits are usually much larger than the transistors in digital circuits. Also, all NMOS transistors are implemented by using radiation tolerant layout. This explains the smaller number of transistors in a front-end chip compared to the number of transistors in a processor.

first approximation that the amount of positive trapped charge depends linearly on the oxide thickness. This linear dependence holds down to the oxide thickness of 17 nm which corresponds to the gate-oxide thickness in a semiconductor process with 0.8 μm minimal gate length. Such a 0.8 μm technology was standard ten years ago, when the work on the front-end chip started. The front-end prototypes called FE-A, FE-B, FE-C and FE-D have been implemented in different 0.8 μm technologies, fig. 2.27. To implement the final version of the front-end chip (FE-I) we have chosen a smaller technology with the minimal gate length of 0.24 μm. In this technology the gate-oxide thickness is only 5.5 nm. In the case of so thin gate-oxide layer the annealing due to electron tunnelling becomes very efficient. Virtually no positive charge is trapped in the gate-oxide layer. Also, the effects caused by the activation of the interface traps are negligible [27, 28], presumably due to the high quality of the interface between gate-oxide and bulk. The only relevant radiation effect in the case of the 0.25 μm technology is the trapping of positive charge in the field-oxide. To repeat, the positive charge trapped in the field oxide attracts the electrons from the bulk. If electron density becomes sufficiently large, conducting channels will be formed. Such channels can lead to the leakage currents between neighboring n^+ layers. It is important to note that the positive oxide charge cannot induce the hole channel between two p^+ diffusion layers which are separated by the n-type bulk material.

A technique, which can be used to prevent the leakage currents between two NMOS transistors, is to separate the transistors by p^+ guard rings. The electron density in a p^+ region is so low that the trapped oxide charge cannot collect enough electrons to produce the channel. The parasitic channels are therefore cut by the guard rings. In order to prevent the leakage currents flowing around the gate from drain to source of a single transistor, fig. 2.26 (b), one can use the transistors with the annular gate electrode, fig. 2.28. In the case of such transistors drain and source are separated from all sides by the region covered with the gate-oxide. There is no parasitic field-oxide transistor between drain and source.

The two explained layout techniques[24] have been used on the front-end chip. This makes the front-end chip tolerant to the doses beyond 100 Mrad. It is important to note that the minimum-size annular transistor needs much more space than the minimum-size rectangular transistor. This influences the number of devices that can be integrated on a chip, assuming a reasonable manufacturing yield, see fig. 2.27.

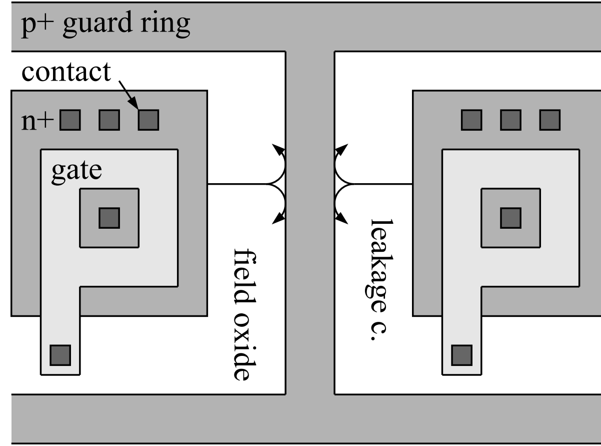


Figure 2.28: Annular NMOS transistors with guard rings.

Electrical Properties of Annular Transistors

Fig. 2.29 shows the layouts of different annular transistors. Transistors with circular gate geometry cannot be realized in the deep-submicron technology used. Mask pattern can contain only strait lines. Despite of that, the circular gate geometry is often a good approximation for the real octagonal geometry. Assuming the circular gate, many important formulas can be analytically derived. For the moment, we will only calculate the size of the equivalent rectangular transistor which generates the same drain-source current under the same bias conditions as a circular transistor of given size (see [27]). Let us consider the circular transistor shown in fig. 2.30. The current flowing through the transistor in the r direction is equal to

$$I = \mu W(r) Q'_{ch}(r) E_r = -\mu W(r) Q'_{ch}(r) \frac{dV_r}{dr} \ . \tag{2.17}$$

μ is the mobility, $W(r)$ is the perimeter: $W(r) = 2\pi r$, $Q'_{ch}(r)$ is the charge density of the carriers in channel per unit area, E_r and V_r are the radial electric field and potential respectively. For the moment, we will assume that the charge density $Q'_{ch}(r)$ depends only on the potential V_r and not explicitly on the coordinate r. We can now separate the variables r and V_r:

$$I \frac{dr}{W(r)} = -\mu Q'_{ch}(V_r) dV_r \ . \tag{2.18}$$

Integration of both sides leads to

[24]more details about the radiation tolerant layout can be found in [26, 27, 28, 29, 30, 66, 67, 68]

$$I \int_{r_{in}}^{r_{out}} \frac{dr}{W(r)} = -\mu \int_{V_r(r_{in})}^{V_r(r_{out})} Q'_{ch}(V_r) dV_r \ . \tag{2.19}$$

From eq. 2.19 we can see that the current changes only its sign if we exchange the electrode potentials. Assuming $Q'_{ch} = Q'_{ch}(V_r)$, a circular transistor behaves symmetrically when we exchange its source and drain.

When we substitute $W(r) = 2\pi r$ in eq. 2.19 and solve the integral, we obtain

$$I = -\frac{2\pi}{\ln(r_{out}/r_{in})} \int_{V_r(r_{in})}^{V_r(r_{out})} \mu Q'_{ch}(V_r) dV_r \ . \tag{2.20}$$

Eq. 2.19 can also be used in the case of a rectangular transistor. The term $W(r)$ corresponds then to the gate width W and the difference $r_{out} - r_{in}$ to the gate length L, fig. 2.30. By substituting $W(r) = const = W$ and $r_{out} - r_{in} = L$ in eq. 2.19 we obtain the formula for the current of a rectangular transistor

$$I = -\frac{W}{L} \int_{V(0)}^{V(L)} \mu Q'_{ch}(V_r) dV_r \ . \tag{2.21}$$

By comparing eq. 2.21 and eq. 2.20 we see that the rectangular transistor with the ratio $W/L = 2\pi/\ln(r_{out}/r_{in})$ conducts the same current as the annular transistor with radii r_{in} and r_{out}. The rectangular transistor with the gate dimensions $W = 2\pi(r_{out} - r_{in})/\ln(r_{out}/r_{in})$ and $L = r_{out} - r_{in}$ is called the equivalent transistor.

Fig. 2.31 shows a transistor with a circular gate and its equivalent rectangular transistor. As we see, the width-length ratio of the equivalent transistor is quite large. Such a transistor has a low resistance in linear region and a large transconductance in saturation. In analog circuits one needs very often the transistors with lower current gain. This is a serious disadvantage of the circular gate geometry. If we want to make a circular transistor that behaves as a rectangular transistor with small W/L ratio, we must make the outer radius r_{out} much larger than the inner radius r_{in}, due to the slow increase of the logarithm function in the denominator (eq. 2.20). This is often not feasible because it leads to an unacceptably large gate area since the inner contact cannot be fabricated too small. Assuming $(r_{out}/r_{in}) = 10$, which already implies quite a large gate area, the effective width-length ratio is 2.61. Considerably smaller values are very often preferable. This makes the chip design based on annular transistors difficult.

2.5 Architecture and Design of the Front-End Chip

2.5.1 Introduction

The work on the ATLAS front-end chip began almost ten years ago. It seems that the idea to build a complex radiation tolerant pixel-readout chip was ahead of that time. The smallest feature size in the semiconductor technologies that was up to date then was 0.8 μm. We should compare these technologies with the contemporary 90 nm technologies - the transistor area today can be almost 100 times smaller than it was achievable when the

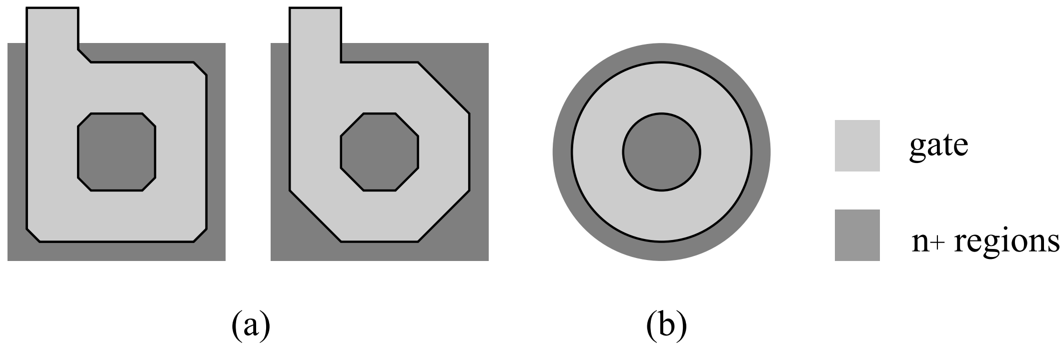


Figure 2.29: (a) - real annular transistors. (c) - circular transistor.

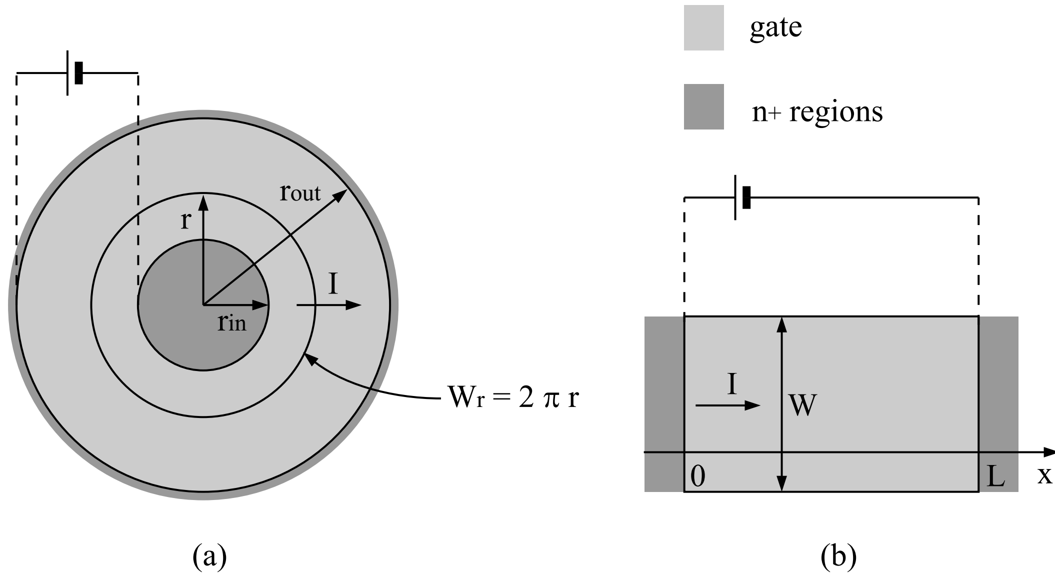


Figure 2.30: (a) - circular transistor. (b) - rectangular transistor.

work on the front-end chip began. Despite of such low integration density, fully functional front-end chips have been produced in 0.8 μm technologies [13, 17, 18, 21, 22]. One chip is called FE-A; it was designed in AMS 0.8 μm BiCMOS technology and submitted in 1997. A similar purely CMOS chip, called FE-C, was submitted in 1998. The design of these chips was done in collaboration between the institutes in Bonn and Marseille. At the same time a front-end chip with different digital architecture, called FE-B, was designed in Berkeley. The chip was implemented in HP 0.8 μm CMOS process. These old front-end chips fulfilled almost all electrical requirements. However, they were not radiation tolerant. In this time the deep-submicron technologies (feature size $< 0.25\,\mu$m) were not yet available. Only special radiation hard technologies which were developed for space applications possessed the radiation tolerance required for the ATLAS application. The next front-end chip was implemented in such a technology. The chip is called FE-D1 [12] and it was designed in DMILL 0.8 μm BiCMOS rad-hard technology. The submission went out in 1999. The chip was functional, unfortunately the manufacturing yield was extremely low. The poor yield was related to technology problems as high leakage currents of NMOS transistors. However, the foundry never succeeded isolating the problems. A new version of DMILL chip (FE-D2) with more conservative design (static RAM cells)

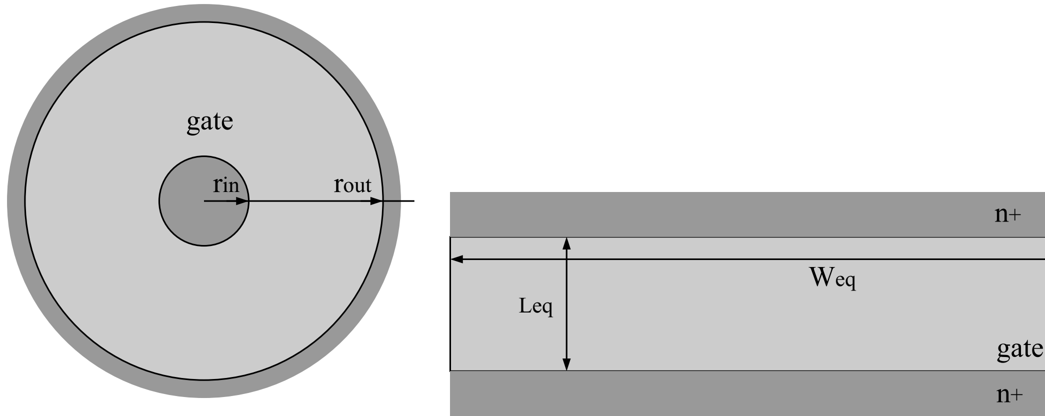


Figure 2.31: Circular transistor with gate radii $r_{in} = 0.5\,\mu$m and $r_{out} = 2\,\mu$m. The equivalent transistor has the following dimensions: $W_{eq} \approx 6.8\,\mu$m and $L_{eq} \approx 1.5\,\mu$m.

was submitted in 2000. The yield was better but still unacceptable. The work with this vendor was terminated.

The appearance of deep-submicron technologies opened new possibilities in the radiation tolerant chip design. The oxide thickness of a few nm makes the tunnel annealing in the gate-oxide very efficient. A small minimal size of the mask patterns allows the use of annular transistors and guard rings without making transistors unacceptably large. The new version of the front-end chip was implemented in a $0.25\,\mu$m deep-submicron CMOS process. The chip is called FE-I. The work on FE-I started in 2000 and the author of the thesis was involved from the beginning in the design process. The chip was designed in collaboration between Berkeley and Bonn. It is not easy to put the strict boundaries between the work done by each institute or designer. Very roughly, Laurent Blanquart (Berkeley) designed the pixel amplifiers and important analog blocks. He did also the final chip integration. Emanuelle Mandelli and Gerrit Meddeler (Berkeley) designed the digital readout system of the chip and did the final chip verification. Bonn designed various analog and digital blocks placed in the pixel matrix and on the chip periphery. The author of the thesis worked in Bonn, together with Peter Fischer and Giacomo Comes. The coordinator of the project was Kevin Einsweiler (Berkeley). In the last three years we submitted three full-scale chips and four test chips. The last chip version, referred to as FE-I3, is the production chip for the ATLAS pixel detector. The chip fulfills all requirements needed. The results given in the following subsections refer to the latest version of the ATLAS pixel front-end chip.

The first subsection gives a short overview of the chip architecture. The second subsection describes the analog pixel cell in its full complexity. A few alternative solutions are also introduced. The third section is devoted to the digital readout system of the chip. The smaller blocks used to test, control and bias the pixel matrix are described in the fourth section. The fifth section gives the results of measurements performed on the front-end chips.

2.5.2 Architecture of the Front-End Chip

The internal architecture of the front-end chip is shown in fig. 2.32. The front-end chip has a pixel matrix with 18 columns and 160 rows.

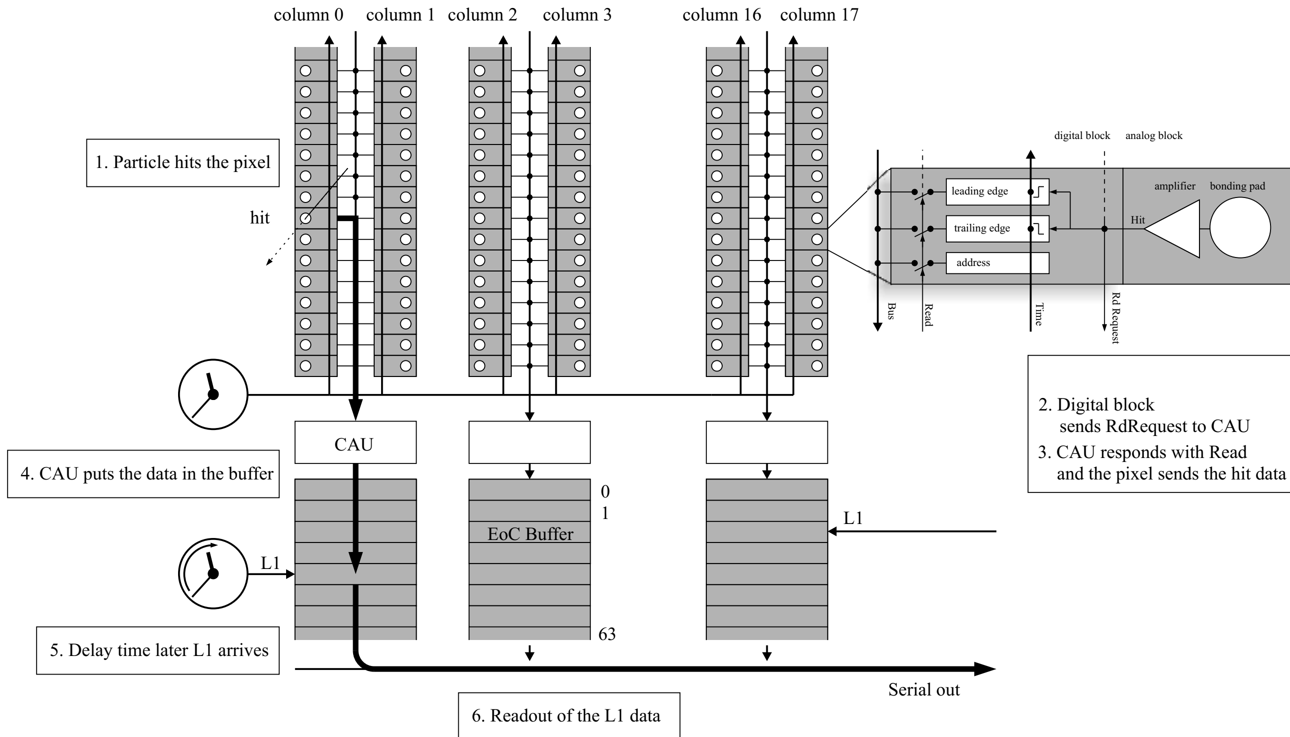


Figure 2.32: Architecture of the front-end chip. Signal flow.

Every pixel contains an analog and a digital block, see fig. 2.32 right. The task of the analog pixel block is to amplify and shape the signal generated by the sensor pixel connected to the pixel of the front-end chip. The output signal of the analog block is a pulse with defined amplitude. We refer to this signal as the *Hit* pulse. The *Hit* pulse length is linearly proportional to the charge deposited by an ionizing particle in the sensor pixel. The leading edge of the *Hit* pulse is coincident with the particle hit. The task of the digital block is to generate a hit information which contains the pixel address and the leading/trailing edge times of the *Hit* pulse. The hit information is then sent to the column arbitration unit (CAU), located at the bottom of the pixel column, fig. 2.32.

To be able to measure the *Hit* pulse time, the pixel electronics needs a time information. Time is measured by using an 8-bit counter which is incremented every 25 ns. The time information is distributed via bus system in all pixels.

CAU transmits hit information from a pixel column-pair to the end-of-column (EoC) buffer, fig. 2.32. The EoC buffer is a memory unit with space for 64 hits. A buffer is shared by two pixel columns.

Let us now explain why do we measure the *Hit* pulse time and why do we need the EoC buffers? The flux of the particles coming from the LHC interaction point is so high, making impossible to read out all of the hit coordinates and to store it for later off-line analysis. Fortunately, the other subcomponents of the ATLAS detector, like the calorimeters and the muon spectrometer, can detect whether an interesting interaction occurred in a bunch-crossing. This gives the possibility to reduce the amount of transmitted data. Every interesting crossing is followed by a trigger signal. The front-end chips in the pixel detector receive the level 1 (L1) trigger signals and transmit only the data that originate from the events to which the received L1 signals are pointing. The other data are deleted. Every

L1 signal is generated with some known delay (not longer than 3.2 μs) relative to the bunch-crossing that causes the trigger generation. The hit data are locally stored on the chip, until the delay time is up. To be able to distinguish between the data that belong to different bunch crossings and to be able to monitor the age of the hits stored in chip memory, *Hit* pulse time is measured and stored as a part of hit information.

A digital block referred to as the readout controller (ROC) does the formatting and serializing of the hits selected by the L1 signals. The hit data that correspond to different L1 signals are sent out in the order those trigger arrive.

2.5.3 Analog Pixel-Electronic

The main goal of this work is the design of various analog circuits on the front-end chip.

Digital Crosstalk

The front-end chip is a mixed-mode integrated circuit - every pixel cell contains digital and analog circuits. A great concern in the design of the analog and digital pixel blocks is the elimination of the crosstalk between them. Let us briefly explain the origins of the crosstalk.

Each change of logic state in a digital circuit implies a change of its node potentials. In the case of a CMOS digital circuit the potentials, used to define logic levels, swing between the ground potential Gnd (logic 0) and the positive power supply potential Vdd (logic 1). Each node has a certain capacitance, see fig. 2.33. In order to charge the node

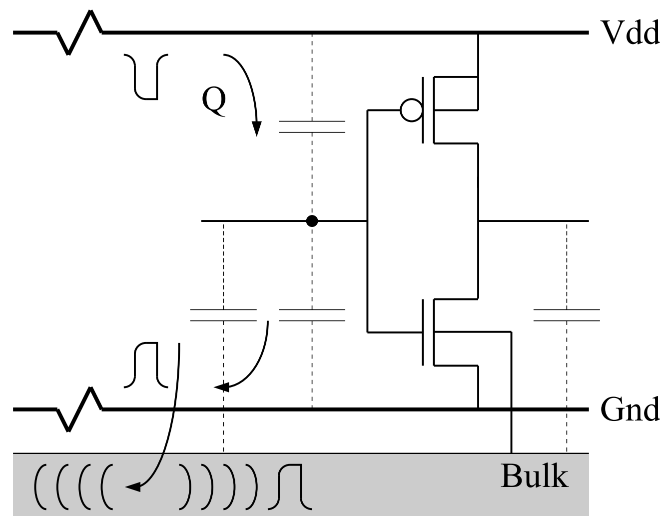


Figure 2.33: Inverter with parasitic capacitances.

with capacitance C, the power supply lines must provide an amount of charge given by the formula: $Q = C \cdot Vdd$. Assume that many logic circuits are triggered by the same clock signal. Many digital circuits changes then its state in the same moment. The transient current in supply and return lines caused by these state transitions can be quite large. Due to finite resistance of the lines, the supply voltage drops for a moment. In this way, a digital activity generates voltage spikes. There is one another unwanted effect. Due to gate-bulk and drain-bulk capacitances of the transistors used to implement the logic circuits some

amount of charge is also injected into the bulk, fig. 2.33. Such short charge pulses generate voltage noise which propagates through the bulk like waves. All effects described above are pronounced by the use of annular transistors since such transistors have considerably larger parasitic capacitances than the minimum size rectangular transistors.

Differential Current-Logic (DCL)

In order to eliminate the voltage spikes due to logic transitions, other types of logic circuits (gates) might be used. The idea is here to reduce the swing between logic 1 and 0 and to use differential signals. A differential signal is encoded with two voltages. During signal transition one voltage increases, while the other decreases. Within the scope of this work, a special type of low swing logic gates has been implemented and tested in a test chip. The structure of the logic gates is published in [46]. Such type of logic circuits is called the differential current logic (DCL). The transistor scheme of a DCL invertor is shown in fig. 2.34.

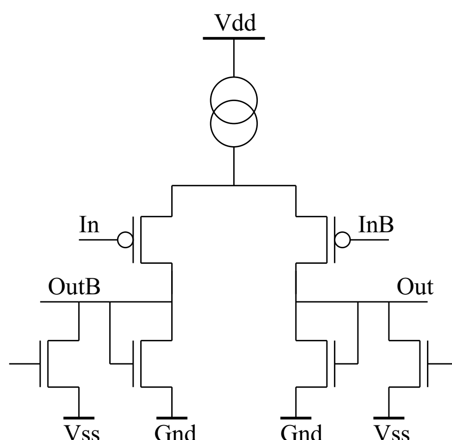


Figure 2.34: DCL inverter.

Pixel Amplifier with NMOS Input-Transistor

In this section we will explain the design of the amplifier used to amplify the sensor signal. In section 2.2.2 we introduced the charge-sensitive amplifier. Fig. 2.35 shows the simplest transistor realization of such an amplifier.

Let us first discuss the crosstalk between the digital circuits and the charge-sensitive amplifier implemented in a pixel of the front-end chip.[25] The need to reduce the crosstalk influences the choice of the optimal amplifier structure. Fig. 2.36 shows the simplified cross-section of a pixel, assuming an amplifier with input NMOS transistor like that shown in fig. 2.35. The input NMOS transistor of the amplifier is shown on the left side of the figure. An NMOS transistor, which is the part of a CMOS logic gate, is shown on the right side. As we explained, logic transitions generate voltage pulses on power supply lines. Assume that both NMOS transistors share the same ground connection, fig. 2.36. Each transition of logic level leads to a positive ΔV pulse at the ground line, see fig. 2.36.

[25]See, also, [33].

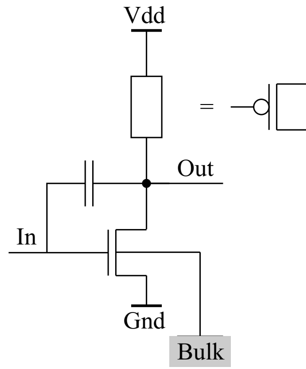


Figure 2.35: Amplifier with NMOS input-transistor.

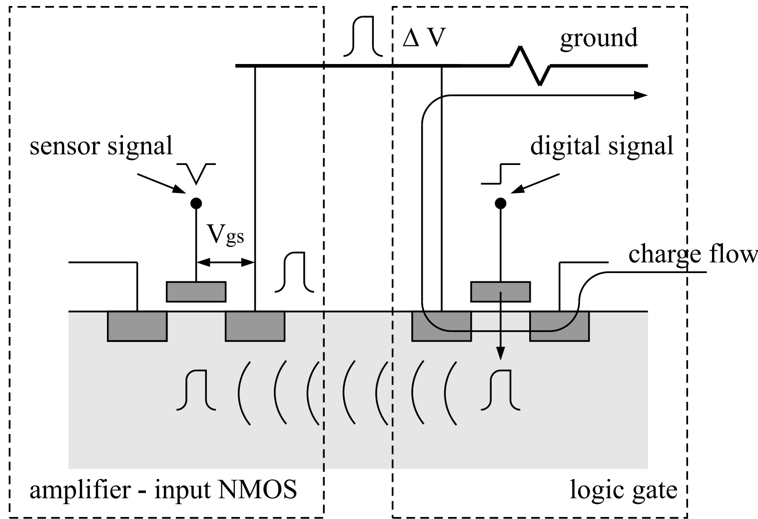


Figure 2.36: Crosstalk.

This pulse is seen also at the source of the input NMOS transistor. Such a pulse causes the same response of the amplifier as a ΔV signal an the input node since it leads to the same change of the gate-source voltage. In this way, a digital pulse can produce a fake noise hit.[26] The voltage pulses, which propagate through the bulk, are also captured and amplified by the input NMOS transistor, fig. 2.36.

Pixel Amplifier with PMOS Input-Transistor

In order to diminish the crosstalk in pixels of the front-end chip, a PMOS transistor can be used as the input device of the charge-sensitive amplifier, fig. 2.38. The source of the input PMOS transistor is connected to a quiet analog supply line referenced to as $VddRef$ which is separated from the digital power line Vdd. Since the input PMOS transistor has its own substrate, n-well, which can be also connected to the quiet analog potential, the pick-up of digital noise-pulses through the p-type bulk is eliminated, see fig. 2.37. Another measure to reduce the crosstalk is to separate the source diffusion contacts of the NMOS transistors in digital circuits from the bulk. The sources of digital NMOS transistors are

[26]Assuming a detector capacitance of 400 fF, a digital pulse of only 0.8 mV is equivalent to a sensor signal of 2000 e.

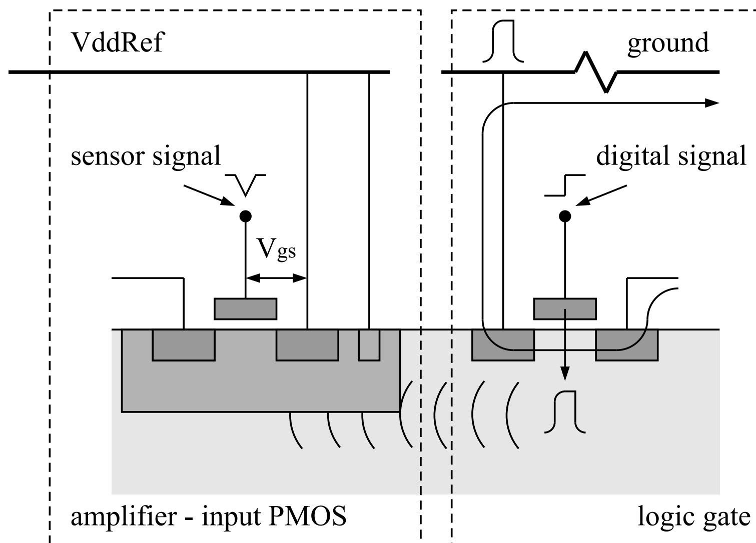


Figure 2.37: Reduced crosstalk in the case of the PMOS input-transistor.

connected to a digital ground line that we reference to as *DGnd*. The bulk connection is called *Gnd*.

As we have seen, the choice of the input device type in the charge-sensitive amplifier is driven by the need to reduce the crosstalk. Beside the input transistor, the amplifier has at least one another transistor used to implement the active load. The choice of the active load transistor is of great importance for the behavior of the amplifier.

The active load is usually implemented by using a transistor which operates in saturation and has a constant gate-source voltage, see NMOS transistor in fig. 2.38. Such transistor can be considered as (almost) perfect current source. For small signals, the load transistor behaves as a high-ohmic resistor. Let us estimate the load resistance.

The small-signal drain-source resistance of a transistor, which operates in saturation, is given by the formula [53, 52]:

$$R_{ds} = V_A/I_{DS} \approx E_{sat}L/I_{DS} \ . \tag{2.22}$$

V_A is the Early voltage at the onset of saturation. The Early voltage is proportional to the length of the transistor gate L. I_{DS} is the drain-source current, E_{sat} is the saturation voltage which is, typically, 6.4 V/μm for PMOS and 2.4 V/μm for NMOS transistor [53, 52]. In the case of an amplifier with PMOS input-transistor, it would be simplest to realize the active load by using an NMOS transistor, see fig. 2.38. However, due to annular gate geometry of NMOS transistors, large values for L cannot be achieved without great increase in transistor area. A small L leads to a decreased small-signal resistance of the active load, see eq. 2.22.

Small load resistance is a serious disadvantage, let us briefly explain why. In chapter 2.2.2, we discussed the action of the capacitive feedback in a charge-sensitive amplifier. Provided that the DC voltage-gain, given by $g_m R_o$, has a large value, the input potential of the amplifier is kept at a constant level (g_m is the transconductance of the input transistor and R_o is the total small-signal resistance of the node *Out*, approximately, equal to the drain-source resistance of the load NMOS transistor). If the DC voltage-gain is not large enough, the input potential of the amplifier changes significantly after the integration of

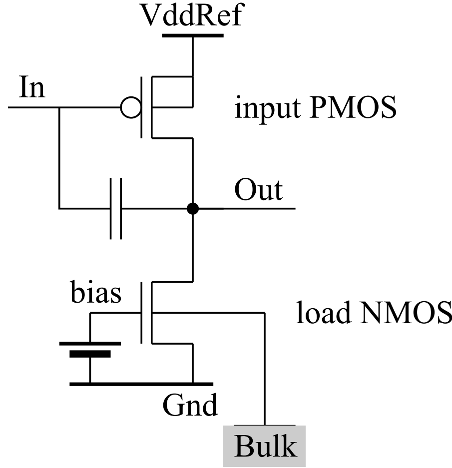


Figure 2.38: Simple amplifier with PMOS input-transistor.

the sensor signal on the feedback capacitor. This can cause a few negative effects. Only a part of the signal charge is collected by the feedback capacitor. This leads to a signal loss. As we derived in chapter 2.2.2, the amplitude of the output pulse is equal to

$$v_{out}(t_p) = \left(\frac{Q}{C_f}\right)\left(\frac{(g_m R_o + 1)C_f}{C_i + (g_m R_o + 1)C_f}\right) .$$

C_f is the feedback capacitance, C_i is the capacitance of the input node, Q is the signal charge. The output signal amplitude becomes $\approx Q/C_f$ only if the condition $g_m R_o C_f \gg C_i$ holds. The second negative effect, caused by a poor DC gain, is the cross-coupling of signals from pixel to pixel. Due to all these reasons, it is of great importance to have a large voltage gain. With an NMOS transistor as active load this is hard to achieve.[27]

Pixel Amplifier with Folded Cascode

As we shown, the main disadvantage of the amplifier shown in fig. 2.38 is a small drain-source resistance of the NMOS active load transistor. A way to overcome this disadvantage is to design an amplifier with a PMOS input and a PMOS load transistor. The basic scheme of such an amplifier is shown in fig. 2.39. The voltage gain of the amplifier is given by the formula $g_m R_o$, where R_o is the total small-signal resistance of the node Out. This resistance is the parallel connection of the drain-source resistances of all transistors connected to the node. The circuit must be designed in the way that the small drain-source resistance of the annular NMOS current source I_n does not dominate the equivalent resistance of the node Out. The following circuit is used on the front-end chip, fig. 2.40.

[27]Let us try to estimate the DC voltage-gain $g_m R_o$ for the circuit shown in fig. 2.38. Assume that R_o is equal to the drain-source resistance of the NMOS load transistor: $R_o \approx E_{sat} L/I_{ds}$, where $E_{sat} = 2.4\ V/\mu$m. The values larger than $L = 1\ \mu$m are hard to achieve. The bias current of the front-end pixel-amplifier is, typically, $8\,\mu$A. We have then $R_o \approx 300\ K\Omega \cdot L[\mu\text{m}]$. We will assume that the input PMOS transistor operates in weak inversion. This leads to the largest g_m value by a given bias current and therefore the largest DC gain. The transconductance is given by: $g_m \approx I_{ds}/(1.5U_T) \approx 200\ \mu Si$, ($U_T = 26\ mV$). The DC gain is only $g_m R_o \approx 60 \cdot L[\mu\text{m}]$. On the front-end chip, the input PMOS has about 0.65 of the its maximal transconductance in the case of weak inversion.

The purpose of the cascode[28] transistor (fig. 2.40) is to keep the potential of the node A at a nearly constant level. Since the potential of the node A does not change, the drain-source resistances of *all* transistors connected to the node A does not matter and these resistances can be neglected. The resistance of the node *Out* is then, approximately, equal the output resistance of the load transistor. The load transistor is of PMOS type,

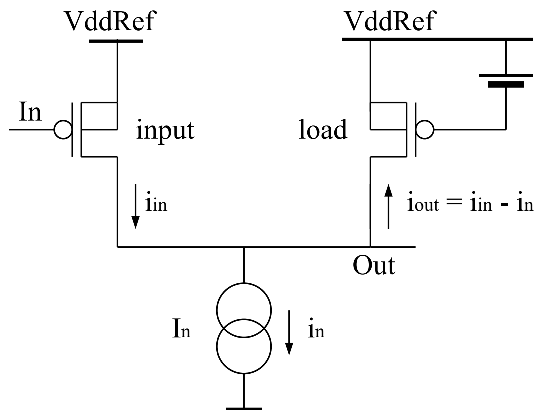


Figure 2.39: Amplifier with PMOS input and PMOS load transistor.

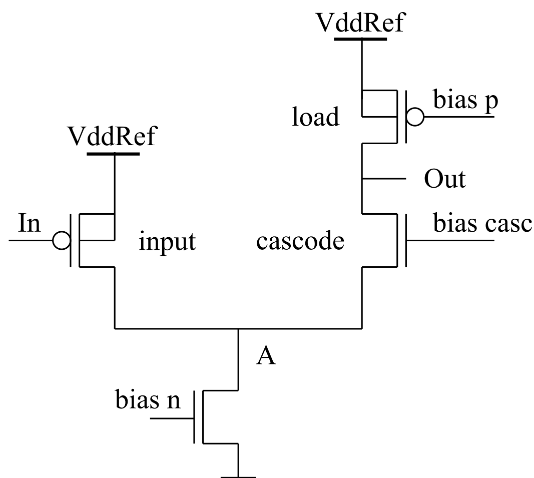


Figure 2.40: The folded-cascode amplifier with PMOS input-transistor.

it gives us the freedom to chose a large L. The second advantage, compared to the circuit in fig. 2.38, is that the current flowing through the active load can be decreased without affecting the current flowing through the input transistor. In this way, we can additionally enhance the output resistance of the load transistor and, at the same time, have a large transconductance of the input transistor.[29] To summarize, by using the folded-cascode configuration (fig. 2.40), much larger voltage gains can be achieved than in the case of

[28]The term cascode originally appeared in a January 1939 paper in the *Review of Scientific Instruments* by Hickman and Hunt entitled "On Electronic Voltage Stabilizers." They explained this therm as a contraction of a phrase cascade to cathode.

[29]In the pixel-amplifier of the front-end chip the bias current of the input transistor is, typically, $8\,\mu$A and the current flowing through the load transistor is $1.5\,\mu$A.

the simple configuration, fig. 2.38.[30]

Feedback Circuit

In this section, we will describe the complex feedback circuit used to stabilize the pixel pre-amplifier. The feedback circuit has been invented and designed by Laurent Blanquart [9, 11].

As already described, a charge-sensitive amplifier can be realized with the feedback circuit consisting of the parallel connection of a capacitor and a resistor, see fig. 2.12. The feedback capacitor in the front-end amplifier has the value of 6 fF. It is implemented as metal-metal capacitor. High-ohmic linear resistors are usually very large. The usual way to make a resistor in an integrated circuit is to use a MOS transistor operating in linear region. In our case, the input signal of the charge-sensitive amplifier has negative polarity. In the case of such signal direction, an NMOS transistor would be the correct choice for the feedback device [12], fig. 2.41. In the stationary state, the input and output poten-

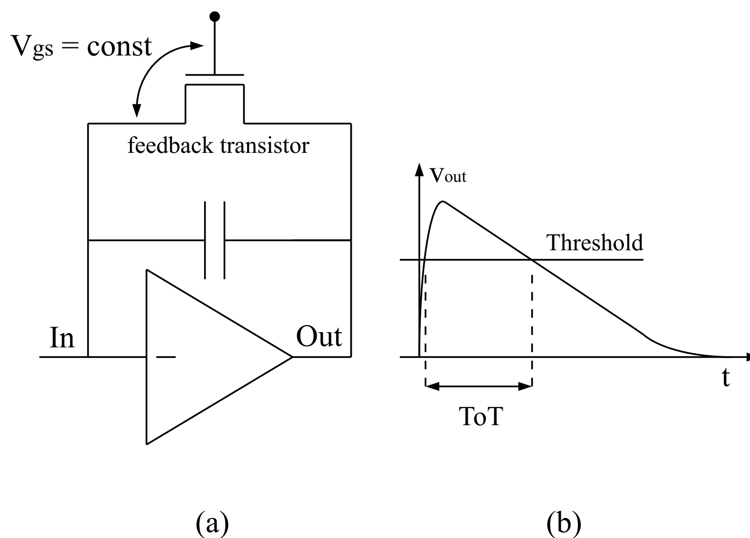


Figure 2.41: (a) - resistive feedback implemented with an NMOS transistor. (b) - shape of a pulse at the output node.

tials of the charge-sensitive amplifier are equal. The feedback transistor behaves then as a linear resistor. A negative sensor signal causes a positive output pulse. In the case of the output pulse amplitude, which is higher than, typically, a few thermal voltages U_T ($U_T = 26mV$), the current of the NMOS feedback transistor saturates. Further increase of the output potential is not accompanied by the increase of the feedback current. Since the feedback capacitor is discharged by a constant current, the output pulse has, approximately, a triangular shape, fig. 2.41 (b). This feature can be used for the digitization of the output signal amplitude by measuring the "time over threshold" (ToT), see fig. 2.41 (b). The second good property of the NMOS feedback is that the feedback current increases more rapidly when the output potential becomes smaller than the input potential. This

[30] R_o is equal to the drain-source resistance of the PMOS load. If we assume a load bias current of 1.5 μA and $L = 3\,\mu$m (the values on the front-end chip) the resistance is: $R_o \approx 6.4L[\mu\text{m}]/I_{ds} = 12.8M\Omega$. This is about 40 times better than in the case of the amplifier in fig. 2.38. Assuming $g_m \approx 200\mu Si$, we have a DC gain of about 2400.

prevents undershoots and makes the output signal unipolar. To summarize, the feedback action is strong when desirable (to keep the baseline on a constant level and to prevent undershoots) and weak when the output potential exceeds some threshold. In order to achieve good performances, the feedback transistor must be carefully dimensioned, *i.e.* it must have a small transconductance. In the case of an annular NMOS, this is not achievable. Because of that, the amplifier in the front-end chip uses a feedback circuit with a transfer characteristic equivalent to that of the NMOS based circuit, but implemented by using only PMOS transistors [9, 11], fig. 2.42. In the stationary state, the action of

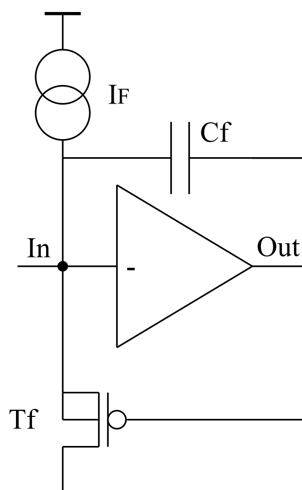


Figure 2.42: PMOS-based feedback circuit.

the feedback transistor Tf can be modelled by a resistor with resistance $1/g_{mf}$ connected between the nodes In and Out. g_{mf} is the transconductance of the transistor Tf. In the case of a high output pulse, the transistor Tf becomes switched off and the capacitor C_f is discharged by the constant current delivered by the source I_F. The feedback current is, typically, 5 nA.

Leakage-Current Compensation (1)

As we explained in section 2.4.4, the radiation induced displacement damage in the sensor bulk material causes an increase of the sensor leakage current. The leakage current per pixel is expected to reach 50 nA. The feedback circuit of the charge-sensitive amplifier must prevent the leakage current to be integrated on the feedback capacitor together with the signals produced by particle hits. The leakage current (I_{leak}) must therefore be compensated by delivering the same amount of current to the input node of the amplifier. In the case of the simple circuit in fig. 2.42, the source I_F should provide the current equal to I_{leak} plus the value needed for the feedback action. The front-end amplifier has an analog circuit that automatically adjusts the current of the source I_F to the needed value. The circuit that performs the active leakage-current compensation is shown in fig. 2.43. The circuit introduces a new feedback path: beside the fast $In - Out - In$ feedback loop, needed to discharge the feedback capacitor after the integration of the sensor signal, there is a slow loop $In - Out - A - In$, fig. 2.43. It is easy to show that the new feedback is negative. In order to check it, we can cut the loop, for example, before

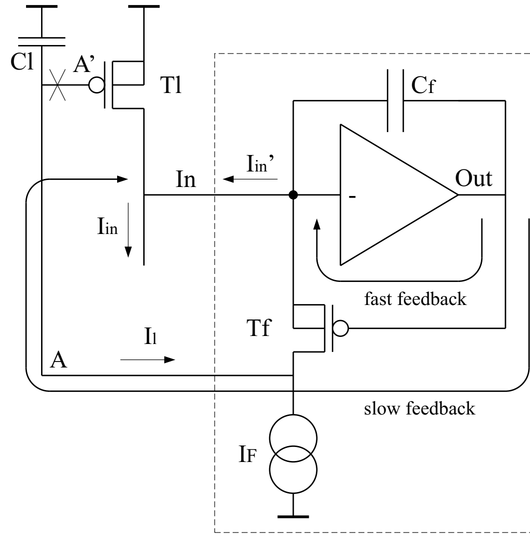


Figure 2.43: Active leakage current compensation.

the gate of the transistor Tl, connect the test generator to A' and calculate the potential at A, fig. 2.43.

Let us now calculate the simplified form of the transfer function: $V_{out}(s)/I_{in}(s)$. Note, the circuit enclosed with the dashed line behaves like the charge-sensitive amplifier that we introduced in section 2.2.2. The input-output transfer function of the circuit is (see eq. 2.9):

$$\frac{V_{out}(s)}{I_{in}(s)'} \approx \frac{R_f(1 - s\tau_z)}{(s\tau_r + 1)(s\tau_f + 1)} \approx \frac{1/g_{mf}}{(s\tau_f + 1)} \equiv H_0(s) \ . \tag{2.23}$$

τ_f is the feedback time-constant: $\tau_f = C_f/g_{mf}$. The AC current $I_l(s)$ (fig. 2.43) is, approximately, equal to $g_{mf}V_{gsf}(s)$, which is equal to $g_{mf}(V_{out}(s) - V_{in}(s)) \approx g_{mf}V_{out}(s)$, since the potential V_{in} does not change significantly. Therefore the circuit enclosed by the dashed line has, approximately, the transfer function: $I_l(s)/I_{in}(s) = H_0(s)g_{mf}$. The open loop gain ($\beta A$) of the slow feedback loop $In - Out - A - In$ is $\beta A \approx -g_{ml}H_0(s)g_{mf}/sC_l$. g_{ml} is the transconductance of the transistor Tl. The forward amplification $V_{out}(s)/I_{in}(s)$ assuming no feedback $In - Out - A - In$ is equal to $H_0(s)$. This can be calculated by connecting a test current generator to node In and calculating $V_{out}(s)$, provided that the feedback loop is cut. The complete transfer function $V_{out}(s)/I_{in}(s)$ assuming feedback can be obtained by using eq. 2.5. We have:

$$\frac{V_{out}(s)}{I_{in}(s)} \approx \frac{H_0(s)}{1 + \beta A} \approx \frac{(1/g_{mf})(sC_l/g_{ml})}{(s\tau_f + 1)sC_l/g_{ml} + 1} \ . \tag{2.24}$$

The time-constant $\tau_l \equiv C_l/g_{ml}$ describes the reaction time of the circuit for the leakage compensation. It should be much larger than the feedback time-constant τ_f otherwise the circuit also "compensates" the sensor signal, which leads to a decreased output amplitude. If $\tau_l \gg \tau_f$, we can write the transfer function in the simplified way:

$$\frac{V_{out}(s)}{I_{in}(s)} \approx \frac{(1/g_{mf})s\tau_l}{(s\tau_f + 1)(s\tau_l + 1)} \ . \tag{2.25}$$

The amplification of the system becomes virtually zero for the input currents with time constants smaller than τ_l. By proper choosing of the time-constants of the feedback

system, we can achieve that the leakage current is not amplified, whereas the signal does not see the damping action of the circuit for the leakage compensation.

The complete scheme of the feedback circuit is shown in fig. 2.44. There is an additional

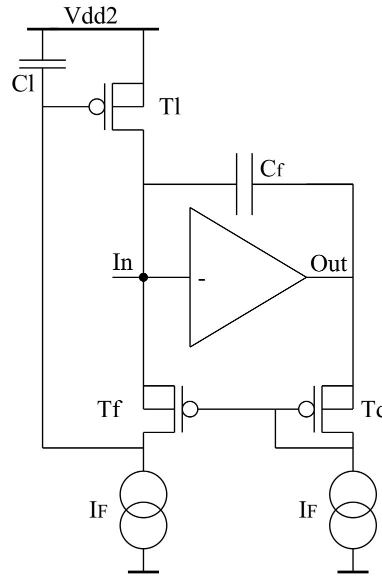


Figure 2.44: Feedback scheme of the charge-sensitive amplifier.

diode connected transistor Td acting as a generator of a constant voltage between the nodes *Out* and the gate of the transistor Tf. We neglected this device in the simplified AC analysis. The purpose of this transistor is to increase the DC potential at the output of the amplifier in order to assure the correct biasing of the cascode transistor in the amplifier, fig. 2.42. Note that the source of the transistor Tl is connected to a special supply potential $Vdd2$. This potential has to be about a threshold voltage V_{th} lower than the supply potential used for the input transistor of the amplifier ($VddRef$), in order to achieve the proper biasing of the transistor Tf. The circuit of fig. 2.44 has the nice property that the node potentials depend only on the $VddRef$ potential and not on the ground level. It makes the circuits insensitive to the variations of the ground potential. The ground rejection-ratio is high.

Leakage-Current Compensation (2)

Within the scope of this work, a slightly different amplifier/feedback system has been designed and tested on a test IC. This amplifier is designed to be the sensor signal amplifier in a pixel-readout chip for an X-ray detector. Although this application is completely different from the application of the ATLAS front-end chip, the electronic requirements for both amplifiers are similar. The amplifier used for the X-ray pixel detector is differential with folded cascode, fig. 4.7.

A differential amplifier has the good property that voltage pulses at supply lines are not amplified. It has good power rejection ratio.[a] Therefore in the X-ray chip it is not necessary to separate analog and digital supply lines. It is, however, of great importance to provide a quiet reference potential for the second amplifier input.

A differential amplifier has two disadvantages with respect to a single input amplifier. Assuming the same type of the input transistor and the same bias current, a differential transconductance-amplifier has two times smaller current gain than the equivalent single-input amplifier. The reason for such a gain deficit of a differential amplifier is that the total bias current is shared by two input transistors. Also, each of the input transistors "sees" only the half of the input voltage. The second disadvantage of a differential amplifier with respect to a single input amplifier is increased output noise power caused by the noise in the channels of the input transistors. A differential amplifier has two input transistors, whose noise contributions quadratically add, whereas a single input amplifier has only one input transistor. These two disadvantages are not critical for the application in the X-ray chip. A small rise time of the amplifier is not required and the sensor signal amplitudes are quite large. Also, the ionizing radiation doses in a typical X-ray application are much lower than the dose experienced by the ATLAS pixel detector. It is therefore not necessary to use annular NMOS transistors in the X-ray chip. Because of that, we are able to choose the transistor type and geometry without any constraint, so that NMOS transistors are chosen to implement the input devices. The feedback circuit for the X-ray sensor amplifier is also differential. It is based on a circuit published in [25], fig. 2.46. The circuit behaves in the way similar to the circuit shown in fig. 2.42. For high output signals, the feedback capacitor C_f is discharged by the constant current I_F. The transistor Tl compensates the leakage current.

[a] As we have seen, in the case of a single input transistor it is not so. The pulses at the source of the input transistor are amplified in the same way as the input voltage signal.

Threshold and the Generation of the Hit Signal

The output signal of the pixel amplifier is electronically compared with a threshold potential. The comparison is done by the circuit called comparator. The comparator is a high gain differential amplifier without feedback.[31] In our case the potential at the output of the comparator becomes low (approximately *Gnd*) when the output potential of the pixel amplifier exceeds the threshold potential. The output signal generated by the comparator is referred to as the *Hit* signal. The difference between the threshold potential and the DC potential at the output of the amplifier (called also the base line potential) will be called the threshold. It is highly desirable to have a constant value of the threshold in the

[31] A positive feedback is sometimes used to achieve a transfer characteristic with hysteresis.

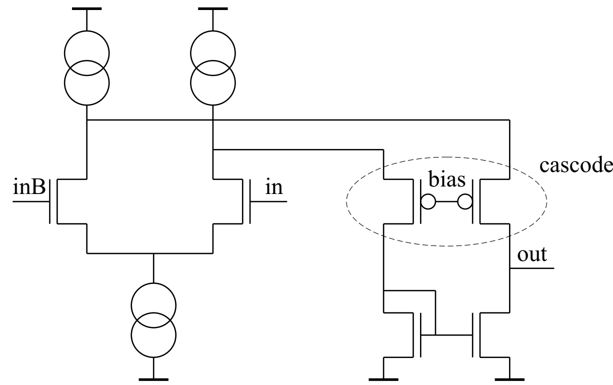


Figure 2.45: Differential amplifier with folded cascode.

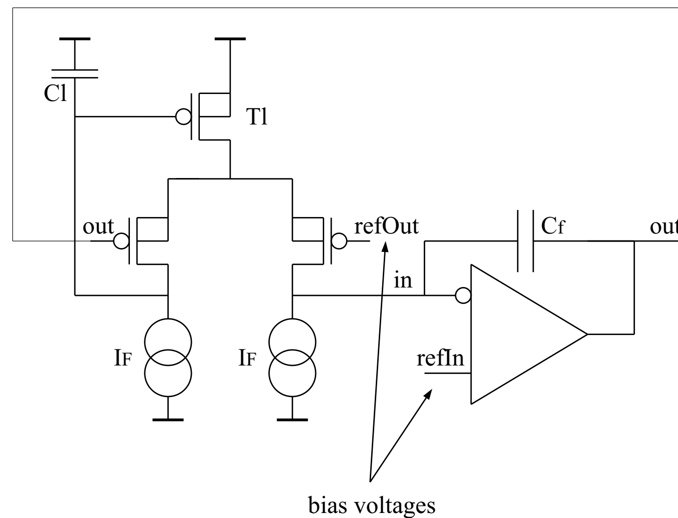


Figure 2.46: Differential feedback circuit and leakage current compensation.

whole pixel matrix. As we have seen, a threshold dispersion limits the minimal threshold value that can be used and therefore the minimal signal that can be detected. In the following chapter we will describe the origins of the threshold dispersion and the design measures used to minimize it.

Systematic Threshold Mismatch

The signal and power supply pads are placed at the bottom of the front-end chip, as shown in fig. 2.20. Although designed as wide as possible, the power supply lines on the chip have resistance which cannot be neglected. Constant currents flowing through the analog power lines produce voltage drops across the lines. The pixels, which are at the largest distance from the supply pads, are especially affected by voltage drops. The analog supply voltage in these areas is considerably lower than in the matrix regions closer to the pads. The non-uniformity of the supply voltage in the pixel matrix influences the design of the comparator and the threshold generator. The simplest implementation of the threshold generator would be to generate a global threshold potential and distribute it in the whole pixel matrix. The problem is here that the DC potential at the output of a pixel amplifier depends on the positive supply potential and the supply potential

is not constant in the matrix due to the potential drop. The threshold defined as the difference between the threshold and base line potentials would be lower in the pixels which are further from the power supply pads. To avoid such and similar problems, the threshold potential is generated in each pixel separately [9, 11]. This is done in two steps. First, a circuit is designed that generates the "replica" potential of the base line potential. Second, an another circuit is implemented that adds a threshold voltage to the "replica" potential. Let us describe these circuits. The "replica" generator is shown in fig. 2.47. The transistors Tf and Tf' are layouted identically, as well as the transistors Td and Td'.

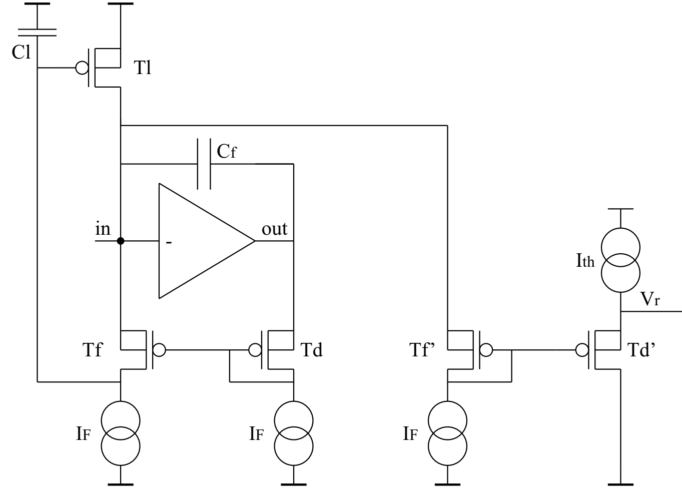


Figure 2.47: Generator of the base line potential.

If we neglect the drain-source resistances of these transistor and assume $I_{th} = I_F$, we have $V_{gsf} = V_{gsf'}$ and $V_{gsd} = V_{gsd'}$. This leads to $V_r = V_{out}$; r is the output of the replica generator. It is important to note that the replica potential V_r does not change when a sensor signal arrives. The action of the feedback keeps the potential of the input node *in* and therefore the potential V_r at a nearly constant level.

As we mentioned, the second step in the threshold potential generation is to add a threshold voltage to the replica of the base line potential. Let us assume that the current source I_{th} (fig. 2.47) generates the current $N \times I_F$, where N is some number that defines the threshold. The gate-source voltage of a MOS transistor in weak inversion, which operates in saturation, is given by the formula [1][32]

$$|V_{gs}| = |V_{th}| + nU_T \ln\left(\frac{LI_{ds}}{WI_{sub}}\right) \ . \tag{2.26}$$

V_{th} is the threshold voltage of MOS transistor (typically 0.5 V), U_T is the thermal voltage (26 mV at room temperature), n is the slope factor ($n \approx 1.5$), I_{sub} is a process dependent constant $I_{sub} = (n-1)\mu C'_{ox}U_T^2$, where μ is the mobility and C'_{ox} is the gate capacitance per unit area. The potential V_r is

$$\begin{aligned} V_r &= V_{in} - |V_{gsf'}| + |V_{gsd'}| = V_{out} + |V_{gsf}| - |V_{gsf'}| + |V_{gsd'}| - |V_{gsd}| = \\ &= V_{out} + nU_T \ln\left(\frac{I_{dsd'}}{I_{dsd}}\right) = V_{out} + nU_T \ln(N) \ . \end{aligned} \tag{2.27}$$

[32]Since the current I_F is very small (typically 5 nA), the transistors Tf, Tf', Td and Td' operate in weak inversion.

If we use V_r as the threshold potential, meaning if we connect the node r to the input of the comparator, the threshold has the value $V_{Thr} = V_r - V_{out} = nU_T \ln(N)$. The threshold V_{Thr} depends only on three parameters: the slope factor n, the thermal voltage U_T and the number N (the number that defines threshold). It does not depend on the supply voltage and on the current I_F.

The current source I_{th} is implemented in the following way. Every pixel contains a matrix with 32 identical PMOS current sources, each generating the current I_F, fig. 2.48. A current source is always connected to the node V_r. The other sources can be disconnected by using PMOS switches. The switches are arranged by using the binary weighted scheme, fig. 2.48. By applying the control logic signal G_{DAC0}, one current source is connected to

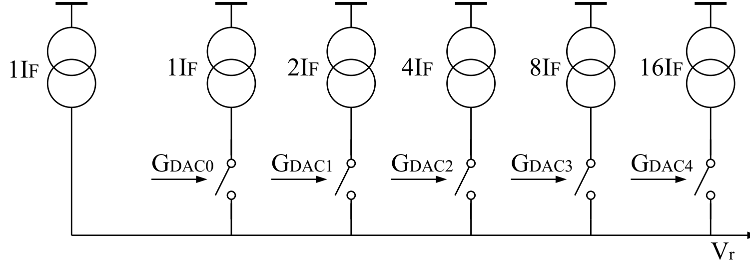


Figure 2.48: Digital-to-analog converter used to generate the threshold V_{Thr}.

the node r. By applying G_{DAC1} - two sources, G_{DAC2} - four and so on. Therefore by the use of the 5 logic signals G_{DAC0} - G_{DAC4}, any number of sources between $N = 1$ and $N = 32$ can be connected to the node r. The threshold value is given by the formula

$$V_{ThrUntuned} = nU_T \ln(N), \;\; N = \sum_{i=0}^{4} 2^i G_{DACi} \; . \tag{2.28}$$

G_{DACi} has either the value 1, meaning high Vdd potential, or the value 0 meaning low $DGnd$ potential. These logic signals are distributed in all pixels by using five metal lines. The values of these signals are stored in a register outside the pixel matrix.

Threshold Dispersion

According to eq. 2.28, we could expect that the threshold has the same value in every pixel. However, to derive this equation we assumed that the transistors, which are designed to be identical, match perfectly. This is not so in reality. The transistors having identical mask layouts slightly differ in their electrical properties when fabricated on the chip. There are a few reasons for this. The gate width and length deviate from the designed values. The gate-oxide growth and the treatment of the interface between the oxide and the semiconductor are not perfectly uniform over the chip area. Also, doping densities in channel areas of different transistors fluctuate. All the processing non-uniformities lead to the mismatch of identically designed transistors.

Let us rewrite the formulas for the drain-source current of a transistor which operates in saturation. In the case of strong inversion

$$I_{ds} = \frac{K}{2n}\left(\frac{W}{L}\right)(V_{gs} - V_{th})^2 \tag{2.29}$$

and weak inversion

$$I_{ds} = I_{sub} \left(\frac{W}{L}\right) \exp\left(\frac{V_{gs} - V_{th}}{nU_T}\right) \ , \tag{2.30}$$

where $K = \mu C'_{ox}$, $I_{sub} = (n-1)KU_T^2$. Mismatch between transistors is usually characterized by means of the following deviations: σ_L, σ_W, σ_K and $\sigma_{V_{th}}$. Assuming a rectangular gate geometry, it can be shown [59] that the terms σ_K/K and $\sigma_{V_{th}}$ are inversely proportional to the square of transistor gate area

$$\frac{\sigma_K}{K} = \frac{A_K}{\sqrt{WL}}; \quad \sigma_{V_{th}} = \frac{A_{V_{th}}}{\sqrt{WL}} \ . \tag{2.31}$$

Semiconductor producer usually provides the values of A_K, $A_{V_{th}}$, σ_L and σ_W, measured in the case of strong inversion.

Let us redraw the complete feedback scheme with the NMOS current sources as they are implemented on the chip, fig. 2.49. We derive the following formula for the variance

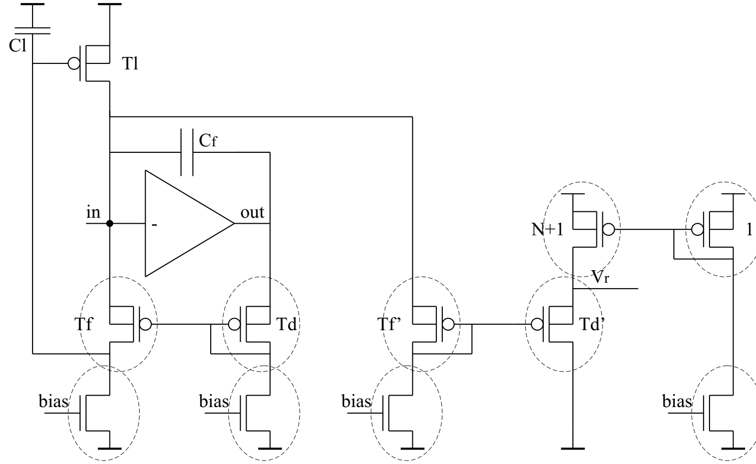


Figure 2.49: Scheme of the feedback circuit.

of the threshold dispersion

$$\sigma_{Th}^2 = \sum_{i=1}^{10} \left(\sigma_{V_{Thi}}^2 + (nU_T)^2 \left(\frac{\sigma_{I_{subi}}}{I_{sub}}\right)^2 + (nU_T)^2 \left(\frac{\sigma_{Wi/Li}}{Wi/Li}\right)^2 \right) \ . \tag{2.32}$$

This formula is derived assuming that all transistors operate in weak inversion. The ten transistors enclosed with the dashed circles (fig. 2.49) contribute to the total threshold mismatch. Because of this we have ten terms in the sum. We can expect that the four NMOS transistors introduce a particularly large mismatch. They have annular gates and the simple mismatch area dependence given by eq. 2.31 is not more valid. It can also be shown that the mismatch of annular transistors becomes worse when the inner diffusion contact is chosen to be drain [26, 28]. However, such a drain orientation has to be used if one wants to reduce the parasitic capacitance of the drain contact. We can, after all, conclude that the threshold dispersion is hard to calculate or simulate. Relatively large values can be expected due to many transistors whose contributions quadratically add, eq. 2.32.

Threshold Scan

In this subsection we will introduce the method used to measure the thresholds in individual pixels. On the front-end chip there is no possibility to measure directly the base line and threshold potentials. The threshold must be measured in indirect way. In every pixel there is a simple circuit that allows injection of a known amount of charge into the amplifier input node. This circuit is called "chopper". In its simplest implementation, chopper is a capacitor with its first terminal connected to the amplifier input, while a voltage pulse of a known amplitude can be applied to the second terminal. Since the potential at the amplifier input node is kept constant, the charge injected in the amplifier is given by the relation $Q_{inj} = C_{inj}\Delta v_{inj}$, C_{inj} is the capacitance of the injection capacitor, Δv_{inj} the pulse amplitude. The threshold is measured in the following way. We do a sequence of charge injections by changing the amplitude of the Δv_{inj} pulse. For every injected charge we measure the probability that the *Hit* signal is generated.[33] Finally, we obtain the probability versus the injected charge. If we had a system without noise, the result of the threshold scan would be a step function. The threshold would be precisely defined. Note, by means of the threshold scan, we obtain the threshold in units of the input charge Q_{Thr}. If we know the gain of the charge-sensitive amplifier (C_{gain}), we can calculate the threshold expressed in volts: $V_{Thr} \equiv Q_{Thr}/C_{gain}$. C_{gain} is, approximately, equal to the feedback capacitance C_f. In reality, noise is superimposed on the output signal. If we assume that the noise amplitude has a Gaussian distribution, the probability that the comparator generates a *Hit* signal is given by the following expression:

$$p_{hit} = \Phi\left(\frac{\Delta v_{out} - V_{Thr}}{\sigma_n}\right) . \tag{2.33}$$

Δv_{out} is the idealized (noiseless) signal amplitude at the output of the amplifier, given by $\Delta v_{out} \equiv \Delta v_{inj} C_{inj}/C_{gain}$; V_{Thr} is the threshold expressed in volts, σ_n is the square root of the voltage noise power at the output of the amplifier.[34] The function Φ is defined as

$$\Phi(x) \equiv \frac{1}{\sqrt{2}} \int_{-\infty}^{x} \exp\left(\frac{-t^2}{2}\right) dt = \frac{1}{2}\left(1 + Erf\left(\frac{x}{\sqrt{2}}\right)\right) .$$

To summarize, the threshold can be found by continuous increasing the chopper pulse amplitude and measuring the *Hit* signal probability. The measured data can be fitted using the function of eq. 2.33.

Threshold Tuning

Since we are able to precisely measure the thresholds in individual pixels, it would be also useful to be able to adjust the thresholds separately and in this way eliminate the threshold dispersion. The circuit that enables the local threshold tuning is shown in fig. 2.50. Two NMOS transistors T1 and T2 operate as source followers. Since biased

[33] The probability can be measured by performing a large number of injections and by dividing the number of received *Hit* signals by the number of injections. A *Hit* signal can be detected either by reading out the hit data transmitted by the chip or by using the hit-bus.

[34] The kTC noise of the chopper contributes to the value of σ_n too. The switching noise of the chopper corresponds to the equivalent input noise charge (ENC) of $\sqrt{kTC_{inj}}$, which is, typically, 36 e assuming $C_{inj} = 8\ fF$, or 80 e for $C_{inj} = 40\ fF$.

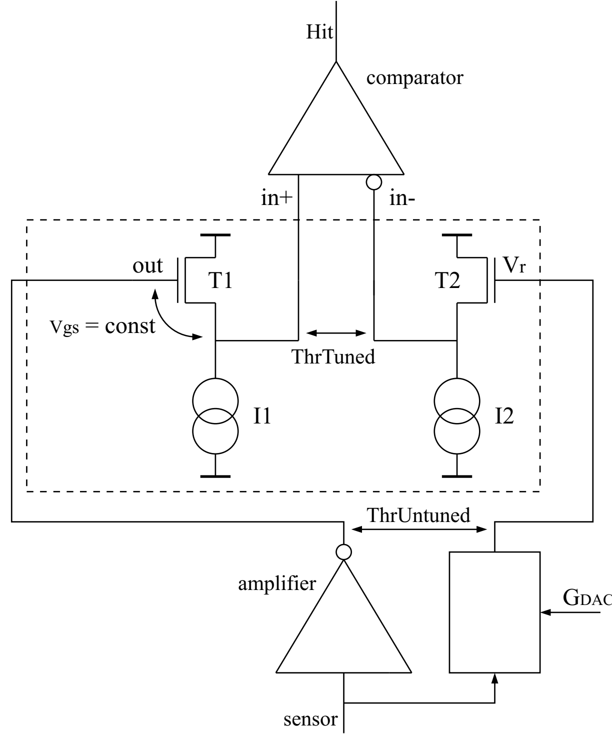


Figure 2.50: Circuit for local threshold tuning (enclosed with dashed line).

with constant currents $I1$ and $I2$, the gate-source voltages of the transistors are constant. The tuned threshold, seen at the input of the comparator is given by the following formula

$$\begin{aligned} V_{ThrTuned} &= V_{in-} - V_{in+} = V_r - |V_{gs2}| - V_{out} + |V_{gs1}| = \\ &= V_{ThrUntuned} + nU_T \ln\left(\frac{I_1}{I_2}\right) . \end{aligned} \tag{2.34}$$

We assumed that the NMOS transistors T1 and T2 operate in weak inversion. In the latest versions of the front-end chip (FE-I2 and FE-I3) the currents I_1 and I_2 are generated by using a special differential digital-to-analog converter (DA converter or shortly DAC) placed in every pixel [10]. The design of the DA converter will be introduced in the following subsection.

Differential DA-Converter for Local Threshold-Tuning

The differential DA-Converter consists of a core-matrix with 63 identical PMOS current sources. The current sources are grouped in seven groups consisting of 1, 2, 4, 8 and three times 16 sources, fig. 2.51. The drains of the transistors in a group are connected together. The current generated by a group of transistors can be directed either into the output terminal 1 or into the output terminal 2. The currents are steered by using the PMOS switch transistors which are controlled by the differential logic signals T_{DACi}, fig. 2.51. The four groups consisting of 1, 2, 4 and 8 sources are directly controlled by the logic signals T_{DAC1}, T_{DAC2}, T_{DAC3}, T_{DAC4} and their complements $T_{DAC1}B$, $T_{DAC2}B$, $T_{DAC3}B$, T_{DAC4B} respectively, fig. 2.51. The current generated by these sources, which flows into the terminal 1, is given by the equation $I_{out1} = I_{TrimTh} \sum_{i=1}^{4} T_{DACi} 2^{i-1}$. The binary-weighted

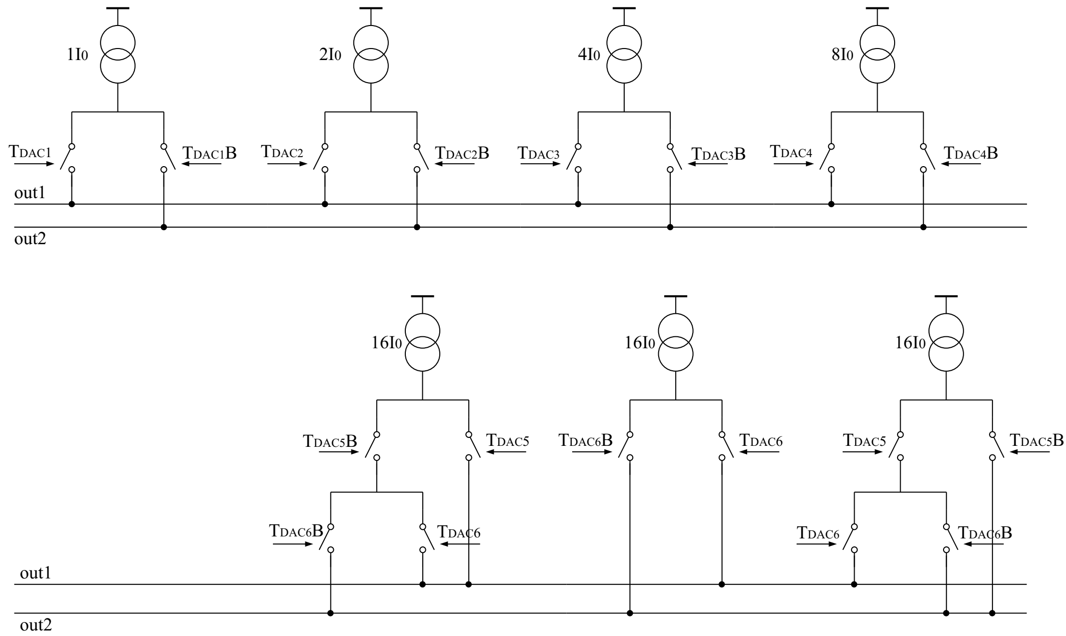


Figure 2.51: Differential DA converter.

switching scheme is not the only possibility to implement a current mode DA converter. The advantage of this scheme is its simplicity. We can directly connect the digital control lines to the corresponding switches. The disadvantage of the binary weighted scheme is the existence of the incremental transitions whereby a large number of sources are simultaneously switched off and on. For example, during the transition between $N = 7$ (binary - 0111) and $N = 8$ (binary - 1000), seven current sources are switched off, while the other eight switched on. Such transitions are sensitive to the mismatch between the transistors in the current source matrix and usually show the largest deviation from the ideal DA converter increment. In order to avoid the possible deviation from linearity, which could occur during the transition 31-32, the remaining 48 sources are not divided into two groups of 16 and 32 sources and then directly controlled by the signals T_{DAC5} and T_{DAC6}, which would be the simplest solution. They are divided into three groups each containing 16 sources, fig. 2.51. During the critical transition 31-32 only one group of 16 sources is additionally switched on and not a group of 32 sources, which would be the case when we had a binary weighted scheme. This is achieved by implementing simple logic functions with PMOS switch transistors. If the bits $(T_{DAC6}, T_{DAC5}, ..., T_{DAC2}, T_{DAC1})$ represent the binary coded number N, the current flowing out of the output terminal 1 is: $N \times I_0$. Since the DA converter is differential, meaning that the sum of the output currents is constant and equal to $63I_0$, the current flowing out of the terminal 2 is $(63 - N)I_0$. The current sources in the matrix are realized as PMOS transistors. Because of that, the currents generated by the DA convertor have the opposite direction than that needed in the scheme of fig. 2.50. The currents I_1 and I_2, fig. 2.50 are generated by the "mirroring" of the currents delivered by the DA converter. This is done by using the two current mirrors shown in fig. 2.52.

The differential DA converter has a few nice properties. If we substitute the current

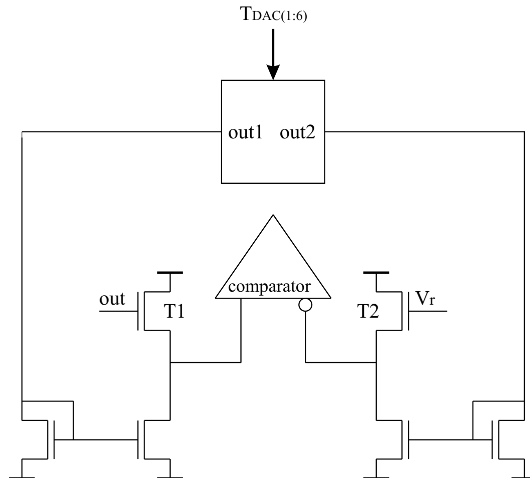


Figure 2.52: Threshold tune DAC with NMOS current mirrors.

values into eq. 2.34, we get the following formula for the tuned threshold

$$V_{ThrTuned} = V_{ThrUntuned} + nU_T \ln\left(\frac{N}{63-N}\right) . \qquad (2.35)$$

The tuned threshold does not depend on the value of the DAC current I_0. This is a fine feature since it eliminates the need for perfect matching between the DA convertors in different pixels. The threshold can be tuned in both directions. For $N > 63/2$, the tuned threshold is larger and for $N < 63/2$, smaller than the untuned. Fig. 2.53 shows the function $nU_TC_f \ln\left(\frac{x}{63-x}\right)$. As can be seen, the function is linear in a large mid-range, symmetrical around $N = 32$. The linearity of the tuning circuit is highly desirable since

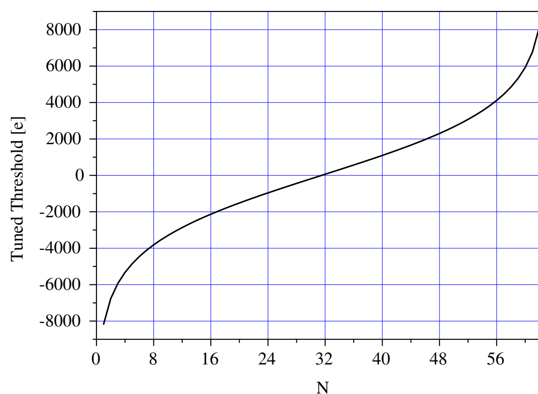


Figure 2.53: Tune DAC characteristic. The untuned threshold is assumed to be zero.

it allows the implementation of fast tuning algorithms. The design of the differential DAC is also explained in [10, 11]

Bit "For Free"

The linearity of the tuning characteristic gives us the following interesting possibility. Usually, if we want to increase the resolution of a high precision DA converter, which

consists of identical current sources, by adding a bit more, we need to double the number of unit cells.[35] In the case of the differential DAC introduced in the last subsection, we can decrease the DAC step size by factor of two by adding only one current source that is identical to the other 63, fig. 2.54. According to fig. 2.54, the current I_1 is equal to

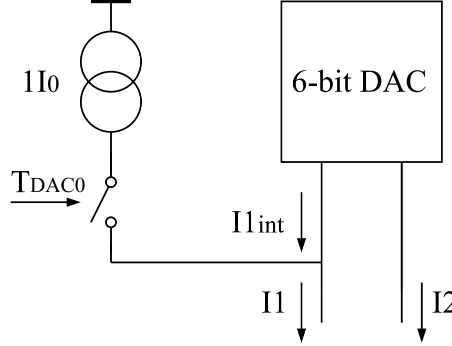


Figure 2.54: Additional 7th bit.

$I_{1int} + T_{DAC0} I_0$. If we substitute the formula for the current I_1 into eq. 2.34, we obtain

$$V_{ThrTuned} = V_{ThrUntuned} + nU_T \ln\left(\frac{N + T_{DAC0}}{63 - N}\right) . \tag{2.36}$$

In the region, where the function is linear, we can approximate eq. 2.36 with the first two terms of its Taylor series:

$$\begin{aligned} V_{ThrTuned} &\approx V_{ThrUntuned} + nU_T \frac{4}{63}(2N + T_{DAC0} - 63) \\ &= nU_T \frac{4}{63}\left(\sum_{i=0}^{6} T_{DACi} 2^i - 63\right) \end{aligned} \tag{2.37}$$

As we see, the 6-bit core DA convertor has the unit step of $nU_T \frac{4}{63} 2N$, whereas the extended converter has two times better resolution $nU_T \frac{4}{63} N$. We made in this way a 7-bit DA converter, which generates 128 discrete voltage levels, by using a matrix that contains only 64 identical current sources. We have the seventh bit practically "for free" [10]. The values of the seven tune bits $T_{DAC0} - T_{DAC6}$ are stored in a pixel register. The cells of this register are implemented as standard static random access memory (RAM) cells.

Scheme of the Analog Pixel-Block

In the previous sections we have introduced the most important analog circuits of a pixel cell: the charge-sensitive amplifier with its feedback, the global threshold DAC and the circuit for threshold tuning. In this section we will give the complete scheme of the pixel analog part. Let us briefly introduce the other smaller blocks that we did not discuss so far.

The scheme of the charge-sensitive amplifier (pre-amplifier) is shown in fig. 2.55 (see also fig. 2.40). The bias potentials denoted with V_P and V_L are generated by two 8-bit DA

[35] We can also add a cell that generates one half of the unit current. However, this introduces a systematic mismatch between the half-current source and the full-current sources. Only in the case when all cells are identical (and identically biased), a high linearity of the DAC characteristic can be achieved.

converters placed outside of the pixel matrix in the area that we refer to as the bottom of column region. These potentials are distributed in the whole matrix via metal lines. The corresponding electrical scheme is shown in fig. 2.56. The transistor Tp, placed in a pixel

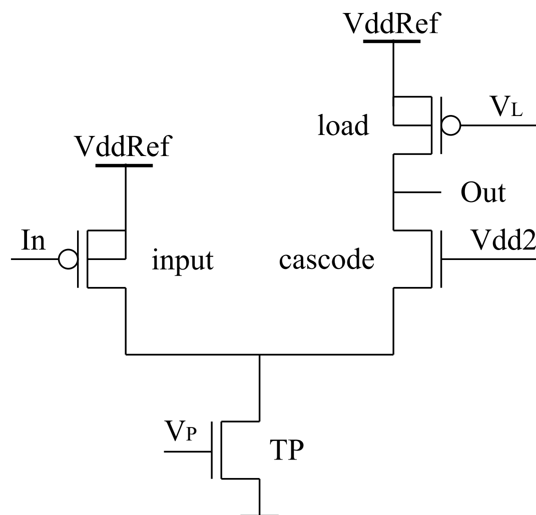


Figure 2.55: Charge sensitive amplifier.

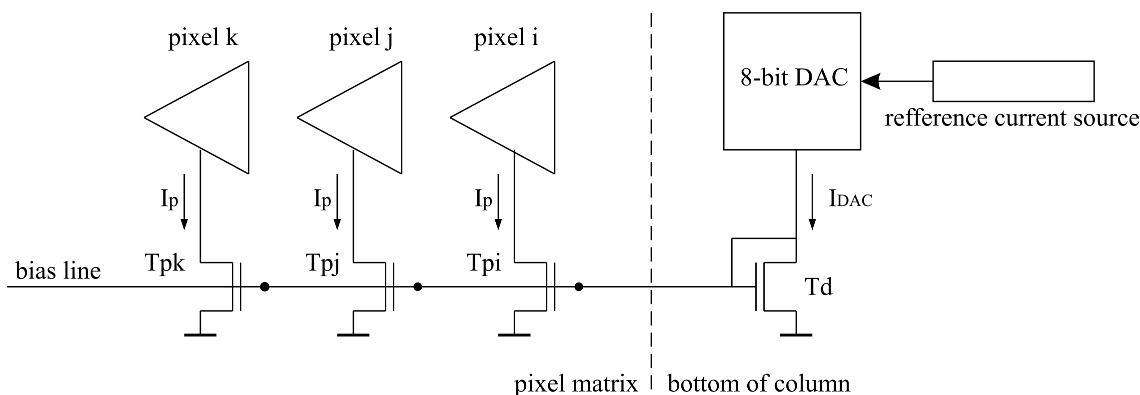


Figure 2.56: Biasing of pixel amplifiers.

amplifier and the diode connected transistor Tdio, which is biased by the output current of an 8-bit DAC, make together a current mirror. The current I_P in the pixel amplifier is therefore proportional to the current produced by the 8-bit DAC. The nominal value of the I_P current is $8\,\mu$A; for I_L 500 nA.

The cascode bias potential *Vdd2* is generated locally in every pixel.[36] It gives us the possibility to disable the pre-amplifier in every pixel. For this purpose a shutdown (called also $Kill$) bit is stored locally in the pixel register. If the *Shutdown* is active, the Vdd2 generator sets the $Vdd2$ potential to the ground level. The cascode transistor (fig. 2.55) acts then as an open connection and the amplification of the charge-sensitive amplifier becomes zero. The output potential of the amplifier becomes *VddRef* regardless of the input potential. The shutdown does not change the current consumption of the amplifier.

[36]The *Vdd2* generator is biased by a potential *Vvdd2*, also generated by an 8-bit DAC in the bottom of column region.

The complete scheme of the feedback/leakage compensation circuit is shown in fig. 2.57. This is almost the same scheme as that shown in fig. 2.49 except for two simple elements

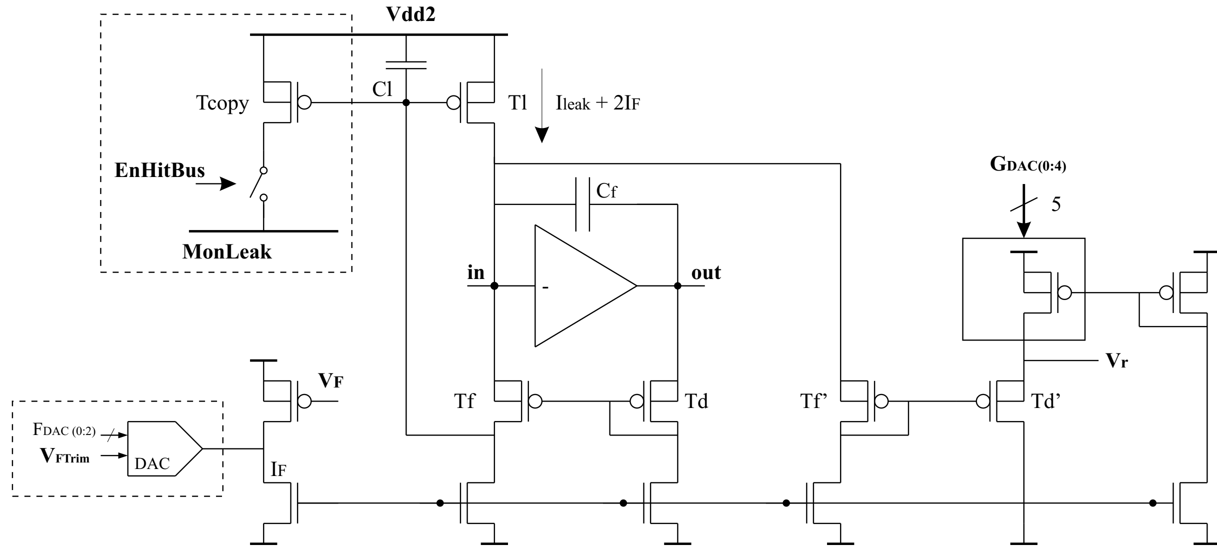


Figure 2.57: Full scheme of the feedback circuit for the charge sensitive amplifier. Input and output nets are denoted by using bold letters.

enclosed by dashed lines.

The feedback current I_F can be tuned locally in every pixel. It is achieved by using a 3-bit PMOS based current DAC. Three control bits F_{DAC0}, F_{DAC1} and F_{DAC2} are stored in the pixel register.

The transistor Tcopy allows the measurement of the sensor leakage current. Current flowing through the Tl is equal to $2I_F + I_{leak}$, where I_{leak} is the sensor leakage current. The transistors Tcopy and Tl make a current mirror. The transistor Tcopy can be connected to a global line referred to as $MonLeak$. The $EnHitBus$ bit must thereby be set to 1. This bit is also stored in the pixel register. The $MonLeak$ line is connected to an analog output pad, called $MonLeak$. This circuit allows multiplexing out the current that flows through the transistor Tcopy. It is also possible to digitize the MonLeak current on the chip by using a simple analog-to-digital converter.

The potential V_F is generated by an 8-bit DAC placed in the bottom of column region, the nominal value of the feedback current I_F is 5 nA.

The electrical scheme of the chopper is shown in fig. 2.58. We have already explained the principle of the circuit when we talked about the threshold scan. The right terminal of the injection capacitor is connected to the input of the amplifier (In). A voltage pulse of known amplitude is applied to the left terminal denoted with A, fig. 2.58. The value of injection capacitance can be changed by using the signal Hi. If $Hi = 1$ both capacitors are connected in parallel, therefore $C_{inj} = C_{lo} + C_{hi}$. When $Hi = 0$, $C_{inj} = C_{lo}$. The signal Hi is common for all pixels. Its value is stored in a register bit outside the pixel matrix. A bit $Select$, locally stored in every pixel, enables the chopper. The voltage pulse at the node A can be generated in two ways. The first way is to apply a pulse from outside through the $ExtInj$ line. The capacitors must thereby be connected to the $ExtInj$ line by setting the enable bit $ExtInjEn$ to the active level. The second way is to generate a voltage step of a known amplitude inside the chip. This approach is called

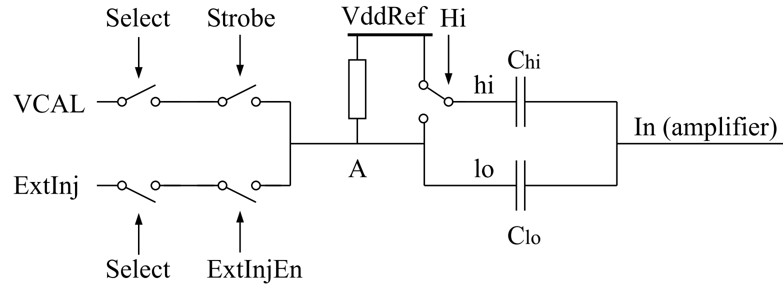


Figure 2.58: Chopper.

internal injection. A potential $VCAL$ is generated inside the chip by a 10-bit voltage DA converter placed in the bottom of column area. In order to do the internal injection, the node A must be disconnected from the line used for external injection, by setting $ExtInjEn$ to the low level. In the stationary state, the node A is "pre-charged" to the $VddRef$ potential by a high-ohmic resistor, fig. 2.58. When we apply the signal $Strobe$, the potential at the node A is pulled down to the $VCAL$ level. In this way, a charge amount $Q_{inj} = C_{inj}(VddRef - VCAL)$ is injected into the amplifier.

Fig. 2.59 shows the scheme of the comparator. The first stage of the comparator

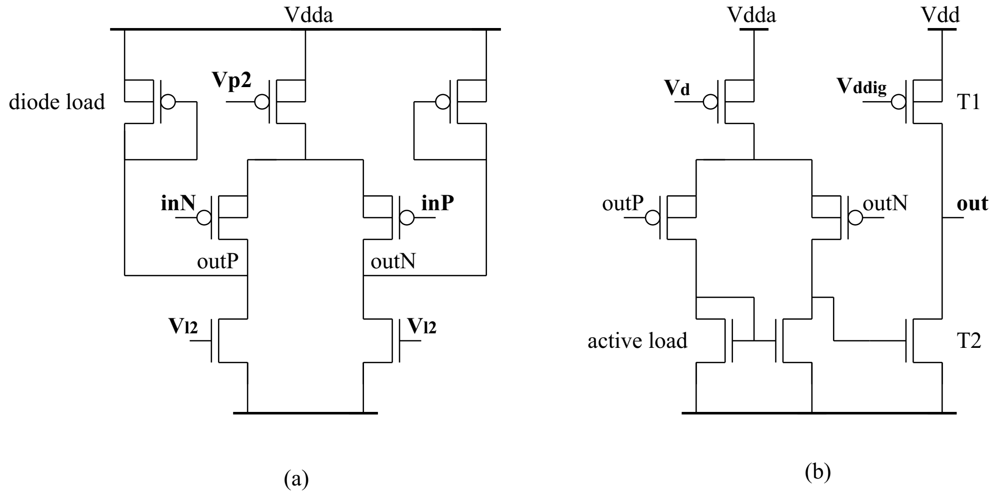


Figure 2.59: Comparator. (a) - the first stage ("second amplifier"). (b) - the second stage.

is a high-speed fully differential amplifier with two diodes as load. We refer to this amplification stage as "second amplifier". The second stage of the comparator is a classical two stage differential amplifier with a current mirror as active load, see also [6]. The differential design of these amplifiers leads to a good power supply rejection ratio. It would be probably possible to use the digital power line Vdd as power supply for the comparator. However, in order to save the power consumption of the chip, a lower analog supply potential $Vdda$ is used. $Vdda$ and $VddRef$ are connected together outside the chip. The nominal value for both potentials with respect to the analog ground potential Gnd is 1.6 V. The nominal value for the digital supply voltage is higher ($\sim$ 2.0 V) in order

to assure the required data processing speed. The comparator output signal is used by CMOS digital gates placed in the digital pixel block. Therefore the output potential of the comparator has to swing in the range between $DGnd$ and Vdd. Because of that, the source of the load transistor Tl (fig. 2.59 (b)) is referenced to Vdd and not to $Vdda$. Note that the CMOS signal produced by the comparator is active low. The signal is inverted to improve the speed and to save the DC power at the same time. The NMOS transistor T2 has a much higher current capacity than the PMOS load Tl. The important leading Hit signal edge is therefore fast by keeping the modest bias current of about $5\,\mu$A.

The Hit signal is derived from the output signal of the comparator. The corresponding logic circuit is shown in fig. 2.60. The multiplexer enables us to bypass the analog pixel

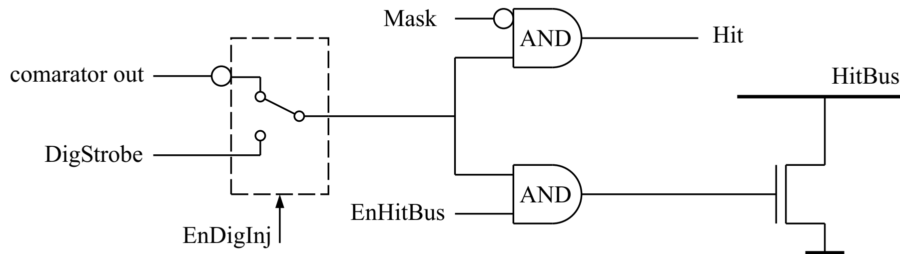


Figure 2.60: Hit logic.

block and test the digital readout blocks directly by generating the digital strobe pulse *DigStrobe*. To allow this operation, an enable signal called $EnDigInj$ must be active. The value of this signal is common for all pixels and stored in a register outside the pixel matrix. The $Mask$ bit is stored in the pixel register. This enables us to disconnect the analog part of a pixel from the digital readout system. The $HitBus$ is a test feature. The Hit signal can be sent through a line, called $HitBus$, to an output pin. In this way, we can receive the Hit signal directly. The $HitBus$ can be also used for the automatic generation of the L1-trigger signal, which is necessary when we want to test the sensor and the external trigger is not available. The $EnHitBus$ bit is stored as a bit in the pixel register.

Pixel Register and the SEU-Tolerant RAM Cell

Every pixel contains a register to keep the configuration bits. These bits are: $Shutdown$,[37] $EnHitbus$, *Select*, $Mask$,[38] T_{DAC0} - T_{DAC6}, F_{DAC0} - F_{DAC2}. The register cells are implemented as static random-access-memory (RAM) cells.

The simplest static RAM cell consists of two inverters, fig. 2.61. The output of one inverter is connected to the input of the other. Such connection leads to a positive feedback. The circuit has two stable states that correspond to the CMOS logic levels and an astable state between them, see fig. 2.61 (b). In order to write into RAM cell, one of the internal nodes has to be connected to a generator of the desired logic level. Provided that the generator can overcome the action of the feedback and bring the RAM cell from a stable operating point into the other, the new logic level will be written in the cell.

The hit of a high-energy particle in the vicinity of a RAM cell can cause the change of the logic state stored in the cell. This effect is called single event upset (SEU), see for

[37]Logic 0 has to be stored in order to shutdown the amplifier.

[38]Logic 0 has to be stored in order to mask the analog part.

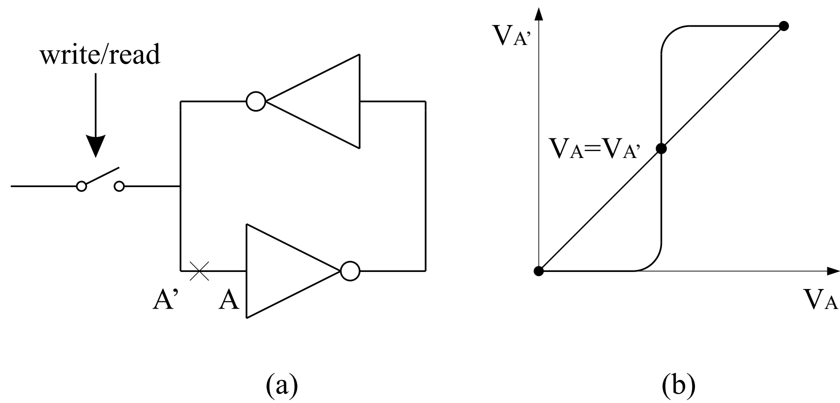


Figure 2.61: (a) - RAM cell. (b) - open loop transfer characteristic.

example [68]. The electrons, generated by the particle hit are attracted by the n^+ doped layers which are at a positive potential with respect to the bulk. Among them also the n^+ doped layers that are connected to the internal RAM node which is at logic 1 level. The electron current flowing into this node can be higher than the regenerative feedback of the RAM cell. The potential at the positive internal node then drops below the astable point and logic 0 is written. To decrease the probability of a SEU, a more complex RAM cell design is used for the pixel storage elements, fig. 2.62. The dual interlock cell (DICE)

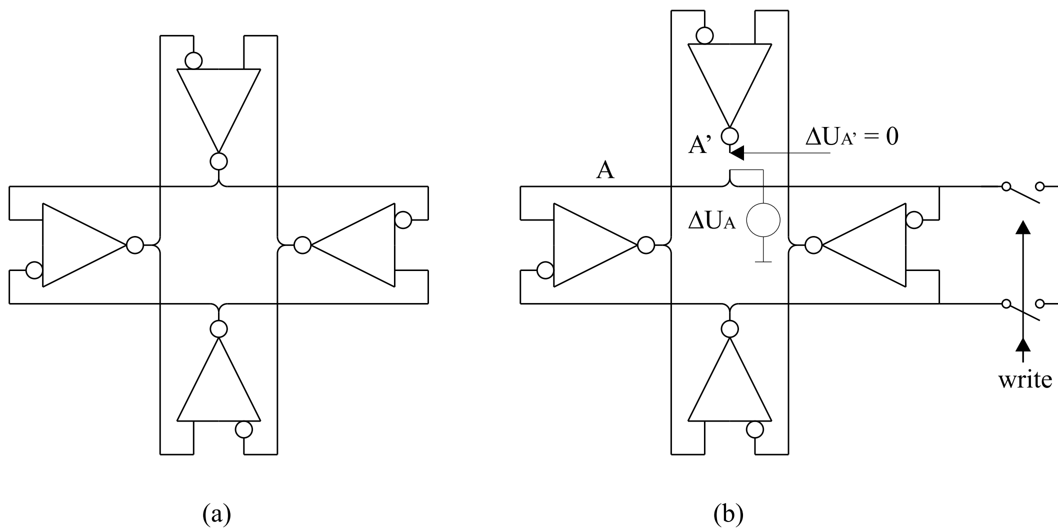


Figure 2.62: (a)- SEU tolerant DICE cell. (b) - test circuit.

[32] consists of two cross coupled standard RAM cells. One logic level is stored at two internal nodes. In order to change the bit stored in the cell, we must force both nodes that are at the same potential to the desired logic level. This is an interesting property of the complicate feedback scheme. It relies on the fact that a PMOS transistor in an invertor cell cannot overcome an NMOS if the latter is on. Generally, a RAM cell accepts a new logic level only if the RAM feedback responds to our attempt to write into the cell. If we only force a RAM node to the potential that we want to have, but there is no feedback response, the old logic level is restored after we disconnect the generator. Such behavior is illustrated in fig. 2.62. We broke the feedback path behind an invertor in

order to examine the feedback action. The generator ΔU_A simulates our attempt to write into the RAM or the potential change caused by a particle hit. It can be shown that the logic level change does not propagate from the node A to the node A'. To verify this, we can simply check the two possibilities $\Delta U_A = \pm Vdd$. In both cases the potential at the node A' does not change. There is no feedback loop and, as consequence of this, the logic level change applied at only one internal node cannot be saved in the DICE RAM cell. The charge generated by a particle hit can cause an upset only if two internal nodes are affected. If these two nodes are far enough from each other in the integrated circuit, the probability of a simultaneous charge injection is small and the SEU rate negligible. This has been verified by the irradiation of front-end chips.

Fig. 2.63 shows the digital circuit that enables writing into the pixel register. The

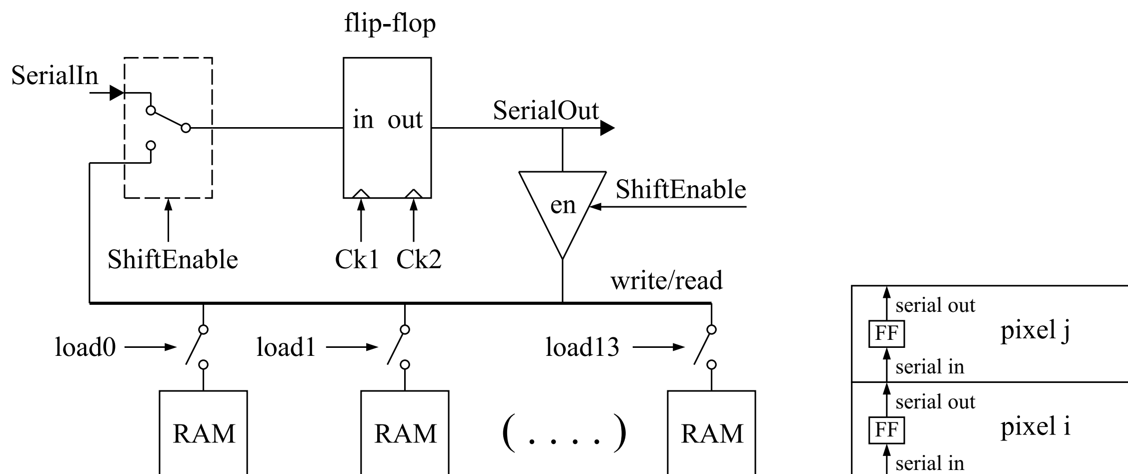


Figure 2.63: Pixel shift register.

circuit contains a memory element, called flip-flop. By applying the $CK1$ and $CK2$ signals, shifted in phase, the logic level at the serial input $SerIn$ is stored in the flip-flop.[39] The flip-flops of neighboring pixels in a column are connected in the form of daisy

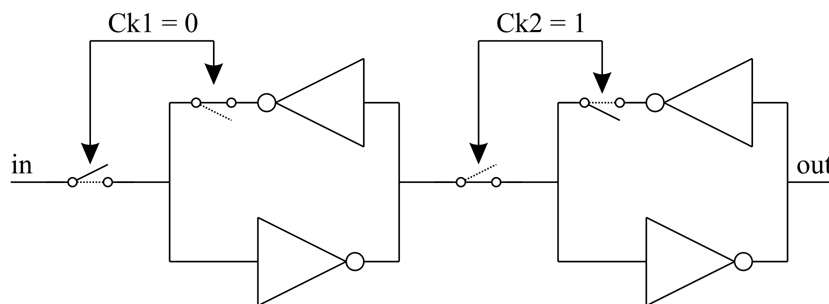


Figure 2.64: Flip-flop.

chain. Such structure is called shift register. After issuing the clock sequence, the bit

[39]The internal structure of a flip-flop is shown in fig. 2.64. A flip-flop usually consists of two RAM cells. The logic level at the node In is written into first cell by applying the $CK1$ pulse. By applying $CK2$ the logic level stored in the first cell is written into the second cell. The positive feedback of the RAM cell is broken during writing, so the write procedure does not require overcoming an inverter.

pattern stored in the shift register is shifted a place forward. For example, if we want to store logic 1 in the flip-flop of the Nth pixel, we must set the input of the shift register in the first pixel to 1 and then issue N clock cycles. The shift enable signal *ShiftEnable* must the be set to 1 during the shift sequence. Once we stored the desired bit in the flip-flop, we can connect a RAM cell to the write/read line and the cell latches the logic level. When we disconnect the RAM cell from the write/read line, the digital information remains stored in the cell. There is also the possibility to change the write direction and to write the bit stored in a RAM cell into the flip-flop. In order to do this *ShiftEnable* has to be set to 0. When we connect a RAM cell to the write/read line, the cell biases the line to the logic level stored. After that, we need only to latch the digital information into a flip-flop by applying a clock sequence.

Fig. 2.65 shows the scheme of the analog pixel block.

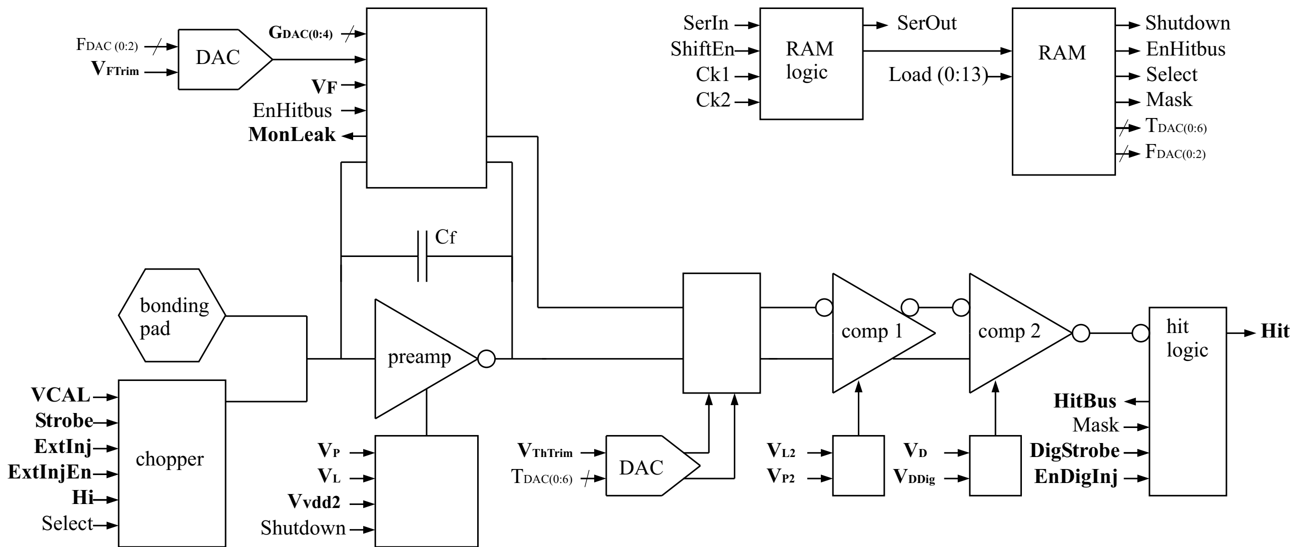


Figure 2.65: Analog pixel block.

Active Bias-Compensation

In this section we will describe a circuit that is designed to eliminate the influence of the supply voltage drop. The circuit is described in [10, 11].

We have already mentioned voltage drops when we talked about the threshold generation. The influence of a voltage drop on the threshold is eliminated by the proper design of the threshold generator. However, not only the threshold can be affected by voltage drops on supply lines. Fig. 2.56 shows the network used to bias the pixel amplifiers. If the source potentials of the transistors Tdio and Tp are not the same, the drain-source currents of these transistors do not match. The bias currents in the pixel amplifiers deviate from their nominal values. The bias transistors with large W/L ratio and therefore large transconductance, are especially affected by a voltage drop. Whereas PMOS transistors can always be designed to be long and have a small transconductance, NMOS transistors, due to their annular geometry, have always quite a large W/L ratio. Note, in the pixel matrix, the sources of NMOS transistors are separated from the bulk contacts. The special analog ground (*AGnd*) line is used to bias the sources of analog NMOS transistors. This line is connected to the bulk ground (*Gnd*) line at the bottom of

the chip. The bias currents generated by NMOS transistors can be substantially smaller in the matrix area far away from the supply pads, where the *AGnd* potential is higher than the *AGnd* potential in the vicinity of the 8-bit DACs.[40]

A simple way to eliminate the influence of the *AGnd* voltage drop would be to provide every bias voltage as the V_{gs} voltage of a diode connected PMOS transistor. When necessary the current direction can be changed by using an NMOS current mirror, fig. 2.66. This solution has the disadvantage that it substantially increases the current consumption

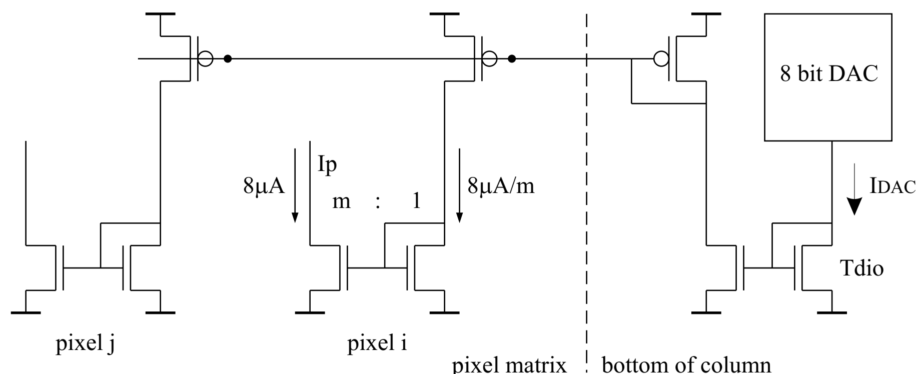


Figure 2.66: PMOS biasing scheme.

of the chip. The reason for this is the following. The most important NMOS bias currents are the I_P bias current for the charge-sensitive amplifier and the I_{L2} bias current for the first stage of the comparator.[41] The NMOS transistors, used to generate these currents, have to be as small as possible in order to minimize the parasitic capacitances. The current multiplication factor m (fig. 2.66) cannot be large - let us therefore assume $m = 1$. Since the nominal values are $I_p = 8\,\mu\text{A}$ and $I_{l2} = 3\,\mu\text{A}$, the simple circuit of fig. 2.66 would consume $11\,\mu\text{A}$. This would yield to $17\,\mu\text{W}$ per pixel, which is about 40% of the maximal power consumption allowed - a lot of power.

In order overcome the disadvantages of the scheme shown in fig. 2.66, an active biasing circuit is designed and implemented on the front-end chip [10, 11]. The block scheme of the circuit is shown in fig. 2.67. An identical circuit is used to provide the I_{L2} current. The pixels in even rows contain the I_{L2} and the pixels in odd rows the I_P compensation circuit. The heart of the circuit is a differential amplifier. The voltage between the output terminals of the amplifier Δ_{out} is nearly equal[42] to the voltage between its input potentials

[40]Let us try to estimate the voltage drop and the bias current change. The typical current consumption of an analog part in a pixel is $25\,\mu\text{A}$. The length of a pixel column is $160 \cdot 50\,\mu\text{m} = 8000\,\mu\text{m}$. The *AGnd* line has particularly little metal, its width is about $100\,\mu\text{m}$. The sheet resistance a metal layer is about $0.1\ \Omega/sq$. It yields to the total *AGnd* line resistance of $R \approx 8\ \Omega$. The voltage drop at the top of the column is: $\Delta V = RI_{mean} = 8\ \Omega \cdot 160 \cdot 25\,\mu\text{A}/2 \approx 16\ mV$. If we assume that an bias NMOS transistor operates in weak inversion, the NMOS bias current is given by the equation $I \propto I_{sub} \exp{(V_{gs}/39\ mV)}$. Using the last formula, we calculate that the NMOS bias currents at the top of a column have only about 66% of their nominal values.

[41]There is an additional NMOS bias current used by the Vdd2 generator, however, a perfect matching of this current with its nominal value is not so important.

[42]The amplifier compensates bulk effect. The voltage between the output terminals Δ_{out} is given by the equation $\Delta_{out} = \Delta_{in} + n(AGnd_{local} - Gnd_{local})$, n is the slope factor. The bulk effect can occur since the source contacts of the analog NMOS transistors are connected to the separate *AGnd* net and not directly to the silicon bulk *Gnd*.

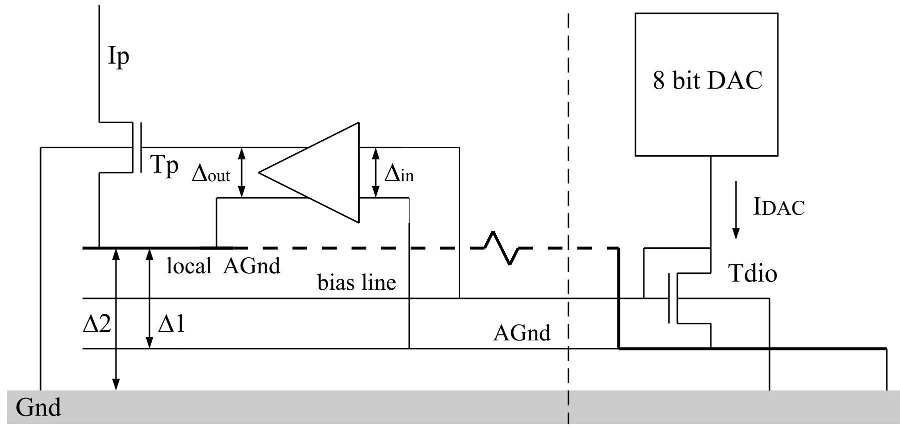


Figure 2.67: Active voltage drop compensation.

Δ_{in}. Two lines are distributed into the pixel matrix. One line is connected to the gate of the transistor Tdio (bias line). The other line is connected to the analog ground *AGnd* in the vicinity of the transistor Tdio. The voltage between these two lines defines the V_{gs} voltage needed for the generation of the current I_P or I_{L2}. The differential amplifier does the correction of the gate-source voltage by taking into account the local analog ground level (*AGnd*) and the local voltage between the *AGnd* and the bulk.

The bias compensation can be switched-off by setting a signal common for all pixels. An I_P and I_{L2} monitoring scheme is implemented. The monitor circuit is similar to that used for the monitoring of the leakage current.

Shielding

In order to decrease the pick-up of the digital activity that propagates through the bulk, the analog and digital blocks in a pixel are separated as much as possible. The floor-plan of a pixel is shown in fig. 2.68. The structural unit of a column-pair contains two pixels,

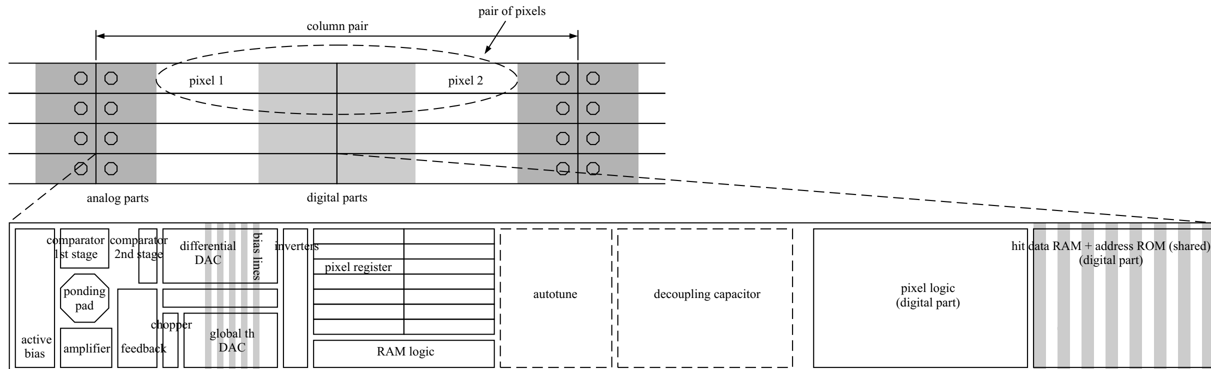


Figure 2.68: Floor-plan of a pixel.

see fig. 2.68. These pixels are mirrored in the way that its digital blocks are close to each other. The digital blocks share the same read only memory (ROM) used to store a pixel address. The pixels of neighboring column pairs touch with their analog sides. In this way, the separation between the digital pixel parts and the sensitive analog amplifiers is maximized.

Another possible cross talk path could be the capacitive coupling of digital activity across the 20 μm narrow gap to the sensor electrodes, fig. 2.69. Potentially dangerous are the long metal lines used to distribute digital signals into digital pixel parts. Some of these lines are permanently active. In order to avoid the propagation of the digital noise to the sensor, two metal layers on the chip are used as a shield between the sensor and the digital circuits, fig. 2.69. The upper shield layer, close to the sensor surface, has to be quiet. Therefore it is connected to the supply line $VddRef$. The lower shield layer is used to distribute the analog supply potentials $Vdda$ and $AGnd$. The shield geometry is quite complex. The used technology has 6 metal layers, however, the metal coverage must be between 30% and 70% of the total chip area. The cross section of a pixel along its longer side is shown in fig. 2.69. The digital lines, which are permanently active, are buried in the the lower metal layers. We can distinguish between three pixel regions.

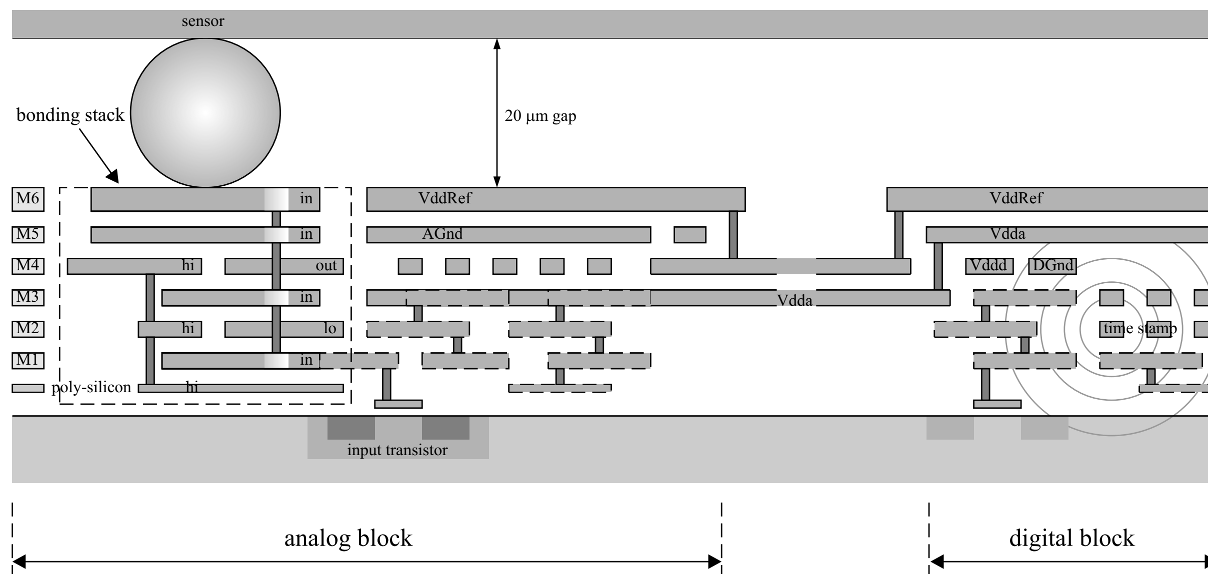


Figure 2.69: Cross section of a pixel. Shield and bonding stack.

The first region accommodates the analog pixel block, fig. 2.69, left. These circuits are connected to the analog supply lines: $VddRef$, $Vdda$ and $AGnd$. In fig. 2.69 we can also see the bonding stack. The bonding pad, implemented in metal 6, is connected to the input of the charge-sensitive amplifier (in). The feedback capacitor, as well as two injection capacitors are implemented inside the stack. The digital pixel block is placed in the second region, fig. 2.69, right. The digital block is supplied by the potentials Vdd and $Dgnd$. In the third area, between analog and digital block, the shied is implemented by using the third and fourth metal layer. The pixel register with the write/read circuit and the Hit logic are placed in this region. These blocks are connected to the digital supply lines. The remaining area in the middle of the pixel is used for two interesting circuits. We will describe these circuits in the next section.

Decoupling

In the design process of the front-end chip a lot of attention has been paid to separate the digital and analog circuits and in this way eliminate the cross talk between them. The other approach used in mixed-mode ICs is to reduce the digital noise at its source

i.e. to make the digital circuits as quiet as possible. This can be done by a careful design of digital gates, for example, by reducing node capacitances or by using differential signals. The amplitude of the voltage pulses generated during logic transitions at supply lines can be also reduced by using on-chip decoupling capacitors. The influence of a decoupling capacitor is illustrated in fig. 2.70. Assume that a charge ΔQ is needed for a

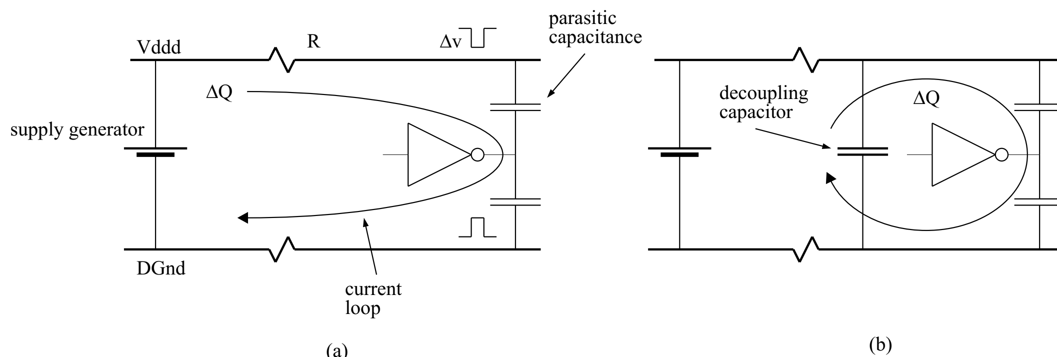


Figure 2.70: Current loop caused by change of logic level. (a) - without decoupling capacitor. (b) - with decoupling capacitor.

logic transition. Suppose that the logic transition occurs is a time Δt. Assume also that the charge ΔQ has to be provided by the distant voltage source as shown in fig. 2.70, (a). The amplitude of the voltage pulse seen at the supply line is then $\Delta v = R\Delta Q/\Delta t$; R is the resistance of the supply line. The typical transition time in the case of the logic gates used on the front-end chip is $\Delta t = 100\ ps$. The charge ΔQ is, typically, $\Delta Q = Vdd \cdot C_{load} \approx 2\ V \cdot 10\ fF = 20\ fC$; C_{load} is the total parasitic capacitance at the output of the logic gate. The resistance of the digital supply line is, typically, 100 Ω. Therefore we obtain $\Delta v \approx 20\ mV$. Let us now assume that we have a large decoupling capacitor between the digital supply lines plced close to the logic gate, fig. 2.70 (b). The charge ΔQ necessary for the logic transition is in this case delivered by the capacitor. The generator only needs to slowly refill the decoupling capacitor. The voltage pulse seen at the supply line is now given by the formula: $\Delta v = Vdd \cdot C_{load}/C_{dec}$; C_{dec} is the capacitance of the decoupling capacitor. The ratio between the capacitances C_{dec} and C_{load} can be large, typically, 1000. The voltage pulse at the power lines is much smaller than in the case without decoupling.

We should keep in mind that there is always a certain parasitic capacitance between power lines. For example, large n-wells connected to the positive supply Vdd and the n^+ regions connected to $DGnd$ form parasitic capacitors. Such parasitic capacitors can filter the digital noise in the way similar to that explained when we assumed an ideal decoupling capacitor. To improve the capacitive decoupling we implemented large Vdd - $DGnd$ capacitors. The capacitors are placed in the spare area near the pixel registers, see fig. 2.68. The capacitors are implemented as large poly silicon to n-well structures, as shown in fig. 2.71. The area of the capacitor is, approximately, $400\,\mu\text{m}^2$. The oxide thickness is only 6 nm. This leads to the capacitance of about 1.8 pF. Such a large ratio between the oxide area and the oxide thickness can be potentially dangerous. Damage in the oxide layer could produce a short connection between the Vdd and $DGnd$ lines and lead to chip failure. Because of that, every capacitor has a control circuit which can detect oxide damage and disable the capacitor if necessary. The scheme of the circuit is shown

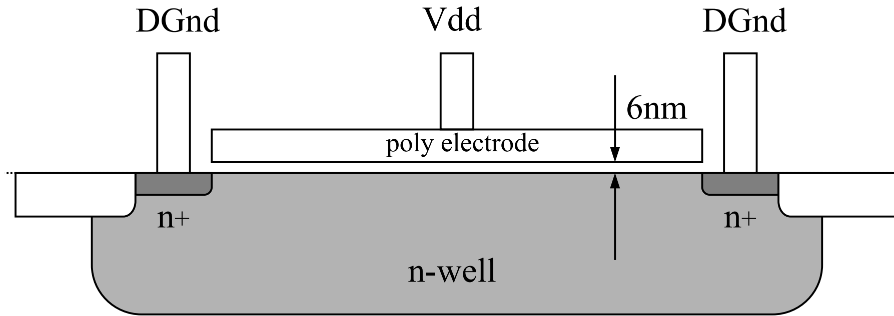


Figure 2.71: Poly-silicon to n-well decoupling capacitor.

in fig. 2.72. During normal operation, the switch S is closed and the decoupling capacitor

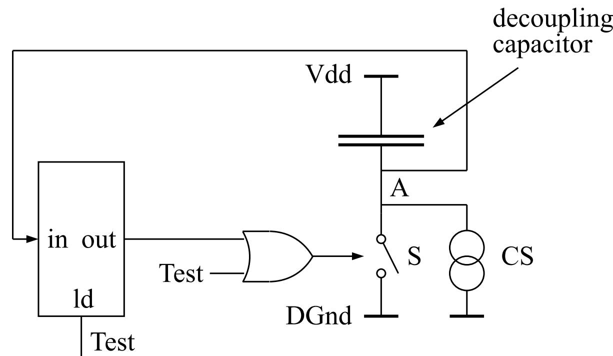


Figure 2.72: Protecting circuit.

connected between Vdd and $DGnd$. By setting the $Test$ signal to logic 1, the switch S becomes open. The current generated by the current source CS is small, typically, 10 nA. If the capacitor is undamaged, this small current is enough large to pull the potential of the node A down to the ground level. If there is a short connection between the capacitor electrodes, the potential at A increases to, approximately, Vdd. By returning the $Test$ signal to 0, the logic level at A is latched in the latch circuit. If the capacitor is damaged, logic 1 is latched and the switch S remains open. If there is no oxide damage logic 0 is stored and the capacitor is connected to the supply lines.

Auto-Tune Circuit

The second circuit placed in the area between the analog and digital part is the auto-tune circuit. This is a circuit which can automatize the fine tuning of the pixel thresholds in order to achieve a lower threshold dispersion. The idea belongs also to Laurent Blanqaurt, [11].

In the previous sections we introduced the threshold scan and the tune DAC. The straightforward way to perform the threshold tuning would be the following. We perform the threshold scan for each pixel, fit the data and calculate the thresholds. After that we calculate for every threshold the deviation from the desired value and the tune DAC setting that eliminates this deviation. Such a procedure is quite long, since it requires the threshold scan of all pixels and the software analysis of the data.

We can follow a different approach. By using the tune DAC, we set the threshold of a pixel at a high value. By using the chopper, we inject an amount of charge Q_{Thr0} which corresponds to the desired threshold $Q_{Thr0} \equiv C_{gain}V_{Thr0}$; C_{gain} is the gain of the charge-sensitive amplifier. If the pixel does not generate the Hit signal, we decrease tune DAC setting and inject the charge again. This procedure should be repeated until the first Hit signal is received. The last tune DAC value is a good approximation of the desired threshold value. The accuracy of the method is limited by the noise. If a Hit signal is generated, it does not necessarily mean that the pixel threshold $Q_{Thr} \equiv C_{gain}V_{Thr}$ is below the desired one. The probability of the Hit generation is:

$$p_{hit} = \Phi\left(\frac{Q_{Thr0} - Q_{Thr}}{ENC}\right) . \tag{2.38}$$

Q_{Thr0} is the injected charge that corresponds to the desired threshold. Q_{Thr} is the actual threshold expressed as charge. ENC (equivalent noise charge) and the RMS of the voltage noise at the output of the amplifier are related as: $ENC = C_{gain}\sqrt{\langle v_{nout}^2\rangle}$. The function Φ is defined as:

$$\Phi(x) \equiv \frac{1}{\sqrt{2}} \int_{-\infty}^{x} \exp\left(\frac{-t^2}{2}\right) dt = \frac{1}{2}\left(1 + Erf\left(\frac{x}{\sqrt{2}}\right)\right) .$$

Fig. 2.73 shows the Hit signal probability (eq. 2.38) assuming an ENC of 300 e. The

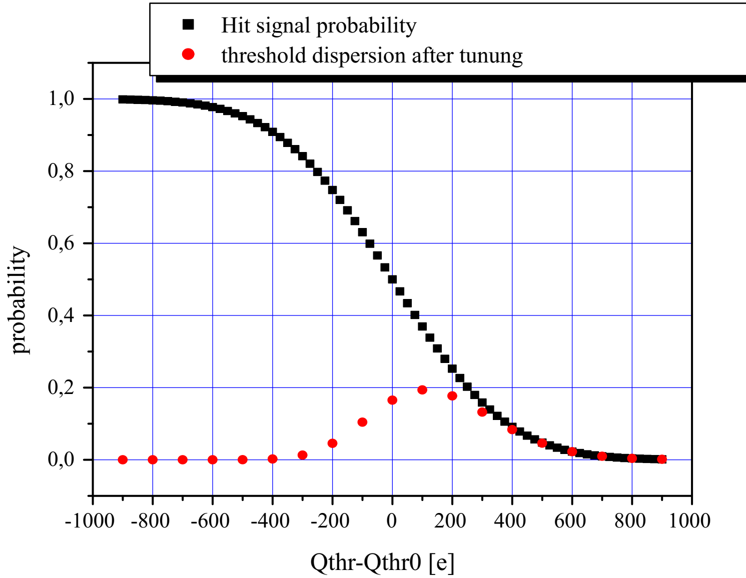


Figure 2.73: Hit signal probability, Q_{Thr} is the threshold expressed as charge, Q_{Thr0} is the injected charge. Threshold dispersion after tuning can be calculated starting from the Hit signal probability.

estimated threshold dispersion in the pixel matrix after the tune procedure is also shown. In order to calculate the dispersion of the tuned threshold, we assumed that the tuning procedure starts at $Q_{Thr} - Q_{Thr0} = 3ENC = 900\ e$. After that the threshold is decreased in steps of 100 e. The probability that the threshold value is frozen in the i-th step of the tune procedure is equal to the probability that the Hit signal was not generated in

all previous steps multiplied by the probability that the Hit signal is generated in the i-th step. As we can see, the standard deviation of the threshold dispersion after tuning has the value similar to the equivalent noise charge ENC. The mean value of the tuned threshold is higher than the desired value Q_{Thr0} since the tune procedure is assumed to start from higher threshold values.

Let us now discuss a similar tuning procedure. We start the tuning of the pixel threshold by setting a high threshold value. By using the chopper we inject now N times the same charge Q_{Thr0}. If the number of generated Hit signals does not exceed $N/2$, we decrease the threshold and do the injections again. When we receive more than $N/2$ hits, the tune procedure is finished and the threshold value frozen. The advantage of the repeated injections and the Hit signal counting is that the probability $n \geq N/2$ (n is the number of Hit signals after N injections) has much steeper transition between zero and one than the single Hit signal probability. Fig. 2.74 shows the probability functions.[43] According to fig. 2.74, we can expect a much lower threshold dispersion in the case of

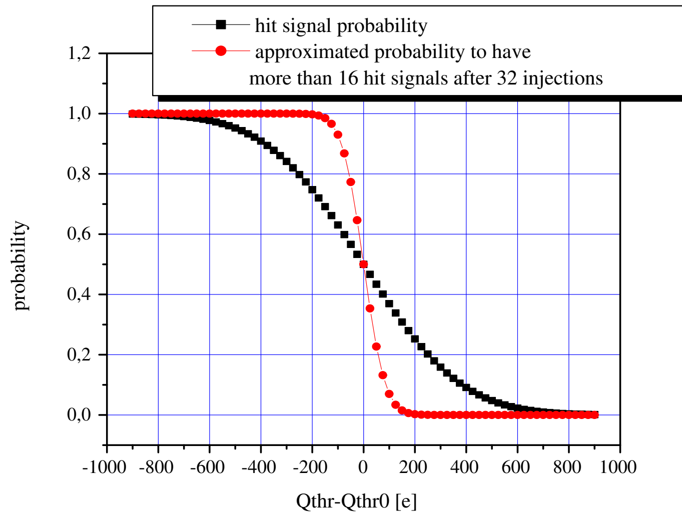


Figure 2.74: Hit signal probability and the probability to receive more than 16 Hit signals after 32 injections. Q_{Thr} is the threshold expressed as charge, Q_{Thr0} is the injected charge.

tune procedure with repeated injections than in the case when we inject only once.

The circuit implemented in pixel is based on repeated injections and Hit signal counting. The used counter can count in the both directions. If the Hit signal is generated, the counter is incremented. If the comparator does not respond to an injection, the counter is decremented. The counter can count up until it reaches the maximal state. Further increment takes no action. Similarly, the decrement of the minimal counter state does not lead to the change of its state. In the front-end pixel, there is enough space for a 5-bit counter (32 states). The idea is to do much more injections than the counter has states. If the pixel threshold Q_{Thr} has a higher value than the injected charge Q_{Thr0}, we

[43]The probability $n \geq N/2$ can be estimated analytically. The final counter state n has a binomial distribution with the mean value $p_{hit}N$ and the variance $Np_{hit}(1-p_{hit})$; N is the number of injections. The probability $n > N/2$ can be calculated as a function of p_{hit}. Here, we approximated the binomial distribution with the Gaussian function. p_{hit} is given by eq. 2.38. N is chosen to be 32.

can, with great probability, expect that the final counter state exceeds the middle state $n \geq 16$. On the other hand, if the threshold is lower we, most probably, obtain $n < 16$. To calculate the probability $n \geq 16$, we make the following assumption. The number of injections is so high that the probability of each particular end-state does not depend on the number of injections *i.e.* the probability does not change when we do an injection more. The probabilities reach their asymptotic values. Let us express it mathematically. As first, we write the probabilities after $N+1$ injections as functions of the probabilities after N injections

$$\begin{aligned} p_i(N+1) &= p_{i+1}(N)(1-p_{hit}) + p_{i-1}(N)p_{hit} \\ p_{31}(N+1) &= p_{31}(N)p_{hit} + p_{30}(N)p_{hit} \\ p_0(N+1) &= p_0(N)(1-p_{hit}) + p_1(N)(1-p_{hit}) \; . \end{aligned} \quad (2.39)$$

$p_i(N)$ is the probability that the counter hast the state i after N injections. 0 and 31 are the minimal and the maximal counter states respectively. As second we have the condition

$$p_i(N+1) = p_i(N) \; . \quad (2.40)$$

When we substitute eq. 2.40 in eq. 2.39, we obtain the following simple result

$$\begin{aligned} p_{31} &= \frac{1-\alpha}{1-\alpha^{32}} \\ p_{30} &= \alpha p_{31} \\ p_{29} &= \alpha^2 p_{31} \; \ldots \\ p_0 &= \alpha^{31} p_{31} \; . \end{aligned}$$

$\alpha \equiv (1-p_{hit})/p_{hit}$. The probability that the counter end state is higher or equal than 16 is then

$$p_{n \geq 16} = \sum_{i=16}^{31} p_i = \frac{1}{1 + \left(\frac{1-p_{hit}}{p_{hit}}\right)^{16}} \; . \quad (2.41)$$

The corresponding probability as a function of $Q_{Thr} - Q_{Thr0}$ can be calculated by substituting of p_{hit}, given by eq. 2.38, into eq. 2.41. The function is plotted in fig. 2.75, together with p_{hit} probability and the probability in the case of the incremental *Hit* counter. As we see the probability given by eq. 2.41 has a very steep transition between 0 and 1. The condition $n \geq 16$ can be easily checked by looking at the most significant bit of the 5-bit counter.

The implementation of the auto-tune circuit is shown in fig. 2.76. If the signal *TuneEnable* is active, all pixels are in tune modus. In the tune modus the logic circuit that controls writing into pixel register (fig. 2.63) works in a little bit different way, see fig. 2.76. The tune procedure is quite simple. First, we set the tune DACs high enough so that the thresholds in all pixels have, for sure, higher values than the desired value. After that, we repeat the following sequence. We do a large number of injections. We apply then the *LoadShutdown* signal. At the end, we send the bits which correspond to the tune DAC setting decreased by one in all pixel flip-flops and apply the Load Tune DAC signals. The whole sequence repeats until the minimal DAC setting is reached. As long as a pixel threshold Q_{Thr} has a value which is higher than the desired value Q_{Thr0}, the tune DAC setting is decremented in each cycle. When the threshold Q_{Thr} becomes

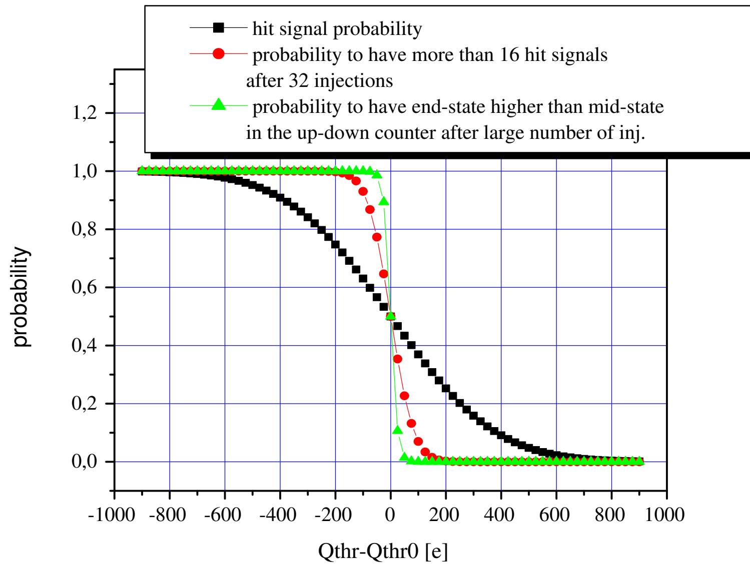


Figure 2.75: *Hit* signal probability and probabilities in the case of hit counting.

close to Q_{Thr0}, the probability to have the most significant bit M equal to 1 after injections increases steeply form almost zero to almost one. Provided the $M = 1$, by applying *LoadShutdown*, logic 1 will be stored into Shutdown RAM cell. The active *Shutdown* level blocks any subsequent writing into the RAM of this pixel. The tune DAC setting, which is currently in the RAM, is frozen. The next decrement has no effect in this pixel.

2.5.4 Readout System

The digital readout system of the front-end chip has been designed by Emanuelle Mandelli and by Gerrit Meddeler [7].

Digital Pixel-Block

The block scheme of the digital pixel block is shown in fig. 2.77. The digital pixel block contains the following important components. Two 8-bit RAM blocks used to store the leading and trailing edge times of the *Hit* signal. One 8-bit address ROM used to store the pixel address. Timing bus, referred to as *TSI*, used to distribute 8-bit Gray coded time into the pixels. The time is incremented every 25 ns. Data bus lines used to read out the time stamps stored in the RAM and the pixel address stored in the ROM.[44] At the input of the block there is a digital differentiator. The differentiator receives the *Hit* signal. Coincident with the *Hit* leading edge, a short pulse is generated at the first output of the differentiator (*LE* output). Coincident with the trailing *Hit* edge, a short pulse is generated at the second output (*TE* output). The *LE* and *TE* signals control writing into the *LE* and *TE* RAM blocks, fig. 2.77. In this way, the *TSI* time that coincides with the leading *Hit* edge is stored in the *LE* RAM block and the trailing edge time is stored into the *TE* RAM. Coincident with the trailing edge of the *Hit* signal, a RS flip-flop,

[44] The address ROM and the data bus lines are shared by two pixels in a column-pair.

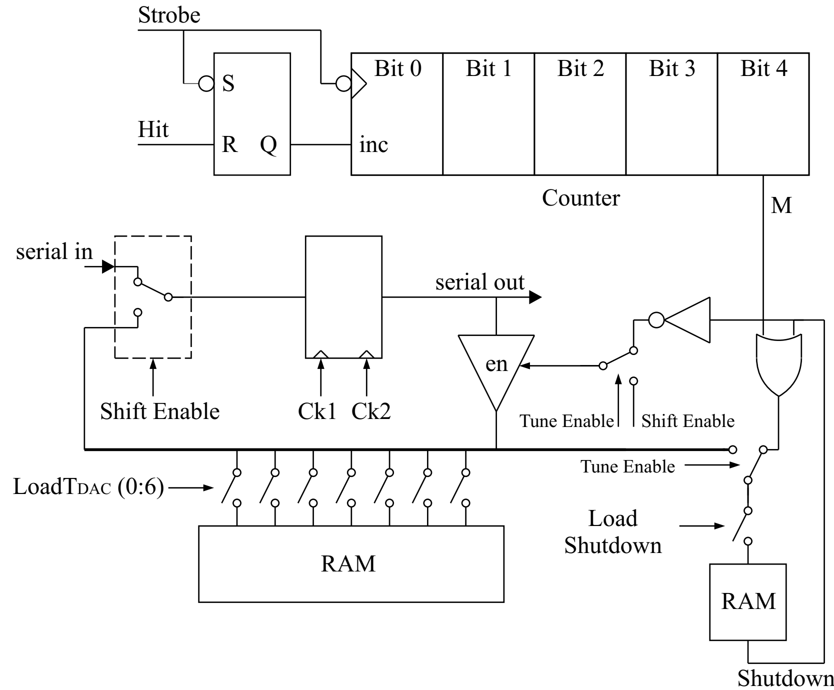


Figure 2.76: Autotune circuit.

called the hit flag, is set to 1. The output of the hit flag is connected to the chain of OR gates that extends through the whole pixel column, from its top to the bottom. The output signal of the OR chain (*ReadRequest*) is received by the column arbitration unit (CAU). In this way, the CAU receives the OR function of all hit flags in the column. When the *ReadRequest* signal becomes active, the CAU starts a readout sequence. Let us explain the sequence. The CAU generates the $Freeze$ and the *Read* signals. Although the *Read* signal is received by all pixels in the column, if there are many hits, only the pixel with the highest place in hierarchy responds to the *Read* signal. This pixel opens the RAM/ROM gates and puts the data at the bus lines. The CAU copies the information into its local register. The *Read* signal clears the hit flag. The circuit that encodes the column hierarchy is implemented by using two AND gates, fig. 2.77. When the new *Read* signal is issued, the next pixel in the hierarchy transmits its data. The $Freeze$ signal is needed to prevent a pixel that has just been hit by a particle to send its data through the bus, while some other pixel with lower hierarchy is communicating with the CAU.

Column Arbitration-Unit

The block scheme of CAU is shown in fig. 2.78. CAU does the readout of the hit data from a pixel column-pair. The CAU block contains the following sub-circuits. First, the control block. The task of this circuit is to communicate with the two pixel columns of a column-pair by generating the $Freeze$ and *Read* signals and to control the other sub-circuits in the CAU. The sense amplifier can reconstruct the correct logic levels by measuring the voltages at the data bus even if the voltages does not correspond to the full swing CMOS levels. The Gray to binary code converter and the subtractor are used to calculate the difference between the LE and TE time stamps. This difference corresponds to the time over threshold (ToT) which is proportional to the charge deposited by a particle. The

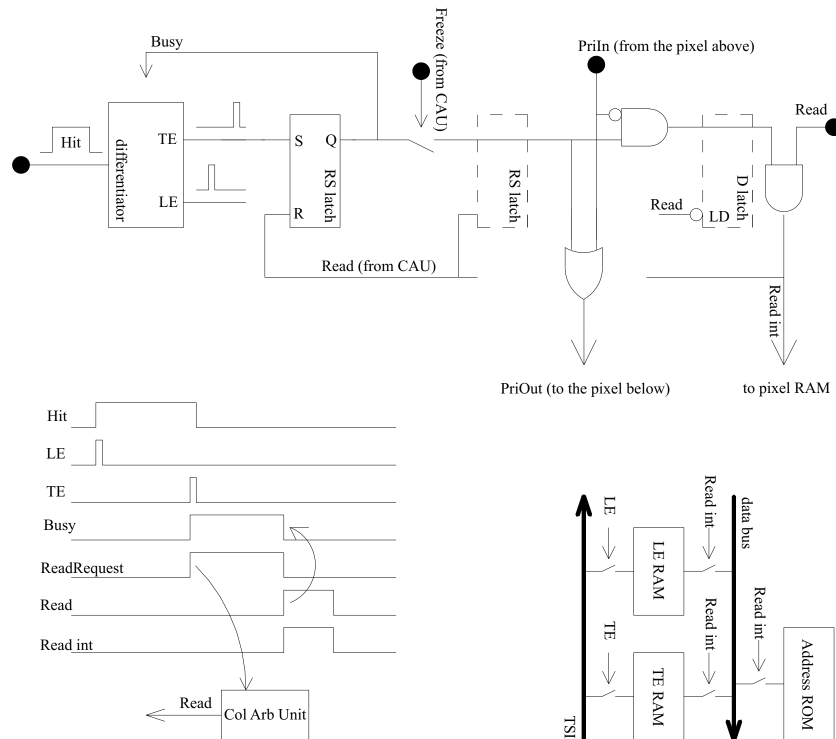


Figure 2.77: Digital block in a pixel cell. Timing diagram of the most important signals.

last block in the CAU is the ToT processor. The task of the processor is to perform the time walk correction.

Control Block in CAU

The generation of the read sequence is the main task of the CAU control block. We have already talked about this sequence in the section about the digital pixel block. The control block receives the *ReadRequest* signal from a pixel column[45] and generates, first, the *Freeze* and, some time later, the *Read* signal. After issuing the *Read* signal, the pixel RAM alone would need quite a long time to set the data lines to the clear CMOS levels. In order to increase the readout speed, the control block waits only a short time and then activates the sense amplifier, by generating the *Sense* signal. The sense amplifier can recognize a logic level at the bus even it does not have a full CMOS amplitude. The *Sense* signal activates also the ToT processor. When the ToT processing is finished, the read sequence is over. If the *ReadRequest* signal is still 1, which means that there are more hits in the pixel column waiting to be read out, a new read sequence is started. If both columns send the request, the read procedure is repeated alternatively for the left and right column. Fig. 2.78 shows the timing diagram of the relevant signals.

ToT Processor in CAU

The ToT processor can work in four modi. In the first modus, the processor copies the hit information delivered by the sense amplifier in its internal register and generates the

[45]There are two read request inputs, one for the left and one for the right column.

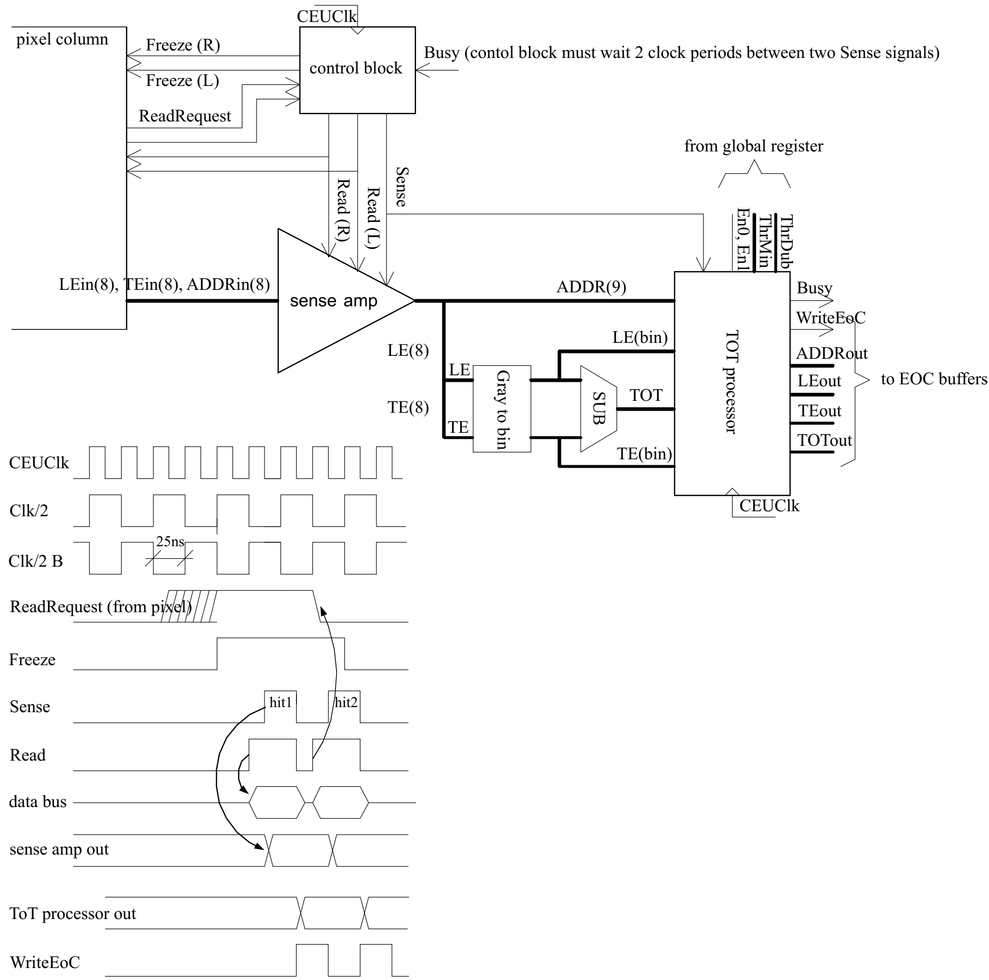


Figure 2.78: Column arbitration unit (CAU). Timing diagram of the most important signals.

$WriteEoC$ signal. This signal is received by the end-of-column buffer. The hit information contains a 9-bit pixel/column address, ToT and the binary coded LE time-stamp.

In the second modus, the ToT processor compares the ToT of the received hit with a programmable threshold, referred to as $ThrMin$. If the ToT is smaller than the threshold, the hit data will be rejected. In the case of the hits with small ToT, the leading edge time measurement may be wrong. Sometimes, it is better to disregard such data.

In the third modus, the ToT is compared with an another threshold, referred to as $ThrDub$. If the ToT exceeds the $ThrDub$, the hit data are transferred to the EoC buffer in the original form, as it is the case in the first modus. If the ToT is smaller than the threshold, the hit information is sent two times, first with the original LE time-stamp and after that with the LE time-stamp decremented by one (corresponds to the previous bunch-crossing). The doubling of the hit can be useful. In the case of the hits with small charge deposition (low ToT) the leading edge time-stamp can be by one higher than it should be. Such hits cold be assigned to the wrong crossing and lost if the L1 coincidence does not match. The time walk correction should prevent the loss of the hits with small charge deposition.

The fourth modus is the combination of the previous two.

State Machines in CAU

The control block and the ToT processor perform their tasks in a few steps with precisely defined timing between them. To synchronize these circuits a clock signal is needed. The clock signal used in CAU ($CEUclk$) is derived from the 40 MHz bunch-crossing clock XCK. The original 40 MHz frequency can be used, as well as decreased to 20 MHz, or 10 MHz. As we can see in the timing diagram of fig. 2.78 the rate of the hit data readout is two times smaller than the $CEUclk$ frequency. The maximal readout speed is therefore a hit per 50 ns.

Fig. 2.79 shows the simplified scheme of the ToT processor. This circuit has an inter-

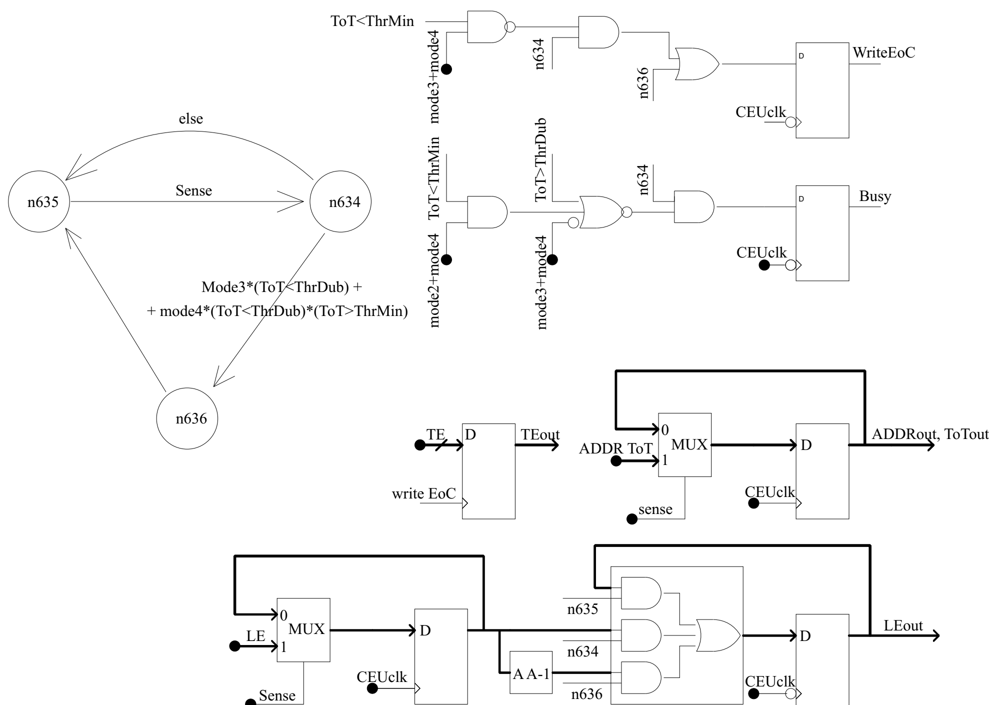


Figure 2.79: Block scheme of the ToT processor and its state machine.

nal control part that generates the control signals needed for the time walk processing. The control block is realized as a state machine. The state machine contains the memory elements, flip-flops, used to store the state code and the combinatorial circuit that calculates the next state. The next state is calculated by taking into account the current state and the input signals. Writing into flip-flops is synchronized with the clock signal. Synchronous with the active clock edge, the next state is stored in the flip-flops. The state machines are usually defined with a graph, as shown in fig. 2.79.

Readout of EoC-Buffer

The next large block in the readout system of the front-end chip is the end-of-column (EoC) buffer. A buffer is shared by two pixel columns in a column-pair and has 64 memory cells. A hit information is kept in a buffer cell until the L1 delay for the corresponding crossing is up. If the L1 arrives, the hit is read out. The readout sequence is controlled by the readout controller (ROC). The structure of the EoC buffer cannot be easily explained

without knowing how the readout controller works. Because of that, we will first introduce the ROC block.

Readout Controller

The task of the ROC block is to read out the L1 data from all nine EoC buffers and to send the data to the MCC chip. The L1 data are sent in the order their triggers arrive. The transmitting of the data is done by using a serial protocol. The main parts of the ROC are: the control block (state machine), two 4-bit counters, a memory block with sixteen 13-bit[46] cells (trigger FIFO)[47] and a shift register, fig. 2.80. The number stored

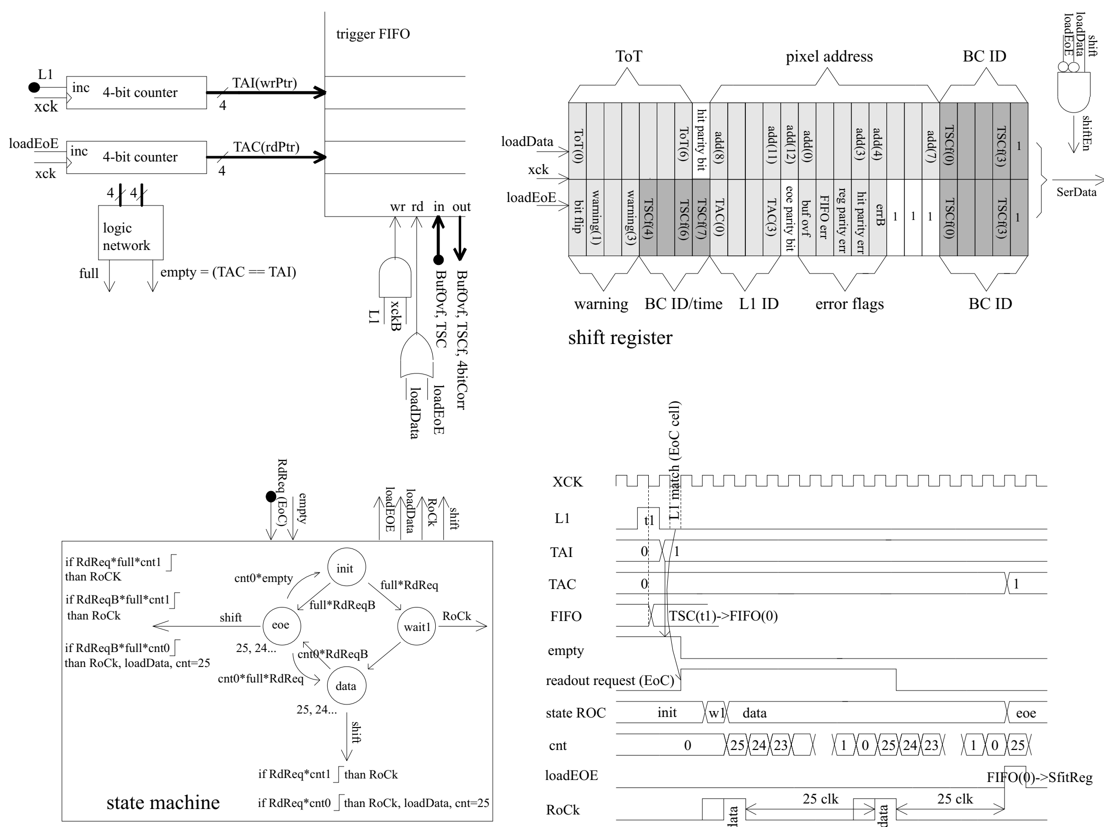


Figure 2.80: Read out Controller (ROC), its state machine and timing diagram of the most important signals.

in the first counter, referred to as trigger address increment (TAI), is the identification (ID) of the next L1 signal. This number is also the address of the next free cell in the trigger FIFO. The binary coded bunch-crossing time will be stored in this cell when the next L1 arrives. Every new L1 signal increments the TAI counter. The number stored in the second counter, referred to as trigger address counter (TAC), is the ID of one of the previous L1 triggers which is currently processed by the ROC. This ID is also the address of the memory place with the crossing time of the processed L1. This time is the identification of the bunch-crossing and it is transmitted together with the hit data as a

[46]Eight bits for bunch-crossing time, one parity bit and four bits for error correction.
[47]FIFO means first in, first out. This is the way how the L1 signals are processed.

part of the event word. The TAC counter is incremented at the end of the L1 processing sequence.

Let us briefly explain the L1 processing sequence. The processing of a L1-trigger can be more easily explained if we suppose that the ROC is just finishing the processing of the previous L1. The last hit information that corresponds to the previous L1 has just been transmitted from the EoC buffer. The ROC state machine generates the *loadEoE* (load end-of-event word) signal. The EoE word is formatted[48] and copied into a shift register. The content of the register is then shifted out of the chip, see fig. 2.80. The *loadEoE* signal increments the TAC counter. The EoC buffer cells notice the change of the TAC signal. The cells that contain the L1 hits with their L1 ID equal to the new TAC value send the readout request signal (*RdReq*). The request is received by the state machine in the ROC block. What happens after that is similar to the communication between a pixel column and CAU. ROC generates a read signal, referred to as $RoCk$.[49] The $RoCk$ signal is received by all EoC buffer cells. Only the cell with the highest place in hierarchy puts its data at the bus lines. The ROC then copies the data into the shift register[50] and wait until the shift register content is serially emptied out. If the readout request is still active, a new RoCk signal is generated and the next data word serialized. The ROC state machine repeats the $RoCk$ sequence until the readout request is active. When the all hit data belonging to this particular L1 are out, the readout request becomes zero. The ROC generates then the *loadEoE* signal. The sending of end-of-event word was explained at the beginning. If there is no hits that correspond to a L1-trigger, only an EoE word will be transmitted. The ROC keeps looping until $TAC = TAI$. This means that all of the L1 data stored on the chip have been transmitted and each trigger that arrived has been processed.

End-of-Column Buffer

An end-of-column buffer consists of 64 cells. Let us first introduce the data and signal lines which are received by the buffer cells (fig. 2.81).

EoC buffer receives from CAU binary coded leading edge time-stamp, 9-bit pixel address[51] and ToT.[52] A $WriteEoC$ signal is generated by CAU when a hit data is ready. In order to measure L1 coincidence, an EoC buffer receives the L1-trigger signal and the bunch-crossing time. The crossing time received in the buffer is binary coded and delayed relative to the TSI time, used in the pixel matrix. The delay can be varied. The delayed crossing time is referred to as TSC - time-stamp counter. In the moment when TSC and a hit leading edge time-stamp match, the "age" of the hit is equal to the L1 delay.

EoC buffer receives from ROC the following signals: TAI - the ID of the next L1 pulse; TAC - the ID of the L1-trigger which is processed by the ROC for that moment

[48] The format of the EoE word is shown in fig. 2.80. The EoE word contains the bunch-crossing ID (the lower four bits of the crossing time), the higher four bits of the crossing time, overflow, parity error and the bit flip flags.

[49] Readout clock.

[50] The format of the data word is shown in fig. 2.80. Beside the hit data, received from the EoC cell (7+1 bit ToT and 9-bit pixel/column address), the data word contains an additional four bit column address and the bunch-crossing ID obtained from the trigger FIFO.

[51] 8-bit address that encodes the position of the pixel in a column and an additional bit that determines whether the hit arrives from the left or from the right column of a column-pair.

[52] There is the possibility to send TE or LE time-stamp instead of ToT.

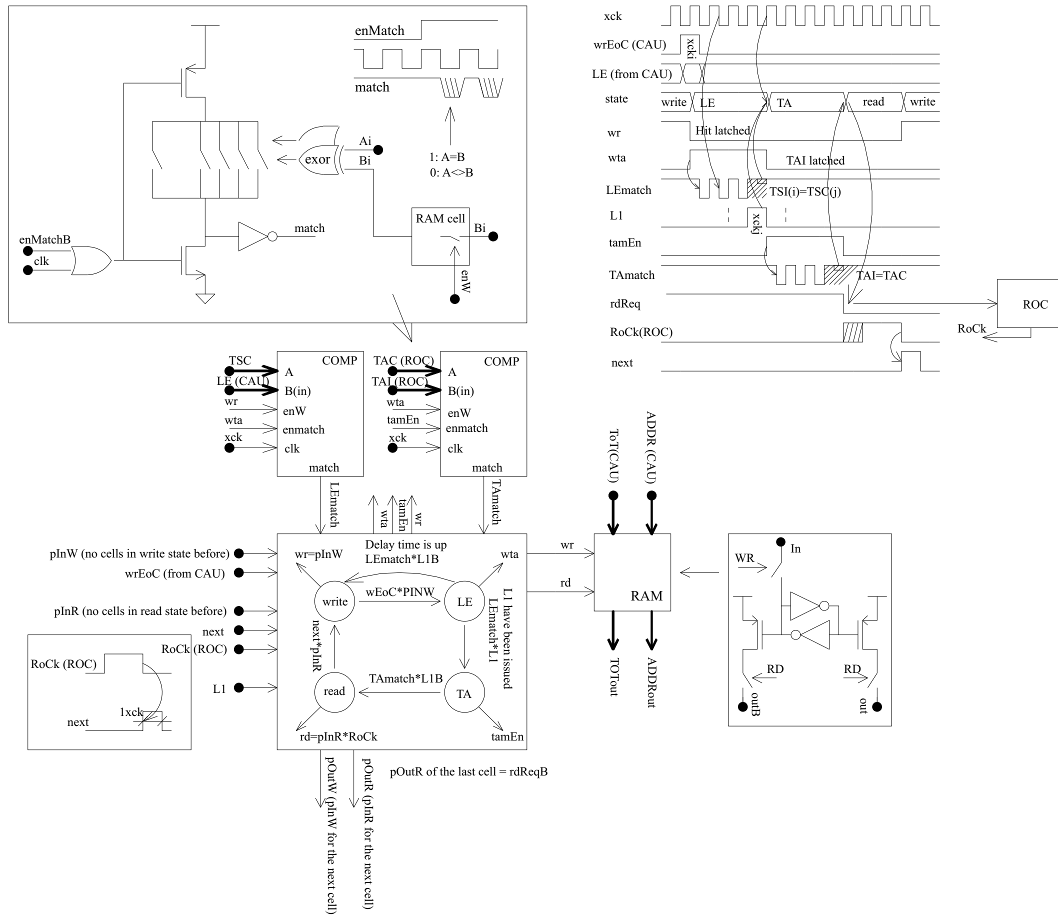


Figure 2.81: End-of-column buffer, state machine and timing diagram of the most important signals.

and $RoCk$ - the readout signal.

Every buffer cell contains a state machine with four states, an 8-bit and a 4-bit comparator, RAM for the hit data and a simple logic circuit which encodes hierarchy of the cells.

An EoC buffer cell is declared to be empty provided its state machine is in the first write-state. An empty cell is ready to accept new hit data. When the hit data are ready at the data bus and a $WriteEoC$ signal is generated by CAU, the first empty cell in hierarchy takes the data and stores them into its memory elements. The leading edge time-stamp is stored into the register of the 8-bit comparator. The remaining hit data (ToT and pixel address) are stored in the data RAM. After the data are stored, the cell goes in the second LE-state. The 8-bit comparator compares the stored LE time-stamp with the bunch-crossing time TSI. As we already mentioned, a delay period after the hit generation the two numbers LE and TSI become equal. In this moment, a L1 signal is expected. If the L1 does not arrive, the hit is not worth to be read out and the buffer cell goes in the empty-state. The uninteresting hit can be overwritten by a new one. If a L1 arrives in the moment when the 8-bit comparator fires, the the buffer cell copies the L1 ID - TAI into the register of the 4-bit comparator and goes in the third TA-state. All cells that are in the third state contain important L1 hits. A cell waits in the TA-state until ROC is ready to start the readout of the particular L1 event. The 4-bit comparator

has the task to detect when TAC, updated by the controller, becomes equal to the L1 ID of the stored hit. When the comparator responds, the cell goes in the fourth read-state and sends the readout request to ROC. The request signals of all cells are OR-ed. ROC answers with $RoCk$ signal. This signal is received by all 9 buffers and all of the cells that sent the readout request. The first cell in hierarchy transmits the hit data from the data RAM through the data bus. After the data transmitting a cell goes in the first empty-state.

2.5.5 Small Circuits

In the following sections we will describe the circuits used to set the bias and configuration parameters needed for the proper operation of the analog and digital blocks on the front-end chip. Potentials used to bias the analog pixel circuits are generated by 8-bit DA converters. Let us briefly introduce the design of the 8-bit DA converter.

8-bit DA Converter

The analog pixel cells are biased by ten potentials. These are: V_{VDD2}, V_{ID}, V_{IDDig}, V_{IP2}, V_{IP}, V_{TrimTh}, V_{IF}, V_{TrimIF}, V_{IL}, V_{IL2}, see fig. 2.68. These ten bias potentials are distributed as the gate potentials of diode connected transistors which conduct the currents generated by 8-bit DACs, see fig. 2.56. There are nine such DACs on the chip. The potentials V_{ID}, V_{IDDig} are generated by using the same DAC. Fig. 2.82 shows the block scheme of the 8-bit DAC. The output current can be adjusted in the range between 0 and 255 μA, with the least significant bit (LSB) of 1 μA. The straightforward design approach would be to use a matrix with 255 identical 1 μA cells. However, in the chip region underneath the pixel matrix, there were not enough space for nine so large matrices. It is necessary to make a compromise between the DAC size and its accuracy. The DAC contains a very precise 8×8 matrix consisting of 4 μA current sources and two smaller current sources that generate 2 μA and 1 μA currents. A similar DAC has been published in [34]. The matching between the smaller sources and the sources in the matrix is not so good as the matching between identical 4 μA sources. However, by proper sizing of the transistors, a very linear characteristic can be achieved. The 6-bit core is designed very carefully. The block diagram of a 4 μA unit cell is shown in fig. 2.82. The current flows into the output terminal either when the Xon signal is active, or when both $Ysel$ and $Xsel$ intersect in the particular row and column. Eight Yon signals are generated by the Y decoder which receives as input the three DAC control bits $Dac(2:4)$. Eight $Xsel$ and eight Xon signals are derived from the three most significant bits $Dac(5:7)$. The signal patterns are shown in fig. 2.82, bottom. If we keep increasing the DAC setting starting from zero, the cells incrementally switch on, one after the other. There are no steps when some larger block of cells is switched off, while some other group is switched on. An incremental scheme, as that shown in fig. 2.82, has always a monotonic characteristic regardless to the mismatch between transistors. The pattern of the current cells that are switched on is radially symmetrical with respect to the matrix center, see fig. 2.82 bottom right. In this way, the influence of possible gradients in the material characteristics is eliminated.

The design of the switches in the unit cell assures that the current flows either into the output terminal or into ground. To assure that the drains of identical transistors are

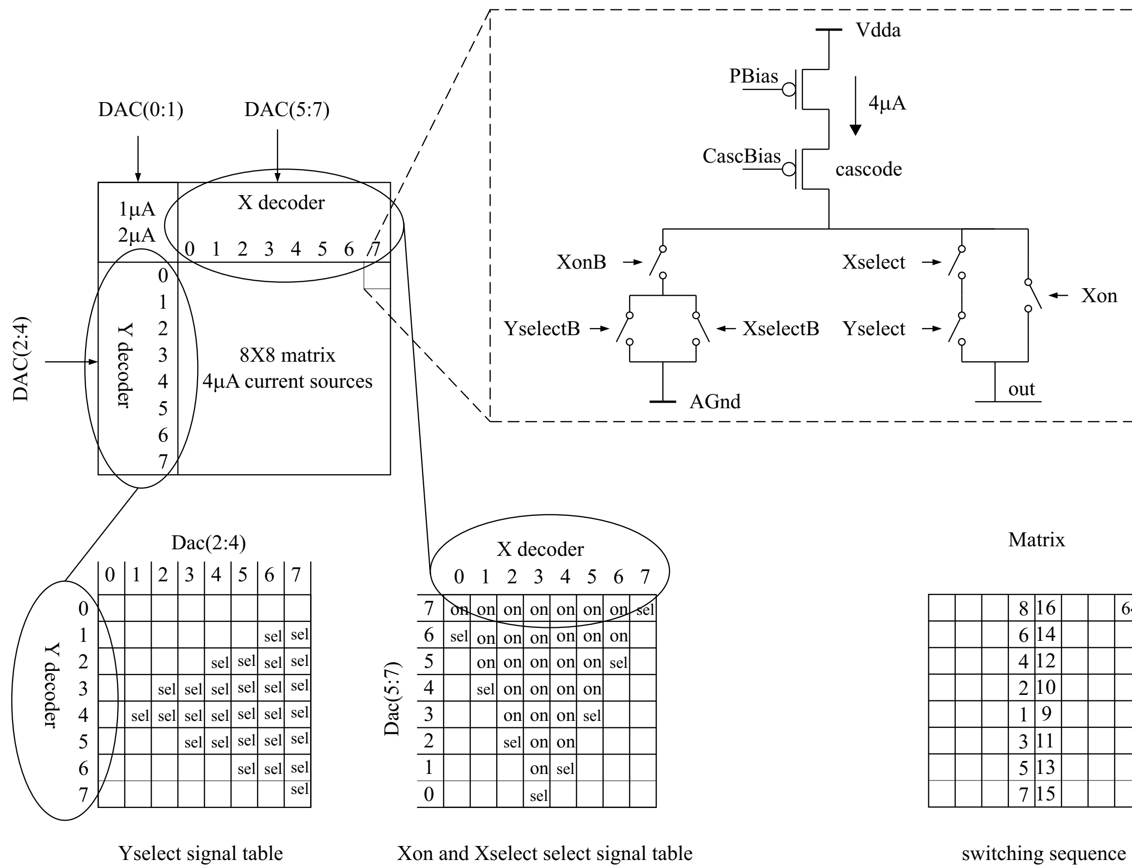


Figure 2.82: 8-bit current-mode DA converter.

always at the same potential, a cascode transistor is used, fig. 2.82. Only when identical transistors are biased in the same way, a good current matching can be obtained.

The injection potential ($VCAL$) is generated by a 10-bit DA converter. The 10-bit DAC is similar to the 8-bit converter. The current sources in the 8×8 matrix of the 10-bit DAC generate $8\,\mu$A current. The PMOS current sources are designed as parallel connections of two smaller transistors, each conducting $4\,\mu$A. A single $4\,\mu$A transistor is placed in the spare matrix field (64). The remaining three binary weighted current sources (0.5, 1 and $2\,\mu$A) are outside of the matrix. By the use of such arrangement we improve the matching between the $8\,\mu$A and $4\,\mu$A sources. The output current can be steered in the range between 0 and $511\,\mu$A, with a $0.5\,\mu$A step. The output current is converted to the $VCAL$ potential by using a resistor and a cascoded current mirror.

Reference-Current Generator

The current mode DAC introduced in the last subsection is biased by the voltage between $Vdda$ and $PBias$, fig. 2.82. This voltage defines the unit current of the DAC. Since the 8-bit DACs generate the bias potentials for the whole pixel matrix, all DC currents on the chip depend indirectly on the PBias level. The PBias potential is generated by a circuit that we refer to as current reference. The scheme of the current reference is shown in fig. 2.83. In order to understand the principle of the current regulation, it is enough to consider only the core of the circuit. The core consists of eight transistors and one resistor, fig. 2.83. The transistors T5, T6, T7 and T8 are the casode transistors. Their purpose is

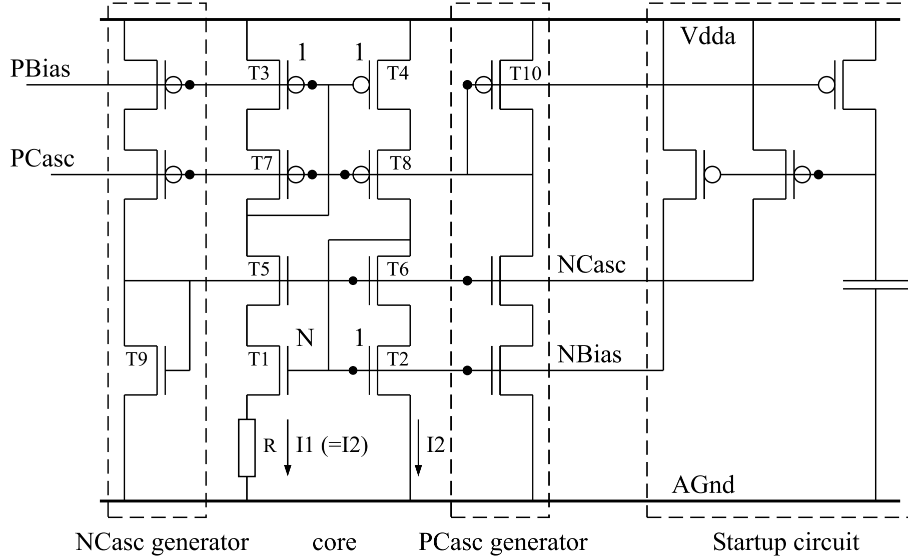


Figure 2.83: Current reference.

to improve the current matching between the PMOS mirror transistors T3 and T4 and the NMOS mirror transistors T1 and T2.[53] We can disregard the cascode transistors in the simplified analysis which assumes the perfect current matching.

The PMOS current mirror keeps the currents I1 and I2 equal. Let us now write the equations for the gate potentials of the NMOS transistors T1 and T2 (we assume that the transistors operate in weak inversion, see eq. 2.30)

$$V_{g1} = RI1 + V_{th} + nU_T \ln\left(\frac{I1}{(W/L)_1 \, I_{sub}}\right); \quad V_{g2} = V_{th} + nU_T \ln\left(\frac{I2}{(W/L)_2 \, I_{sub}}\right) .$$

V_{th} is the threshold voltage ($V_{th} \approx 0.5\ V$), n is the slope factor ($n \approx 1.5$), U_T is the thermal voltage ($U_T \approx 26mV$), $I_{sub} = 2n\mu C'_{ox}U_T^2$.

If we take into account I1 = I2, $(W/L)_1 = N\,(W/L)_2$ and $V_{g1} = V_{g2}$, the equation for the current I1 can be easily obtained:

$$I1 = \frac{nU_T}{R}\ln(N) \ . \tag{2.42}$$

The current is independent of the supply voltage. By proper choosing of the resistance R and the width multiplication factor N (see fig. 2.83), a nominal current of $1\,\mu$A is achieved.

Shift Registers in the Front-End Chip

The analog and digital blocks on the front-end chip use a lot of configuration bits. These bits are stored in SEU tolerant latch/RAM cells placed in the vicinity of the controlled circuits. The latch cells are written by using shift registers. A shift register allows us to send the desired bit through the chip to the area, where the latch is placed. In the section about the analog pixel electronic we have introduced the pixel shift register. The

[53]The cascode transistors keep the drain potentials of the mirror transistors at the same level.

front-end chip has an additional shift register, called the global register, used to write the bits that control the global chip functions and circuits like the 8-bit bias DACs, 10-bit injection DAC, readout system, chopper modes, autotune, *etc.* The shifting of the bit patterns as well as latching of the bits in the local latch cells is controlled by a command register. The block scheme of the front-end registers and the logic circuit that implements the interface to these registers is shown in fig. 2.84. The communication with the front-

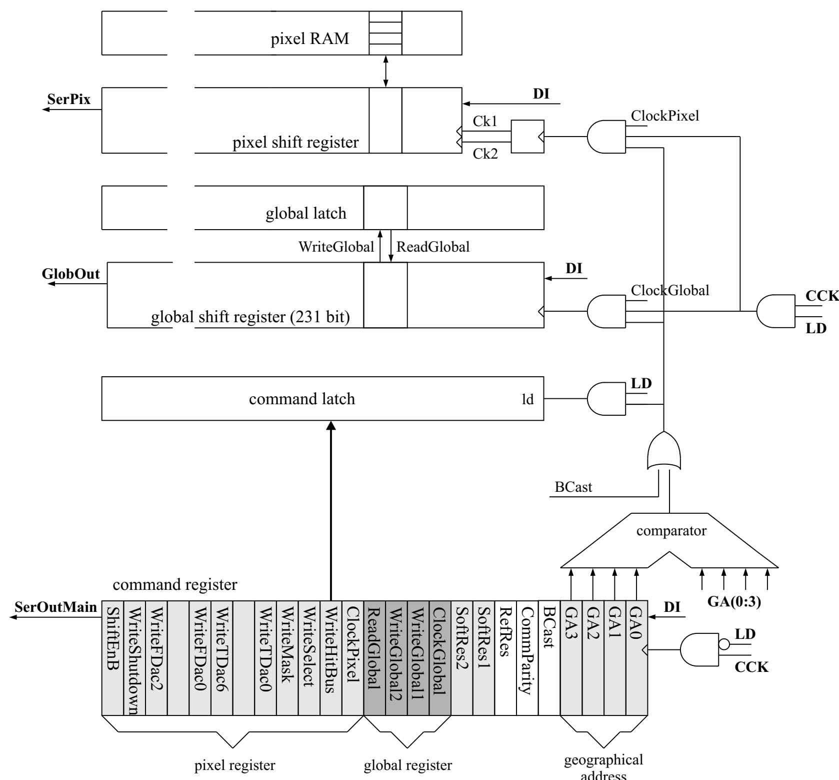


Figure 2.84: Registers on the front-end chip.

end chip uses a simple serial protocol [16]. There are three input signals: DI (data in), LD (load) and CCK (command clock). When LD is low, a command can be written into the command register. The last four bits in a command encode the geographical address of the front-end chip, $GA3$,...,$GA0$, fig. 2.84. Each chip in the sixteen chip module has its own address. If the address in a command matches with the chip address, the command is accepted and stored in the command latch, fig. 2.84. In order to latch a command LD must be set high.[54] Note that a command becomes active only if it is written into the command latch.

The meaning of command bits is shown in fig. 2.84. There are two bit blocks that control the access to the global and pixel registers. By setting the $ClokGlobal$ or $ClokcPixel$ bits at the high level, we can shift a bit pattern into the global or pixel register, respectively. LD must be high during the shifting.[55] If we set the $WriteGlobal$ bits to 1, the bits stored in the global shift register are written into the global latch cells.[56] A bit from

[54]There is also a broadcast bit that enables the MCC chip to send a command to all 16 chips at the same time.

[55]The signal $ShiftEnB$ must be zero while we shift into the pixel register.

[56]Both bits $WriteGlobal1$ and $WriteGlobal2$ must be set to 1, otherwise the global latch cannot be

the pixel shift register can be stored into one of the fourteen pixel RAM cells, fig. 2.63. We need therefore fourteen different write bits for the pixel register, fig. 2.84. There is also the possibility to read back the bits from the global latch or from the pixel RAM into the shift registers. In this way, we can test the SEU tolerance of the latch cells. In order to read back the content of the global latch into the global shift register, we need only to activate the *ReadGlobal* command. In the case of the pixel register, the read back procedure is more complicate. It has been already explained, fig. 2.63. First, we need to connect a RAM cell to the input of the flip-flop. It is done by setting the $ShiftEnB$ and a $Write$ command to 1. The RAM content can be then stored in the flip-flop by applying of a clock pixel sequence. Finally, the bits have to be read out by doing the clock pixel sequence. The $ShifEnB$ signal is thereby 0.

Triple-Redundant Global Latch-Register

The global shift register has 231 bits. The first 38 bits are placed at the bottom of the chip, where a lot of free space is available. The remaining 193 bits are close to the bias DACs underneath the pixel matrix. In this region the layout density is high. Because of that, the register cells are somewhat simpler than those at the bottom of the chip. The design concept of the latch cells is similar for the both segments of the global register. Triple redundant scheme are used. A bit is stored in three SEU tolerant latch cells, similar to that used in pixels, fig. 2.62. A majority logic calculates the output of the triple redundant block. Fig. 2.85 shows the implementation of the global latch cell used in the region close to the pixel matrix (193 bits). As we already mentioned, both $WriteGlobal$ signals must

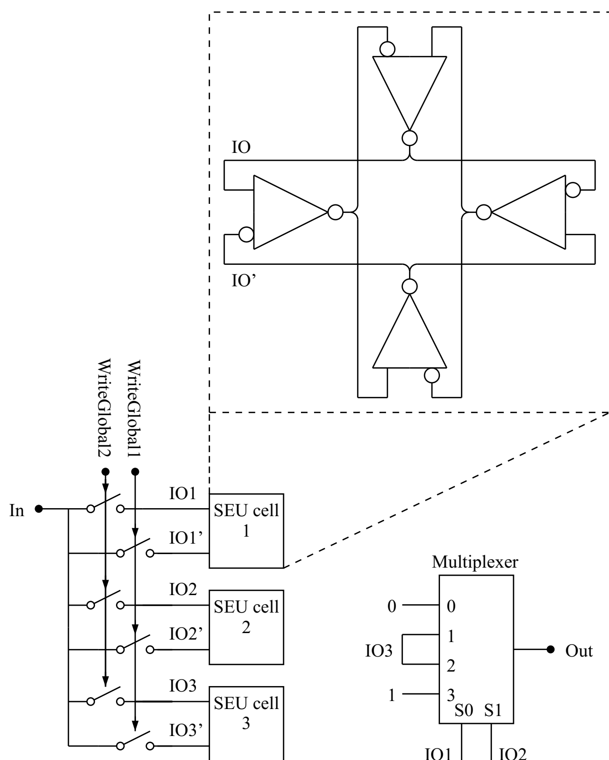


Figure 2.85: Triple redundant latch cell and the majority logic.

written. The purpose of this redundancy is to improve the SEU tolerance.

be active in order to write into the SEU cells.

Description of Global Bits

DI

Bits	Name	Group
0	GlobalParity	
1:8	L1 Latency (0:7)	
9:12	SelfTriggerLatency (0:3)	self trigger
13:16	SelfTriggerWidth (0:3)	self trigger
17	EnableSelfTrigger	
18	EnableHitParity	
19:22	SelectDO (0:3) (CCK, **HitBus**, SoftRes1B, **SerData**, MonSel, PixCk1, PixCk2, HardRes1B, **SerData**, SerOutMain, ReadFIFO, **SerPix**, L1, RoCk, SerOutReg, **GlobalOut**)	output MUX
23:26	SelectMonHit (0:3)	
27	TSI/TSC Enable	
28	Spare 0	
29	Spare 1	
30:37	HitBusScaler (0:7)	

Bits	Name	Group
38	TestLeakReference	leakage current ADC
39:48	Leak DAC (0:9)	leakage current ADC
49	Test Leak DAC	leakage current ADC
50	EnableLeakComparator	leakage current ADC
51	LeakComparatorOut	leakage current ADC
52:53	ARegulatorTrim (0:1) (1.5V, 1.6V, 1.7V, 1.8V)	Vdda regulator
54	EnableARegulatorMeas	Vdda regulator
55:56	ARegulatorMeas (0:1) (Out, V_{in}, V_{BG}, gnd)	Vdda regulator
57	EnableARegulator	Vdda regulator
58	Test LVDS Reference	
59:60	DRegulatorTrim (0:1) (1.8V, 2.0V, 2.2V, 2.4V)	Vdd regulator
61	EnableDRegulatorMeas	Vdd regulator
62:63	DRegulatorMeas (0:1) (Out, V_{in}, V_{BG}, gnd)	Vdd regulator
64:67	CapMeasDUT (0:3) (NC, 4lo, 2lo, 1lo, NC, 0lo, 4fb, 2fb, NC, 1fb, 0fb, 4hi, NC, 2hi, 1hi, 0hi)	capacitance measurement
68:69	CapMeasFrequency (0:1) (f_{CK} = 5, 10, 20, 40MHz)	capacitance measurement

Bits	Name	Group
70	EnDecouplingCapTest	
71	EnTestPixelBuffer	analog buffer
72:73	TestPixelMUX (0:1)	analog buffer
74	EnableVCALMeas	analog buffer
75	EnableLeakMeas	
76	BoostTestPixelBuffer	
77	EnableColumn8	
78	Test IVDD2 DAC	
79:86	IVDD2 DAC (0:7)	
87:94	ID DAC (0:7)	
95	Test ID DAC	
96	EnableColumn7	
97	Test IP2 DAC	
98:105	IP2 DAC (0:7)	
105:113	IP DAC (0:7)	
114	Test IP DAC	

Bits	Name
115	EnableColumn6
116	Test ITrimTh DAC
117:124	ITrimTh DAC (0:7)
125:132	IF DAC (0:7)
133	Test IF DAC
134	EnableColumn5
135	Test ITrimIF DAC
136:143	ITrimIF DAC (0:7)
144:153	VCAL DAC (0:9)
154	Test VCAL DAC
155	EnableColumn4
156	Hi (Injection)
157	EnExternalInjection
158	Test Analog Reference
159:160	EOC MUX Control (0:1) (ToT, LE, ToT, TE)
161:162	CEU Clock Control (0:1) (0, 10, 20, 40MHz)

Bits	Name	Group
163	EnableDigitalInjection	
164	EnableColumn3	
165	Test ITh1 DAC	
166:173	ITh1 DAC (0:7)	
174:181	ITh2 DAC (0:7)	
182	Test ITh2 DAC	
183	EnableColumn2	
184	Test IL DAC	
185:192	IL DAC (0:7)	
193:200	IL2 DAC (0:7)	
201	Test IL2 DAC	
202	EnableColumn1	
203:210	THRMIN (0:7)	time walk correction
211:218	THRDUB (0:7)	time walk correction
219:220	ReadMode (0:1) (NoCorr, ThrMin, ThrDub, both)	time walk correction
221	EnableColumn0	
222	HitBusEnable	
223:227	GDAC (0:4)	
228	AutoTune Enable	
229	VoltDropCompensation	
230	IP/IL2 Monitor En	

GlobalOut

Figure 2.86: Bits in the global register.

Fig. 2.86 shows the list of all bits in the global register. This long list illustrates the complexity of the front-end chip. Although we have discussed the most important parts of the chip, there are many other smaller circuits that we did not mention. The purpose of these blocks is mainly the testing of the other electronic parts. In this chapter we will shortly introduce the small circuits.

Self-Trigger Circuit

If we start to examine the global register from the bit 0, after the parity bit and the L1 latency block, we come to the nine-bit block that controls the self-trigger circuit. The self-trigger is a feature that allows the testing of the front-end chip when an external L1 signal is not available. The self-trigger circuit receives the *HitBus* signal and generates, after some programmable delay, an internal L1 signal. The trigger delay and the width of the internally generated L1 are programmable. The values are stored in the global register. Using the self-trigger mechanism we can, for example, test a front-end chip bonded with the sensor by irradiation with a radioactive γ source, *e.g.* ^{241}Am (6 KeV). Every photon hit activates the self-trigger and the chip automatically transmits the hit data.

Output Multiplexer

The front-end chip has only two digital output pins, *DO* (data output) and *MonHit* (monitor hit). There are many internal digital signal of importance, for example, the hit data output (*SerData*), the *HitBus* signal, the shift register outputs (*GlobalOut* and *SerPix*), *etc.* All these signals can be multiplexed out.[57]

AD Converter for the Leakage Current

Fig. 2.87 shows the simplified scheme of the analog-to-digital converter (ADC) used to automatize the measurement of the leakage current. We explained the leakage current

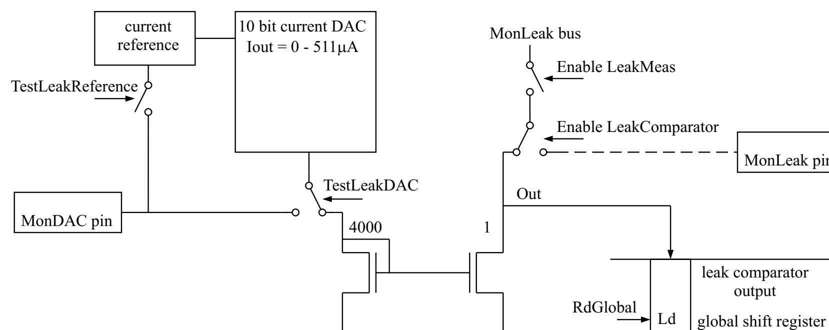


Figure 2.87: ADC for the digital measurement of the leakage current.

measurement when we introduced the complete scheme of the feedback circuit for the charge-sensitive amplifier, fig. 2.57. The current flowing through the *MonLeak* bus can be either directed to the analog output *MonLeak* pin and measured externally, or compared with the current generated by a 10-bit DAC, fig. 2.87. The result of the comparison is stored in the *LeakComparatorOut* bit of the global register when *ReadGlobal* is issued. Note that the 10-bit DAC has its own reference. The DAC and the reference can be calibrated by directing their output currents to the *MonDAC* pin.

On-Chip Power-Supply Regulators and the Serial Powering

The bits 52-57 in the global register are reserved for the *Vdda* regulator. In this section we will explain the purpose of this regulator.

There are two approaches in powering of front-end chips and multi-chip modules. The first possibility is to distribute the analog and digital supply voltages in parallel, fig. 2.88 (a). The supply generators are placed quite far from the modules, typically, 100 m away. The power line resistance is therefore quite large. If we have n front-end chips connected in parallel to a power line, the current that flows through the line is n times the supply current of a device. A lot of power is lost across the wire resistance. This produces heat and makes the cooling of modules difficult. The second possibility is to connect the front-end chips (or modules) serially, fig. 2.88 (b) [8]. The same supply current flows then through all devices. This current is only a little bit larger than the maximal supply

[57] The multiplexer channels 3 and 8 are reserved for *SerData*. The output of the global register (*GlobalOut*) is at the channel 15, the pixel register output (*SerPix*) is at the channel 11.

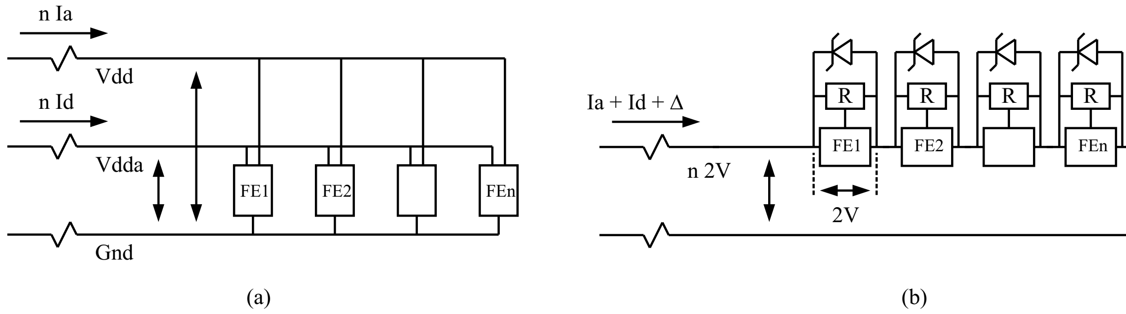


Figure 2.88: (a) - parallel powering of the front-end chips. (b) - serial powering scheme.

current of a device. This current is generated by a supply current source.[58] The digital and analog supply voltages are generated by on-chip regulators. The digital supply voltage (2 V nominal value) is defined for each chip (module) by a shunt regulator denoted in fig. 2.88 with a Zener diode symbol. The lower analog supply voltage (1.6 V) is generated for each chip or module by a linear voltage regulator denoted with R. The input voltage of a regulator is the digital supply voltage. A clear advantage of the serial powering versus the parallel powering scheme is that much less power is lost in the long supply lines. High-voltage transmission lines use a similar idea. If we connect n devices (front-end chips or modules) serially, the total supply voltage is n times higher than that in the case of the parallel powering. However, the current flowing through supply lines is n times smaller. The disadvantage of the serial concept is that the failure in a device could affect the whole chain.

Shunt Regulator

Fig. 2.89 shows the implementation of the shunt regulator. The current-voltage charac-

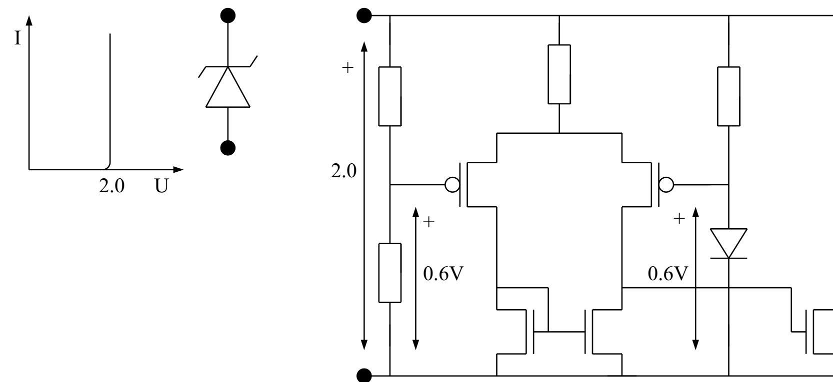


Figure 2.89: Shunt regulator.

teristic of the circuit is also shown, fig. 2.89. The characteristic corresponds to that of a Zener diode with 2.0 V breakdown voltage.

[58]There are two approaches in the serial powering concept. The first is to connect some number of modules serially, while the chips on a module are powered in parallel. The second is to power the chips on a module serially, while the modules are in parallel.

Voltage Regulator

The block scheme of the voltage regulator used to generate the analog supply voltage is shown in fig. 2.90. The input of the circuit is connected to a separate pin *RegIn*. The

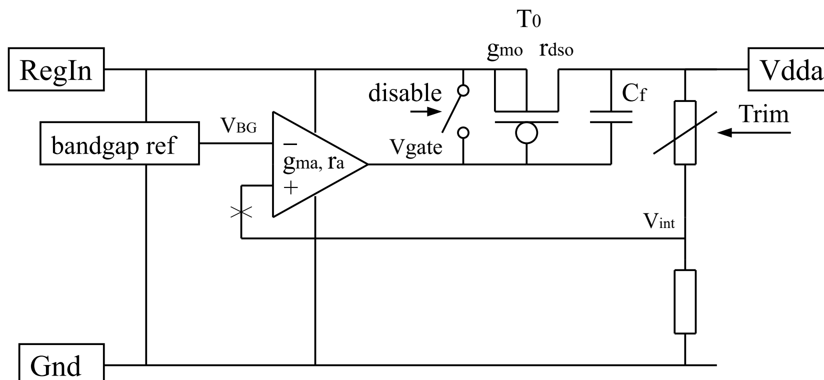


Figure 2.90: Voltage regulator.

output is connected to the $Vdda$ line. The ground of the regulator is connected to the analog ground Gnd. The output node is connected to a voltage divider. The potential V_{int} is compared with the reference potential V_{BG} which is about 1.1 V calculated with respect to ground. The high gain transconductance amplifier, which drives a large PMOS transistor T_O, detects any deviation of V_{int} from V_{BG} and corrects it. Due to the high-voltage gain V_{gate}/V_{int}, the potential V_{int} is always almost equal to V_{BG}. The resistors in the voltage divider are sized in the way to assure that the potential of the $Vdda$ node with respect to ground has the value $1.6V$ when $V_{int} = V_{BG}$. There is the possibility to trim the output potential by setting of two register bits, fig. 2.86.[59]

The capacitor C_f is important for the stability of the circuit. The open loop gain has two dominant poles.[60] A pole arises due to the large parasitic Miller capacitance C_{Mill} seen at the node V_{gate},[61] and a high output resistance of the transconductance amplifier r_a. The pole frequency is $1/(r_o C_{Mill})$. The second pole arises at the frequency when the Miller capacitance drops below the large gate-source capacitance of the PMOS transistor C_{gs}. The decrease of the Miller capacitance is caused by the decrease of the voltage amplification of the transistor T_O. The gain decrease is result of the large capacitive impedance of the Vdda net C_o. If the two poles are close to each other the differential equations of the system with feedback have oscillatory solutions with low damping. The additional capacitor C_f splits the poles and stabilize the circuit. The pole splitting is result of increase of the Miller capacitance. The pole-frequencies are given by the equations: $\omega_{p1} = g_{ma} C_{Mill} = g_{ma} g_{mo} r_o C_f$ and $\omega_{p2} = g_{mo} C_f/(C_{gs} C_o)$.

[59] The potentials V_{BG}, V_{int}, as well as the output potential and the ground potential in the vicinity of the regulator can be measured. These nets can be connected to the analog output pad MonVCAL.

[60] The X symbol in fig. 2.89 shows the point, where we can break the feedback loop in order to calculate the open loop gain.

[61] Since a small change of V_{gate} is followed by a large opposite change of $Vdda$, the parasitic gate-drain capacitance of the PMOS (C_{gd}) is effectively amplified. The amplification factor is equal to the voltage gain of the PMOS transistor T0: $C_{Mill} = g_{mo} r_o C_{gd}$; g_{mo} is the transconductance of the transistor TO, r_o is the equivalent small-signal resistance of the $Vdda$ node, C_{gs} is the gate-source capacitance of the transistor T_O.

Band-Gap Reference

The reference voltage generator used to generate V_{BG} (see fig. 2.90) has a scheme similar to that of the current reference shown in fig. 2.83. The transistor scheme of the voltage reference is shown in fig. 2.45. The PMOS current mirror (transistors T3 and T4) keeps

Figure 2.91: Generator of the reference voltage (bandgap reference).

the currents I1 and I2 equal. The source potentials of the identical NMOS transistors T1 and T2 are therefore equal.[62] A forward biased diode has the following characteristic [2]

$$I_d = \alpha {n_i}^2 A \exp\left(\frac{eV_d}{kT}\right) = \beta T^3 A \exp\left(\frac{eV_d - E_g}{kT}\right) . \tag{2.43}$$

n_i is the intrinsic carrier density in silicon, A is the diode area, E_g is the band-gap in silicon, T is temperature and α and β are constants independent of the diode area and temperature. A typical forward bias voltage is $V_d \approx 0.6\ V$ therefore smaller than E_g/e. Therefore the following condition holds

$$I_d < \beta T^3 A . \tag{2.44}$$

We can now write the equations for the source potentials of the NMOS transistors T1 and T2

$$V_{s1} = RI1 + \frac{E_g}{e} + \frac{kT}{e}\ln\left(\frac{I1}{\beta T^3 A_1}\right); \quad V_{s2} = \frac{E_g}{e} + \frac{kT}{e}\ln\left(\frac{I2}{\beta T^3 A_2}\right) .$$

If we take into account that $I1 = I2$, $A_1 = NA_2$ and $V_{s1} = V_{s2}$, the following equation can be derived

$$I1 = \frac{kT}{eR1}\ln(N) . \tag{2.45}$$

The equation eq. 2.45 is very similar to the equation for the current generated by the current reference, eq. 2.42

[62] We assume a large drain-source resistance.

The output potential V_{out} calculated with respect to ground is

$$V_{out} = R2 \cdot I1 + \frac{E_g}{e} + \frac{kT}{e} \ln\left(\frac{I1}{\beta T^3 A_3}\right) = \frac{kTR2}{eR1} \ln(N) + \frac{E_g}{e} + \frac{kT}{e} \ln\left(\frac{kT \ln(N)}{eR1\beta T^3 A_3}\right) . \quad (2.46)$$

The logarithm in the last term is negative, this is a consequence of the condition given by eq. 2.44. By proper choosing of the resistance R, the potential V_{out} can be made almost independent of temperature in a large temperature range. Note that the third term decreases with temperature raise (due to its negative sign) while the first term increases. It can be shown that the first and third term both vanish when the temperature comes close to the absolute zero. The output voltage becomes than E_g/e, or 1.1 V. If the output voltage is made to be temperature independent, nearly the same value (E_g/e) holds also for the higher temperatures. Because of the interesting property that the output voltage is equal to the semiconductor bangap expressed in volts, the circuit is usually called bangap reference.

Capacitance-Measurement Circuit

The register block that controls the $Vdda$ regulator is followed by a similar bit block used for the Vdd regulator. The Vdd regulator is identical to the $Vdda$ regulator, with the exception that it generates a higher nominal output voltage (2.0V). The regulator allows us to power the chip with an external voltage generator which provides a higher voltage than the nominal digital supply voltage.

The bits 64:69 control the circuit for the capacitance measurement, see also [35]. The principle of capacitance measurement is illustrated in fig. 2.92. During a clock period the transistor T2 is switched on. The capacitor under test C_{ut} is charged through the pin $CapMeas$ which is kept at a constant potential (usually Vdd). In the next clock period the transistors T1 and T2 are switched off. After that, the transistor T1 is switched on and the charge from the capacitor C_{ut} flows into ground. In the fourth clock period the both transistors are again disconnected. The whole sequence repeats. The mean current flowing through the pin $CapMeas$ is $I_{CM} = V_{CM} C_{ut} f_{Ck}/4$, V_{CM} is the potential of the $CapMaes$ pin with respect to ground, f_{Ck} is the clock frequency. If we measure the current I_{CM} with an instrument sensitive to the mean value, the capacitance C_{ut} can be calculated according to the last formula. The circuit measures the capacitances of the bonding stacks identical to these used in pixels, fig. 2.69. The bonding stack contains a feedback capacitor and two injection capacitors. The exact values of these capacitors are of great importance for the calibration of the front-end amplifier.

In order to eliminate the offset from the measurement, four channels are reserved for the measurement of one capacitor. This is illustrated in fig. 2.92, bottom. The first channel measures only the stray capacitance of the metal frame used to connect the capacitors under test. The second channel measures one capacitor (and the stray capacitance of the frame). The third channel measures two and the fourth four capacitors. The capacitance can be found by linear fit. Note that the measured values are not exactly the same as the values "seen" by the charge-sensitive amplifier, fig. 2.92. This is illustrated in fig. 2.92 for the case of the feedback capacitor measurement. The amplifier sees only the value defined as $\partial Q_{in}/\partial V_{out}$. However, the test circuit measures $\partial(Q_{in} + Q_{hi})/\partial V_{out}$. An additional small out-hi capacitance is measured although it is not a feedback capacitance. The bonding stack is designed in the way to minimize the out-hi capacitance.

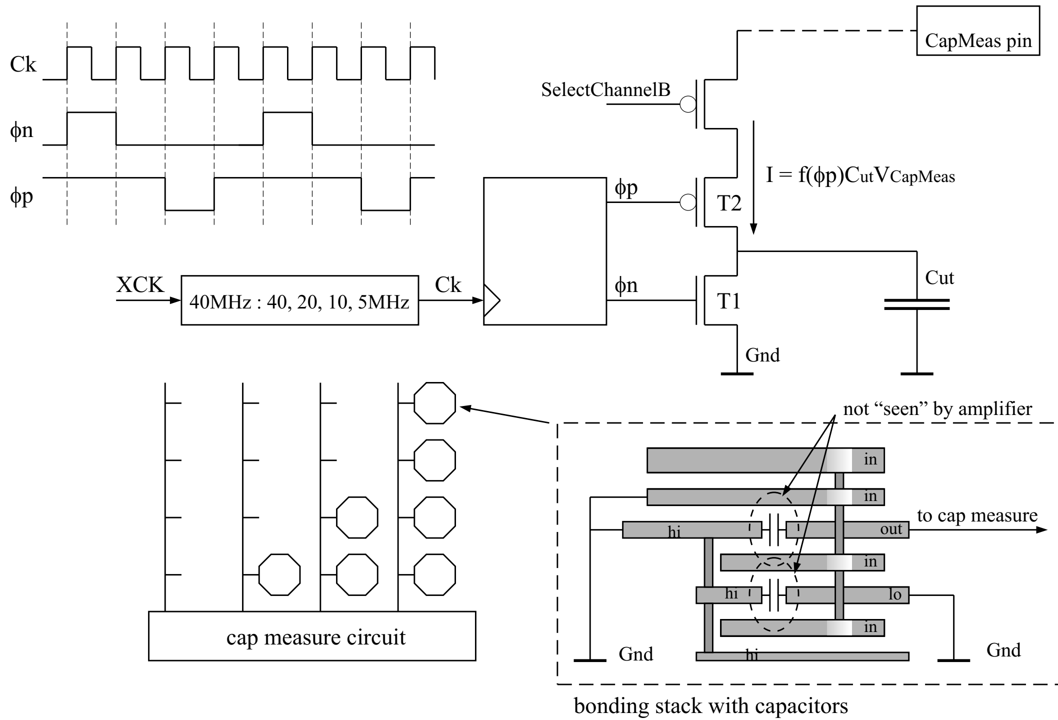


Figure 2.92: Scheme of the circuit for capacitance measurement and timing diagram, top. Bottom, left, four channels used for the measurement of the feedback capacitor. Bottom, right, bonding stack.

The clock frequency can be adjusted to be 40, 20, 10 and 5MHz. The frequency can be set by two bits in the global register. A channel can be selected by four additional bits, fig. 2.86.

The bit block that controls the capacitance measurement is followed by the bits used by the test pixel buffer. The buffer allows the measurement of the important voltage signals in a test pixel. The test pixel is placed in the bottom right corner of the pixel matrix. The buffer has a very small input impedance. It is capable to drive a large load, for example an oscilloscope probe. The voltage amplification of the buffer is one. Analog multiplexer, placed in front of the buffer enables us to select one of four channels. Channel 0 is the step input signal for the chopper, channel 1 is the output of the charge-sensitive amplifier, channel 2 is the noninverted and channel 3 inverted output of the second amplifier.

The remaining bits in the global register have been either already explained in the sections about the analog and digital blocks, or their meaning is intuitively clear. It is worth to mention that the current of each 8-bit DAC or reference can be directed into a test pad, called $MonDAC$. It is done by setting the corresponding test bits, see also fig. 2.87. The injection potential $VCAL$, used in the internal chopper mode, can be measured by using the $MonVCAL$ pad.

2.5.6 Measurements

In this section we will very briefly introduce the results of a few measurements that verify the functionality of the circuits on the front-end chip. Fig. 2.93 shows the output voltage of a voltage regulator as the function of its input voltage. Both voltages are

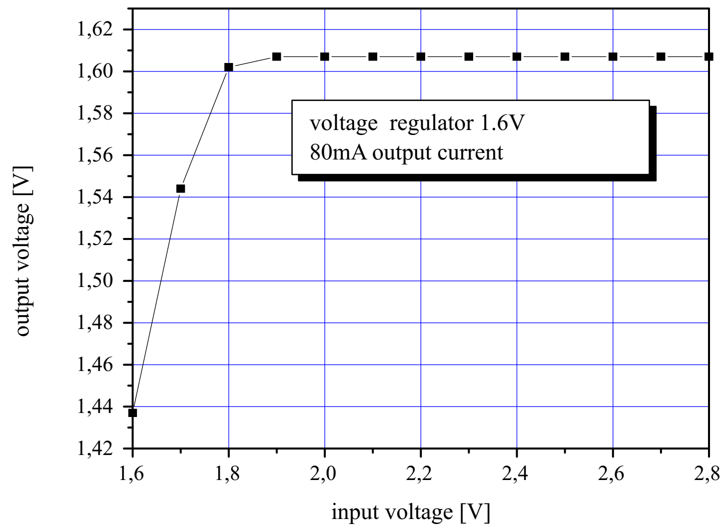


Figure 2.93: Regulated output potential of the analog voltage regulator versus the input potential. Output current was 80 mA.

defined with respect to ground. The output voltage should be constant, which is verified by this measurement. Note that the input voltage must be, typically, 200 - 300 mV higher than the nominal output voltage, otherwise the circuit does not work properly. The minimal input-output voltage corresponds to the minimal V_{ds} voltage of the big pass transistor T_O (fig. 2.90) needed for the conduction of the given output current. The measurement illustrates good performances of the bandgap voltage reference, too. Although the bandgap reference is supplied by the input voltage (fig. 2.90) the output voltage of the reference obviously does not change.

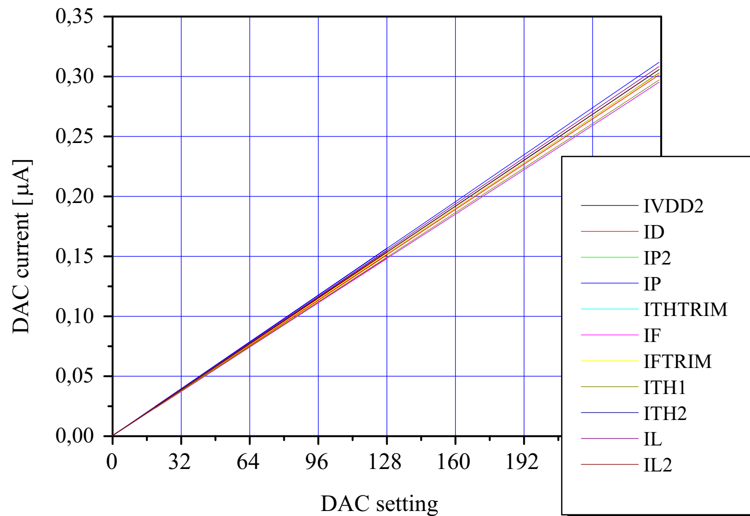


Figure 2.94: Measured characteristics of all 8-bit DA converters in a chip.

Fig. 2.94 shows the output currents generated by 9+2(spare) 8-bit DA converters versus their DAC settings. The currents was measured by using the $MonDAC$ pin. DAC slopes fluctuate due to the mismatch between the current mirrors used to generate the

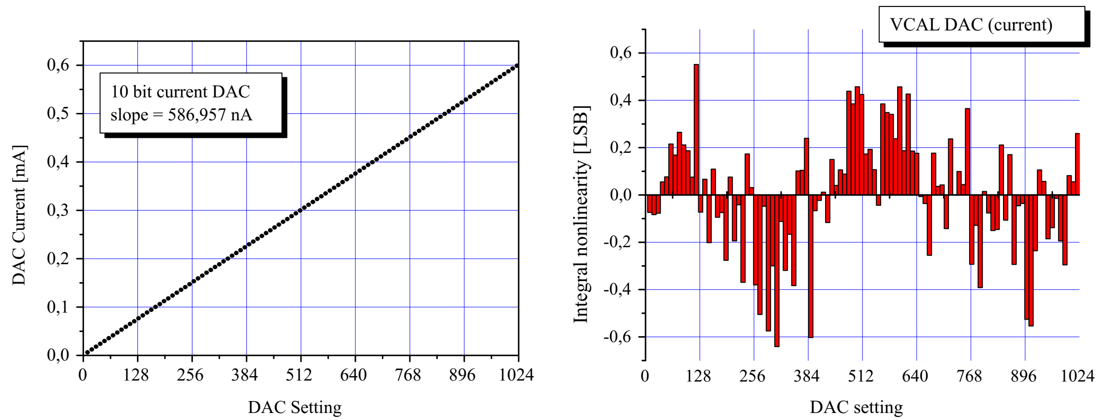


Figure 2.95: Characteristic of the 10-bit VCAL DAC (current input), left. Integral linearity, right.

bias voltages $PBias$, see fig. 2.82 (the mirrors are not shown). The current mirrors are placed in each DAC and they are biased by the voltage generated by the current reference. The slope fluctuation is not problem.

Fig. 2.95 shows the characteristic of the 10-bit VCAL DAC. The output current of this converter has been measured by using the $MonDAC$ pin. In the same figure, left, integral linearity has been shown. The integral linearity is defined as the deviation of the measured values from the linear fit. The difference is divided by the slope of the fit. It should not be higher than 0.75 DAC steps.

Fig. 2.96 shows the mean value of the pixel thresholds as function of the 7-bit tune DAC setting. The curves was measured for different GDAC values. Compare with the theoretical function in fig. 2.53.

Fig. 2.97 illustrates the functionality of the circuit for the active voltage drop compensation. The I_P and I_{L2} currents was measured in a pixel column. In fig. 2.97, left, the compensation ia switched off. We can clearly see that the bias currents drop when we move further from the bottom of the top of the column. When the compensation is switched on, the systematic change of the bias current disappears. A statistical bias current mismatch occurs due to the random offset of the amplifiers used in the compensation circuit, fig. 2.67.

Fig. 2.98 shows the results of a threshold scan. Before tuning, the chip had a threshold dispersion of $\sigma = 562$ e. The threshold map is homogenous - the threshold dispersion is the result of the statistical transistor mismatch. After threshold tuning performed by the autotune circuit, the threshold dispersion has been reduced to $\sigma = 49$e.

Fig. 2.99 shows the measured ENC in a front-end chip. The noise has been determined by the error function fit.

Fig. 2.100 shows a measurement with ^{241}Am radioactive source. A chip bonded with a pixel sensor has been irradiated. The spectrum, left, has been obtained by using the ToT data. From the hit frequency map we can recognize the area, where the radioactive source was placed.

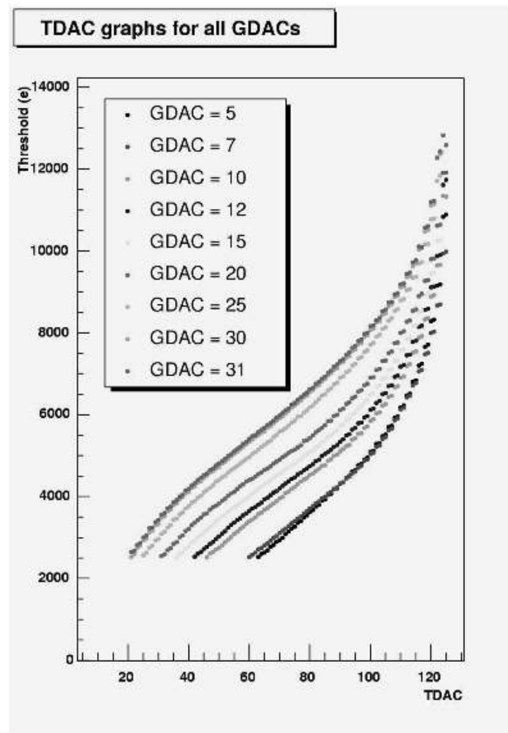


Figure 2.96: Mean pixel threshold versus TDAC setting.

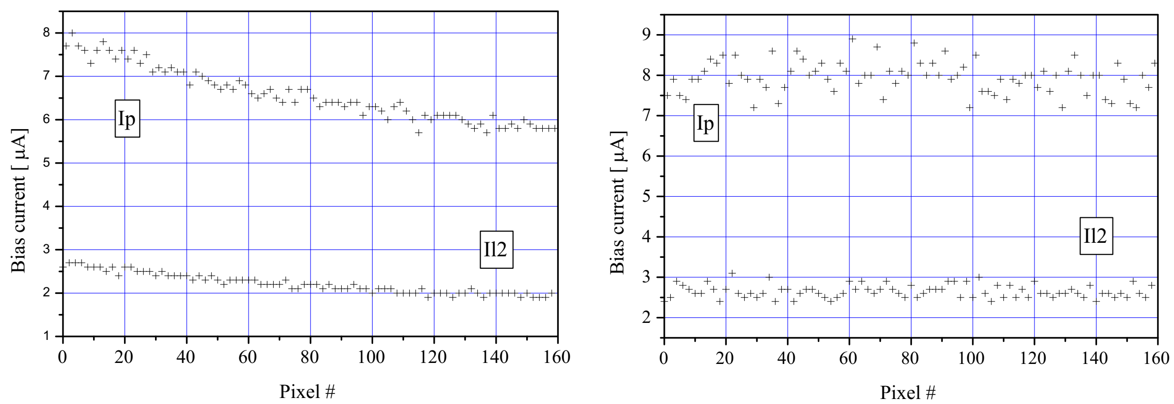


Figure 2.97: I_P and I_{L2} bias currents measured in a pixel column. Voltage drop compensation is switched off, left. Voltage drop compensation is on, right.

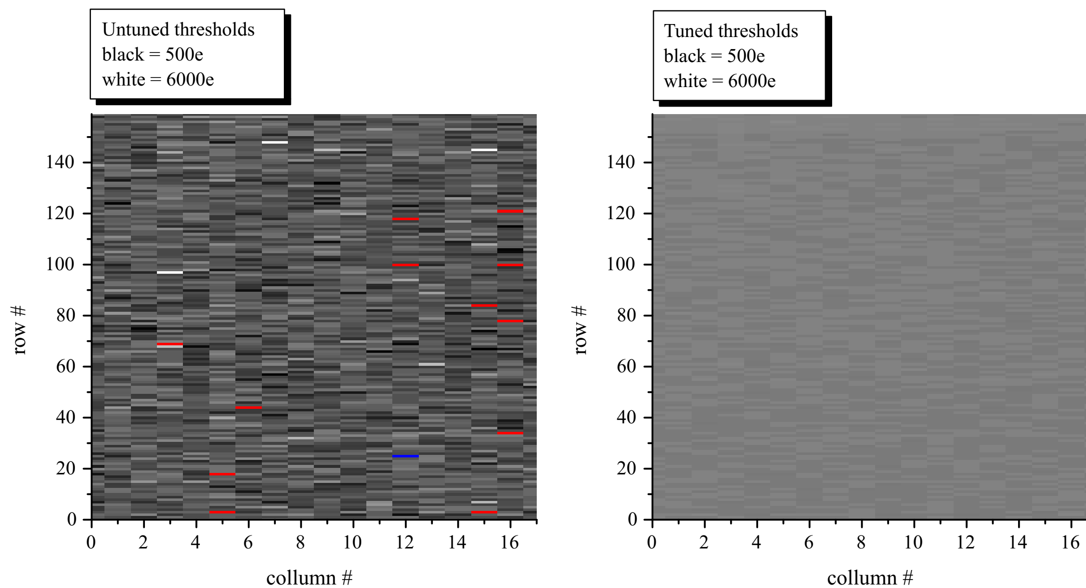


Figure 2.98: Pixel thresholds obtained by the threshold scan. Right, untuned chip. Mean threshold was 2327 e and the threshold dispersion was $\sigma = 562$ e. Right, chip tuned by the autotune circuit. Mean threshold is 3196 e and the threshold dispersion is $\sigma = 49$ e.

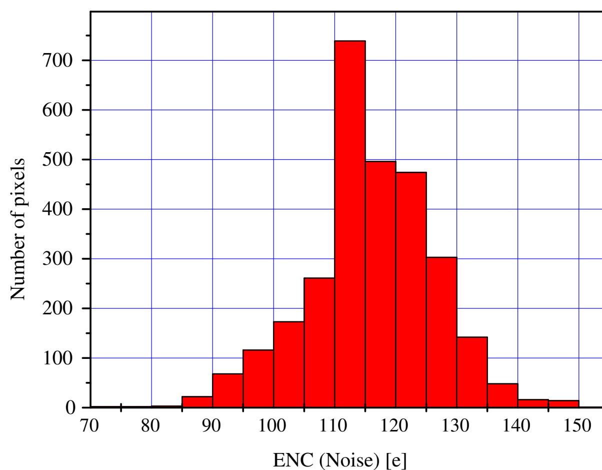


Figure 2.99: Noise of a tuned chip without sensor.

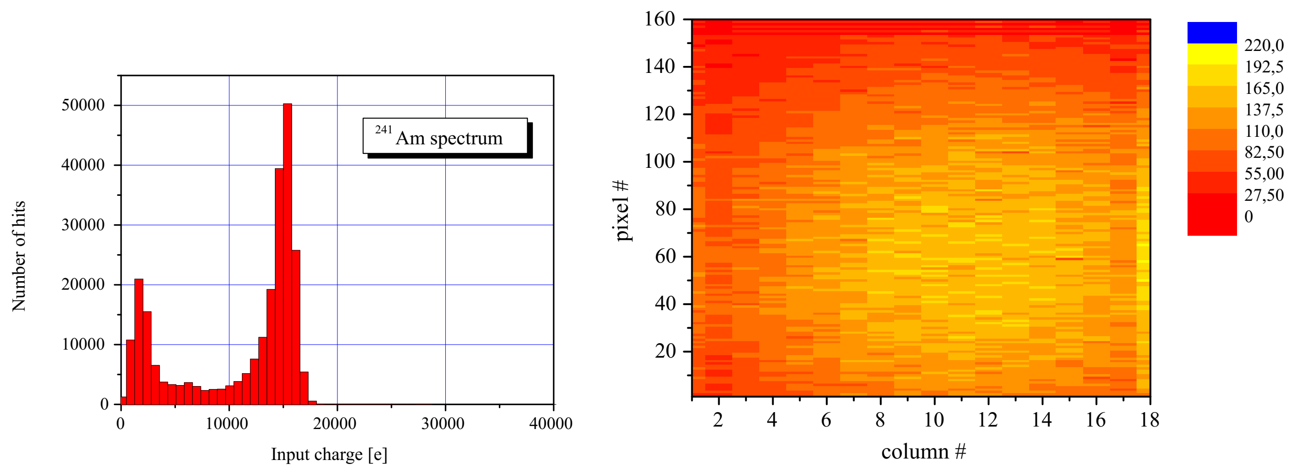


Figure 2.100: A measurement with ^{241}Am radioactive source. Spectrum is measured by using the ToT information, left. Hit frequency, right.

Chapter 3

Fast Sequencer-Chip for a DEPFET Pixel-Sensor

In this chapter we will explain the design of a high-voltage digital chip used to generate the fast steering sequence necessary for the readout of the DEPFET sensor matrix.

3.1 DEPFET Pixel-Sensor

DEPFET [5, 36, 37, 38, 39, 40] is a silicon pixel sensor with simple amplifiers integrated on the sensor substrate. A pixel amplifier consists of a single transistor. The integration of amplifiers on the sensor substrate leads to a very good signal-to-noise ratio. The active material is n^--type low-doped bulk that can be totally depleted by applying of negative voltages between the p-type regions located on both sides of the bulk and the bulk [5], see fig. 3.1 (b). Such depletion technique is called sidewards depletion. The sidewards depletion leads to the formation of a potential minimum for electrons inside the bulk. PMOS pixel transistors are implemented on the bulk, see fig. 3.1 (b). In every DEPFET pixel, typically, $1\,\mu$m below the gate-oxide, there is a deep n implantation that shapes the potential minimum for electrons. This region is called the internal gate. The electrons generated by a particle hit drift to the internal gate of the closest pixel and stay there. The principle of the signal charge amplification can be explained in the following way. The potential of the transistor gate relative to the potential of the interface between the gate-oxide and bulk can be expressed by the formula for the plate capacitor $V_{GI} = -|Q_G|/C'_{ox}$. Q_G is the gate charge, C'_{ox} is the gate capacitance. The gate charge Q_G is the mirror charge of the positive channel charge Q'_{ch} and the positive depleted region charge Q'_{dep}. Therefore $V_{GI} = -(Q'_{ch} + Q'_{dep})/C'_{ox}$. The collection of negative charge in the internal gate leads to the effective decrease of the positive bulk charge Q'_{dep}. If we suppose that the potential of the interface between the gate-oxide and bulk does not change significantly after generation of the negative charge, the decrease of Q'_{dep} is accompanied by the equivalent increase of the positive channel charge Q'_{ch}. The absolute value of the transistor threshold voltage is in this way effectively decreased. If we apply a drain-source voltage, the transistor current increases.

The internal gate has to be freed from electrons before every measurement. It can be done by applying a positive voltage to a specially designed n^+-type *clear* electrode, fig. 3.1 (c). Recall that any n^+-doped area in an n^--doped bulk forms a potential minimum for

electrons. However, it should be avoided that the electrons generated by radiation always flow into the *clear* contact instead into the internal gate. To avoid the unwanted clearing, the n^+-doped *clear* region is separated from the surrounding bulk by a p^+ region. Above the p^+ region there is a $clear - gate$ electrode that alleviates the clear procedure, fig. 3.1 (c). Let us assume that the *clear* electrode and the $clear - gate$ contact are at low potential with respect to the surrounding n^--type bulk. The p^+ region, which surrounds the clear electrode, acts then as a potential barrier and prevents that the electrons drift into the *clear* contact and get lost. If we, however, apply a large positive voltage between the $clear - gate$ and the bulk, a conducting n-type channel is formed between the n^+ clear region and the internal gate. Similarly if the *clear* potential becomes high with respect to the bulk potential, the p^+ region gets totaly depleted. Its insulating action then vanishes as the result of the punch-through effect. Clearing of the internal gate by using only the $clear - gate$ voltage would not be perfect since the electrons can flow in both direction, from and to the internal gate. On the other hand, clearing by using only the *clear* voltage requires quite a high-voltage. Because of that, the both voltages are simultaneously used for effective clearing.

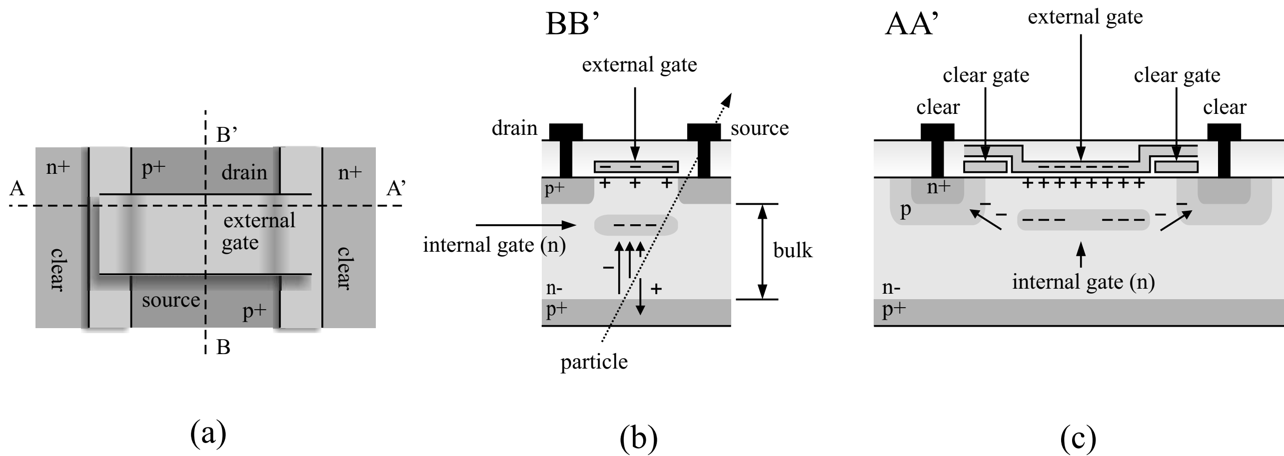


Figure 3.1: (a) - Pixel of the DEPFET detector, top view. (b) - BB' cross section; signal generation and charging of the internal gate. (c) - AA' cross section; clear mechanism.

Readout of the DEPFET Matrix

In this subsection we will discuss the readout of a DEPFET matrix. We will consider the drain readout. The drains of all transistors in a pixel column are connected together, see fig. 3.2. The gates and $clear/clear - gate$ contacts are connected row-wise. The vertical drain lines are received by the current readout chip, called CURO. The horizontal voltages are generated by two digital sequencer chips - switchers. The block scheme of a DEPFET matrix, connected to CURO and two switcher chips is shown in fig. 3.2. A possible readout sequence looks as follows. The DEPFET sensor is irradiated while the *gate*, *clear* and $clear - gate$ voltages are zero. The electrons, generated by radiation, drift to the closest potential minima and are collected in the internal gates. The thresholds of the PMOS transistors with charged internal gates decrease. After irradiation, the DEPFET matrix is read out row-wise. A switcher chip selects all PMOS transistors in a row by setting their *gate* potentials low with respect to the bulk potential, see fig. 3.2. It is important

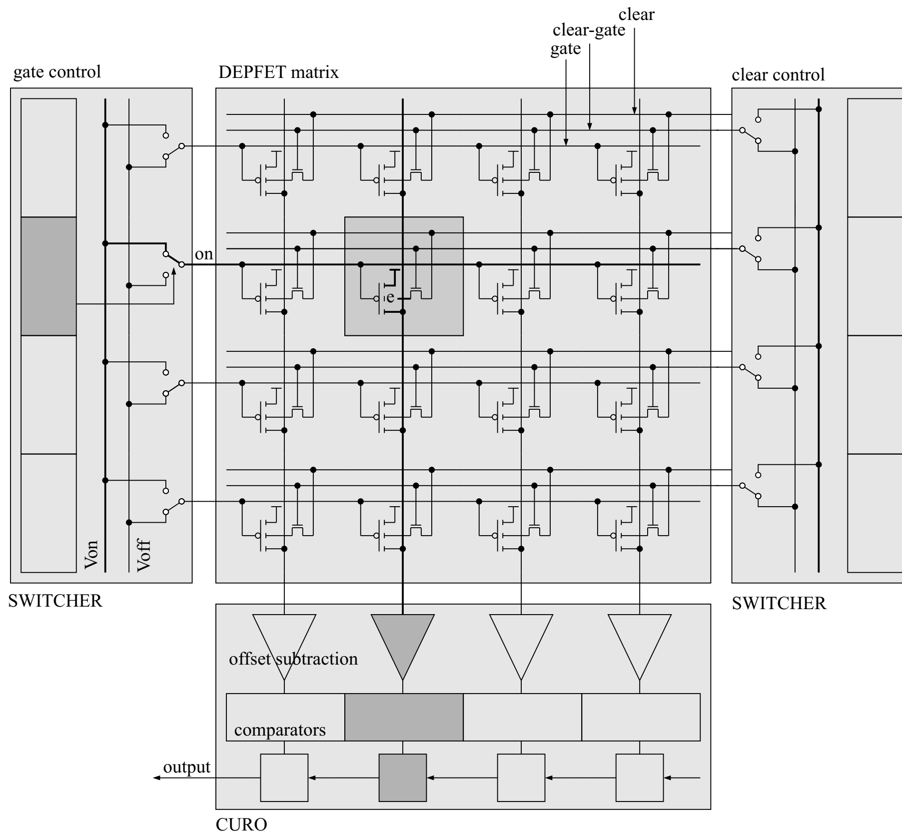


Figure 3.2: DEPFET sensor matrix connected to readout chips.

to note that the charging of an internal gate cannot switch on the PMOS transistor having external gate potential zero. The threshold change caused by the collection of electrons is not large enough to activate a transistor. Therefore only the drain currents of the PMOS transistors in the selected row flow through the vertical DEPFET lines. The PMOS transistors in unselected rows are deep in subthreshold region and their currents are negligible. The vertical drain lines are received by the CURO chip. The current values are sampled on the chip as V_{GS} voltages of diode connected transistors. The threshold change induced by charging of an internal gate is generally very small and a drain current of a transistor with full internal gate has a large offset part, called the pedestal current and a much smaller signal current. Also, the pixel to pixel offset variation can be larger than the signal. Because of that, the next step in the DEPFET row readout is measurement of the offset currents. First, the internal gates have to be emptied. A switcher sets the $clear$ and the $clear - gate$ lines at a high potential. When the internal gates become free from electrons, only the offset currents flow through the vertical lines. The CURO chip receives the offset currents and subtract them from the sampled total currents. The current differences correspond to pure signals. These signals are then compared with a threshold and a digital hit pattern is stored on the chip. The whole sequence repeats for the other pixel rows until the whole DEPFET matrix is readout.

DEPFET Detector for the TESLA Linear Collider

The DEPFET detector would be a good choice for the vertex detector in the new linear collider experiment [37]. In order to meet the required resolution of primary and secondary vertices, the innermost detector layer has to be placed at a radius of 15 mm from the interaction point. The pixel detector at this place has to provide a spatial resolution of less than $5\,\mu$m. It implies the pixel size of about $25 \times 25\,\mu\text{m}^2$. To avoid the multiple scatterings, the material introduced in detector has to be minimized ($\approx 0.11\% X_0$). This means no extra material for cooling pipes *etc.* is allowed and, additionally, the sensor substrate must be thinned to $50\,\mu$m thickness. This results in reduced signal charges compared to that generated in standard $300\,\mu$m thick silicon detectors. Due to its inherent low noise, the DEPFET detector can cope with such small signals. The new TeV-energy superconducting linear accelerator, called TESLA, will generate, approximately, 1 ms long "trains" of electron-positron bunches. The train separation will be 200 ms. A DEPFET detector module would contain a pixel matrix with 4000 rows and 520 columns. The readout chips should be able to take 25 frames per train crossing time (1 ms) in order to suppress the high occupancy caused by e^+ e^- particles from bremsstrahlung. To increase the readout speed, the DEPFET detector is read out from the both sides in parallel by using two CURO chips. There would be 20 ns time for the readout of a pixel row. It puts tough requirements on the design of the CURO and switcher chip. Assuming a simple readout sequence, shown in fig. 3.3, CURO must be able to sample the currents with 10 ns period, whereas the switcher must be able to set the correct $gate$, $clear$ and $clear - gate$ voltages in less than 10 ns. The large parasitic capacitances of DEPFET row lines limit the speed of the steering sequence. The DEPFET detector utilizing PMOS transistors is a new device and the optimal values for control voltages are still subject to research. The amplitudes are assumed to be lower than 20 V. The best possible clear sequence is still not determined, hence the switcher chip has to be as versatile as possible. Since standard CMOS chips operate with 5 V supply, the amplitudes of the $gate$, $clear$ and $clear - gate$ pulses can be considered as high-voltage (HV). The chip must be implemented using a special HV technology.

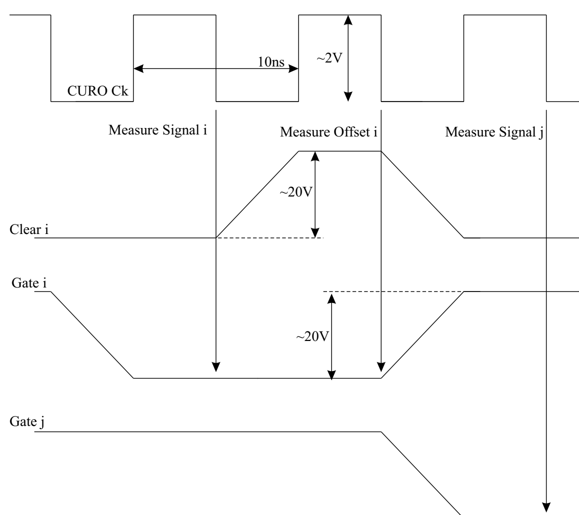


Figure 3.3: Signal Diagram.

3.2 Architecture and Design of the Switcher Chip

In this section we will explain the design of the switcher chip.

3.2.1 Main Blocks of the Switcher Chip

The architecture of the switcher is shown in fig. 3.4. The chip contains 64 high-voltage

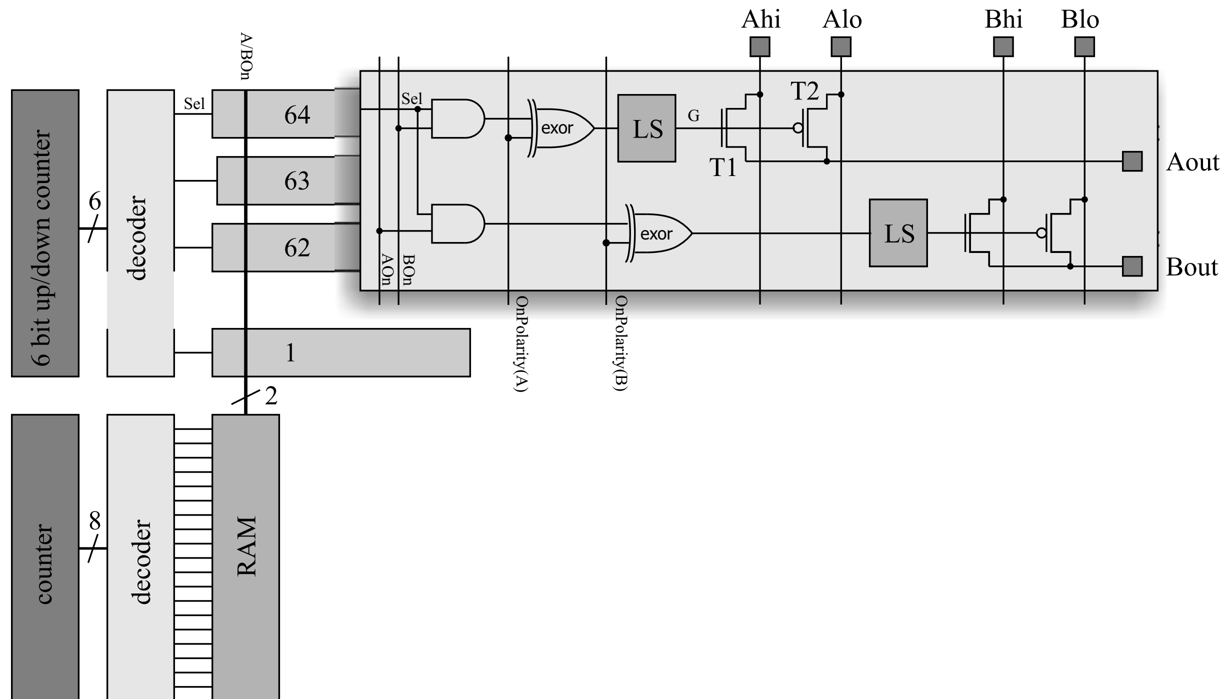


Figure 3.4: Block scheme of the switcher chip.

channels used to control 64 DEPFET rows. Every channel has two identical voltage outputs, denoted in fig. 3.4 as *Aout* and *Bout*. The output *Aout* can be used, for example, to generate the *gate* signal and the output *Bout* to generate the *clear* voltage.[1] The switcher channels can be activated sequentially by using a 6-bit counter and a decoder. Only one channel can be activated at a time. Every time when a channel is selected, a programmable switching sequence is repeated. An example of such a sequence is shown in fig. 3.3: after channel activation the *gate* output is low while the *clear* output is high. A clock period later the *clear* output becomes high. Such, or a more complicate sequence, can be stored in the random access memory (RAM). A RAM cell contains two bits which control the polarities of both outputs. These bits are referred to as *Aon* and *Bon*. There are 256 cells in the RAM, which determines the maximal length of a sequence. The RAM is addressed with an 8-bit counter and a decoder. In order to meet the TESLA requirements, it must be possible to readout the RAM at 100 MHz frequency, see *clear* pulse in fig. 3.3.

The detailed block scheme of a switcher channel is also shown in fig. 3.4. The off and on high-voltage levels, used to steer the DEPFET matrix, are generated outside the switcher

[1] In order to control all three signals independently (*gate*, *clear* and *clear* − *gate*) two switchers have to be used.

chip and distributed on the chip via four vertical lines. The high and low potentials for the *Aout* outputs are denoted as *Ahi* and *Alo* and the high and low potentials for the *Bon* outputs *Bhi* and *Blo*. If the output *Aout* should be set at the low potential *Alo*, the block denoted as LS (level shifter) sets the control signal G high. This switches on the NMOS transistor T1 and switches off the PMOS T2, fig. 3.3. Similarly if *Aout* should be high, the signal G is set low. This connects *Aout* to *Ahi*. The transistors in the high-voltage parts of a channel are designed in a spacial way. They can withstand 0-20 V potential variations at their drain and gate terminals with respect to the bulk potential. The maximal potential difference between the high- and low-voltage inputs (for example *Ahi* and *Alo*) is therefore 20 V. Note that the G signal must have an amplitude at least equal to the high-low difference in order to assure that only one transistor, T1 or T2, is on at a time. A level shifter (LS) converts an input low-voltage signal into a high-voltage output signal that drives the gates of the transistors T1 and T2. The purpose of the EXOR gate is to define the "on-polarity". For example, the *gate* signal is active low, whereas the *clear* and $clear - gate$ signals are active high. At the beginning of this section we have mentioned that only a channel can be activated at a time. The outputs of the channels that are not active are automatically set at their off levels. The polarity signals are stored as two bits in a chip register.

3.2.2 Synchronization of many Switcher Chips

A switcher chip contains 64 channels. For the readout of DEPFET matrices with larger number of rows, many switchers must be used. These chips must be synchronized. This is achieved by using an output $NextChip$ signal and two input signals called $StartFromAbove$ and $StartFromBelow$, fig. 3.5. In a column of switcher chips we can recognize a bottom chip a top chip and middle chips. All these chip types behave in slightly different ways. The chip position is internally coded by setting the corresponding $IsBottomChip$ and $IsTopChip$ bits, see fig. 3.5. Counting direction can be defined for each chip separately.

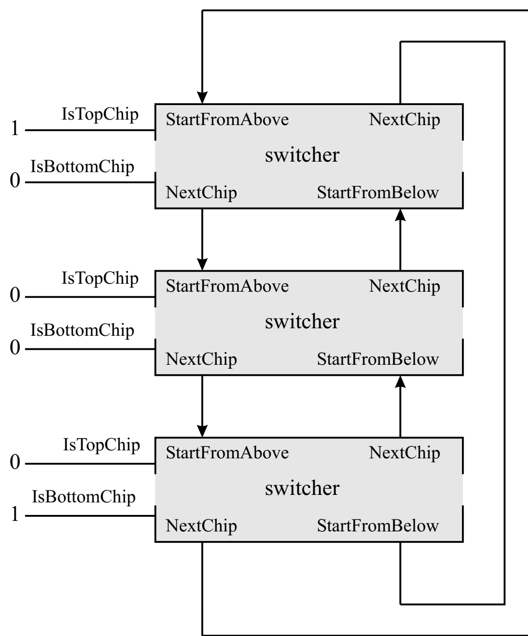


Figure 3.5: Connection of many switcher chips.

This feature is mandatory since a switcher placed on the right side of the DEPFET matrix is rotated by 180^0 with respect to the switcher on the left side, fig. 3.2. These chips should count in the opposite directions. If the $CountUp$ bit is set to 1, the channels are sequentially selected from the bottom to the top.

The control logic circuit used to synchronize the 8-bit RAM address counter with the 6-bit channel select counter is shown in fig. 3.6. Let us explain this circuit. The

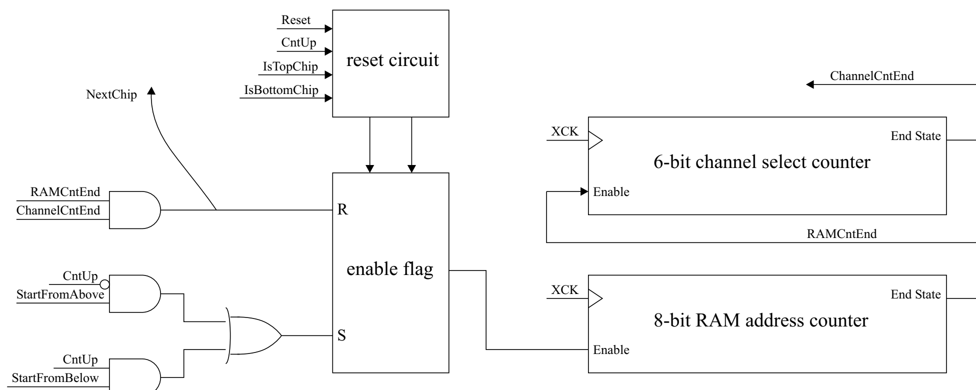


Figure 3.6: Counter control.

central part of the control circuit is a RS flip-flop - enable flag. A reset circuit, shown in the figure as a block, is used to define the initial conditions. If the counting direction "bottom-top" is chosen, an external signal *Reset* clears the enable flags in all chips except in the switcher with $IsBottomChip = 1$, *i.e.* the chip at the bottom of the chip row. The 6-bit and 8-bit counters are also reseted. The counters are driven by an external fast clock signal, called XCK. The RAM address counter counts in loop from 0 to n; n is the programmable length of the sequence. Every time when a sequence is over, the RAM address counter sends an enable signal ($RAMCntEnd$) to the channel counter and a new channel is selected. When the sequence in the last row is finished, both counters are in end-state: $RAMCntEnd = ChannelCntEnd = 1$. The enable flag is then cleared and the chip becomes idle. At the same time the $NextChip$ signal is sent to neighboring chips. The switcher placed above receives the $NextChip$ signal as $StartFromBelow$. The switcher placed below receives $NextChip$ as $StartFromAbove$, see fig. 3.5. Assuming the counting from bottom to top, the $NextChip$ signal sets only the enable flag of the switcher placed above. The whole counting sequence repeats on that chip. The block scheme of the counters is shown in fig. 3.7.

3.2.3 High-Voltage Part

High-Voltage Transistors

The transistors in the high-voltage parts must be high-voltage tolerant. The switcher chip has been implemented in AMS 0.8 μm CMOS technology that offers special high-voltage devices beside standard 5 V transistors. Fig. 3.8 shows the cross sections of a standard and a high-voltage NMOS transistor as comparison. High-voltage transistors are usually much larger than standard transistors. High-voltage tolerance is achieved in the following way. First, drain and source n^+ layers are embedded in lower doped n-wells, fig. 3.8 (a). The n-well at the drain side increases the resistance between the drain contact and

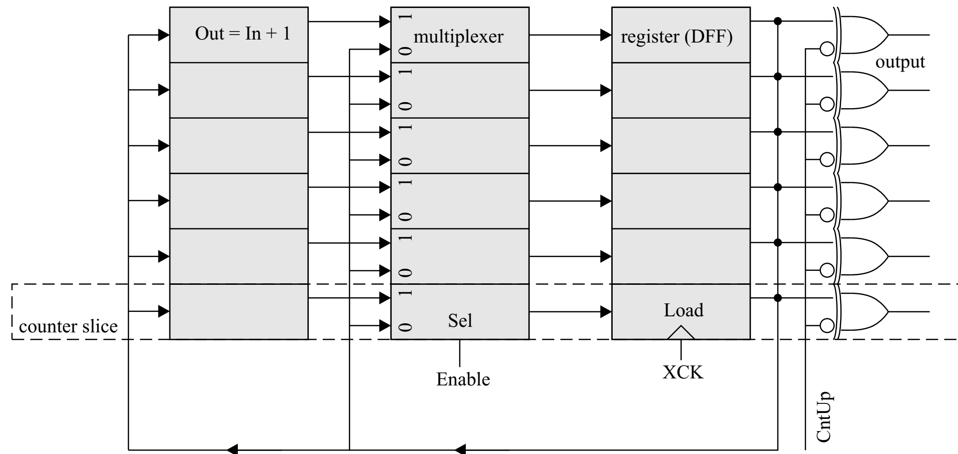


Figure 3.7: 6-bit counter. A counter bit (counter slice) is the unit cell of the counter. The slice based architecture allows an easy layout design and a simple increase of the bit number if necessary.

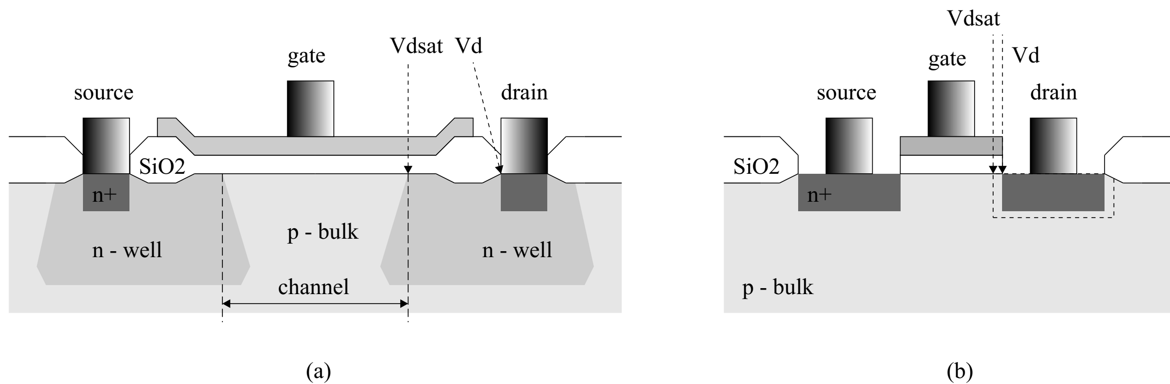


Figure 3.8: (a) - high-voltage NMOS transistor. (a) - standard NMOS transistor.

the channel area. This introduces a voltage drop and reduces the longitudinal electric field. In the case of a standard NMOS transistor, potential changes from V_d to V_{dsat} in a very narrow depleted region (typically $150nm$) close to drain, fig. 3.8. If a standard transistor operates deep in saturation ($V_d \gg V_{dsat}$), the longitudinal electric field in the drain depleted region and the vertical electric field in the thin oxide above it are high. Scattered high-energy electrons penetrate in the $Si0_2$ and damage it. This effect is called the hot-carrier effect. The use of thicker gate-oxide in the areas close to the drain/source n^+ regions (fig. 3.8) reduces the vertical electric field in the oxide. The breakdown voltage of n-well to p-substrate junctions is higher than the breakdown voltage of n^+ to p-substrate diodes.

Transconductance of High-Voltage Transistors

High-voltage transistors have a smaller current gain (transconductance) when compared with the standard transistors having the same gate width. This makes HV transistors slower if they are used to charge a capacitance. Let us explain this.

The formula for drain-source current in saturation is [1]

$$I_{DS} = \frac{\mu C'_{ox}}{2n}\frac{W}{L}(V_{GS} - V_{th})^2 \ . \tag{3.1}$$

μ is the channel carrier mobility, C'_{ox} is the oxide capacitance per unit area, n is the slope factor, W and L are the gate width and length respectively and V_{th} is the threshold voltage. The transconductance is defined by the equation

$$g_m \equiv \frac{\partial I_{DS}}{\partial V_{GS}} = \frac{\mu C'_{ox}}{n}\frac{W}{L}(V_{GS} - V_{th}) \ . \tag{3.2}$$

The minimal gate length in the case of a high-voltage transistor is always larger than the minimal length of a standard gate. The reason for this becomes clear when we look at fig. 3.8. Due to transversal dopant diffusion, the actual n-well area is always larger than the n-well mask size. The n-well mask separation and therefore the gate length, must be large enough in order to avoid the short between the drain and source wells. According to eq. 3.2, a larger L leads to a smaller current gain. Also, larger dimensions of high-voltage transistors imply larger parasitic capacitances.

The technology used to implement the switcher chip offers many variations of high-voltage transistors. The transistor structure shown in fig. 3.8 allows biasing of both terminals, source and drain, at a high potential with respect to the bulk and gate potentials. Sometimes it is enough to have only one high-voltage terminal. In that case we can use a device which is a combination between the fully high-voltage transistor and the standard one. For example, we can have only an n-well on the drain side, whereas source is implemented as a standard n^+ layer. Also, the gate-oxide thickness above the channel can be larger to allow higher gate-bulk voltages.

It is interesting to compare two high-voltage transistors with different gate-oxide thicknesses in terms of their current gains and current capacities. The oxide capacitance is given by the the formula $C'_{ox} = \epsilon_{SiO2}/t_{ox}$, where tox is the oxide thickness. A transistor with greater oxide thickness has a smaller oxide capacitance. From eq. 3.2 we can see that the transconductance g_m depends linearly on C'_{ox}. Clearly, assuming the same $V_{GS} - V_{th}$, a transistor with thinner oxide has a higher gain that that having thicker oxide. It is, however, more useful to compare the maximal current values that can be achieved by using transistors with different gate-oxide thicknesses. The vertical electric field in gate-oxide can be approximated by the formula[2] $E_y \approx (V_{GS} - V_{th})/t_{ox}$. Using the last result, we obtain $(V_{GS} - V_{th})_{max} \approx E_{crit} t_{ox}$. The maximal transconductance is therefore, approximately, independent on t_{ox} since $(V_{GS} - V_{th})_{max}$ increases linearly with t_{ox} and the capacitance C'_{ox} is inversely proportional to the oxide thickness. The maximal saturation current given by eq. 3.1 is, however, higher in the case of a transistor with a thicker oxide. Due to its higher maximal current, a HV transistor with thick gate-oxide layer is more suitable for the fast charging of large capacitances. Such transistor type is used to implement the transistors T1 and T2 in switcher channels, fig. 3.4. To repeat, these transistors charge the control lines of a DEPFET detector which represent high capacitive loads.

[2] The correct expression is $t_{ox} E_y = (Q'_{ch} + Q'_{dep})/C'_{ox}$. If the gate-source voltage is large enough, we can neglect the depleted zone charge Q'_{dep} with respect to the channel charge Q'_{ch}. Since we have $Q'_{ch} = C'_{ox}(V_{GS} - V_{th})$ [1], the electric field is $E_y \approx (V_{GS} - V_{th})/t_{ox}$.

Level Shifter

The basic scheme of a level shifter is shown in fig. 3.9. The transistor terminals that are denoted with thick lines are high-voltage tolerant. We should keep in mind that the high-voltage drain and source contacts have large parasitic capacitances with respect to the bulk. The output NMOS and PMOS transistors (T1 and T2 in fig. 3.4) have thick gate-oxide. The other transistors, which make the interface to low-voltage circuits, have thin gate-oxide.

The differential pair, fig. 3.9, left, is controlled by two differential low-voltage signals *Up* and *Down*. The signals, which are generated by standard CMOS circuits, swing between the low-voltage supply potentials $Vddd$ and *Gndd*. The voltage between $Vddd$ and *Gndd* is not higher than 5 V. The value of the *Gndd* potential with respect to the bulk potential *Gnd* does not influence the functionality of the level shifter. Signal between the differential pair and the high-voltage part of the level shifter are distributed as currents Iup and $Idown$. Free choice of the low-voltage supply potentials with respect to the bulk potential was a design requirement. The functional principle of the circuit shown

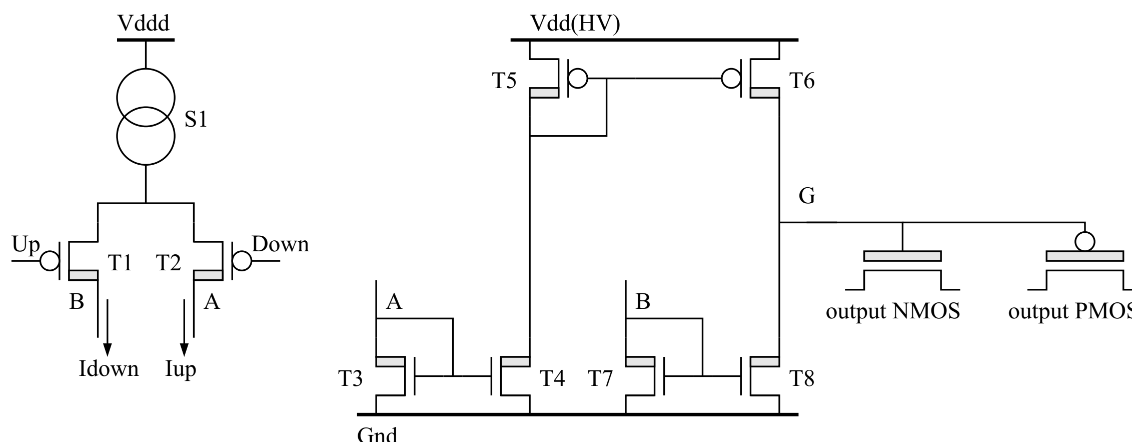


Figure 3.9: Level Shifter.

in fig. 3.9 is quite simple. If the *Up* signal becomes high and the *Down* signal low, the current generated by the source S1 flows through the net A and the transistors T2, T3, T4, T5 and T6 become switched on. The node G is charged by the transistor T6 until the potential of the node (V_G) reaches the $Vdd(HV)$ value. If *Up* is low and *Down* high, the current generated by the source $S1$ flows through the transistor T1 in the node B. The transistors T7 and T8 become switched on. The node G is then discharged through the transistor T8.

A disadvantage of the simple level shifter (fig. 3.9) is that the circuit needs quite a long time to charge the large parasitic capacitance of the node G. Recall that the switcher must generate the *clear*, *gate* and $clear - gate$ pulses with a rise time less than 10 ns. The output NMOS and PMOS transistors (T1 and T2 in fig. 3.4) must be large enough in order to achieve high enough current capacity needed for the fast charging of the DEPFET electrodes. The gate capacitances of the output transistors represent a large capacitive load for the transistors T6 and T8. We can make the circuit faster if we increase the current generated by the source S1 and/or if we increase the current multiplication factors between the transistors pairs (T7, T8) and (T5, T6). Here, we are

limited by two things. The transistors T6 and T8 should not be too large, otherwise their drain-bulk capacitances dominate the total parasitic capacitance of the node G. The current generated by the source S1 must not exceed the value when the gate-source voltages of the transistors T3 and T7 become higher than 5 V (the maximal allowed gate-source voltage of a transistor with thin oxide). To increase the speed of the level shifter, we can use that fact that the transistors with thick oxide can generate a larger maximal current. The modified circuit is shown in fig. 3.10. We added an additional

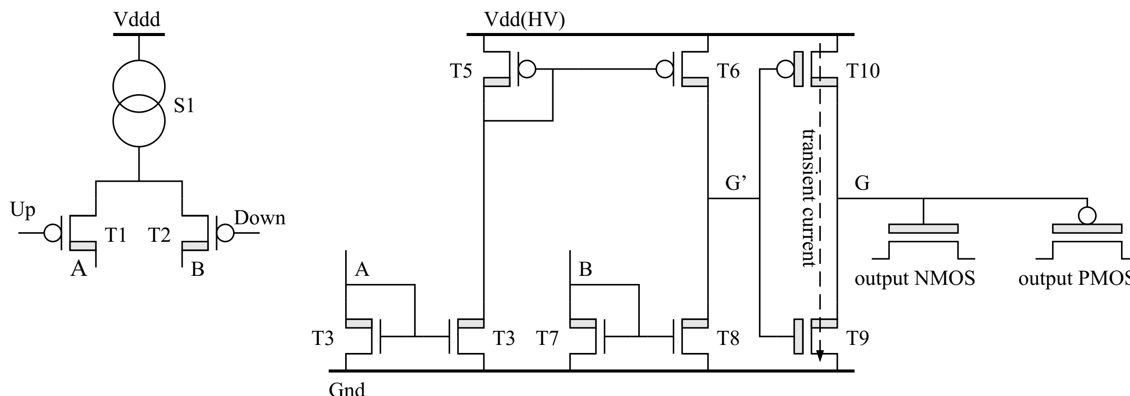


Figure 3.10: Level Shifter with an additional inverting stage.

inverting stage that is implemented using the thick-oxide transistors T9 and T10. This inverter drives the gates of the large output transistors. The transistors T9 and T10 are much smaller than the output transistors, which decreases the load driven by T6 and T8. The circuit shown in fig. 3.10 is considerably faster than that shown in fig. 3.9. However, there is a problem. Assume that the potential of the node G is Gnd. The NMOS T9 is then switched on whereas the PMOS T10 is off. Suppose now that we want to charge the node G from the Gnd to the $Vdd(HV)$ potential. It would be ideal, first, to switch off the NMOS transistor T9 and after that to switch on the PMOS T10 in order to avoid the transient short connection between $Vdd(HV)$ and Gnd. However, the circuit in fig. 3.10 does the opposite. The PMOS T10 is first switched on and the NMOS T9 becomes off later when the voltage between G' and Gnd drops below the NMOS threshold. A part of the current generated by the PMOS T10 flows unnecessary into the NMOS T9 instead to charge the gate capacitance G. This transient effect increases current consumption and decreases speed. To avoid these problems, a new scheme of level shifter is made. The new circuit prevents that both transistors T9 and T10 conduct current at the same time. The block scheme of the circuit is shown in fig. 3.11 and its MOSFET implementation in fig. 3.12. In the case of the classical inverter scheme (fig. 3.10) at least one of the inverter transistors T9 or T10 is always on. In the case of the circuit in fig. 3.11, both inverter transistors T1 and T2 become switched off when the charging of the node G is finished. A transistor needed for the charging is switched on only for a short time. When the potential V_G reaches the desired value, which is detected by the amplifiers (see fig. 3.11), the transistor that charged the node is switched off. Because of that, in the case of the circuit in fig. 3.11, there is no possibility that both transistors T1 and T2 are at the same time on. The signal diagram is shown in fig. 3.11.

The circuit works in the following way. Assume that the signal *Down* is for a longer time high. The current generated by the source S1 flows through the net A and supplies

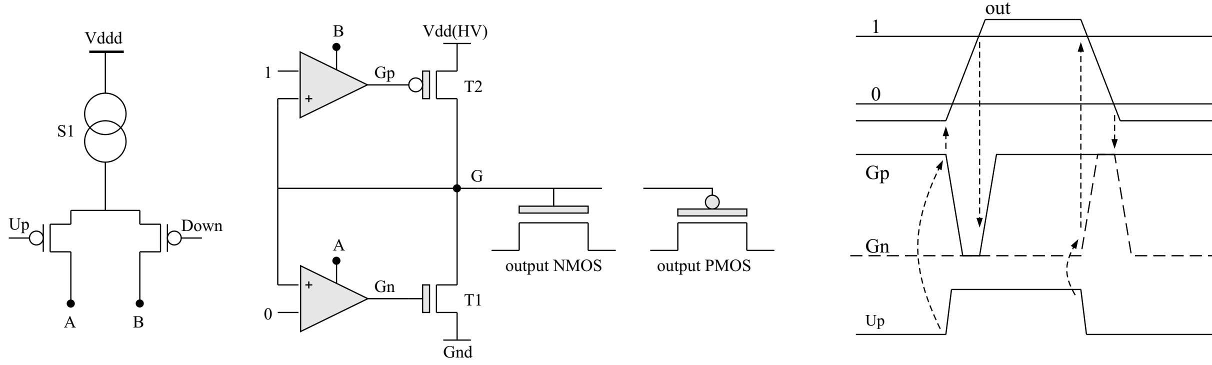


Figure 3.11: Circuit that prevents a transient short between $Vdd(HV)$ and Gnd.

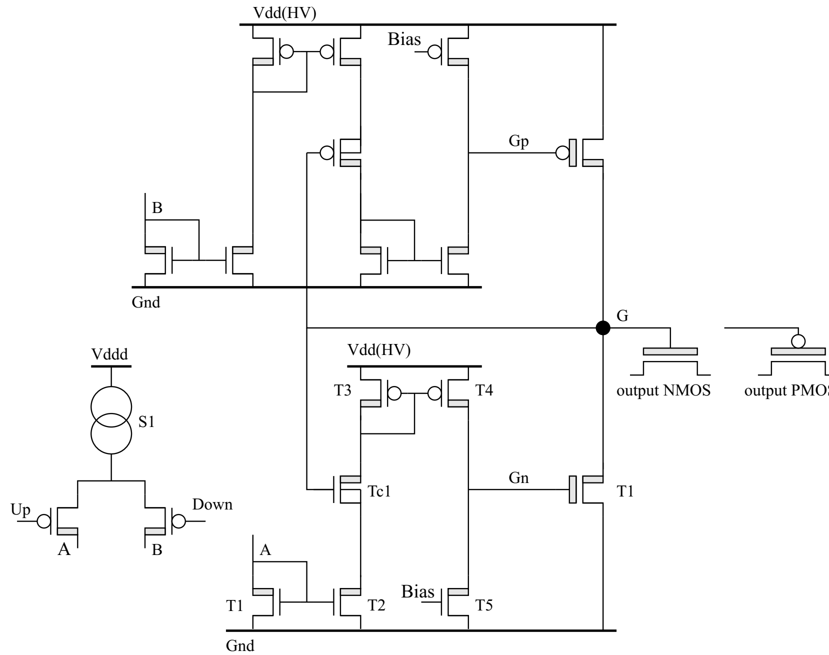


Figure 3.12: MOSFET implementation of the circuit of fig. 3.11.

the amplifier connected to the NMOS transistor T1 (node Gn). The other amplifier is off and its output is in high impedance state. The potentials of nodes denoted as 1 and 0 correspond to the V_G levels needed to safely switch on and off the output transistors. The potential of the node 0 is a little bit higher than the Gnd potential and the potential of the node 1 is a little bit lower than the $Vdd(HV)$ potential, see signal diagrams in fig. 3.11. Provided $Down$ is for a longer time high, we can assume that V_G is equal to V_{Gnd}. The node G is therefore at a lower potential than the node 0. The amplifier connected to T1 sees a negative input voltage and it pulls the gate of T1 (node Gn) at the Gnd level. Note, both amplifiers are supplied by $Vdd(HV)$ and Gnd. The amplifiers are assumed to have a large gain. If we set $Down$ low and Up high, the differential pair biases the amplifier connected to T2 (node Gp). The other amplifier becomes switched off and its output goes into high impedance state. The potential of the node Gn does not change since it is "stored" on a large gate capacitance. At the first time V_G is lower than V_1. The amplifier connected to T2 sees a negative input voltage and pulls the node Gp at

the Gnd potential. The transistor T2 becomes switched on and it charges the node G. When the V_G potential exceeds the potential of the node 1, the input voltage seen by the amplifier connected to T2 changes its polarity and the node Gp is pulled at the $Vdd(HV)$ potential. The PMOS T2 is then switched off.

The NMOS implementation of the amplifiers is quite simple, fig. 3.12. We will explain only the amplifier used to bias the NMOS transistor T1. If the potential V_G is high with respect to ground, the transistor Tc1 operates as cascode transistor. In the first approximation, Tc1 can be considered as a short connection. The current that flows into the node A is then simply mirrored two times through the transistors T1, T2, T3 and T4. The transistor T4 charges the node Gn until V_{Gn} becomes nearly equal to the $Vdd(HV)$ potential. The transistor T1 is then switched on and the node G is discharged. When the voltage between the nodes G and Gnd drops below the threshold of the transistor Tc1, the transistor T2 becomes switched off.[3] The current stops flowing through the transistors T2, Tc1, T3 and T4 and the transistor T5, biased by a constant voltage, discharges the node Gn. The transistor T1 becomes switched off.

Design of Low-Voltage Digital Transistors

The low-voltage parts in the switcher chip have separate supply nets $Gndd$ and $Vddd$. The voltage between $Vddd$ and $Gndd$ must be lower than 5 V. However, the relative potential level of these supply lines with respect to the *p*-substrate potential can be between 0 and 20V - the current coupling between the low-voltage part and the level shifter assures the proper chip functionality. All low-voltage transistors are placed in a large common *n*-well that is biased at the $Vddd$ potential. The NMOS transistors are inside *p*-wells, biased at the $Gndd$ potential. The *p*-wells are placed in the large *n*-well. The *n*-well is processed in the special way (deep *n*-well). It can withstand a high well-substrate voltage.

3.2.4 Sequencer

The block scheme of the RAM sequencer and the address decoder is shown in fig. 3.13. Precharge logic elements are used to implement the NAND gates in the address decoder, fig. 3.13. Such logic gates are also called domino logic gates. Domino logic gate can generate only a low impedance logic 0 at its output. If the result of logic function should be 1, a domino gate puts its output in high impedance state. Such type of logic circuit needs a precharge cycle, where the output is set at a high potential. The precharge must be performed before the calculation of logic function. The advantage of domino gates in comparison with CMOS gates is a smaller number of transistors and hence a smaller output parasitic capacitance. This increases the switching speed. The RAM output bus and the domino logic gates are precharged to the $Vddd$ level when the clock signal XCK becomes active. The RAM sequencer has been successfully tested up to 90 MHz readout speed.

[3] The transistor T2 is switched off in a rather unusual way. Usually the drain-source current is controlled by the gate potential. Here we control the transistor current by changing the drain potential by using the other transistor in common gate configuration.

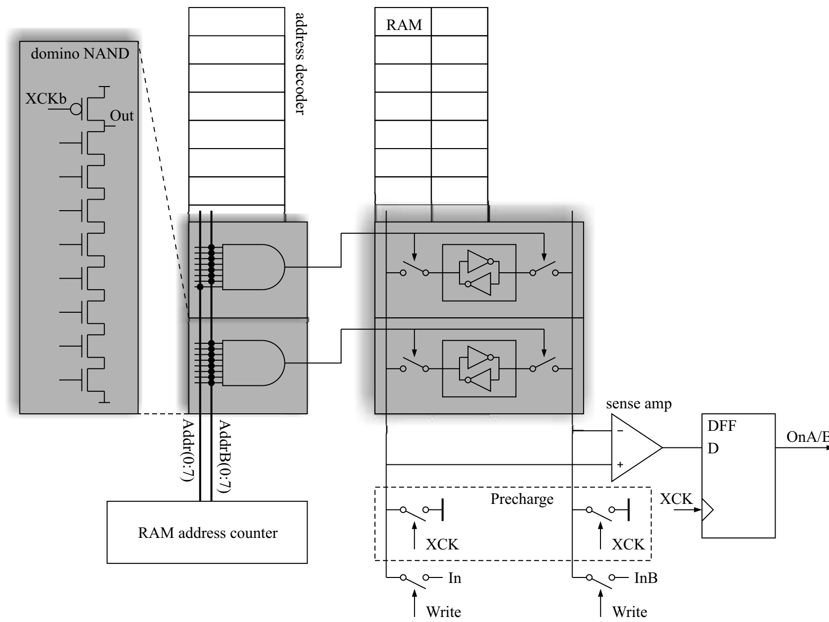


Figure 3.13: RAM sequencer and address decoder.

3.2.5 Measurements

Fig. 3.14 shows a *clear* pulse generated by the switcher channel 0 and a *gate* pulse generated by the channel 1. The RAM test sequence was: on, off, on, on. The measurement verifies the functionality of the channel select counter, RAM and the control circuits that synchronize the main switcher blocks. Note the opposite polarities of the two signals. The oscilloscope probe had a capacitance of about 10 pF. The rising time from 0 to 20 V is slightly less than 20 ns. Fig. 3.15 shows the layout and photomicrograph of the switcher

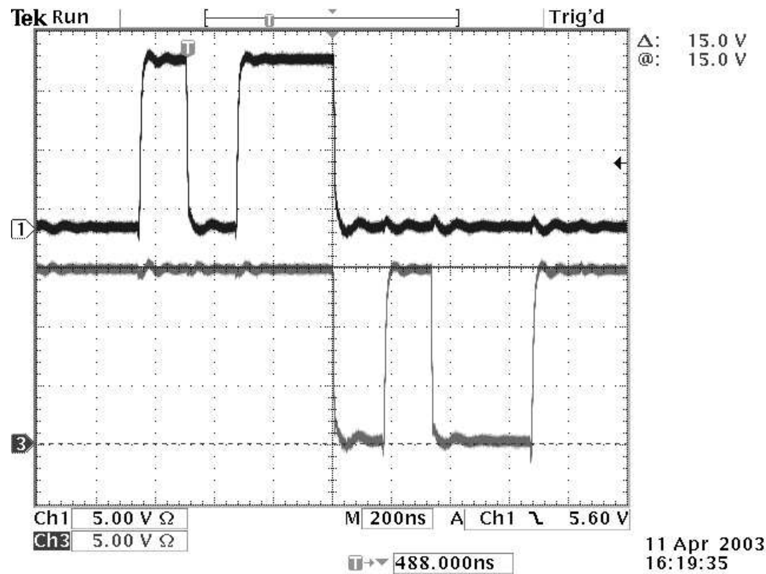
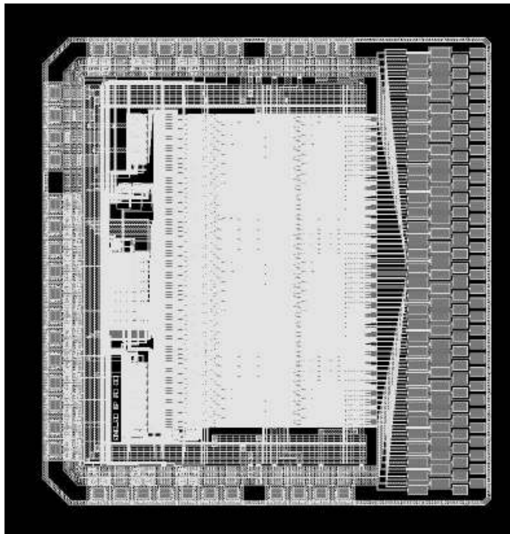


Figure 3.14: Outputs of two neighboring channels.

chip.



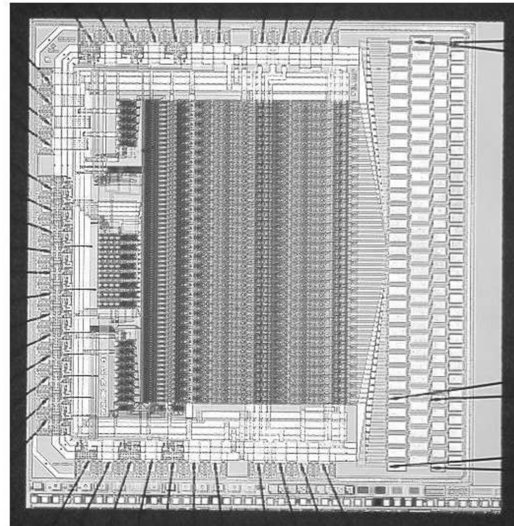


Figure 3.15: Layout and photomicrograph of the switcher chip.

Chapter 4

X-Ray Imaging Pixel-Chip with Joint Counting and Integrating Capabilities

4.1 Counting and Integrating

In the chapter 2 we described the design of a hybrid pixel detector which has application as tracking particle detector. A similar hybrid pixel detector can be also used as imaging device [18, 22].

In the case of X-ray imaging (radiography), an object is irradiated with an X-ray source. The task of an imaging detector is to measure the transmitted radiation in the way that allows identification of object's internal structure. X-ray sensitive film is an example of classical imaging detector. Transmitted X-ray energy is usually first converted into fluorescent light by using an intensifying screen. The fluorescent light causes ionization of a light sensitive substance in a film emulsion, for example, silver-iodo-bromide. The electrons, generated in this way, initiate certain local chemical reactions, for example, buildup of metallic silver. After a development process, the chemically changed areas become optically dark.

An alternative implementation of X-ray imaging detectors involves the use of semiconductor radiation detectors segmented into pixels. X-ray photons can be absorbed and converted into light by using a fluorescent material and the light detected by using semiconductor sensors. The other possibility is to convert the X-ray photons directly into electric signals by using the semiconductors with a high atomic number. The electric signals produced in every sensor pixel can be, for example, counted by using digital counters. In this way, an image with digital contrast information can be obtained.

We will discuss a hybrid pixel X-ray detector. Pixel-readout chips with sensor-signal amplifiers and sensors are produced separately. Pixel pads on the chip- and on the sensor-side are bonded using flip-chip technique. The size of basic electronic components - transistors - is today much smaller than the typical size of a pixels used in X-ray imaging. It gives us the possibility to implement different types of signal processing circuits in pixels of a readout chip and test their performances.

Integration of Sensor Signal

The classical approach in the design of a pixel-readout chip for imaging applications is to implement in every pixel a circuit that collects and measures the charge generated in a sensor pixel during the exposition time. Note, the charge collecting is equivalent to the mathematical integration of the current input signal. A possible implementation of the electronic circuit that performs integration of the input current is shown in fig. 4.1. The

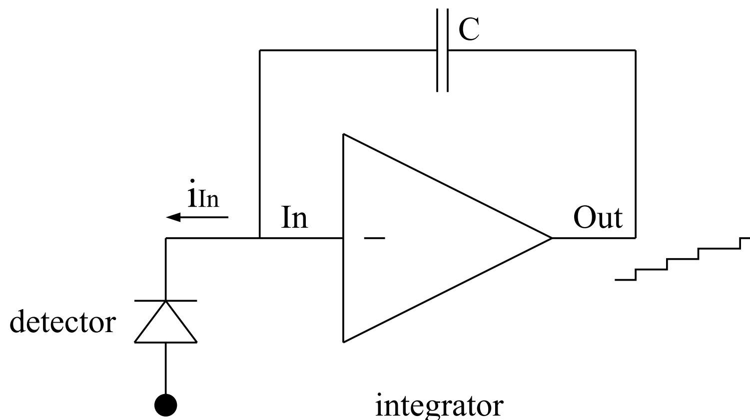


Figure 4.1: Electronic integrator.

output potential of the active integrator (fig. 4.1) is equal to:

$$v_{Out}(t) = \frac{1}{C}\int_0^T i_{In}(t)dt + const\ .$$

C is the feedback capacitance, T is the frame period. The input potential is kept at a constant level by action of the capacitive feedback. Therefore the circuit performs the integration of the input current and, at the same time, acts as a bias electrode for the sensor pixel. The total collected charge and hence the output voltage signal are proportional to the radiation energy deposited in a pixel. Note, in the case of an X-ray sensitive film, blackening of the sensitive emulsion is also proportional to the total deposited energy. An advantage of an electronic detector in comparison with film is much larger dynamic range of the signal that can be precisely measured. The blackening of film is nearly linearly proportional to the absorbed energy only in a small dose range, maximally, two orders of magnitude [62]. A higher absorbed dose leads to overexposure and a lower dose to underexposure of the film. Contrast is in both cases lost. On the other hand, the electric signal generated by the integrator shown in fig. 4.1 can be measured by using the techniques that allow high precious measurement over a wide signal range, typically, more than four orders of magnitude, see [48, 47]. Picture contrast will be therefore preserved in a much larger dose range than in the case of the X-ray film.

Photon Counting

An interesting alternative to the sensor-current integration is the counting of the charge packets produced in a sensor by photon hits, [41, 42, 43, 44, 45]. We call this technique "photon counting". Fig. 4.2 shows a possible implementation of the pixel electronic for

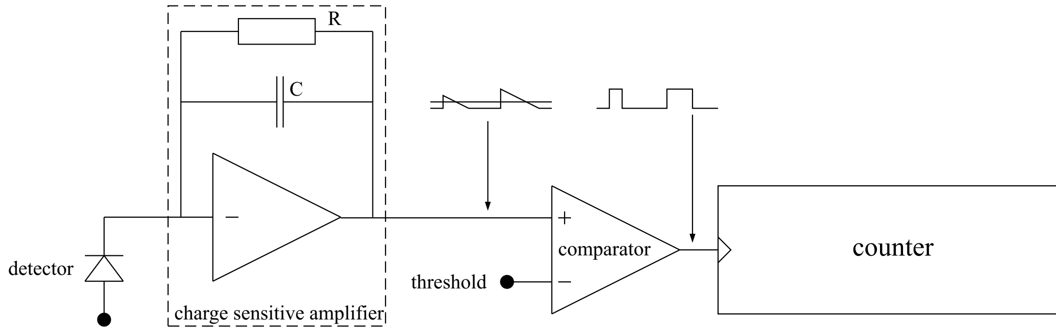


Figure 4.2: Photon counting.

photon counting. A charge-sensitive amplifier has the structure similar to that of the integrator shown in fig. 4.1. The feedback capacitance of a charge-sensitive amplifier is usually very small. This leads to a high charge gain, a charge pulse generated by a single photon hit produces a clear detectable signal at the output of the amplifier. The resistor R clears the feedback capacitor C after integration of a charge pulse, fig. 4.2. The signal generated at the output of the amplifier is compared with a threshold. If the threshold is exceeded, a counter is incremented. The dynamic range of a photon counting detector is practically infinite since the counter can be read out during exposition. In the case of a small radiation flux per pixel, when single photon signals can be easily recognized and counted, counting detectors have better accuracy than the detectors with integrator. One of the reasons for this is that photon counting has already at the input of the system a better signal-to-noise ratio than the measurement of deposited energy. Let us briefly explain this interesting fact.

The number of photons that hit a pixel during a certain exposition time can be calculated as

$$N = \int_{E_{min}}^{E_{max}} N(E)dE \ . \tag{4.1}$$

The corresponding mean value is

$$\overline{N} = \int_{E_{min}}^{E_{max}} \overline{N(E)}dE \ . \tag{4.2}$$

$N(E)dE$ and $\overline{N(E)}dE$ are the actual and mean numbers of the incident photons that have an energy between E and $E + dE$, respectively. E_{min} and E_{max} are the minimal and maximal energies of incident photons. If we assume that the actual number of photon hits follows a Poisson statistics, the variance of the number of photon hits is[1]

$$Var(N) = Var\left(\int_{E_{min}}^{E_{max}} N(E)dE\right) = \int_{E_{min}}^{E_{max}} Var(N(E)dE) = \int_{E_{min}}^{E_{max}} \overline{N(E)}dE = \overline{N} \ . \tag{4.3}$$

The noise-to-signal ratio for photon counting is

$$(n_{cnt}/s_{cnt})^2 = \frac{Var(N)}{\overline{N}^2} = \frac{1}{\overline{N}} \ , \tag{4.4}$$

[1]The variance of a Poisson random variable is equal to its mean value. Therefore $Var(N(E)dE) = \overline{N(E)}dE$.

where we used the result given by eq. 4.3.

The energy deposited by *all* photons, which hit a pixel, is given by the following equation

$$E_{all} = \int\limits_{E_{min}}^{E_{max}} EN(E)dE \ . \qquad (4.5)$$

The corresponding mean value is

$$\overline{E_{all}} = \int\limits_{E_{min}}^{E_{max}} \overline{EN(E)dE} = \int\limits_{E_{min}}^{E_{max}} E\overline{N(E)}dE \ . \qquad (4.6)$$

The variance of the energy given by eq. 4.5 is

$$Var(E_{all}) = Var\left(\int\limits_{E_{min}}^{E_{max}} EN(E)dE\right) = \int\limits_{E_{min}}^{E_{max}} E^2 Var(N(E)dE) = \int\limits_{E_{min}}^{E_{max}} E^2\overline{N(E)}dE \ . \quad (4.7)$$

We assumed a Poisson statistics. The noise-to-signal ratio for energy measurement is

$$(n_{int}/s_{int})^2 = \frac{Var(E_{all})}{\overline{E_{all}}^2} = \frac{\int_{E_{min}}^{E_{max}} E^2\overline{N(E)}dE}{\overline{E_{all}}^2} \ . \qquad (4.8)$$

In order to compare the noise-to-signal ratios in the case of photon counting and energy measurement let us first derive a few auxiliary formulas. The mean energy of a single incident photon is

$$\overline{E_{sing}} = \int\limits_{E_{min}}^{E_{max}} Ew(E)dE \ . \qquad (4.9)$$

$w(E)dE$ is the probability that an incident photon has an energy between E and $E + dE$, which is $w(E) = \overline{N(E)}/\overline{N}$. Therefore

$$\overline{E_{sing}} = \int\limits_{E_{min}}^{E_{max}} Ew(E)dE = \frac{1}{\overline{N}} \int\limits_{E_{min}}^{E_{max}} E\overline{N(E)}dE = \frac{\overline{E_{all}}}{\overline{N}} \ . \qquad (4.10)$$

The variance of the single photon energy is

$$Var(E_{sing}) = \int\limits_{E_{min}}^{E_{max}} (E - \overline{E_{sing}})^2 w(E)dE = \int\limits_{E_{min}}^{E_{max}} E^2 w(E)dE - \overline{E_{sing}}^2 \ . \qquad (4.11)$$

By substituting of the result $w(E) = \overline{N(E)}/\overline{N}$ in the last equation, we obtain

$$Var(E_{sing}) = \frac{1}{\overline{N}} \int\limits_{E_{min}}^{E_{max}} E^2\overline{N(E)}dE - \overline{E_{sing}}^2 \ , \qquad (4.12)$$

or after some rearrangement

$$\left(\frac{Var(E_{sing})}{\overline{E_{sing}}^2} + 1\right)\frac{1}{\overline{N}} = \frac{\int_{E_{min}}^{E_{max}} E^2 N(E)dE}{(\overline{NE_{sing}})^2} = \frac{\int_{E_{min}}^{E_{max}} E^2\overline{N(E)}dE}{(\overline{E_{all}})^2} \ . \qquad (4.13)$$

By comparing of last equation with eq. 4.4 and eq. 4.8, we have

$$\left(\frac{Var(E_{sing})}{\overline{E_{sing}}^2}+1\right)(n_{cnt}/s_{cnt})^2=(n_{int}/s_{int})^2\ . \tag{4.14}$$

The variance $Var(E_{sing})$ is a positive number. Therefore the noise-to-signal ratio in the case of energy measurement must be larger than the noise-to-signal ratio in the case of photon counting, see [62].

Counting and Integrating

Assuming a low photon flux, photon counting is a very precise method for the measurement of the transmitted radiation. However, in the case of a high flux per pixel, where individual photons cannot be distinguished any more, the integration of the sensor current remains the only choice. An interesting way to increase the dynamic range of pixel readout electronics is to implement a circuit that integrates and counts at the same time [49]. The counting part works better in the case of lower photon flux, while the integrating part covers the region, where the counting is not possible. The enhanced dynamic range is not the only benefit of the simultaneous charge collection and the photon detection. The region, where both techniques work, is especially interesting. The sensor current integration gives an output signal proportional to the total deposited energy in a pixel. If we divide this information by the number of photons, we get the measure of the mean photon energy. This can simplify the recognition of different structures obtained by X-ray radiography. Two different objects can absorb the same amount of X-ray energy, which leads to the same response of an energy sensitive detector. However, the spectral distributions of transmitted radiation can be different, which can be detected by a counting-integrating detector. In the next few sections we will describe the design of a counting-integrating pixel-readout chip.

4.2 Architecture and Design of the CIX Chip

CIX is an acronym for counting-integrating X-ray. In the first subsection we will briefly introduce the design of the integrating part.

4.2.1 Integrating Block

To simplify the discussion which follows, we will assume that an electronic integrator receives a constant current signal. In this case, the integrator output potential increases as a linear time function. In order to simplify the readout of measured data, the analog signal generated at the output of the integrator (fig. 4.1) should be digitalized *i.e.* converted into a number. The analog-to-digital (AD) conversion can be performed in a few different ways, see [47]. From the theoretical point of view, the simplest solution would be to freeze the the output voltage of the integrator after integration in some period T_{int} and after that perform the AD conversion. The digitization can be understood as comparison of the output potential with a scale containing equidistant levels, fig. 4.3. The absolute error due to output voltage quantization is then constant and equal to the distance between levels. The relative error of the output voltage measurement increases when the signal

becomes smaller. This is illustrated in fig. 4.3. The figure shows the integrator output signals that correspond exactly to the equidistant voltage levels used in the analog-to-digital conversion. As can be seen, the slope difference between the neighboring lines is the same regardless of the input signal amplitude. In the CIX chip we have used a

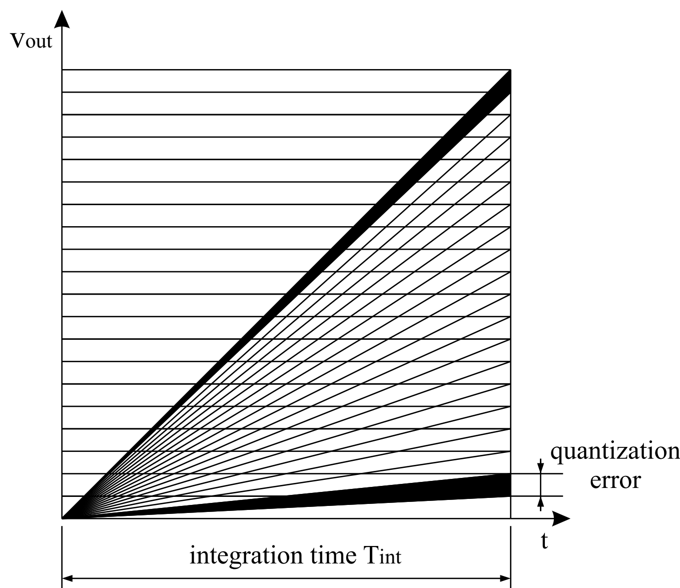


Figure 4.3: AD conversion of the output signal of integrator. The output signals that correspond to the discrete voltage levels are shown.

different technique of the analog-to-digital conversion having the property that a relative quantization error remains nearly constant over the whole signal range. Let us explain this briefly.

The basic idea used on the CIX chip is to perform quantization two-dimensionally, which means to divide the amplitude and the time axis into equidistant segments as shown in fig. 4.4. Analog to digital conversion is performed in the following way. An electronic device determines the highest discrete amplitude level that is reached by the integrator output during a certain integration period T_{int}. This device is denoted as device 1 in fig. 4.5 and it can be implemented as follows. A control signal is generated every time when the potential at the output of the integrator exceeds a discrete level. This is done by the device denoted as multi-threshold comparator in fig. 4.5. The control signals are used to increment a counter, denoted as counter 1 in fig. 4.5. The counter state corresponds then to the number of the highest discrete level exceeded by the integrator output potential. Another electronic device, denoted as device 2, measures the moment when the last discrete level has been exceeded. The time is measured in equidistant steps T. The time measuring device can be implemented as a counter, whereby the counter state is incremented by an external clock signal Ck of the period T. Every time a control signal is generated by the first device, the state of the time counter is stored in a memory element, called time latch, fig. 4.5. The last stored number is then the time information that we need. As result of the described analog-to-digital conversion, we obtain an amplitude and a time information. From the ratio between the amplitude and the time, we can calculate the input current signal. Fig. 4.4 shows the integrator output signals that correspond to all combinations of the discrete amplitude and time values.

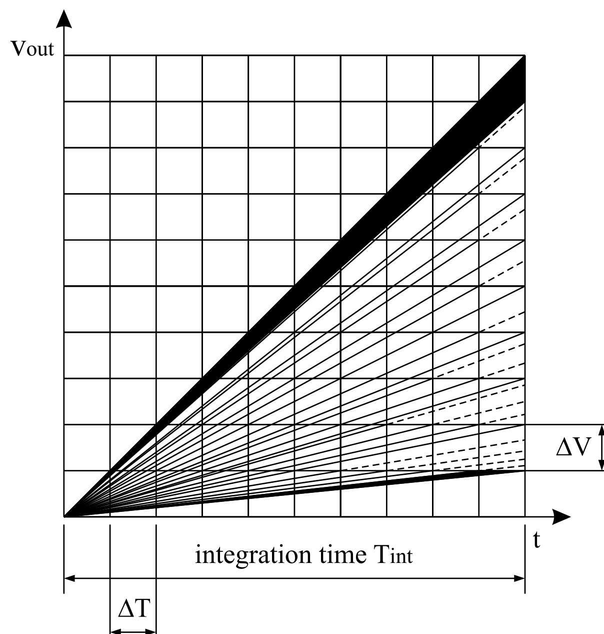


Figure 4.4: Two-dimensional AD conversion. The output signals that correspond to the discrete amplitude and time values.

The maximal signal slope is assumed to be $\Delta V/T$. As we can see in fig. 4.4, the slope difference between neighboring lines becomes smaller when the line slope decreases. The precision of the quantization is therefore higher for weaker signals. The exact calculation shows that the relative quantization error remains in average constant over the whole signal range. The smallest signal that can be measured by using this technique is equal to $\Delta V/T_{int}$. This disadvantage can be overcome if we add some known offset to the signal current.

The quantization of the integrator output potential could be performed by implementing a multi-threshold comparator, as shown in fig. 4.5. We have used a much simpler solution (fig. 4.6) which is published as patent in [48]. There is only one threshold. When the integrator output potential exceeds the threshold value, synchronous with a clock signal Ck, a control signal is generated. The capacitor of the integrator is discharged by injecting a certain known charge amount Q_0 into the input node. The control signal is used to increment the charge packet counter and to store the time information into a memory element, see fig. 4.6. At the end of the integration period we obtain two stored numbers. The number in the charge packet counter n corresponds to the number of charge packets that was injected into the integrator. The number stored in the time memory (latch) m corresponds to the moment of the last injection. If we assure that the integrating capacitor contains the same amounts of charge before and after a measurement, the mean input current can be calculated by using the following formula

$$I = \frac{nQ_0}{mT} \ . \tag{4.15}$$

Q_0 is the value of a charge packet, T is the clock period. Assuming a perfect system with constant Q_0, the only error in the measurement is the result of time quantization. The relative time-quantization error is in average constant and, approximately, equal to

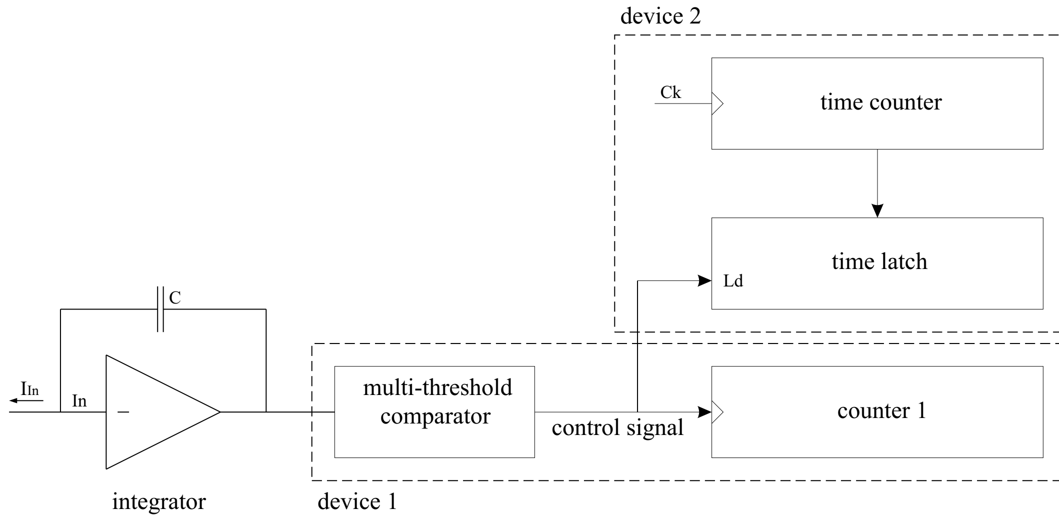


Figure 4.5: Block scheme of the circuit which performs the two-dimensional AD conversion.

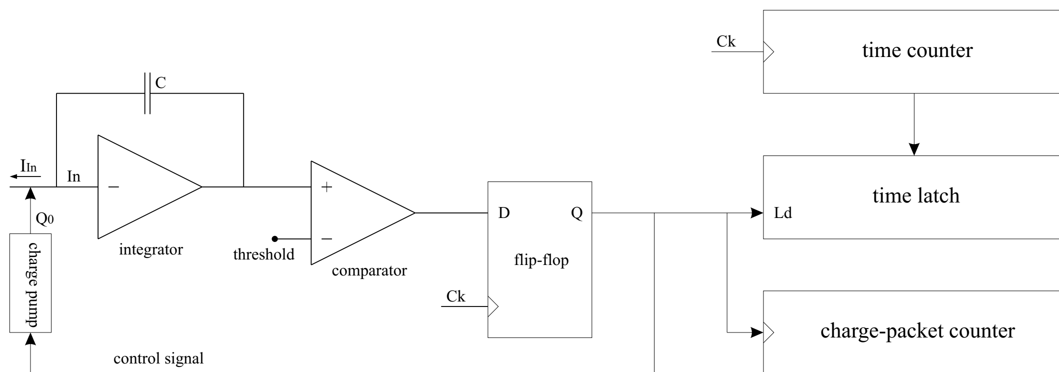


Figure 4.6: AD conversion on the CIX chip.

$1/2^N$, N is the number of bits in the time counter. The signal range is determined by the size of the charge packet counter.

4.2.2 Counting Block

The photon counting part of the CIX chip is based on the scheme shown in fig. 4.2. The transistor implementation of the amplifier is shown in fig. 4.7. We have chosen a differential amplifier with folded cascode. This type of amplifier has been introduced in the section 3.6.3. The feedback circuit is based on the circuit published in [25]. We will explain the principle of the feedback circuit considering the simplified circuit shown in fig. 4.8, although the feedback implementation used on the CIX chip slightly differs from this simplified scheme. Let us neglect for the moment the capacitor Cl and the transistor Tl which are of importance only for the leakage/offset current compensation. In the stationary state, assuming ideal conditions, the currents i2 and i1 are zero and the both transistors making the differential pair (T1 and T2) conduct the same current If. The output potential of the amplifier (node Out) is equal to the reference potential $RefOut$. Assume now that we have a very fast negative charge injection in the input node In. The

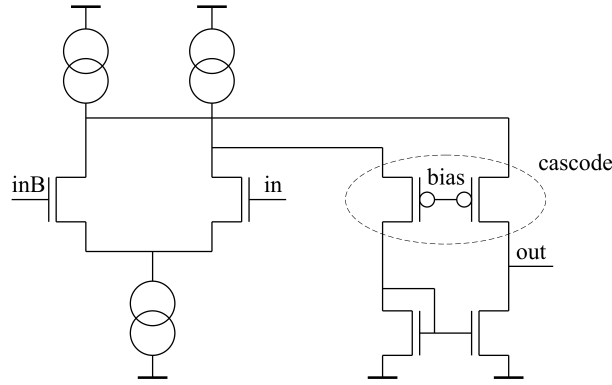


Figure 4.7: Differential folded-cascode amplifier.

charge-sensitive amplifier stabilized with a feedback containing the transistors T1 and T2 has a two-pole transfer function. The transfer function has a smaller time-constant that describes the response time of the amplifier and a larger time-constant that describes the response speed of the the feedback transistors T1 and T2.[2] Assume that the time constants are widely separated, the response time of the amplifier is much smaller than the response time of the feedback transistors. As a result of the charge injection, the output potential (node *Out*) increases until the charge flowing trough the capacitor C_f cancels the charge injected by the detector. The output potential increases by the value $\approx Q_{sig}/C_f$, $-Q_{sig}$ is the charge injected by the sensor. Since the node *Out* is connected to the gate of the PMOS T1, the current flowing through this transistor decreases. The transistors T1 and T2 are connected to the same constant current source. The decrease of the current flowing through T1 leads to an equal increase of the current that flows through T2 and an increase of the current i2. The current i2 has the direction opposite to the signal current i_{sig}. The current i2 flows so long until the effect of the signal current injection becomes cancelled, which means the current flows so long until all node voltages return to their stationary levels. If the sensor injected $-Q_{sig}$ in the node *In*, the same charge amount must be delivered by the feedback (current i2) in order to restore the stationary conditions.

Let us now explain the circuit for the sensor leakage current compensation. In the moment when the sensor bias voltage is switched on, the leakage current I_{leak} starts to flow. The potential of the node *Out* increases until the current i2 becomes equal to I_{leak}.[3] Let us look at the loop 1, fig. 4.8. There is a constant current $2If$ that flows into the loop (source S1) and two constant currents, each If, that flow out of the loop (sources S2 and S3). Assuming that the current sources are ideal, the sum current of all current sources flowing into the loop 1 is zero. The sum of the currents i1 and i2 must be also zero, assuming that the total charge inside the loop 1 does not change. Therefore $i1 = -i2 = -I_{leak}$. The current i1 charges the large capacitor Cl. The gate potential of the transistor Tl slowly decreases and the current flowing through Tl increases. The increase of the current flowing through Tl leads to an increase of the potential of the node *In* and decrease of the potential of the node *Out*. After some time the currents i2 and i1 become zero and the charging of the capacitor Cl stops. The potential of the node *Out*

[2]The exact analysis of a similar charge-sensitive amplifier is given in section 2.2.2.

[3]If the leakage current I_{leak} exceeds the maximal i2 value (If), the potential of the node *Out* increases until the amplifier saturates. The leakage current compensation works also in that case.

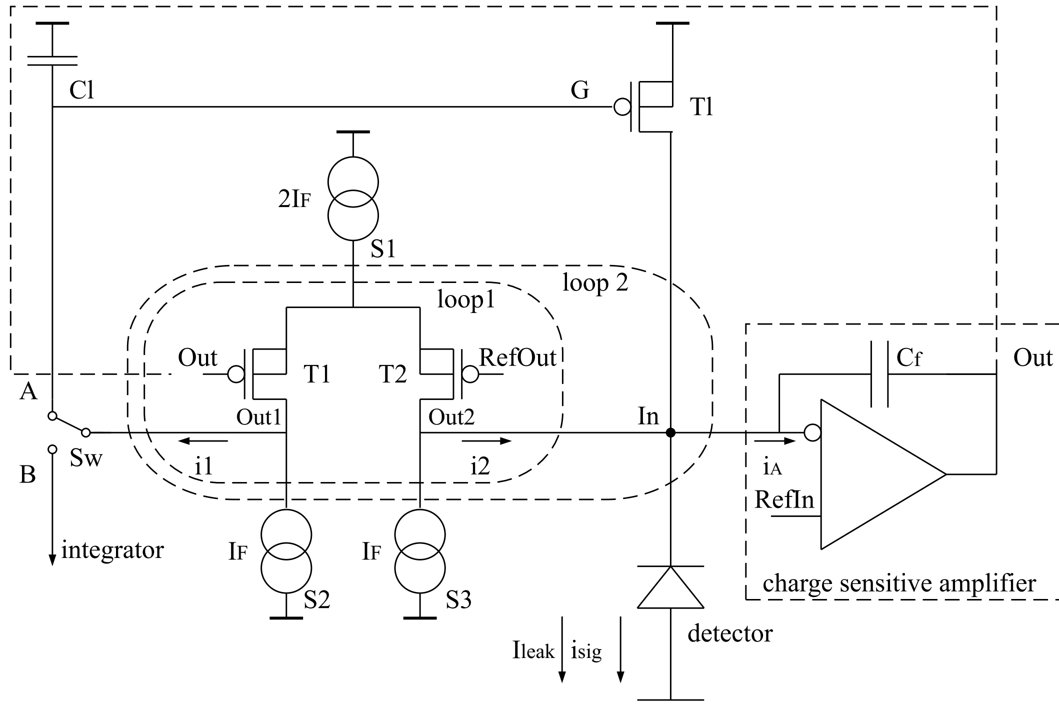


Figure 4.8: Feedback circuit of a charge sensitive amplifier.

becomes equal to $OutRef$ and the current flowing through Tl reaches the value I_{leak}. The charging of the capacitor Cl occurs with quite a long characteristic time.[4] Any possible mismatch between sources S1, S2 and S3 has the same effect as the leakage current. The current flowing through transistor Tl will be adjusted in the way that the stationary current flowing into capacitor Cl becomes zero.

4.2.3 Interface between Counting and Integrating Block

The simultaneous counting-integrating measurement can be performed as follows. The switch Sw is first set into position *A*. The circuit for the leakage current compensation compensates the leakage and offset currents. After some settling time, the current i1 becomes zero. The switch Sw is then set into position B. Such switch position allows the simultaneous use of the counting amplifier and the integrator. The potential needed for the leakage/offset compensation is "stored" on the large capacitance Cl (node G). The detector can be now irradiated. Assume that a photon hits the sensor pixel. The negative charge signal $-Q_{sig}$ is injected into the node In. The output potential (Out) increases, the feedback responds and after some time determined by the feedback time-constant, the output potential returns to its stationary value. The voltage pulse at the node Out can be used for photon counting. Let us now consider the loop 2, fig. 4.8. The currents generated by the sources S1, S2, S3 are constant as well as the current flowing through the transistor Tl. The algebraic sum of these currents plus the leakage current I_{leak} is zero. If we compare the two stationary states before and after a photon hit, the

[4]In the section about the ATLAS front-end chip chip we have calculated the time-constant of a similar circuit for the leakage current compensation. The time-constant is C_l/g_{ml}. g_{ml} is the transconductance of the transistor Tl.

charge inside the loop 2 is the same. Therefore the charge $-Q_{sig}$ injected into the loop 2 due to photon hit must leave the loop through an another terminal. The mean value of the current that flows into the amplifier i_A calculated for the whole transient period is zero. The only way for the charge $-Q_{sig}$ to go out of the loop is therefore to flow into the integrator. The current signal that flows into the integrator has a much larger time-constant than the signal generated by the sensor, however, the both signals have exactly the same integrals. Note, the only assumption we made is that the current sources have large output resistances.

The mean current that flows into the integrator cannot be higher than If. The circuit in fig. 4.8 does not work properly when the mean value of the current i_{sig} exceeds the value If. A way to fix this problem is to connect a diode between the nodes $Out1$ and $Out2$ with the anode side on $Out1$. When the mean sensor current $\overline{i_{sig}}$ exceeds the value If, the potential of the node In drops since the feedback cannot keep the In potential at a constant level. The sensor current then directly flows through the diode into the integrator. Photon counting is in that case not possible, however, the sensor current integration works.

4.2.4 Pixel Cell

The counting amplifier and the integrator with AD converter are implemented in the form of pixel cells. The block scheme of the pixel cell is shown in fig. 4.9. A CIX test chip is de-

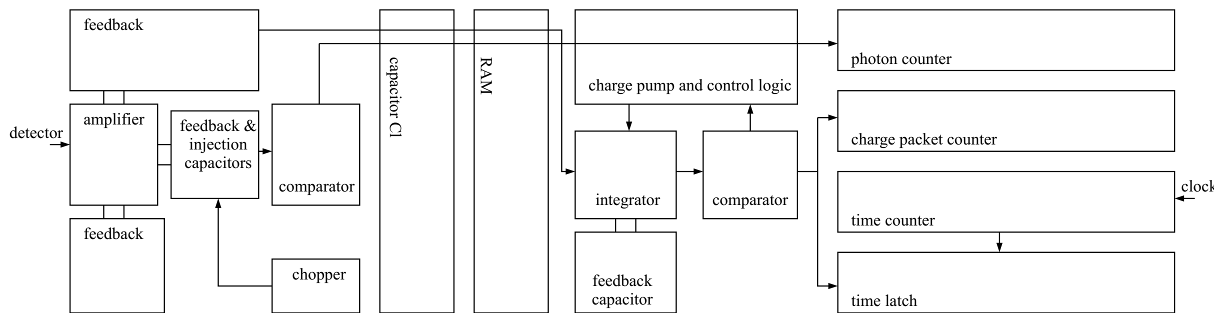


Figure 4.9: Pixel cell in the CIX chip.

signed in AMS 0.35 μm CMOS technology. The chip contains 18 pixels, 18 DA converters which generate the bias potentials, current reference circuit and an analog buffer which allows the measurement of the internal pixel signals. A special injection circuit (chopper), which simulates the sensor signal, is designed. The chopper is implemented as switched current source. The radiation tolerant layout technique has been partially used on the CIX chip. Important NMOS transistors, which generate small currents, are placed inside guard rings and have annular gate geometry. The counters in pixels are implemented by using the differential low-voltage logic circuits (DCL), published in [46]. The DCL logic has been introduced in the section about the ATLAS front-end chip. Fig. 4.14 shows the layout of the CIX chip.

4.2.5 Measurements

In this section we will present a few measurement results.

The output potential of the charge-sensitive amplifier (fig. 4.8) has been measured by using an oscilloscope. The charge-sensitive amplifier has been designed to drive quite a small capacitance of the comparator input node, typically, 30 fF. The capacitance of the oscilloscope probe is much larger, about 10 pF. Therefore we need a unity gain amplifier between the charge-sensitive amplifier and the oscilloscope that has a large input and a small output impedance. Such an amplifier is able to drive a large probe capacitance and, at the same time, it represents a small load for the charge-sensitive amplifier. The unity gain buffer is implemented on the chip.

Fig. 4.10 shows the signal at the output of the charge-sensitive amplifier. The sensor pulse has been simulated by using the chopper. The strength of the feedback circuit has been varied by changing the current If, fig. 4.8. The circuit behaves as expected. Fig. 4.11 shows the signal at the output of the charge-sensitive amplifier. The charge

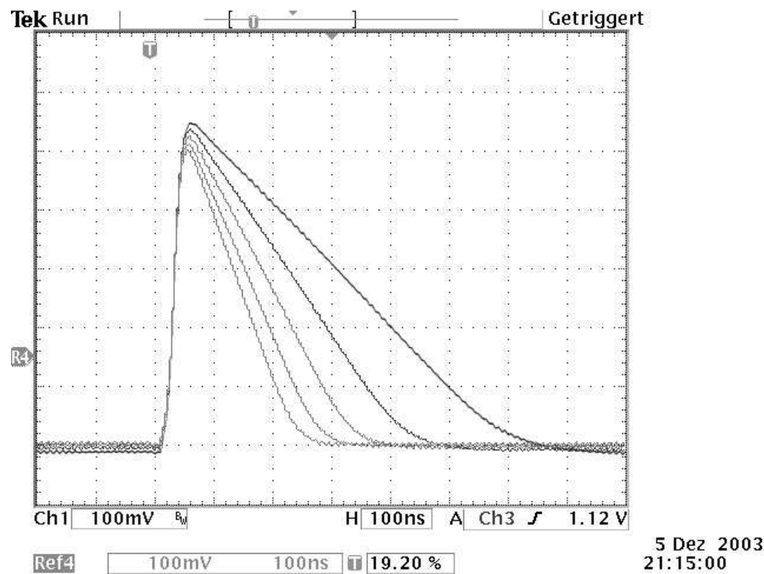


Figure 4.10: Preamplifier output signal for different feedback current values (current If, fig. 4.8).

injected by the chopper has been varied.

The chopper based on the current source allows us to inject a burst of charge packets and in this way test the capability of the counting system to count at high photon rates. The signal at the output of the charge-sensitive amplifier is shown in fig. 4.12. The switch Sw (fig. 4.8) was in the position A. The charge injection period has been varied between 100 ns and 500 ns. The counting with 10 MHz frequency is still possible. The action of the circuit for the leakage current compensation leads to a base line shift. This effect makes the threshold dependent on the counting rate.

A similar measurement has been done in the integration mode (switch Sw in position B, fig. 4.8). The potential needed for the leakage current is then stored on the capacitor Cl. The maximal counting frequency is 5 MHz. There is no base line shift.

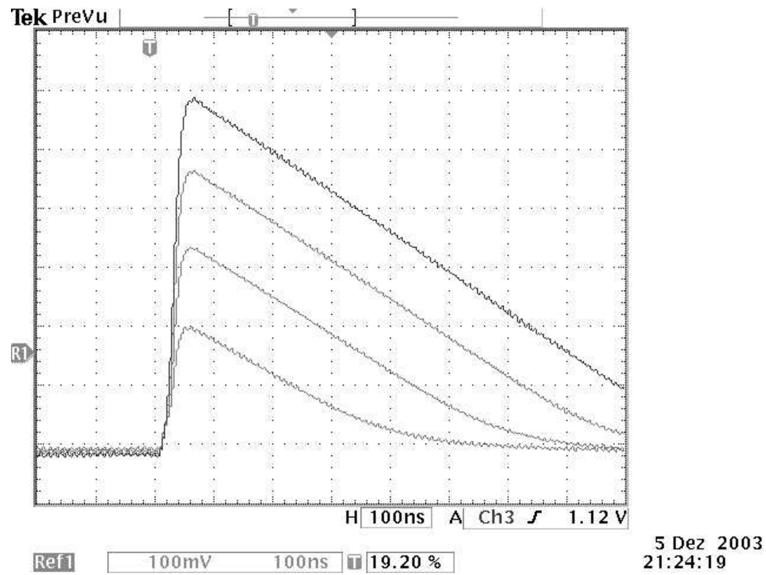


Figure 4.11: Preamplifier output signal for different values of injected charge.

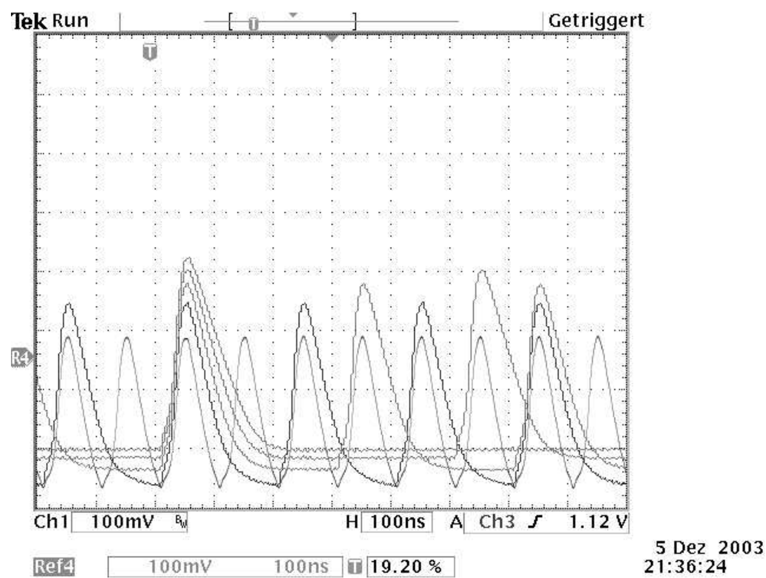


Figure 4.12: Preamplifier output signal for different chopper frequencies. Leakage current compensation is switched on (Switch Sw in position *A*, fig. 4.8). A maximal photon counting frequency of 10 MHz is theoretically possible.

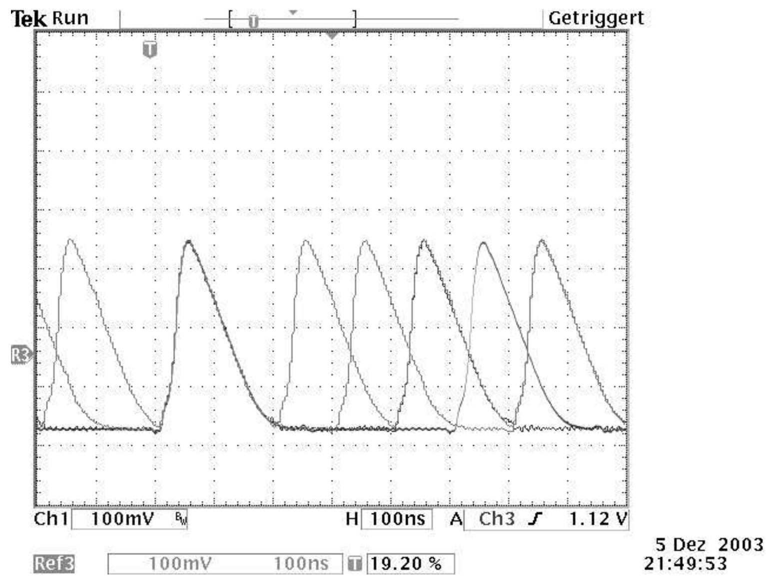


Figure 4.13: Preamplifier output signal for different chopper frequencies. The potential needed for the leakage current compensation is sampled on the capacitor Cl, (fig. 4.8) and the switch Sw is in position B. The maximal photon counting frequency per pixel is 5 MHz. There is no base line shift.

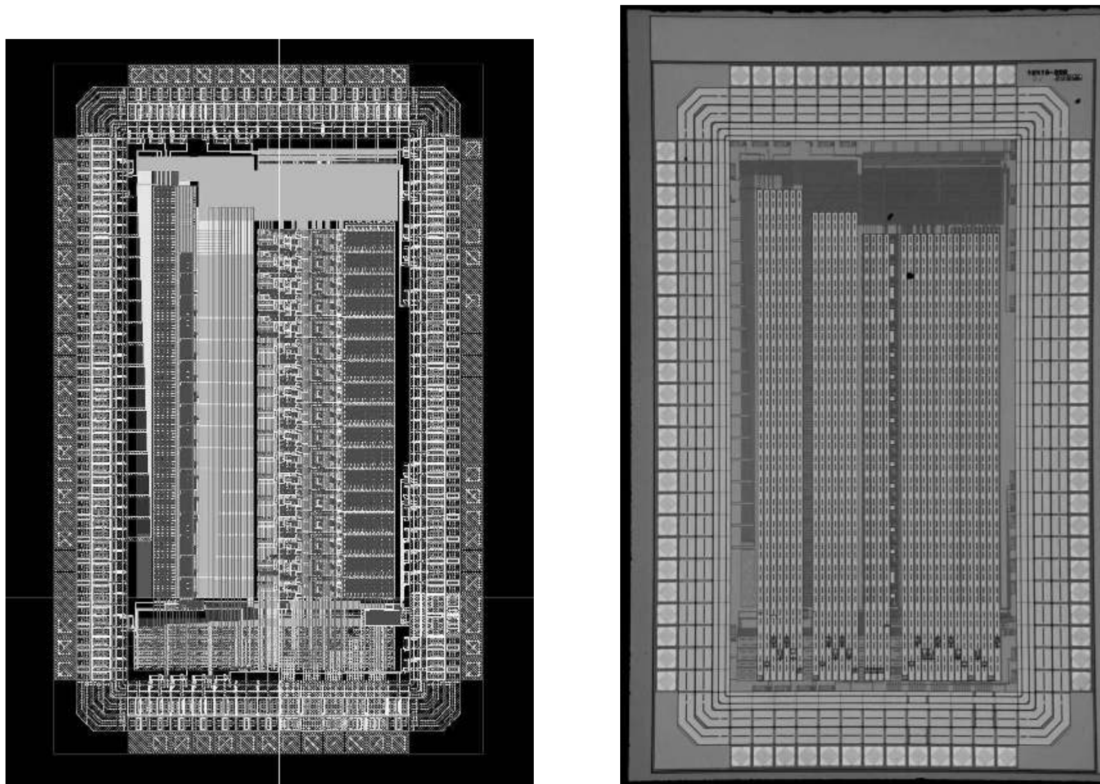


Figure 4.14: Layout and photomicrograph of the CIX chip.

Chapter 5

Summary and Outlook

1. Three versions of the pixel-readout chip for the ATLAS pixel detector have been designed. The last chip, called FE-I3, is the production chip used to build the pixel detector. It fulfills all design requirements needed for the application in the ATLAS pixel detector. The chips are implemented in a deep-submicron CMOS technology, using radiation-tolerant layout techniques. FE-I3 comprises about 3.5 million transistors. It has about 3000 pixels containing complex analog and digital circuits for sensor signal processing. The chip is able to detect signals of 2000 e, delivered by sensor pixels, with a time resolution of 20 ns. All signals generated by ionizing particles traversing sensor pixels are detected, their amplitudes are measured, digitized and stored on the chip together with time stamps and pixel addresses. The stored data are transmitted if they are requested by the ATLAS trigger system. The chip is radiation tolerant, it survives a dose of 100 Mrad. The production of the detector modules containing FE-I3 chips has already begun.

2. A high-voltage sequencer chip for a DEPFET pixel-sensor has been designed. The chip is called SWITCHER, it is implemented in AMS 0.8 μm HV CMOS technology. The chip comprises 64 channels. Every channel generates two signals of, maximally, 20 V amplitude. The signals are needed for the activation and clearing of DEPFET pixel rows. The signal sequence is programmable and stored in a memory block on the chip. The chip operates at 50 MHz frequency. High-voltage circuits in SWITCHER channels are designed to be fast, by keeping power consumption low. A system comprising a DEPFET detector and SWITCHER chips has been successfully tested. Since designed in a 0.8 μm technology, the SWITCHER chip is not radiation tolerant. This can be a problem for the application in the future linear collider. Because of that, the new version of the sequencer chip will be designed in a new AMS 0.35 μm HV CMOS technology. A small gate-oxide thickness in this technology promises better radiation-tolerance.

3. A pixel-readout test-chip for the application in an X-ray imaging system has been designed. The chip is called CIX, it is designed in AMS 0.35 μm CMOS technology. The CIX chip contains electronic circuits, layouted in pixel geometry, which can be used for the readout of an X-ray sensitive semiconductor pixel sensor. A CIX pixel can simultaneously integrate the current generated by a sensor pixel and count the charge pulses generated by individual photons in the sensor pixel. This allows the measurement of the mean photon energy. This information cannot be obtained by

the pixel-readout chips designed so far. The CIX chip contains the circuits which simulate sensor signals. Using these circuits, the functionality of the chip has been verified. The chip does not contain the pads which are necessary for the bonding of a pixel sensor. The next version of the counting-integrating chip will contain such pads. We hope then to be able to produce a small hybrid pixel detector and test it using an X-ray source.

Appendix A

Noise in Charge-Sensitive Amplifier

The noise is unwanted random obscuring signal. It can be produced by external sources (pickup noise), for example by radio transmitters, or by the digital circuits integrated together with an amplifier on the same chip. This type of noise can be eliminated by proper design. The noise is also an intrinsic property of real electronic components, like radiation sensors, transistors or diodes. Such intrinsic noise is a statistical phenomenon related to thermal motion of charge carriers, random nature of charge generation and the carrier number/mobility fluctuations.

A.1 Types of Noise

In the following sections we will analyze the intrinsic noise in the electronic system consisting of a radiation sensor and a charge-sensitive amplifier. Such a system has been already discussed in section 2.2.2. The small-signal circuit and the transfer function have been derived, see fig. 2.13.

We will take into account the three most important sources of noise.

First, we have the shot noise of the sensor leakage current. The sensor leakage current is caused by the generation of electron-hole pairs, the current is equal to the total generation rate. The electron-hole generation is a random process and generation rate fluctuates in time leading to current noise. This type of noise is called shot noise.

The second noise source that we take into account is the thermal noise in the channel of the input NMOS transistor of the charge sensitive amplifier. The transistor channel can be treated as a resistor with nonuniform resistance.[1] In the case of a resistor, the current is directly proportional to the electric field. The thermal motion of the charge carriers modulates the electric field, which influences through resistance the current flowing through the resistor. Such fluctuation is called thermal (or Johnson's) noise.

Generation and recombination of charge carriers do not play significant roles in ohmic resistors as well as in the channel of a MOS transistor operating in strong inversion. However, the charge trapping in the gate insulator leads to fluctuations of the transistor current. This is the third noise source of importance. The trapping causes fluctuations of the oxide/bulk charge, which influences the number of carriers and/or carrier mobility.

[1] We assume the strong inversion of the bulk region close to gate, which occur at high enough gate-source voltages. In weak inversion the diffusion charge transport is dominant and the origin of transistor current is thermal generation.

Due to exponential decay of trapped charge, the noise frequency spectrum has characteristic $1/f$ dependence. Such noise is called $1/f$, or flicker noise.

A.1.1 Mathematical Modelling of Noise

The knowing of the noise level at the output of an amplifier is of great importance. The output noise magnitude is the measure of the smallest signal that can be amplified. The calculation of output noise can be done by using a small-signal circuit since the noise amplitudes are usually small. Noise sources must be modelled with random signal generators. Clearly, we are not able to calculate the output noise as a deterministic time function, we can only calculate the statistical quantities like mean value or variance.

The small-signal circuit of the charge-sensitive amplifier with noise sources is shown in fig. A.1.

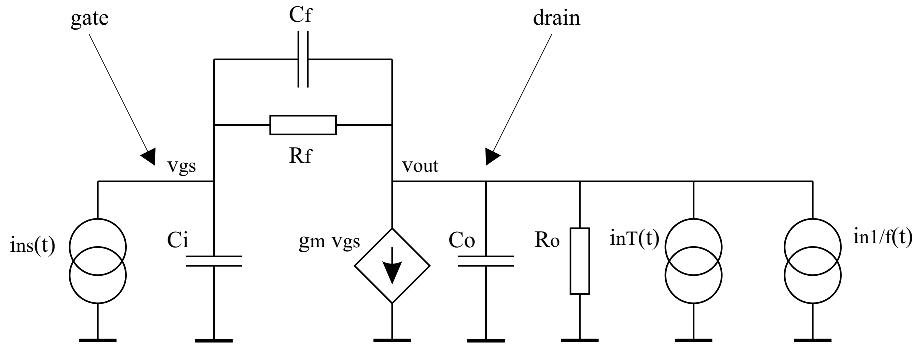


Figure A.1: Small signal circuit of the radiation sensor and the charge-sensitive amplifier with noise sources. i_{ns} corresponds to the sensor shot noise, i_{nT} and $i_{n1/f}$ are the thermal and 1/f noise in the channel of the transistor.

Let us briefly derive a few important formulas. As we already mentioned, the noise is modelled with random signals generators. Suppose that a current source generates the following random signal

$$i_n(t) = \sum_{t_i=-\infty}^{t} q_i p(t - t_i) \ . \tag{A.1}$$

q_i is a random number characterized by the mean value $\langle q \rangle = 0$ and the variance $\langle q^2 \rangle$. $p(t)$ is a pulse having the time integral equal to 1 and the following property $p(t) = 0$ for $t < 0$. For simplicity, let us assume that the time steps $t_j - t_i$ are equidistant $t_j - t_i \equiv \Delta t$.

Due to the linearity of the small-signal circuit, the output noise can be obtained by the superposition

$$v_{nout}(t) = \sum_{t_i=-\infty}^{t} q_i p_{out}(t, t_i) \ . \tag{A.2}$$

$p_{out}(t, t_i)$ is the response of the circuit to the unity input pulse $p(t - t_i)$. The response can be calculated if the transfer function $V_{out}(s)/I_n(s)$ is known.

The output noise is characterized by the mean value of its square amplitude, the quantity called also the noise power. Let us calculate this mean value. The square of the

output noise amplitude is

$$v_{nout}(t)^2 = \left(\sum_{t_i=-\infty}^{t} q_i p_{out}(t,t_i) \right)^2 = \sum_{t_i,t_j=-\infty}^{t} q_i q_j p_{out}(t,t_i) p_{out}(t,t_j) \ .$$

It is of great importance in the following analysis to assume that noise is an ergodic stationary process. If we had an ensemble of identical noisy circuits, we could measure the noise amplitudes in all circuits in a moment and calculate the mean value. The result would be the same as obtained by time integration. Therefore

$$\langle v_{nout}(t)^2 \rangle_t = \langle v_{nout}(t)^2 \rangle_e = \sum_{t_i,t_j=-\infty}^{t} \langle q_i q_j \rangle_e p_{out}(t,t_i) p_{out}(t,t_j) \ .$$

The symbol $\langle f_i \rangle_e$ describes the mean value of all members of the ensemble. $\langle f_i(t) \rangle_t$ is the time mean value.

Since q_i and q_j are non-correlated random numbers, it holds $\langle q_i q_j \rangle_e = \langle q^2 \rangle \delta_{ij}$. By taking this into account, the mean square amplitude gets the following simpler form

$$\langle v_{nout}(t)^2 \rangle = \sum_{t_i=-\infty}^{t} \langle q^2 \rangle (p_{out}(t,t_i))^2 \ ; \tag{A.3}$$

or in integral form (assuming that the time steps $t_j - t_i \equiv \Delta t$ are small)

$$\langle v_{nout}(t)^2 \rangle = \int_{-\infty}^{t} \frac{\langle q^2 \rangle}{\Delta t} (p_{out}(t,\tau))^2 d\tau \ . \tag{A.4}$$

The formula eq. A.4 is very useful, it can be applied in the case of the time dependent transfer functions (see [50]). The term $\frac{\langle q^2 \rangle}{\Delta t}$ has to be known.

Let us derive the term $\frac{\langle q^2 \rangle}{\Delta t}$ in the case of shot noise. Eq. A.1 can be rewritten as

$$i_n(t) = \sum_{t_i=-\infty}^{t} n \cdot e \cdot p(t - t_i) \ , \tag{A.5}$$

where $i_n(t)$ is the deviation of the current from its mean value. Supposing that the time steps Δt are so small that we cannot expect generation of more than one electron in the Δt period, the product $e \cdot p(t-t_i)$ can be understood as the charge pulse produced by the passage of a single electron. The charge pulse $e \cdot p(t-t_i)$ is certainly much faster than any relevant time-constant in an electronic circuit and it can be considered as a delta pulse. The random value n, eq. A.5, is given by the equation $n = N - \langle N \rangle$; N is the number of electrons that was generated in a unit interval Δt. The generation of charge carriers is a random process. We can assume Poisson statistic $\langle n^2 \rangle = Var(N) = \langle N \rangle$. By comparing of eq. A.5 with eq. A.1 and by taking into account the previous results we obtain

$$\frac{\langle q^2 \rangle}{\Delta t} = \frac{\langle e^2 n^2 \rangle}{\Delta t} = e^2 \frac{\langle N \rangle}{\Delta t} = e \langle I \rangle \ . \tag{A.6}$$

I is the mean current.

The formula eq. A.4 can be directly used for the calculation of noise output power. We will, however, derive the corresponding equation in frequency domain.

Let us assume that the transfer function does not change in time. In this case, we have $p_{out}(t,\tau) = p_{out}(t-\tau,0)$. $p_{out}(t,0)$ is the response of the circuit to the input pulse $p(t)$. Starting from eq. A.4 we have

$$\langle v_{nout}(t)^2\rangle = \int_{-\infty}^{t} \frac{\langle q^2\rangle}{\Delta t}(p_{out}(t-\tau,0))^2 d\tau = \int_{0}^{\infty} \frac{\langle q^2\rangle}{\Delta t}(p_{out}(u,0))^2 du \ . \tag{A.7}$$

In order to derive the right hand side of the equation, we made the substitution $t-\tau \to u$.

We can now apply Parseval's[2] theorem

$$\begin{aligned}\langle v_{nout}(t)^2\rangle &= \int_{0}^{\infty} \frac{\langle q^2\rangle}{\Delta t}(p_{out}(t,0))^2 dt = \int_{-\infty}^{\infty} \frac{\langle q^2\rangle}{\Delta t}(p_{out}(t,0))^2 dt = \\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{\langle q^2\rangle}{\Delta t}|P_{out}(i\omega)|^2 d\omega \ . \end{aligned}\tag{A.8}$$

$|P_{out}(i\omega)|^2 \equiv |\mathcal{F}\{p_{out}(t,0)\}|^2 = |H(i\omega)|^2|P(i\omega)|^2$. $H(i\omega)$ is the complex transfer function and $P(i\omega)$ the fourier transform of the noise unit pulse $p(t)$.

Here, we can define the power spectral density (PSD) $S_{i_n}(f)$ of the noise source i_n:

$$\langle v_{nout}(t)^2\rangle = \int_{0}^{\infty} S_{i_n}(f)|H(f)|^2 df \ , \tag{A.9}$$

where $i\omega = \imath 2\pi f$. Comparing eq. A.8 with eq. A.8 leads to

$$S_{i_n}(f) = 2\frac{\langle q^2\rangle}{\Delta t}|P(f)|^2 \ . \tag{A.10}$$

In case of shot noise, see eq. A.6, we get the following well known formula

$$S_{i_{ns}}(f) = 2e\langle I\rangle \ . \tag{A.11}$$

The noise power spectral density of shot noise does not depend on frequency. This is result of our assumption that the noise pulse $p(t)$ has the form of a Dirac delta function (the Fourier transform of the delta function is the constant 1).

A.1.2 Thermal and 1/f Noise

In this chapter we will very briefly introduce thermal and 1/f noise.

The thermal noise of a resistor can be modelled either with a random current source in parallel with the resistor, or with a random voltage source connected serially with the resistor. The power spectral density of the current source is $S_{i_{nT}} = 4kT/R$ and the PSD of the voltage source $S_{v_{nT}} = 4kTR$; k is the Boltzman constant, T the temperature and R the resistance of the resistor. In order to check these formulas, we can consider a simple circuit consisting of a capacitor connected in parallel with a resistor (a real resistor has always a certain capacitance). The mean value of the square noise voltage amplitude between the resistor terminals can be calculated by using eq. A.9, where we substitute

[2] Note that the time integration in eq. A.8 can be extended to $-\infty$, since $p_{out}(t,0) = 0$ for $t < 0$.

the proper transfer function and the noise PSD. The result is $\langle v_n(t)^2 \rangle = kT/C$. The mean energy of the electric field in the capacitor (real resistor) is therefore $\langle E \rangle = \frac{1}{2}C\langle v_n(t)^2 \rangle = \frac{1}{2}kT$. This result is in the agreement with the classical statistical theory. The total field energy corresponds to one degree of freedom (voltage) in the classical system consisting of electrons in a resistor and an electric field in a capacitor. Each degree of freedom contributes with $\frac{1}{2}kT$ to the total energy.

Let us now discuss the transistor current noise caused by fluctuations of the trapped charge in the oxide layer of a MOS transistor. We will consider first the trapping in a small $Si0_2$ volume. The trapping of charge is characterized by two time constants, capture time τ_c and emission time τ_e, [5]. The number of carriers q trapped in some period Δt is proportional to $\Delta t/\tau_c$. The emission of the trapped charge is described by the exponential low

$$q(t) = q \exp\left(-t/\tau_e\right) \ .$$

The transistor current is therefore given by the following equation

$$i_{nout}(t) \propto \sum_{t_i=-\infty}^{t} q_i \exp\left(-(t-t_i)/\tau_e\right) \ . \tag{A.12}$$

By using eq. A.10 we can calculate the power spectral density of the random signal given by eq. A.12

$$S_n \propto 2\frac{\langle q^2 \rangle}{\Delta t}\frac{\tau_e^2}{1+(\omega\tau_e)^2} \ . \tag{A.13}$$

If we assume a Poisson statistics, the variance $\langle q^2 \rangle$ is equal to the number of carriers captured in the period Δt, therefore $\langle q^2 \rangle \propto \Delta t/\tau_c$. We will assume $\tau_c \approx \tau_e$. We have

$$S_n \propto \frac{\tau_e}{1+(\omega\tau_e)^2} \ . \tag{A.14}$$

The last equation gives the power spectral density caused by fluctuations of trapped-charge in a small oxide volume. The trapping time-constant depends exponentially on the distance between the trap and the oxide-channel boundary: $\tau_e = \tau_0 \exp(x/\lambda)$, see [55, 56, 57, 58]. If we assume that the trap centers are uniformly distributed in the oxide, the total spectral density can be found by the integration

$$S_n \propto \int_0^{t_{ox}} \frac{\tau_e}{1+(\omega\tau_e)^2} dx \propto \frac{\lambda}{\omega}[\arctan(\omega\tau(t_{ox})) - \arctan(\omega\tau(0))] \ .$$

t_{ox} is the oxide thickness. In order to derive the right-hand side of the last equation, we have used the result $dx\tau_e = \lambda d\tau_e$. In the frequency range $1/\tau(t_{ox}) \ll \omega \ll 1/\tau(0)$, the noise power spectral density has the characteristic 1/f dependence $S_n \propto \frac{1}{\omega}$. For smaller frequencies $\omega \ll 1/\tau(t_{ox})$, S_n becomes nearly constant.

A.1.3 Output Noise

Let us now calculate the mean value of the square noise amplitude at the output of the charge-sensitive amplifier. Fig. A.1 shows the small-signal circuit. It is convenient to transfer all transistor noise sources to the gate of the input transistor. We replace each

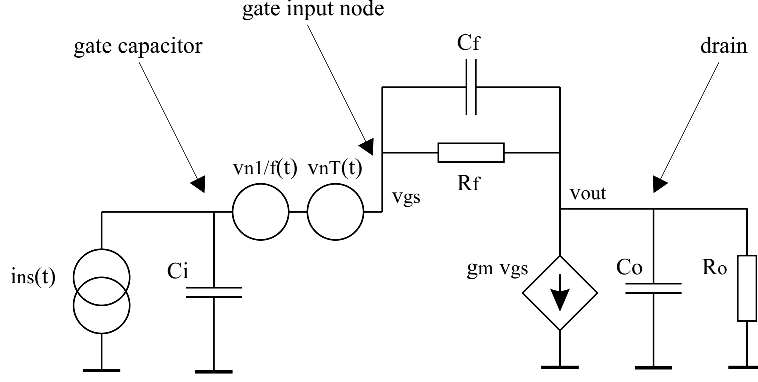


Figure A.2: Small signal circuit with transistor noise sources referred to gate. i_{ns} describes detector shot noise, v_{nT} and $v_{n1/f}$ describe thermal and 1/f noise in the channel of transistor, respectively.

current source connected to drain, fig. A.1, with an equivalent voltage source connected between the gate capacitor and the gate node. The new circuit, equivalent to that in fig. A.1, is shown in fig. A.2.

The voltage amplitudes of the sources v_{nT} and $v_{n1/f}$ are related to the current amplitudes of the corresponding current sources with the following simple relations

$$v_{nT} = i_{nT}/g_m \quad \text{(A.15)}$$
$$v_{n1/f} = i_{n1/f}/g_m \; . \quad \text{(A.16)}$$

g_m is the transconductance of the input transistor. The noise PSD is proportional to the square of the noise amplitude. Therefore the power spectral densities of the voltage and current sources are related as follows

$$S_{v_{nT}} = S_{i_{nT}}/g_m^2 \quad \text{(A.17)}$$
$$S_{v_{n1/f}} = S_{i_{n1/f}}/g_m^2 \; . \quad \text{(A.18)}$$

We have already derived the formula for the PSD in the case of shot noise:

$$S_{i_{ns}} = 2eI_{det} \; . \quad \text{(A.19)}$$

I_{det} is the detector leakage current (mean value), see the formula eq. A.11. We have also discussed the spectral density of the resistor thermal noise. The transistor channel can be treated as a nonuniform resistor. The following formula can be derived $S_{i_{nT}} = 4kTn\gamma g_m$ [1]. The constant γ is 2/3 in strong inversion and 1/2 in weak inversion. The constant n (slope factor) is, typically, 1.5. The power spectral density of the corresponding voltage source connected to gate is

$$S_{v_{nT}} = 4kTn\gamma/g_m \; . \quad \text{(A.20)}$$

In literature [1] we can find a few similar formulas that describe the spectral density of the voltage source modelling 1/f noise. We will use the following formula [55] that is also used within the scope of the BSIM4 transistor model

$$S_{v_{n1/f}} = \frac{e^2 N_T \mu I_{DS}}{L^2 n C'_{ox} g_m^2 f} \ln\left(1 + \frac{L g_m}{W \mu C'_{ox} U_T}\right) \equiv \frac{k_f}{f} \; . \quad \text{(A.21)}$$

N_T is the effective number of traps $N_T = kTN'_T/10^{10}$; $N'_T \approx 6.25 \cdot 10^{41}[1/(Jm^2)]$ for an NMOS transistor and $N'_T \approx 6.188 \cdot 10^{40}[1/(Jm^2)]$ for an PMOS [52]. I_{DS} is the DC drain-source current, C'_{ox} is the gate capacitance per unit area, L and W are the gate length and width, μ is the channel charge mobility, e is the elementary charge, g_m is transconductance and the U_T is the thermal voltage kT/e. The formula is valid in strong and in weak inversion, assuming current saturation.

According to eq. A.9, the voltage noise power at the output of the amplifier can be calculated as follows

$$\langle v_{nout}(t)^2 \rangle = \int_0^\infty S_{i_{ns}} \left| \frac{V_{out}(\imath\omega)}{I_{ns}(i\omega)} \right|^2 df + \int_0^\infty S_{v_{nT}} \left| \frac{V_{out}(\imath\omega)}{V_{nT}(i\omega)} \right|^2 df + \int_0^\infty S_{v_{n1/f}} \left| \frac{V_{out}(\imath\omega)}{V_{n1/f}(i\omega)} \right|^2 df \ . \tag{A.22}$$

The transfer functions can be obtained by using the procedure that we described in section 2.2.2. The results are

$$\frac{V_{out}(\imath\omega)}{I_{ns}(i\omega)} \approx \frac{-R_f(1 - i\omega\tau_z)}{(\imath\omega\tau_r + 1)(\imath\omega\tau_f + 1)} \ , \tag{A.23}$$

and

$$\frac{V_{out}(\imath\omega)}{V_{nT}(i\omega)} = \frac{V_{out}(\imath\omega)}{V_{n1/f}(i\omega)} \approx -\frac{1 - i\omega\tau_{z2}}{(\imath\omega\tau_r + 1)(\imath\omega\tau_f + 1)} \ . \tag{A.24}$$

The constants in equations eq. A.23 and eq. A.24 are: the feedback time $\tau_f \approx R_f C_f$, the rise time $\tau_r \approx (C_o C_f + C_o C_i + C_i C_f)/(g_m C_f)$, the time-constant $\tau_z \approx C_f/g_m$ and $\tau_{z2} \approx \tau_f(C_i + C_f)/C_f$.

The transfer function in eq. A.23 is the same function that we derived in section 2.2.2. (eq. 2.9) in order to calculate the output voltage pulse.

For the noise calculation we can neglect the term $i\omega\tau_z$ in the numerator of eq. A.23. This term becomes considerably large only at very high frequencies. The integration in this frequency range does not contribute to the total noise power because of large noise attenuation due to the factors in the denominator. It can be also shown that the term 1 in the numerator of eq. A.24 does not play important role in the calculation of the noise power.[3] The following result can be obtained by analytic integration:

$$\langle v_{nout}(t)^2 \rangle = \frac{1}{4}\frac{1}{C_f^2}\frac{\tau_f}{\tau_f + \tau_r}\left(\tau_f S_{i_{ns}} + \frac{C_i^2}{\tau_r} S_{v_{nT}} + 4C_i^2 k_f \frac{t_p}{\tau_r}\right) \ . \tag{A.25}$$

t_p is the peaking time (time when the response to the delta function has the maximal amplitude) that we calculated in the section 2.2.2, eq. 2.11:

$$t_p = \frac{\tau_r \tau_f}{\tau_f - \tau_r} \ln\left(\frac{\tau_f}{\tau_r}\right) \ . \tag{A.26}$$

It is convenient to define the value a as the ratio between the rise and feedback time constants $a \equiv \tau_r/\tau_f$. Note that $a \leq 1$. We can rewrite the formula for the output noise power, eq. A.25, in terms of a

$$\langle v_{nout}(t)^2 \rangle = \frac{1}{4}\frac{1}{C_f^2}\frac{1}{1 + a}\left(\tau_f S_{i_{ns}} + \frac{C_i^2}{a\tau_f} S_{v_{nT}} + 4C_i^2 k_f \frac{1}{a - 1} \ln{(a)}\right) \ . \tag{A.27}$$

[3] Even in the case of 1/f noise.

For a given value of a, the output noise power is minimal when the time constants fulfill the following condition

$$\tau_f \tau_r = \left(C_i \sqrt{\frac{S_{v_{nT}}}{S_{i_{ns}}}} \right)^2 \equiv (\tau_{opt})^2 \ . \tag{A.28}$$

If we want to calculate the optimal value for the number a, we must take into account that the output pulse amplitude depends also on the ratio between time constants, see eq. 2.12. The maximal amplitude, in terms of a is

$$v_{out}(t_p) = \frac{Q}{C_f} a^{\left(\frac{a}{1-a}\right)} \ . \tag{A.29}$$

It is useful to define the equivalent noise charge (ENC) as the input charge signal which produces an output pulse with an amplitude equal to the root-mean-square (RMS) noise value at the output. According to this definition, ENC is given by the equation

$$(ENC)^2 \equiv \frac{\langle v_{nout}(t)^2 \rangle}{v_{out}(t_p)^2} Q^2 \ . \tag{A.30}$$

By using the results eq. A.27 and eq. A.29, we obtain

$$(ENC)^2 = \frac{1}{4} \frac{a^{\left(\frac{2a}{a-1}\right)}}{1+a} \left(\tau_f S_{i_{ns}} + \frac{C_i^2}{a\tau_f} S_{v_{nT}} + 4C_i^2 k_f \frac{1}{a-1} \ln(a) \right) \ . \tag{A.31}$$

If we substitute the optimal value for the feedback time $\tau_f = \tau_{opt}/\sqrt{a}$ (eq. A.28) into eq. A.31, the analysis of the function shows that the ENC becomes smaller when a increases. We will therefore have the best signal-to-noise ratio for $a = 1$; i.e. $\tau_r = \tau_f$. For more details see, also [50, 64].

Appendix B

Modelling of MOS Transistors

B.1 DC-Current Equations

In the sections that follow we will express the stationary (DC) currents of a MOS transistor as functions of the voltages between electrodes. A MOSFET has four terminals: drain (D), source (S), gate (G) and bulk (B), fig. B.1. We can define three independent currents

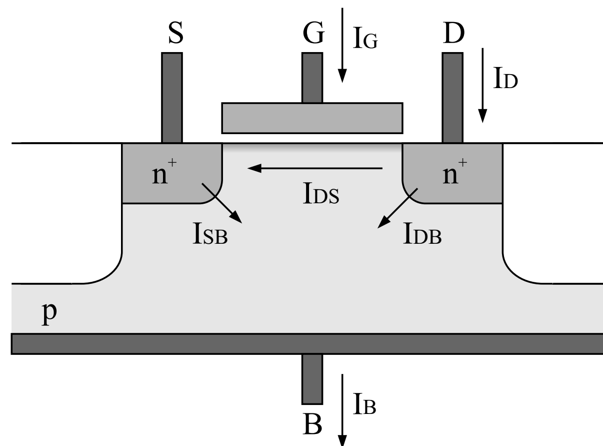


Figure B.1: MOS transistor and terminal currents.

and and three voltages. The fourth current/voltage can be related to the other three currents/voltages by using the Kirchoff laws. Usually, the source terminal is assumed to be reference. Each voltage is calculated between a terminal and source. The transistor is characterized by the following equations

$$\begin{aligned} I_D &= I_D(V_{GS}, V_{DS}, V_{BS}) \\ I_G &= I_G(V_{GS}, V_{DS}, V_{BS}) \\ I_B &= I_B(V_{GS}, V_{DS}, V_{BS}) \ . \end{aligned} \tag{B.1}$$

The DC gate current I_G is equal to zero since no constant current can flow through an insulator. The bulk current I_B is usually very small. This current is the sum of the currents that flow through the inversely biased source-bulk and drain-bulk junctions ($I_B = I_{DB} + I_{SB}$). The drain current is of greatest importance. This current can be decomposed into two parts: $I_D = I_{DS} + I_{DB}$. The current I_{DB} is under normal conditions

negligible. We will therefore restrict us to the calculation of the drain-source current I_{DS} as function of the voltages between nodes:

$$I_{DS} = I_{DS}(V_{GS}, V_{DS}, V_{BS}) \ . \tag{B.2}$$

The transistor currents can be calculated by using the drift-diffusion equations. A MOS transistor is a three-dimensional semiconductor structure that can be assumed to be close to thermodynamical equilibrium. Under such conditions the drift-diffusion model is valid. However, the three-dimensional drift-diffusion equations can be solved only numerically and numerical solutions are not suitable for the simulation of complex electronic circuits. We will here derive approximative, but analytical, formulas.

B.1.1 Gradual-Channel Approximation

A voltage applied between gate and bulk produces a vertical (transversal) electric field in the area below gate. This field component (E_y) leads to the formation of a channel made by attracted electrons. A voltage applied between drain and source produces a horizontal (longitudinal) field component E_x. The drain-source transistor current is the result of this field component.

We assume that the transversal field component (E_y) is much greater then the longitudinal component (E_x). This approximation is referred to as the gradual channel approximation [1] and it holds when the separation between drain and source is large. Note also that the longitudinal field is small only if we are far enough from drain and source. In the closest vicinity of the n^+ junctions we have high field regions. The field in these areas is induced by the uncompensated donor ions which are present in the n^+ material.

In order to simplify the following analysis, we will cut the MOS structure into thin vertical slices, fig. B.2. In every thin slice the longitudinal electric field (E_x) and the

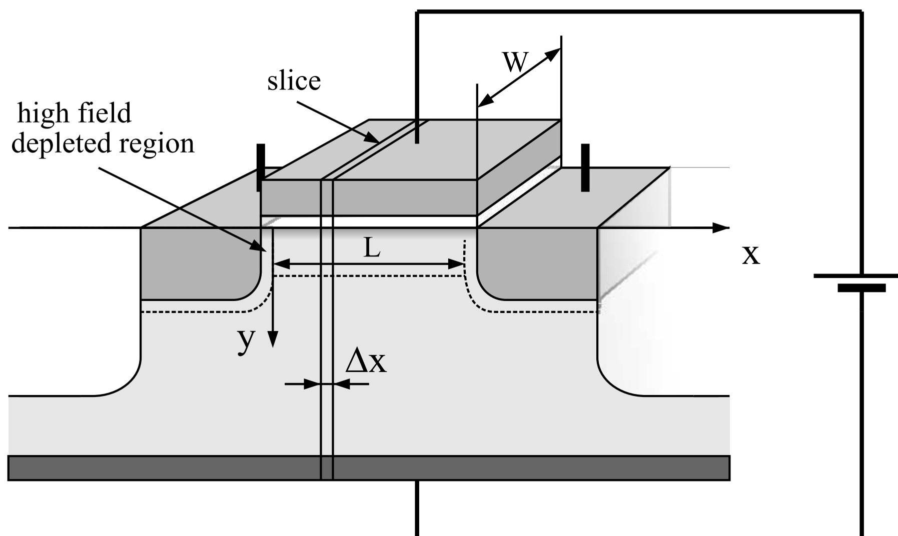


Figure B.2: MOS slice.

potential change in the x direction are negligible. In a thin slice nothing changes in the longitudinal direction. We can therefore calculate the carrier densities in a slice by using one-dimensional equations.

B.1.2 Channel-Charge - Exact Solution

Our first task is to calculate the charge in the channel as function of the positive gate-bulk potential. Fig. B.3 shows the cross section of a MOS slice. We can write

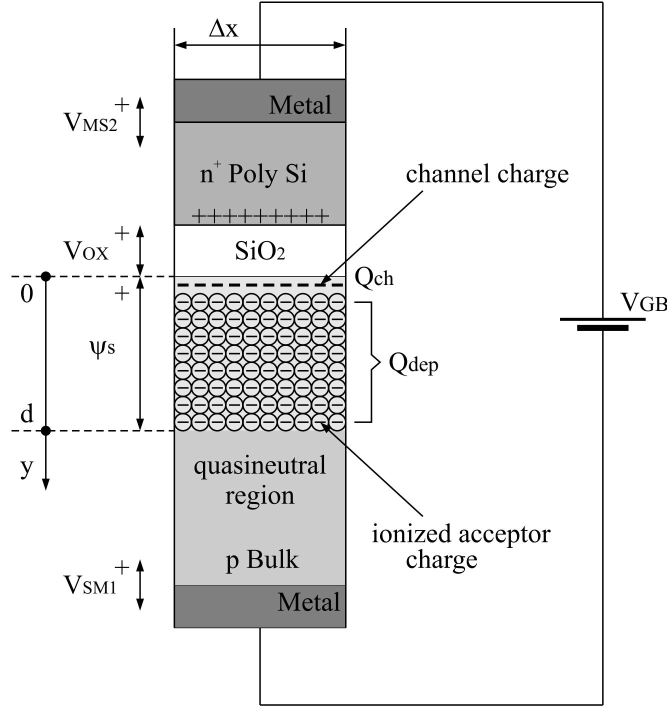


Figure B.3: MOS slice and potentials

$$V_{GB} = \phi_{SM1} + \psi_S(V_{GB}) + V_{OX}(V_{GB}) + \phi_{MS2} \equiv V_{pn} + \psi_S(V_{GB}) + V_{OX}(V_{GB}) \ . \qquad \text{(B.3)}$$

V_{SM1} is the contact potential difference between the p-doped silicon bulk and a metal contact. V_{MS2} is the potential difference between a metal contact and the n^+-doped poly-crystalline (poly) silicon gate electrode. We define $V_{pn} \equiv V_{SM1} + V_{MS2}$. The parameter V_{pn} corresponds to the contact potential difference between the p-doped bulk and the n^+-doped poly gate, the V_{pn} value is close to the silicon band-gap expressed in volts ($V_{pn} \approx -1V$). ψ_S is the potential difference between the surface of the bulk and the deep quasi-neutral field-free bulk region. In the bulk area close to gate there is a lack of holes. In this area we have a negative space charge due to uncompensated acceptor ions. The voltage ψ_S is usually called surface potential. V_{OX} is the voltage drop in the silicon-dioxide insulating layer.

The insulting layer will prevent any current flow from gate to the bulk. In the bulk material, the current flows mainly in horizontal direction, provided we have a voltage between drain and source. Also, a smaller current flows from the inversely polarized n^+ junctions (source and drain) and the channel into the bulk, see fig. B.4. A current can distort distribution functions of charge carriers and cause that they deviate from equilibrium functions. We will, however, assume that the carrier densities in for us interesting space charge region are well approximated with the equilibrium distribution functions. The temperature is homogenous and the Fermi energy constant in a thin slice. The

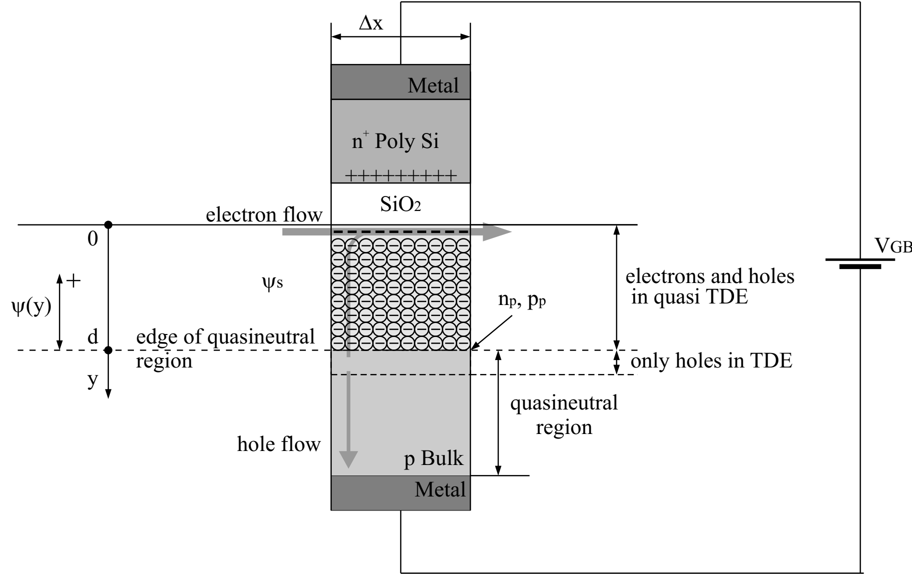


Figure B.4: MOS slice and currents. TDE - thermodynamic equilibrium.

electron and hole densities in a transistor slice are given by Maxwell-Boltzman functions:

$$\begin{aligned} n(y) &= n_p \exp\left(\frac{e\psi(y)}{kT}\right) \\ p(y) &= p_p \exp\left(-\frac{e\psi(y)}{kT}\right) \ . \end{aligned} \tag{B.4}$$

n_p and p_p are the electron and hole densities at the edge of the quasineutral bulk region, see fig. B.4. In the case of equilibrium we have $n_p \approx n_i^2/N_a$ and $p_p \approx N_a$, n_i is the intrinsic carrier density in silicon ($n_i \approx 1.45 \cdot 10^{10} cm^{-3}$ at room temperature) and N_a is the acceptor density in the bulk, typically, $1.7 \cdot 10^{17} cm^{-3}$ [52]. We assume that the holes in a slice are always in equilibrium since their density is high and cannot be easily disturbed by a current flow. Therefore it holds $p_p \approx N_a$. However, the electrons in the quasineutral bulk area are in equilibrium only if there is no current flowing through the bulk, which means only in the case when the drain-bulk and source-bulk voltages are zero. $\psi(y)$ is the electric potential of a point in the bulk with respect to the potential in the quasineutral bulk area. It holds $\psi(0) = \psi_S$ and $\psi(d) = 0$.

If we define the thermal voltage as $U_T \equiv kT/e$ (26mV at room temperature), we have

$$\begin{aligned} n(y) = n_p \exp\left(\frac{\psi(y)}{U_T}\right) \\ p(y) = p_p \exp\left(-\frac{\psi(y)}{U_T}\right) \ . \end{aligned} \tag{B.5}$$

Let us write the one-dimensional Poisson equation:

$$\frac{d^2\psi(y)}{dy^2} = -\frac{e}{\epsilon_{Si}}(p(y) - n(y) - N_a^-) \ . \tag{B.6}$$

N_a^- is the effective density of the ionized acceptors in the bulk. It holds $N_a^- \approx p_p - n_p$ since the semiconductor bulk must be electrically neutral at the edge of its quasineutral

region. Therefore if we substitute eq. B.5 in eq. B.6, we have

$$\frac{d^2\psi(y)}{dy^2} = -\frac{e}{\epsilon_{Si}}\left[p_p\left[\exp\left(-\frac{\psi(y)}{U_T}\right)-1\right]-n_p\left[\exp\left(\frac{\psi(y)}{U_T}\right)-1\right]\right] . \tag{B.7}$$

ϵ_{Si} (silicon dielectric constant) is 11 - 12 times the vacuum value ϵ_0. ($\epsilon_0 \approx 8.854 \cdot 10^{-12}F/m$) . The function $\psi(y)$ cannot be calculated analytically. We can, however, calculate the electric field in the bulk $\left(E(\psi) \equiv -\frac{d\psi(y)}{dy}\right)$ as a function of the potential ψ.[1] The result [2, 1] is

$$E(\psi)^2 = \frac{2eU_Tp_p}{\epsilon_{Si}}\left[\left[\exp\left(-\frac{\psi}{U_T}\right)+\frac{\psi}{U_T}-1\right]+\frac{n_p}{p_p}\left[\exp\left(\frac{\psi}{U_T}\right)-\frac{\psi}{U_T}-1\right]\right] . \tag{B.8}$$

The total amount of negative charge in the bulk per unit area (Q'_{tot}) can be obtained by using the Gauss law:

$$Q'_{tot} = \epsilon_{Si}[E(\psi(0)) - E(\psi(d))] = \epsilon_{Si}[E(\psi=\psi_S) - E(\psi=0)] = \epsilon_{Si}E(\psi_S) . \tag{B.9}$$

If we substitute eq. B.8 in eq. B.9, we obtain

$$Q'_{tot} = \sqrt{2eU_T\epsilon_{Si}n_p}\left[\left[\exp\left(-\frac{\psi_S}{U_T}\right)+\frac{\psi_S}{U_T}-1\right]+\frac{n_p}{p_p}\left[\exp\left(\frac{\psi_S}{U_T}\right)-\frac{\psi_S}{U_T}-1\right]\right]^{1/2} . \tag{B.10}$$

Note, we defined the value Q'_{tot} to be positive when the charge is negative.

The charge, given by eq. B.10, is the sum of two components. First, we have the charge produced by the uncompensated acceptor ions in the depleted bulk region Q'_{dep} and second, we have the charge in the electron channel Q'_{ch}, see fig. B.3,

$$Q'_{tot} = Q'_{dep} + Q'_{ch} .$$

It is of great importance to know the exact value of the channel charge Q'_{ch}, since the channel resistance, and therefore the transistor current, depend on Q'_{ch}. Let us try to estimate Q'_{ch}. The electrons gather in a very narrow area close to the silicon-insulator interface. We can assume that the channel thickness is negligible. This has as consequence that the channel charge Q'_{ch} does not influence the electric potential $\psi(y)$. This approximation is known as the charge sheet approximation [1]. We can now write the Poisson equation for the part of the electric potential $\psi(y)$ that is generated only by the acceptor ions ($\psi(y)|_{dep}$):

$$\frac{d^2\psi(y)|_{dep}}{dy^2} = -\frac{ep_p}{\epsilon_{Si}}\left[\exp\left(-\frac{\psi(y)|_{dep}}{U_T}\right)-1\right] . \tag{B.11}$$

We can calculate the electric field produced only by acceptor ions $E(\psi)|_{dep} \equiv -\frac{d\psi(y)|_{dep}}{dy}$ and the charge $Q'_{dep} = \epsilon_{Si}E(\psi_S)|_{dep}$:

$$Q'_{dep} = \sqrt{2eU_T\epsilon_{Si}p_p}\left[\exp\left(-\frac{\psi_S|_{dep}}{U_T}\right)+\frac{\psi_S|_{dep}}{U_T}-1\right]^{1/2} . \tag{B.12}$$

[1] In order to do it, we can multiply eq. B.7 with $\left(\frac{d\psi(y)}{dy}\right)$ and then perform integration of both sides.

Since we assumed that the channel charge does not influence the electric potential in the bulk, the actual potential $\psi(y)$ is equal to the potential generated only by the ionized acceptors: $\psi(y) \approx \psi(y)|_{dep}$. Therefore $\psi_S \approx \psi_S|_{dep}$. If we substitute this into eq. B.12 we obtain[2]

$$Q'_{dep} \approx \sqrt{2eU_T\epsilon_{Si}p_p}\left[\exp\left(-\frac{\psi_S}{U_T}\right)+\frac{\psi_S}{U_T}-1\right]^{1/2} . \tag{B.13}$$

The channel charge Q'_{ch} can be calculated when we subtract Q'_{dep} given by eq. B.13 from Q'_{tot} given by eq. B.10:

$$\begin{aligned} Q'_{ch} &\approx \sqrt{2eU_T\epsilon_{Si}p_p}\left[\left[\exp\left(-\frac{\psi_S}{U_T}\right)+\frac{\psi_S}{U_T}-1\right]+\frac{n_p}{p_p}\left[\exp\left(\frac{\psi_S}{U_T}\right)-\frac{\psi_S}{U_T}-1\right]\right]^{1/2} - \\ &- \sqrt{2eU_T\epsilon_{Si}p_p}\left[\exp\left(-\frac{\psi_S}{U_T}\right)+\frac{\psi_S}{U_T}-1\right]^{1/2} . \end{aligned} \tag{B.14}$$

Similar formulas can be found in the literature, see for example [1]. Note that the hole density at the edge of the quasi-neutral bulk region - p_p - can be assumed to be equal to the acceptor density N_a.

The charge densities Q'_{dep}, Q'_{ch} and Q'_{tot} versus the surface potential ψ_S are shown in fig. B.5. As we see in fig. B.5, the channel charge Q'_{ch} starts to increase very fast after

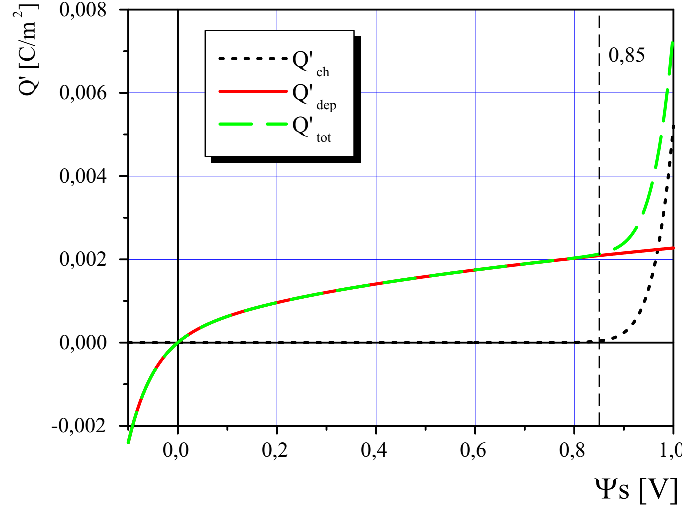


Figure B.5: Charge densities according to eq. B.14, eq. B.13 and eq. B.10. $p_p = N_a = 1.7 \cdot 10^{17} cm^{-3}$, $n_p = n_i^2/N_a$, $\epsilon_{Si} = 11 \cdot \epsilon_0$.

the surface potential reaches a threshold value. The threshold is, approximately, equal to $U_T \ln(N_a/n_p)$. Assuming the equilibrium values $n_p = n_i^2/N_a$ and $N_a = 1.7 \cdot 10^{17} cm^{-3}$, we have $U_T \ln(N_a^2/n_i^2) = 0.85V$.

Eq. B.14 expresses the channel charge Q'_{ch} as a function of the surface potential ψ_S. However, the surface potential ψ_S cannot be directly controlled. We can directly control only the gate-bulk voltage V_{GB}. Because of that, our next step is to find the relation between ψ_S and the gate-bulk voltage V_{GB}. We start from eq. B.3

$$V_{GB} = \phi_{SM1} + \psi_S(V_{GB}) + V_{OX}(V_{GB}) + \phi_{MS2} \equiv V_{pn} + \psi_S(V_{GB}) + V_{OX}(V_{GB}) . \tag{B.15}$$

[2] Note that $E(\psi)$ is not equal to $E(\psi)|_{dep}$ and therefore Q'_{tot} is not equal to Q'_{dep}!

A MOS structure can be treated as a parallel plate capacitor. The negative charge Q'_{tot} is induced by an equal amount of positive charge on gate. Therefore we have

$$V_{OX} = \frac{Q'_{tot}}{C'_{ox}} = \frac{Q'_{ch} + Q'_{dep}}{C'_{ox}} \ . \tag{B.16}$$

C'_{ox} is the oxide capacitance per unit area $C'_{ox} = \epsilon_{SiO2}/t_{ox}$. The dielectric constant $\epsilon_{SiO2} \approx 3.9 \cdot \epsilon_0$, t_{ox} is the oxide layer thickness. A typical value of the oxide thickness is $t_{ox} = 11.7nm$ [52]. This yields the typical value of oxide capacitance of 2.95 $fF/\mu\text{m}^2$.

Eq. B.16 is valid under the assumption that the total negative charge below the gate electrode is equal to the sum of the channel charge and the depleted area charge. In reality we have also the charged states at the bulk-Si0_2 interface and the fixed charge in the oxide layer [1, 2]. This additional charge contributes to the total potential drop in the vertical direction.

The interface states are the result of the imperfections in the silicon crystal lattice at the Si-Si0_2 boundary. A missing Si atom in a lattice tetrahedron (dangling bond) leads to the creation of two trap layers. A layer is placed close to the valence band and the other close to the conduction band. These layers can freely exchange electrons and holes with the bulk. When the energetically higher layer is filled with electrons is becomes negatively charged. On the other hand, when the energetically lower layer is filled with holes (releases electrons) it becomes positive. The contribution of these interface states to the total charge is expressed by the term $Q'_{it}(\psi_S)$ (we define $Q'_{it}(\psi_S)$ to be positive when the charge is negative). Therefore we have

$$V_{OX} = \frac{Q'_{ch} + Q'_{dep} + Q'_{it}}{C'_{ox}} \ . \tag{B.17}$$

The existence of the fixed charge in the oxide is mainly result of two effects. First, we have a contamination with the sodium ions (Na^+). The sodium ions are mobile in Si0_2 at higher temperatures. At lower temperatures they become "frozen". Second, we have positively charged oxide vacancies, mainly in the region close to the Si-Si0_2 interface. If we express the fixed oxide charge with the charge density $\rho(y)$, the additional voltage drop in Si0_2 can be calculated by integration of the formula for a plate capacitor:

$$\Delta V_{OX} = -\int_0^{t_{ox}} \frac{\rho(y)}{\epsilon_{SiO2}} \, y \, dy \ , \tag{B.18}$$

(we define the y direction in the way that $y = 0$ at the gate-oxide boundary and $y = t_{ox}$ at the Si-Si0_2 interface).

Now, we can write

$$V_{OX} = \frac{Q'_{ch} + Q'_{dep} + Q'_{it}}{C'_{ox}} + \Delta V_{OX} \tag{B.19}$$

and, finally (see eq. B.15):

$$V_{GB} = V_{fb} + \psi_S + \frac{Q'_{ch}(\psi_S) + Q'_{dep}(\psi_S) + Q'_{it}(\psi_S)}{C'_{ox}} \ . \tag{B.20}$$

We defined here the flat band voltage V_{fb} as $V_{fb} \equiv V_{pn} + \Delta V_{OX}$. When $V_{GB} = V_{fb}$, the potential change in the bulk ψ_S becomes zero and the energy bands are flat.

B.1.3 Channel Charge - Approximative Solutions

Now, we have all equations that we need for the calculation of $Q'_{ch}(V_{GB})$. Eq. B.13 and eq. B.14 express the functions $Q'_{dep}(\psi_S)$ and $Q'_{ch}(\psi_S)$. When we substitute these equations into eq. B.20, $\psi_S(V_{GB})$ can be obtained.[3] Once we have $\psi_S(V_{GB})$, the exact solution $Q'_{ch}(V_{GB})$ can be obtained. Note that the value n_p must be also known. We will see that the n_p value depends on the drain and source potentials and the x position of the slice. p_p is assumed to be equal to the acceptor density N_a.

In the next sections we will derive a few approximative formulas for $Q'_{ch}(V_{GB})$. In order to simplify the results, we will consider separately two different bias regions. First, we consider the case when the electron density in the channel is higher than n_i, but it still holds $Q'_{ch} \ll Q'_{dep}$. Assuming such conditions, the channel charge does not influence the surface potential. The transistor is said to be in "weak inversion". When the function $Q'_{ch}(\psi_S)$ starts to increase rapidly, we enter the second bias region called "strong inversion" (fig. B.5). In this region the surface potential $\psi_S(V_{GB})$ is assumed to stay nearly constant $\psi_S \approx \psi_0$ and independent of the gate-bulk voltage V_{GB}.

Weak Inversion

The weak inversion region is usually defined by the following condition

$$U_T \ln(n_i/n_p) < \psi_S < U_T \ln(N_a/n_p) \ . \tag{B.21}$$

In weak inversion region the channel charge is negligible in comparison with the depleted zone charge ($Q'_{ch} \ll Q'_{dep}$), but still, there are enough electrons in the channel so that a transistor working in this region has practical application.[4] Assuming the equilibrium value of n_p: $n_p = {n_i}^2/N_a$ and the standard parameter values $n_i = 1.45 \cdot 10^{10} cm^{-3}$ and $N_a = 1.7 \cdot 10^{17} cm^{-3}$, the condition in eq. B.21 becomes $0.42\ V < \psi_S < 0.85\ V$.

In weak inversion region, defined by eq. B.21, the formula for Q'_{dep} (eq. B.13) can be approximated as

$$Q'_{dep} \approx \sqrt{2e\epsilon_{Si} N_a \psi_S} \ . \tag{B.22}$$

Note, this approximation holds also in strong inversion.

Eq. B.20 becomes

$$V_{GB} = V_{fb} + \psi_S + \frac{Q'_{dep}(\psi_S)}{C'_{ox}} + \frac{Q'_{it}(\psi_S)}{C'_{ox}} \ . \tag{B.23}$$

If we neglect the interface charge Q'_{it} and substitute eq. B.22 into eq. B.23 we get the quadratic equation for ψ_S:

$$V_{GB} = V_{fb} + \psi_S + \frac{\sqrt{2e\epsilon_{Si} N_a \psi_S}}{C'_{ox}} \ . \tag{B.24}$$

Fig. B.6 shows the function $\psi_S(V_{GB})$ calculated solving the quadratic equation eq. B.24. We assumed the standard parameter values. The exact numerical solution is also shown.

[3] Here, we neglect $Q'_{it}(\psi_S)$, although the corresponding formula that takes this term into account can be easily derived. $Q'_{it}(\psi_S)$ is a function of the trap state density.

[4] If $\psi_S > U_T \ln(N_a/n_p)$, the electron density at the semiconductor-insulator boundary exceeds the hole density. We have the inversion of the majority carrier type.

This solution is obtained by combining eq. B.20 and eq. B.10. Note, the surface potential is positive for $V_{GB} = 0$. This is result of the negative flat band voltage ($V_{fb} = -1\ V$). In weak inversion region, defined by ($0.42\ V < \psi_S < 0.85\ V$), the surface potential depends nearly linearly on V_{GB} , which can be seen in fig. B.6. Assuming the condition in eq. B.21,

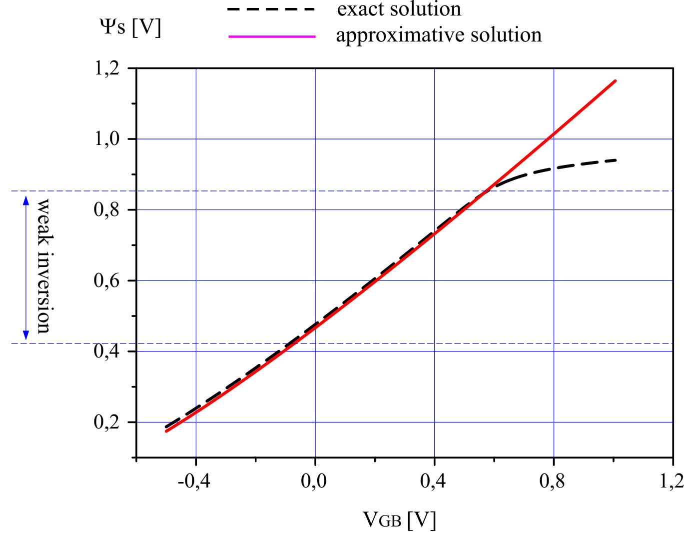


Figure B.6: Surface potential versus the gate-bulk voltage. Solution of the quadratic equation and the exact solution eq. B.23. Standard parameters are used, see fig. B.5.

the expression for the channel charge (eq. B.14) can be approximated as[5]

$$Q'_{ch} \approx eU_T \sqrt{\frac{\epsilon_{Si}}{2eN_a\psi_S}} n_p \exp\left(\frac{\psi_S}{U_T}\right) \quad . \tag{B.25}$$

In weak inversion the square root in eq. B.25 is nearly constant and the channel charge increases exponentially a few orders of magnitude as function of V_{GB}.

Strong Inversion

In strong inversion region large changes in V_{GB} result in small changes of ψ_S. For simplicity we assume that the surface potential stays constant ($\psi_S \approx \psi_0$) when we enter strong inversion. However, if we look at fig. B.6, we see that the assumption ($\psi_S \approx \psi_0$) holds only approximately. The best value for ψ_0 depends on the bias conditions which we want to model accurately. The commonly used value is[6]

$$\psi_0 = U_T \ln\left(n_i/n_p\right) \quad . \tag{B.26}$$

The negative channel charge screens the the depleted bulk area from the influence of the positive gate charge. A result of the screening effect is that the channel charge Q'_{ch} depends much weaker on the gate potential V_{GB} than it is the case in weak inversion. Let us verify this. Eq. B.20 can be rewritten as

$$V_{GB} \approx V_{fb} + \psi 0 + \frac{Q'_{ch} + Q'_{dep}(\psi 0) + Q'_{it}(\psi 0)}{C'_{ox}} \tag{B.27}$$

[5]We neglect the terms $\exp\left(-\frac{\psi_S}{U_T}\right) - 1$ and use the approximation $\sqrt{1+\Delta} \approx 1 + \Delta/2$.

[6]For the standard parameters, we have $\psi_0 \approx U_T \ln\left(N_a/n_p\right) = 0.85\ V$ and $V_{GB}(\psi_0) = 0.57\ V$.

or

$$Q'_{ch} \approx C'_{ox} \left(V_{GB} - V_{fb} - \psi 0 - \frac{Q'_{dep}(\psi 0)}{C'_{ox}} - \frac{Q'_{it}(\psi 0)}{C'_{ox}} \right) . \tag{B.28}$$

As we see from eq. B.28, the channel charge Q'_{ch} depends nearly linearly on V_{GB}. Eq. B.28 is valid assuming strong inversion $V_{GB} > V_{fb} + \psi 0 + \frac{Q'_{dep}(\psi 0)}{C'_{ox}} + \frac{Q'_{it}(\psi 0)}{C'_{ox}} = V_{GB}(\psi_0)$.

Fig. B.7 shows the channel charge Q'_{ch} versus the gate-bulk voltage V_{GB}. The solid line is the exact solution obtained by solving eq. B.20 numerically, where we used the formulas eq. B.14 and eq. B.13. The interface charge was neglected. The dashed lines are the approximative solutions eq. B.25 in weak inversion and eq. B.28 in strong inversion. The vertical line represents the value $V_{fb} + \psi 0 + Q'_{dep}(\psi 0)/C'_{ox} = 0.57\ V$. As we see, the accuracy of eq. B.28 is limited since the assumption ($\psi_S = \psi_0 = U_T \ln(N_a/n_p)$) holds only approximately. Because of that, the threshold value ψ_0 is sometimes defined to be several, typically 6, times higher than the thermal voltage $U_T \ln(N_a/n_p)$ ($U_T = 26\ mV$ at room temperature) [1]. It would be more accurate to take into account the real ψ_S dependence on V_{GB}, see fig. B.6.

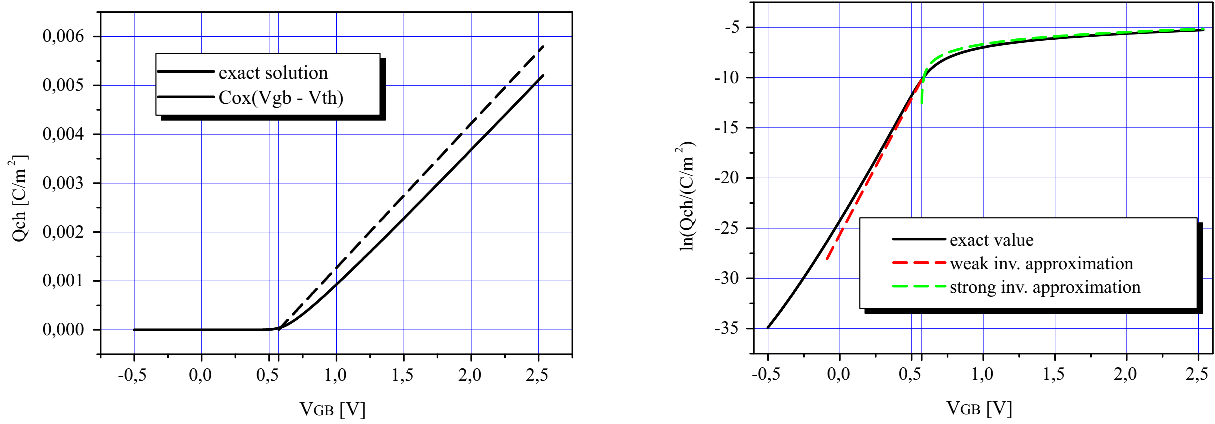


Figure B.7: Channel charge as function of gate potential. Solid lines represent the exact numerical solutions (eq. B.20, eq. B.14 and eq. B.13). Dashed lines are the approximative solutions in weak and strong inversion, eq. B.25 and eq. B.28 respectively. Standard parameters have been used.

Drain and Source Bias

Until now, we did not consider the influence of the drain and source bias voltages on the channel charge. Here, we should distinguish between two channel regions. First, we have the bulk areas in the closest vicinities of the drain/source n^+ diffusions, where the longitudinal field is strong, see fig. B.2. Second, we have a larger middle bulk region, where the longitudinal field is much weaker than the transversal and the gradual channel approximation holds. The regions with high horizontal field are called source and drain depleted regions. The strong horizontal field in these areas is produced by the uncompensated donor charge built in the neighboring n^+ layers. The region below the gate with a negligible longitudinal field is called gate depletion region. The vertical field in this channel area is generated by the charge in gate and bulk. The longitudinal field is induced by the channel itself.

If the drain and source electrodes are shortly connected with the bulk electrode, there are no currents. The system is in thermodynamical equilibrium. The charge-carrier densities at the edge of the quasineutral region n_p and p_p, see eq. B.4, are then equal to their equilibrium values $n_p \approx n_{p0} = n_i^2/N_a$ and $p_p \approx p_{p0} = N_a$.

What does happen when we apply positive source-bulk and drain-bulk voltages? The situation is similar to that in the case of an inversely biased $p - n$ diode. Let us recall the most important results for a diode. First, the depleted area of a diode grows when we increase the inverse bias voltage $V(bias)$, see fig. B.8. This can be related to the fact that the voltage drop in the depleted region increases from the value that holds in equilibrium, $V(built - in)$, to the value $V(built - in) + V(bias)$. A higher voltage drop requires a larger depleted zone. Second, an inverse saturation current starts to flow. One usually assumes that the current flow does not influence the carrier distributions in the depleted area. The net current can be neglected with respect to the drift and diffusion currents separately.[7] The carrier densities can be calculated by using Maxwell-Boltzman functions. However, a minority carrier distribution deviates from its equilibrium form in the regions, where carrier concentration is particularly low - especially in the parts of quasi-neutral regions close to the depleted area, see fig. B.8. The minority carrier concentration at the boundary between the depleted and the quasineutral area is decreased by the factor $\exp(V(bias)/U_T)$ with respect to the equilibrium value. All results that we mentioned are illustrated in fig. B.8.

Making analogy with the previous discussion, we can expect that the drain and source depletion areas grow when we bias the drain and source contacts positively with respect to bulk. Also, the electron density at the edge of the source depleted area decreases by $\exp(V_{SB}/U_T)$ with respect to its value for $V_{SB} = 0$. Similarly, the electron concentration at the edge of the drain depletion area decreases by $\exp(V_{DB}/U_T)$ with respect to its equilibrium value, see fig. B.9.

Let us now extend the formulas for the channel charge in order to take into account the influence of the drain and source bias voltages. In the case of weak inversion it is quite simple. We can start from eq. B.25. The only term that depends on the slice position is n_p, electron density at the edge of the quasineutral region. Close to source, we have $n_p = (n_i/N_a)\exp(-V_{SB}/U_T)$ and close to drain $n_p = (n_i/N_a)\exp(-V_{DB}/U_T)$. We can express the channel charge with the following function

$$Q'_{ch} \approx eU_T\sqrt{\frac{\epsilon_{Si}}{2eN_a\psi_S}}\frac{n_i^2}{N_a}\exp\left(\frac{-V_{SB}-V_x}{U_T}\right)\exp\left(\frac{\psi_S}{U_T}\right) . \tag{B.29}$$

The parameter V_x is equal to 0 at the source end and equal to V_{DS} at the drain end of the gate depleted region. The value $e(V_x + V_{SB})$ is the difference between the Fermi level in equilibrium and the electron quasi-Fermi level, see fig. B.10. The electron channel density is small in weak inversion and diffusion current dominates.

In the case of strong inversion, the density of electrons in the channel is large. There is a low-ohmic connection between the majority electrons in n^+ areas (source and drain) and the electrons in the channel. If we apply the voltage V_{SB}, the channel potential in the vicinity of source increases nearly by the value V_{SB} with respect to its equilibrium potential. Similarly, the channel potential close to drain has a value higher by V_{DB} than

[7] Note that the Boltzman distribution for charge carriers implies the equality between drift and diffusion currents.

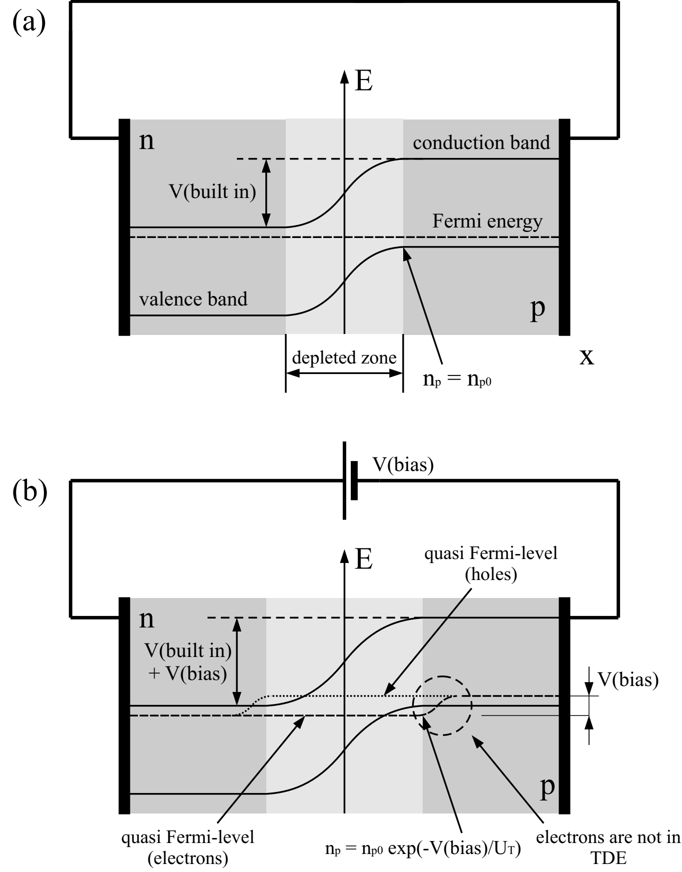


Figure B.8: Energy levels in a $p-n$ diode. (a) - $p-n$ diode in equilibrium. (b) - inversely polarized $p-n$ diode.

in equilibrium. The energy levels in the channel region are shown in fig. B.11. The surface potential is, approximately, equal to

$$\psi_S \approx \psi_0 = U_T \ln(N_a/n_p) = 2U_T \ln(N_a/n_i) + V_{SB} + V_x \ . \tag{B.30}$$

The voltage V_x is equal to 0 at the source end of the channel and equal to V_{DS} at the drain channel end. $V_x + V_{SB}$ is the channel potential with respect to the potential in equilibrium.

The value $2U_T \ln(N_a/n_i)$ is usually denoted by 2Φ. We will rewrite eq. B.31 as

$$\psi_S \approx \psi_0 = 2\Phi^* + V_{SB} + V_x \ . \tag{B.31}$$

The star symbol $(2\Phi^*)$ remind us to add a few thermal voltages to the value $2U_T \ln(N_a/n_i)$, depending on the model we use. In the following text, the exact value $2U_T \ln(N_a/n_i)$ will be denoted with 2Φ and the value increased by a few thermal voltages with $2\Phi^*$. For the typical values $N_a = 1.7 \cdot 10^{17} cm^{-3}$ and $n_i = 1.45 \cdot 10^{10} cm^{-3}$, we have $2\Phi = 0.85\ V$

We can now substitute $\psi_0 = 2\Phi^* + V_{SB} + V_x$ in the formula for the channel charge in strong inversion (eq. B.28). It follows

$$\begin{aligned} Q'_{ch} &= C'_{ox}(V_{GB} - V_{fb} - 2\Phi^* - V_{SB} - V_x) - \\ &- Q'_{dep}(2\Phi^* + V_{SB} + V_x) - Q'_{it}(2\Phi^* + V_{SB} + V_x) \ . \end{aligned} \tag{B.32}$$

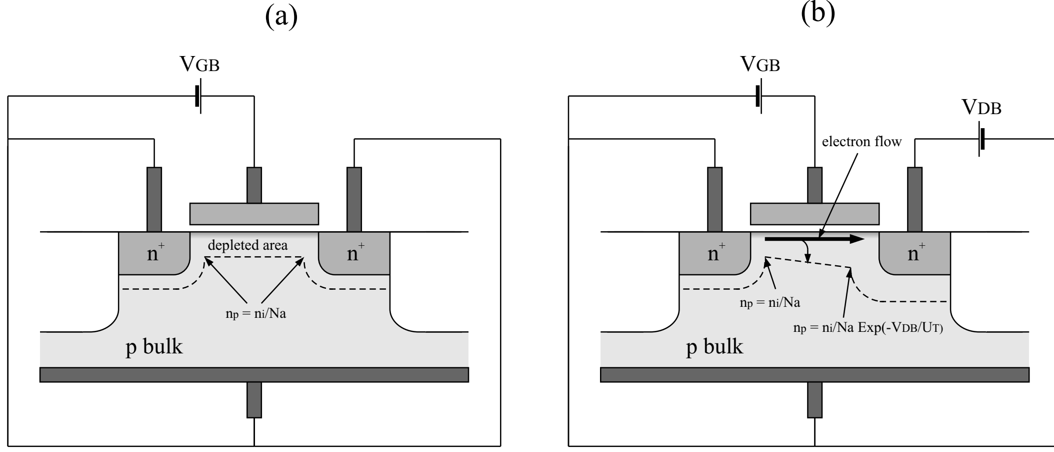


Figure B.9: (a) - MOS transistor in equilibrium. (b) - MOS transistor after the application of the drain bias.

If we define the threshold V_{th} as

$$V_{th} \equiv V_{fb} + 2\Phi^* + \frac{Q'_{dep}(2\Phi^* + V_{SB}) + Q'_{it}(2\Phi^* + V_{SB})}{C'_{ox}} , \tag{B.33}$$

we have

$$Q'_{ch} = C'_{ox}\left[V_{GS} - V_{th} - \left(V_x + \frac{(Q'_{dep} + Q'_{it})_{2\Phi^*+V_{SB}+V_x} - (Q'_{dep} + Q'_{it})_{2\Phi^*+V_{SB}}}{C'_{ox}}\right)\right] . \tag{B.34}$$

If we use the result eq. B.22, we can write

$$V_{th} \equiv V_{fb} + 2\Phi^* + \gamma\sqrt{2\Phi^* + V_{SB}} + \frac{Q'_{it}(2\Phi^* + V_{SB})}{C'_{ox}}; \quad \gamma \equiv \frac{\sqrt{2e\epsilon_{Si}N_a}}{C'_{ox}} \tag{B.35}$$

and

$$Q'_{ch} = C'_{ox}\left[V_{GS} - V_{th} - \left(V_x + \gamma\sqrt{2\Phi^* + V_{SB} + V_x} - \gamma\sqrt{2\Phi^* + V_{SB}} + \frac{\Delta Q'_{it}}{C'_{ox}}\right)\right] , \tag{B.36}$$

where $\Delta Q'_{it} \equiv (Q'_{it})_{2\Phi^*+V_{SB}+V_x} - (Q'_{it})_{2\Phi^*+V_{SB}}$. Eq. B.36 is valid only if the term in square brackets is considerably larger than zero, otherwise we do not have strong inversion.

For the typical values, listed above $\gamma = 0.78\ V^{1/2}$. If we assume $V_{fb} = V_{pn} = -1\ V$, $V_{SB} = 0\ V$, $2\Phi^* = 2\Phi = 0.8\ V$ and neglect Q'_{it}, we obtain $V_{th} = 0.57\ V$.

B.1.4 Drain-Current Equations

In the last section we derived the formulas for the channel charge Q'_{ch} in the case of weak (eq. B.29) and strong inversion (eq. B.36). The next step is to express the drain-source current as function of the terminal voltages V_{GS}, V_{DS} and V_{BS}. We will neglect the recombination and generation in the channel. Due to charge conservation, the conditon $I_{DS}(x) = \text{const} \equiv I_{DS}$ holds.

In the case of weak inversion, diffusion current dominates, therefore

$$I_{DS} = -WD_n\frac{dQ'_{ch}}{dx} . \tag{B.37}$$

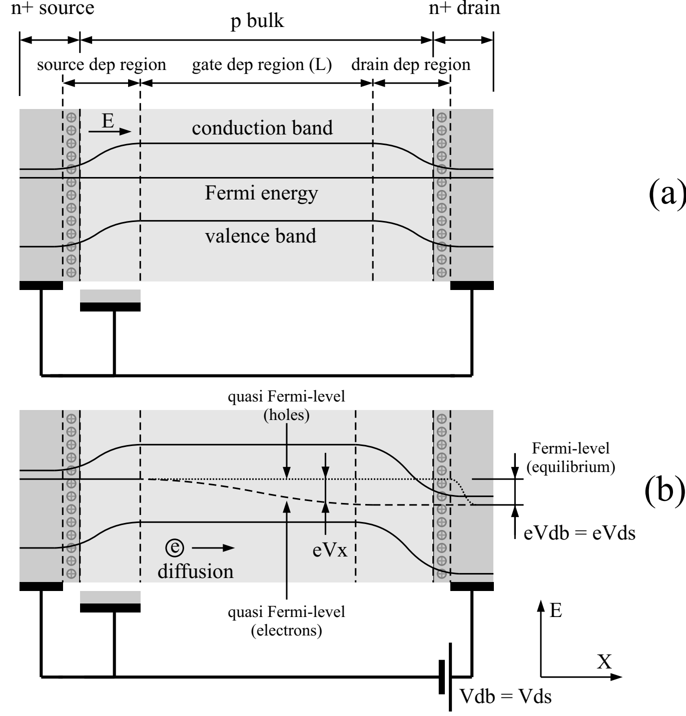


Figure B.10: Transistor in weak inversion, energy levels are shown. Figure (a) shows the channel region in equilibrium. Figure (b) shows the channel region after application of a positive drain-bulk voltage. The source-bulk voltage is zero.

W is the gate width (fig. B.2) and D_n is the diffusion constant.

When we substitute the formula for the channel charge (eq. B.29) into eq. B.37 and differentiate the term Q'_{ch}, we obtain

$$I_{DS} = \frac{W D_n}{U_T} Q'_{ch} \frac{dV_x}{dx} = W \mu_n Q'_{ch} \frac{dV_x}{dx} \ . \tag{B.38}$$

We used Einstein's equation $(D/\mu) = kT/e \equiv U_T$, where D is the diffusion constant and μ mobility.

Eq. B.38 can be multiplied by dx and then integrated (recall that $I_{DS}(x) = \text{const}$)

$$I_{DS} dx = W \mu_n Q'_{ch} dV_x \tag{B.39}$$

$$\int\limits_0^L I_{DS} dx = I_{DS} L = \int\limits_0^{V_{DS}} W \mu_n Q'_{ch} dV_x \ . \tag{B.40}$$

Let us now derive the formula for the drain-source current in the case of strong inversion. In strong inversion, the electron density in the channel is high and drift current dominates. Therefore

$$I_{DS} = -W \mu_n Q'_{ch} E_x = W \mu_n Q'_{ch} \frac{dV_x}{dx} \ . \tag{B.41}$$

The channel charge Q'_{ch} is given by eq. B.36. Note, the formula for the drain-source current in strong inversion eq. B.41 has the form same as the formula in the case of

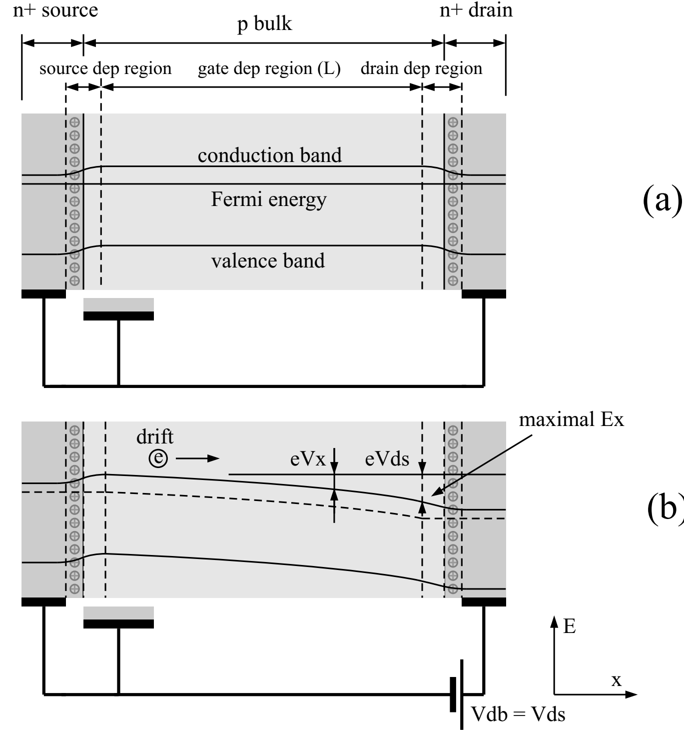


Figure B.11: Transistor in strong inversion. Figure (a) shows energy levels in the bulk assuming equilibrium. Figure (b) shows bulk energy levels after application of a positive drain-bulk voltage. The source-bulk voltage is zero.

weak inversion eq. B.38. Eq. B.40 is therefore valid in both regions of operation. Typical electron mobility is $\mu_n = 670\ cm^2/(Vs)$ and typical hole mobility $\mu_p = 250\ cm^2/(Vs)$ [52].

Drain Current in Weak Inversion

We will now solve the integral in eq. B.40 assuming weak inversion. If we substitute the formula for Q'_{ch} (eq. B.29) into eq. B.40, we get

$$I_{DS} = \mu_n e U_T \sqrt{\frac{\epsilon_{Si}}{2eN_a\psi_S}} \frac{n_i^2}{N_a} \frac{W}{L} \exp\left(\frac{\psi_S - V_{SB}}{U_T}\right) \int_0^{V_{DS}} \exp\left(\frac{-V_x}{U_T}\right) dV_x \tag{B.42}$$

and (see, for example, [1])

$$I_{DS} = \mu_n e U_T^2 \sqrt{\frac{\epsilon_{Si}}{2eN_a\psi_S}} \frac{n_i^2}{N_a} \frac{W}{L} \exp\left(\frac{\psi_S - V_{SB}}{U_T}\right) \left(1 - \exp\left(\frac{-V_{DS}}{U_T}\right)\right) \ . \tag{B.43}$$

Let us now express the surface potential ψ_S as a function of the gate potential V_{GB}. We start from eq. B.24. The solution of this quadratic equation is plotted in fig. B.6. As we see ψ_S versus V_{GB} is a nearly linear function in weak inversion region which is defined by the condition in eq. B.21 ($0.42\ V < \psi_S < 0.85\,\mathrm{V}$). Because of this, we will linearize eq. B.23 and solve it approximately. The terms that depend nonlinearly on ψ_S are usually linearized in the vicinity of the surface voltage ψ_S which corresponds to the upper limit

of weak inversion at the source channel end:

$$\psi_S = 2\Phi + V_{SB}; \quad 2\Phi \equiv 2U_T \ln\left(N_a/n_i\right) \ . \tag{B.44}$$

Eq. B.23 then becomes

$$\begin{aligned} V_{GB} &\approx V_{fb} + 2\Phi + V_{SB} + \frac{Q'_{dep}(2\Phi + V_{SB})}{C'_{ox}} + \frac{Q'_{it}(2\Phi + V_{SB})}{C'_{ox}} + \\ &+ \frac{1}{C'_{ox}}\left[C'_{ox} + \left(\frac{dQ'_{dep}}{d\psi_S}\right)_{2\Phi+V_{SB}} + \left(\frac{dQ'_{it}}{d\psi_S}\right)_{2\Phi+V_{SB}}\right](\psi_S - 2\Phi - V_{SB}) \ . \end{aligned} \tag{B.45}$$

The term $\left(\frac{dQ'_{dep}}{d\psi_S}\right)_{2\Phi+V_{SB}}$ is equal to the small-signal capacitance of the depletion zone:

$$C'_{dep} \equiv C'_{dep}(2\Phi + V_{SB}) = \left(\frac{dQ'_{dep}}{d\psi_S}\right)_{2\Phi+V_{SB}} \ . \tag{B.46}$$

The term $\left(\frac{dQ'_{it}}{d\psi_S}\right)_{2\Phi+V_{SB}}$ is the small-signal interface-trap capacitance:

$$C'_{it} \equiv C'_{it}(2\Phi + V_{SB}) = \left(\frac{dQ'_{it}}{d\psi_S}\right)_{2\Phi+V_{SB}} \ . \tag{B.47}$$

We defined the threshold voltage as (eq. B.33)

$$V_{th} \equiv V_{fb} + 2\Phi^* + \frac{Q'_{dep}(2\Phi^* + V_{SB})}{C'_{ox}} + \frac{Q'_{it}(2\Phi^* + V_{SB})}{C'_{ox}} \ . \tag{B.48}$$

It holds approximatively

$$\begin{aligned} V_{th} &= V_{fb} + 2\Phi + \frac{Q'_{dep}(2\Phi + V_{SB})}{C'_{ox}} + \frac{Q'_{it}(2\Phi + V_{SB})}{C'_{ox}} \\ &+ \frac{1}{C'_{ox}}\left[C'_{ox} + \left(\frac{dQ'_{dep}}{d\psi_S}\right)_{2\Phi+V_{SB}} + \left(\frac{dQ'_{it}}{d\psi_S}\right)_{2\Phi+V_{SB}}\right](2\Phi^* - 2\Phi) \ . \end{aligned}$$

If we substitute the last equation into eq. B.45, we obtain

$$\psi_S = 2\Phi + \Delta_\Phi + V_{SB} + \frac{C'_{ox}}{C'_{ox} + C'_{dep} + C'_{it}}(V_{GS} - V_{th}) \ , \tag{B.49}$$

where $\Delta_\Phi \equiv 2\Phi^* - 2\Phi$. It is useful to define the slope factor n as

$$n \equiv \frac{C'_{ox} + C'_{dep} + C'_{it}}{C'_{ox}} = 1 + \frac{C'_{dep} + C'_{it}}{C'_{ox}} \ . \tag{B.50}$$

Therefore we have

$$\psi_S = 2\Phi + \Delta_\Phi + V_{SB} + \frac{V_{GS} - V_{th}}{n} \ . \tag{B.51}$$

The capacitance $C'_{dep}(2\Phi + V_{SB}) = \left(\frac{dQ'_{dep}}{d\psi_S}\right)_{2\Phi+V_{SB}}$ can be calculated by using the approximative formula for Q'_{dep}, eq. B.22. Assuming the standard BSIM4 values [52], we

obtain $C'_{dep}(2\Phi + V_{SB}) \approx 0.42 \cdot C'_{ox} = 1.25\ fF/\mu\text{m}^2$. If we neglect the capacitance due to interface traps C'_{it}, we have $n = 1.42$.

The slope factor n is very important parameter. Its value depends on the value of the surface voltage, where we linearize $Q'_{dep}(\psi_S)$ (and $Q'_{it}(\psi_S)$). We have linearized the function $Q'_{dep}(\psi_S)$ in the point $\psi_S = 2\Phi + V_{SB}$. This point corresponds to the upper limit of weak inversion. As illustration we give the n value for $\psi_S = U_T \ln(N_a/n_i) = 0.42\ V$ (lower limit of weak inversion). The capacitance $C'_{dep}(0.42V) = \left(\frac{dQ'_{dep}}{d\psi_S}\right)_{0.42V}$ is $C'_{dep}(0.42V) = 0.60 \cdot C'_{ox} = 1.77\ fF/\mu\text{m}^2$. The slope factor is $n(0.42V) = 1.60$, about 13% larger than calculated before.

When we substitute ψ_S, eq. B.51 into the equation for the drain-source current eq. B.43, we obtain (see, for example [1])

$$I_{DS} = \mu_n U_T^2 \sqrt{\frac{\epsilon_{Si} e N_a}{2\psi_S}} \frac{W}{L} \exp\left(\frac{\Delta_\Phi}{U_T}\right) \exp\left(\frac{V_{GS} - V_{th}}{nU_T}\right) \left(1 - \exp\left(\frac{-V_{DS}}{U_T}\right)\right) . \quad \text{(B.52)}$$

The factor in the square root is equal to the capacitance of the depleted area assuming the surface potential equal to ψ_S

$$\sqrt{\frac{\epsilon_{Si} e N_a}{2\psi_S}} = \frac{d}{d\psi_S}\left[\sqrt{\epsilon_{Si} e N_a 2\psi_S}\right] = \frac{dQ'_{dep}(\psi_S)}{d\psi_S} = C'_{dep}(\psi_S) . \quad \text{(B.53)}$$

We used here the result given by eq. B.22. We can therefore write

$$I_{DS} = U_T^2 \mu_n C'_{dep}(\psi_S) \frac{W}{L} \exp\left(\frac{\Delta_\Phi}{U_T}\right) \exp\left(\frac{V_{GS} - V_{th}}{nU_T}\right) \left(1 - \exp\left(\frac{-V_{DS}}{U_T}\right)\right) . \quad \text{(B.54)}$$

If we assume that $C'_{dep}(\psi_S) \approx C'_{dep}(2\Phi + V_{SB}) = (n-1)C'_{ox}$, we obtain the commonly used formula in weak inversion current [1]:

$$I_{DS} = U_T^2 (n-1) \mu_n C'_{ox} \frac{W}{L} \exp\left(\frac{\Delta_\Phi}{U_T}\right) \exp\left(\frac{V_{GS} - V_{th}}{nU_T}\right) \left(1 - \exp\left(\frac{-V_{DS}}{U_T}\right)\right) . \quad \text{(B.55)}$$

Although the last formula (eq. B.55) has a more compact form than eq. B.52, it can lead to the wrong conclusion that weak inversion current depends on the gate capacitance C'_{ox}. We must, however, keep in mind that the term $n-1$ is inversely proportional to C'_{ox} [1].

For the typical parameter values that we use, we obtain $U_T^2(n-1)\mu_n C'_{ox} \approx 56.1\ nA$ and $U_T^2(n-1)\mu_p C'_{ox} \approx 20.9\ nA$

Fig. B.12 shows the drain-source current versus the drain-source voltage. We see that the current saturates when V_{DS} exceeds several U_T. For higher drain potentials, a transistor in weak inversion behaves like a current source. The drain-source current can be controlled by V_{GS}. Transistors in the current saturation mode are very often used. They are, for example, the main building parts of electronic amplifiers.

B.1.5 Drain Current in Strong Inversion

We will now derive a few formulas for the drain-source current in strong inversion. There are a few approaches how to solve the integral in eq. B.40. The first approach is referred to as the variable depletion layer model.

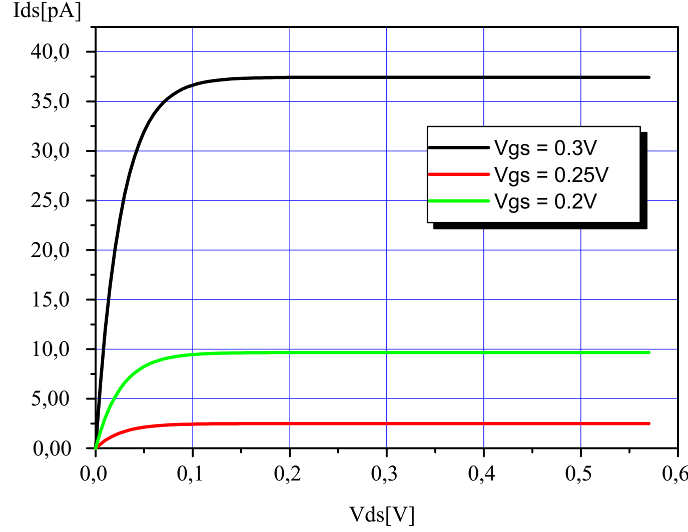


Figure B.12: Drain-source current as function of drain-source voltage. The current is given by eq. B.55. Standard parameter are assumed, $\Delta_\Phi = 0$.

Variable Depletion-Layer Model

We can substitute the nonlinear formula for $Q'_{ch}(V_x)$ (eq. B.36) into eq. B.40 and calculate the integral by neglecting the interface charge term. The result is [1]

$$\begin{aligned} I_{DS} &= \mu_n C'_{ox} \frac{W}{L} \left(V_{GS} - V_{fb} - 2\Phi^* - \frac{V_{DS}}{2} \right) V_{DS} - \\ &- \mu_n C'_{ox} \frac{W}{L} \frac{2}{3} \gamma \left[(V_{DS} + 2\Phi^* + V_{SB})^{3/2} - (2\Phi^* + V_{SB})^{3/2} \right] . \qquad \text{(B.56)} \end{aligned}$$

The drain-source current function $I_{DS}(V_{DS})$ (eq. B.56) increases until a saturation voltage (V_{DSsat}) is reached. The classical definition of the drain-source saturation voltage tells that it is the voltage for which the channel charge at the drain side of the channel becomes zero. The channel charge is given by eq. B.36. According to this definition, V_{DSsat} satisfies the equation $Q'_{ch}(V_x = V_{DS} = V_{DSsat}) = 0$. As we explained, at the drain end of the channel, the voltage V_x is equal to V_{DS}, see fig. B.11.[8] By solving the equation $Q'_{ch}(V_{DSsat}) = 0$, we obtain

$$V_{DSsat} = V_{GS} - V_{fb} - 2\Phi^* + \frac{\gamma^2}{2} \left[1 - \sqrt{1 + \frac{4}{\gamma^2} (V_{GB} - V_{fb})} \right] . \qquad \text{(B.57)}$$

The drain-source saturation current is simply $I_{DSsat} = I_{DS}(V_{DSsat})$. The equations eq. B.56 and eq. B.57 are implemented in the SPICE Level 2 model.

Quadratic Model

If we neglect the square root terms in the formula for the channel charge (eq. B.36), we have

$$Q'_{ch} \approx C'_{ox}(V_{GS} - V_{th}) . \qquad \text{(B.58)}$$

[8] Recall that the formula for the channel charge (eq. B.36) looses accuracy when Q'_{ch} approaches zero. The strong inversion assumption is then not valid. The correctness of the condition $Q'_{ch}(V_{DSsat}) = 0$ is therefore suspect.

Note that the threshold voltage defined by eq. B.35 still contains the square root terms and still depends on V_{SB}. When we substitute eq. B.58 into eq. B.40 and solve the integral, we obtain the well know formula for the transistor current

$$I_{DS} = \mu_n C'_{ox} \frac{W}{L} \left(V_{GST} V_{DS} - \frac{V_{DS}^2}{2} \right); \quad V_{GST} \equiv V_{GS} - V_{th} \ . \tag{B.59}$$

The saturation voltage V_{DSsat} can be calculated by using the equation $Q'_{ch}(V_{DSsat}) = C'_{ox}(V_{GS} - V_{th}) = 0$. Therefore $V_{DSsat} = V_{GST}$. The saturation current is

$$I_{DSsat} \equiv I_{DS}(V_{DSsat}) = \frac{\mu_n C'_{ox}}{2} \frac{W}{L} V_{GST}^2 \ . \tag{B.60}$$

The equations eq. B.59 and eq. B.60 are the basic equations of the SPICE Level 1 model. Fig. B.13 shows the drain-source current versus the drain-source voltage. The solid lines correspond to the quadratic model, eq. B.59. The dashed lines represent the variable depletion layer model, eq. B.56. Supposing that the variable depletion layer model is more accurate, since it takes into account the square root terms in the Q'_{ch} expression, we can conclude that the accuracy of the quadratic model is limited to small drain-source voltages. The I_{DSsat} value, calculated by using eq. B.60, is significantly overestimated.

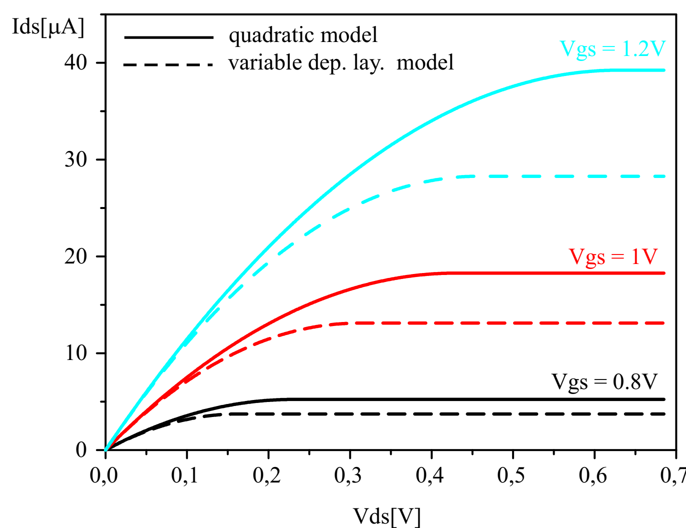


Figure B.13: I_{DS} versus V_{DS}. Dashed lines correspond to eq. B.56 (variable depletion layer model). Solid lines represent eq. B.59 (quadratic model).

BSIM Model

In this section we will calculate the integral in eq. B.40 following the approach used in the Berkeley short-channel IGFET[9] model (BSIM). The BSIM model [52, 53, 54] is the industry standard MOSFET model for the deep-submicron digital and analog circuit design. It takes into account the effects of carrier mobility reduction and drift velocity saturation.

[9]Insulated Gate FET.

Generally, mobility depends on many process and bias parameters. The mobility of electrons in the channel decreases with the increase of the vertical field component E_y. The mobility is reduced due to the fact that the electron wave function extends more into the oxide when the vertical field becomes stronger. The overall mobility is lowered due to lower mobility in the oxide. The mobility reduction is modelled with the following equation [1, 52]:

$$\mu_{eff} = \frac{\mu_0}{1 + \left(\frac{\langle E_y \rangle}{E_0}\right)^{\nu}} \ . \tag{B.61}$$

μ_0 is the low field mobility, typically [52] $\mu_0 = 0.067\ m/Vs$ for n-channel MOSFET (electrons) and $\mu_0 = 0.025\ m/Vs$ for p-channel transistor (holes). E_0 is, typically, 67 V/μm (NMOS) and 70 V/μm (PMOS). $\nu = 1.6$ for NMOS and $\nu = 1.0$ for PMOS. The parameter $\langle E_y \rangle$ is the average vertical electric field in the channel. It is defined by the equation

$$\langle E_y \rangle \equiv \frac{1}{L} \int_0^L \frac{Q'_{ch}(x) + \eta Q'_{dep}(x)}{\epsilon_{SiO2}}\, dx \equiv \frac{\langle Q'_{ch} \rangle + \eta \langle Q'_{dep} \rangle}{\epsilon_{SiO2}} \ . \tag{B.62}$$

η is constant, $\eta = 1/2$ for n-channel device and $\eta = 1/3$ for PMOS [52]. Note, we defined the electric field E_y and the charges Q'_{ch} and Q'_{dep} to be positive when transistor operates under normal conditions (in inversion). The formulas for the mean channel and depletion charge ($\langle Q'_{ch} \rangle$ and $\langle Q'_{ch} \rangle$) will be derived later.

The drift velocity saturation is caused by increased scattering rate of high-energy carriers mainly due to optical phonon emission. The velocity saturation is modelled with the following equations

$$\begin{aligned} v &= \frac{\mu_{eff}}{1 + \frac{|E_x|}{E_{sat}}}; \quad |E_x| < E_{sat} \\ v &= v_{sat}; \quad |E_x| \geq E_{sat} \ . \end{aligned} \tag{B.63}$$

E_x is the longitudinal electric field component $E_x = -\frac{dV_x}{dx}$. E_{sat} is the electric field for which the carrier velocity saturates. From eq. B.63 follows

$$v_{sat} = \frac{\mu_{eff}}{2} E_{sat} \ . \tag{B.64}$$

Typical saturation velocity is $v_{sat} = 8 \cdot 10^4 m/s$ [52]. If we assume $\mu_{eff} \approx \mu_0$, by using the standard values $\mu_0 = 0.067\ m/Vs$ (NMOS) and $\mu_0 = 0.025\ m/Vs$ (PMOS), we obtain $E_{sat} = 2.4\ V/\mu$m for n-channel and $E_{sat} = 6.4\ V/\mu$m for p-channel transistor.

Let us now derive the formula for the drain-source current. When we substitute the formula for Q'_{ch} (eq. B.34 or eq. B.36) and the formula that describes the velocity saturation (eq. B.63) into eq. B.41, we obtain

$$I_{DS} = \frac{\mu_{eff} W C'_{ox}}{1 + \frac{\frac{dV_x}{dx}}{E_{sat}}} \left[V_{GST} - \left(V_x + \frac{(Q'_{dep})_{2\Phi^* + V_{SB} + V_x} - (Q'_{dep})_{2\Phi^* + V_{SB}}}{C'_{ox}} \right) \right] \frac{dV_x}{dx} \tag{B.65}$$

or

$$I_{DS} = \frac{\mu_{eff} W C'_{ox}}{1 + \frac{\frac{dV_x}{dx}}{E_{sat}}} \left[V_{GST} - \left(V_x + \gamma\sqrt{2\Phi^* + V_{SB} + V_x} - \gamma\sqrt{2\Phi^* + V_{SB}} \right) \right] \frac{dV_x}{dx} \ , \tag{B.66}$$

where $V_{GST} \equiv V_{GS} - V_{th}$. We neglected the interface charge Q'_{it}.

If we multiply the both sides of eq. B.66 with $\left(1 + \frac{\frac{dV_x}{dx}}{E_{sat}}\right) dx$ and integrate the left-hand side, we obtain

$$\begin{aligned} I_{DS} &= \frac{\mu_{eff} W C'_{ox}}{L\left(1 + \frac{V_{DS}}{LE_{sat}}\right)} \int_0^{V_{DS}} \left[V_{GST} - \left(V_x + \gamma\sqrt{2\Phi^* + V_{SB} + V_x} - \gamma\sqrt{2\Phi^* + V_{SB}}\right)\right] dV_x \\ &= \frac{\mu_{eff} W}{L\left(1 + \frac{V_{DS}}{LE_{sat}}\right)} \int_0^{V_{DS}} Q'_{ch}(V_x) dV_x \ . \end{aligned} \tag{B.67}$$

Although we can perform the integration of the channel charge without any approximation, we will linearize the square root factor $\sqrt{2\Phi^* + V_{SB} + V_x}$ in the vicinity of $V_x = 0$. Note, by linearizing the channel charge around its value at the source end we favoriz source over drain electrode and in this way break the inherent symmetry of the MOS transistor. The BSIM model is therefore a source referenced model.

Fig. B.14 shows the term in large brackets of eq. B.67 as a function of V_x (the term is equal to $Q'_{ch}(V_x)/C'_{ox}$). As we see, the curvature of the function $Q'_{ch}(V_x)$ is small and the term can be approximated with a linear expression very well. Therefore we have

$$\begin{aligned} Q'_{ch}/C'_{ox} &= V_{GST} - \left(V_x + \frac{(Q'_{dep})_{2\Phi^*+V_{SB}+V_x} - (Q'_{dep})_{2\Phi^*+V_{SB}}}{C'_{ox}}\right) \\ &\approx V_{GST} - V_x \left(1 + \frac{1}{C'_{ox}} \left(\frac{dQ'_{dep}}{dV_x}\right)_{V_x=0}\right) \ . \end{aligned} \tag{B.68}$$

The factor $\left(\frac{dQ'_{dep}}{dV_x}\right)_{V_x=0}$ is equal to the depletion capacitance $C'_{dep}(2\Phi^* + V_{SB}) = \left(\frac{dQ'_{dep}}{d\psi_S}\right)_{2\Phi^*+V_{SB}}$.[10]

Making analogy with the definition of the slope factor in the case of weak inversion (eq. B.50), we can define the n-factor in the case of strong inversion as

$$n^* \equiv 1 + \frac{C'^{\,*}_{dep} + C'^{\,*}_{it}}{C'_{ox}} \ , \tag{B.69}$$

where

$$C'^{\,*}_{dep} \equiv C'_{dep}(2\Phi^* + V_{SB}) = \left(\frac{dQ'_{dep}}{d\psi_S}\right)_{2\Phi^*+V_{SB}} \tag{B.70}$$

and

$$C'^{\,*}_{it} \equiv C'_{it}(2\Phi^* + V_{SB}) = \left(\frac{dQ'_{it}}{d\psi_S}\right)_{2\Phi^*+V_{SB}} \ . \tag{B.71}$$

Note, if $2\Phi = 2\Phi^*$, than $n = n^*$. Finally we can write

$$Q'_{ch}/C'_{ox} \approx V_{GST} - n^* V_x \ . \tag{B.72}$$

When we substitute eq. B.72 into eq. B.67, we have [52]

$$\begin{aligned} I_{DS} &= \frac{\mu_{eff} C'_{ox} W}{L\left(1 + \frac{V_{DS}}{LE_{sat}}\right)} \int_0^{V_{DS}} \left(V_{GST} - n^* V_x\right) dV_x \\ &= \frac{\mu_{eff} C'_{ox} W}{L\left(1 + \frac{V_{DS}}{LE_{sat}}\right)} \left(V_{GST} V_{DS} - n^* \frac{V_{DS}^2}{2}\right) \ . \end{aligned} \tag{B.73}$$

[10] $Q'_{dep}(\psi_S) \equiv C'_{ox}\gamma\sqrt{\psi_S}$ and $Q'_{dep}(V_x) \equiv C'_{ox}\gamma\sqrt{2\Phi^* + V_{SB} + V_x}$.

In the section about weak inversion we have calculated the slope factor n. Assuming the standard parameters and $2\Phi = 2\Phi^*$, we have $n^* = n = 1.42$. If we assume the gate length $L = 0.5\,\mu$m, we have for an NMOS: $E_{sat}L = 1.2\ V$ and for a PMOS: $E_{sat}L = 3.2\ .V$

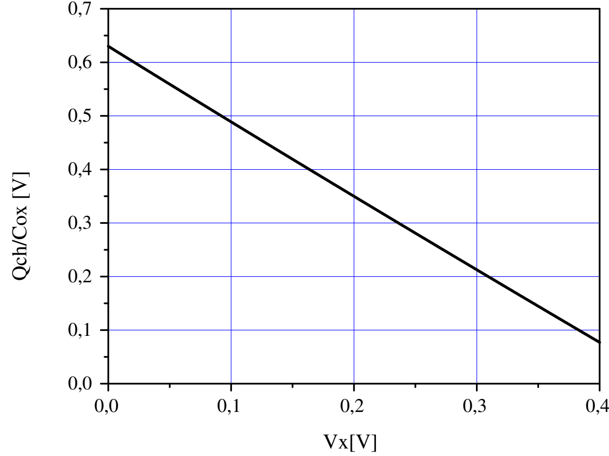


Figure B.14: $Q'_{ch}(V_x)$, given by eq. B.36, versus V_x .

The drain-source current equation (eq. B.73) is valid provided that the condition $|E_x| = \frac{dV_x}{dx} < E_{sat}$ holds all along the channel. The longitudinal field $|E_x|$ has the highest magnitude at the drain end of the channel area, see fig. B.11. When the field at the drain channel end exceeds E_{sat}, electron velocity saturates. We define the saturation voltage V_{DSsat} as the V_{DS} voltage for which the velocity saturation occurs. Let us recall the conventional definition of V_{DSsat}. The drain-source saturation voltage is defined as the voltage when the channel charge at the drain end becomes zero: $Q'_{ch}(V_x = V_{DS} = V_{DSsat}) = 0$ This condition is not realistic for short channel devices [54]. As we already mentioned, Q'_{ch} formulas in strong inversion cannot be applied in the case $Q'_{ch} \approx 0$. Therefore it is more realistic to assume that the drain-source current is limited by the effect of velocity saturation. After onset of saturation, $I_{DS}(V_{DS})$ increases only slightly due to the growth of the drain depleted area and due to the decrease of the channel length L.[11] Fig. B.15 shows the energy levels in the channel area in the case $V_{DS} = V_{DSsat}$ and $V_{DS} > V_{DSsat}$. As we see, the drain depleted area grows when we increase the drain-source voltage. The drain-source saturation voltage V_{DSsat} can be calculated in the following way. We have

$$I_{DSsat} = v_{sat} W Q'_{ch}(V_{DSsat}) = \frac{\mu_{eff}}{2} E_{sat} W C'_{ox} \cdot (V_{GST} - n^* V_{DSsat}) \ . \qquad \text{(B.74)}$$

Here, we used the results given by eq. B.63, eq. B.64 and eq. B.72. The formula eq. B.73 can be also used in order to express the saturation current

$$I_{DSsat} = I_{DS}(V_{DSsat}) = \frac{\mu_{eff} C'_{ox} W}{L\left(1 + \frac{V_{DSsat}}{L E_{sat}}\right)} \left(V_{GST} V_{DSsat} - n^* \frac{V_{DSsat}^2}{2}\right) \ . \qquad \text{(B.75)}$$

[11] As we see from formulas for drain-source current, the current is inversely proportional to the length of the channel.

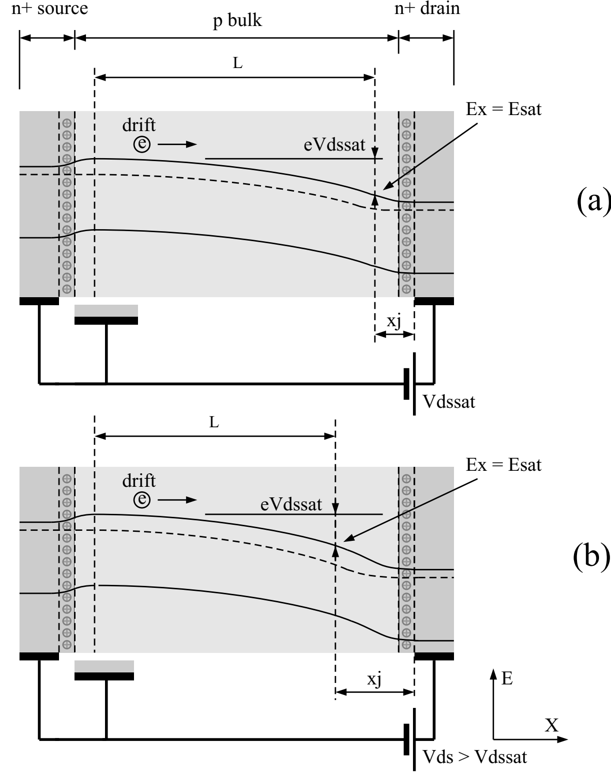


Figure B.15: Energy levels in the channel area (strong inversion). Figure (a) shows the energy levels assuming $V_{DS} = V_{DSsat}$, figure (b) shows the energy levels in saturation.

V_{DSsat} can be calculated by solving the following equation

$$\frac{\mu_{eff}}{2} E_{sat} W C'_{ox} (V_{GST} - n^* V_{DSsat}) = \frac{\mu_{eff} C'_{ox} W}{L\left(1 + \frac{V_{DSsat}}{LE_{sat}}\right)} \left(V_{GST} V_{DSsat} - n^* \frac{V_{DSsat}^2}{2} \right) \ . \quad \text{(B.76)}$$

The result is

$$V_{DSsat} = \frac{V_{GST}}{n^* \left(1 + \frac{V_{GST}}{n^* L E_{sat}}\right)} \ . \quad \text{(B.77)}$$

When we substitute V_{DSsat} (eq. B.77) into the equation for the drain-source current (eq. B.75) we obtain [52]

$$I_{DSsat} = \frac{\mu_{eff} C'_{ox} W}{L\left(1 + \frac{V_{GST}}{n^* L E_{sat}}\right)} \frac{V_{GST}^2}{2n^*} \ . \quad \text{(B.78)}$$

Channel-Length Modulation

In this section we will estimate the increase of the drain-source current when V_{DS} exceeds V_{DSsat} which is result of channel length modulation. The decrease of the channel length (L) is illustrated in fig. B.15. In the case of asymmetrically doped $p - n$ diode with a sharp doping profile, the depleted layer thickness is given by the equation

$$x_{pn} = \sqrt{\frac{2\epsilon_{Si}}{eN_a}} \sqrt{V_{bi} + V_{bias}} \ . \quad \text{(B.79)}$$

$V_{bi} + V_{bias}$ is the voltage drop in the depleted zone. This formula has been derived under assumption that the electric field at the edge of depleted area is equal to zero. In the case od drain depleted region, the electric field is, however, equal to E_{sat} (assuming current saturation). By using Poisson's equation a formula for drain depletion thickness (x_j) similar to eq. B.79 can be derived:

$$x_j = \sqrt{\frac{2\epsilon_{Si}}{eN_a}} \left[\sqrt{V_0 + V_{sat} + (V_{DS} - V_{DSsat})} - \sqrt{V_{sat}} \right] . \tag{B.80}$$

V_0 is a parameter that describes the thickness x_j when $V_{DS} = V_{DSsat}$.[12] $\mathrm{V}_{sat} = \frac{\epsilon_{Si} E_{sat}^2}{2eN_a}$.

Let us first linearize the drain-source current equation close to the point $V_{DS} = V_{DSsat}$

$$I_{DS} = I_{DSsat} + \frac{\partial I_{DSsat}}{\partial L} \frac{\partial L}{\partial V_{DS}} (V_{DS} - V_{DSsat}) . \tag{B.81}$$

This equation is valid for $V_{DS} \geq V_{DSsat}$.

It is useful to define the parameter V_A (Early voltage):

$$I_{DS} = I_{DSsat} \left(1 + \frac{V_{DS} - V_{DSsat}}{V_A} \right) . \tag{B.82}$$

By comparing with eq. B.81 we have

$$V_A = \frac{I_{DSsat}}{\frac{\partial I_{DSsat}}{\partial L} \frac{\partial L}{\partial V_{DS}}} . \tag{B.83}$$

When we substitute the equation for the drain-source current in saturation (eq. B.78) into eq. B.83 and calculate its derivative, we obtain

$$V_A = L \left(1 + \frac{V_{GST}}{n^* L E_{sat}} \right) \frac{1}{\left| \frac{\partial L}{\partial V_{DS}} \right|} . \tag{B.84}$$

It is very often desirable to have a transistor with a high Early voltage. Provided V_A is high, the drain-source current is virtually insensitive to the V_{DS} variation and the transistor behaves nearly as an ideal current source. By using such transistors, we can make a high gain voltage amplifier. We see (eq. B.84) that the Early voltage depends linearly on L. If we want to design a transistor with high output resistance, its gate must be made long.

The derivative $\left| \frac{\partial L}{\partial V_{DS}} \right|$ can be estimated by using eq. B.80:

$$\begin{aligned} \frac{\partial L}{\partial V_{DS}} &= -\frac{\partial x_j}{\partial V_{DS}} = \frac{-\sqrt{\frac{2\epsilon_{Si}}{eN_a}}}{2\sqrt{V_0 + V_{sat} + (V_{DS} - V_{DSsat})}} = \frac{-\sqrt{\frac{2\epsilon_{Si}}{eN_a}} \sqrt{V_0 + V_{sat} + (V_{DS} - V_{DSsat})}}{2(V_0 + V_{sat} + (V_{DS} - V_{DSsat}))} \\ &= \approx \frac{-x_j^*}{2(V_0 + \mathrm{V}_{sat} + (V_{DS} - V_{DSsat}))} , \end{aligned} \tag{B.85}$$

where $x_j^* \equiv -\sqrt{\frac{2\epsilon_{Si}}{eN_a}} \sqrt{V_0 + V_{sat}}$. Therefore

$$\begin{aligned} V_A &= L \left(1 + \frac{V_{GST}}{n^* L E_{sat}} \right) \sqrt{\frac{2eN_a}{\epsilon_{Si}} (V_0 + V_{sat})} + L \left(1 + \frac{V_{GST}}{n^* L E_{sat}} \right) \frac{V_{DS} - V_{DSsat}}{x_j^*/2} \\ &= V_A(V_{DSsat}) + L \left(1 + \frac{V_{GST}}{n^* L E_{sat}} \right) \frac{V_{DS} - V_{DSsat}}{x_j^*/2} . \end{aligned} \tag{B.86}$$

[12] In [1] $V_0 = 0$.

$V_A(V_{DSsat})$ is the Early voltage for $V_{DS} = V_{DSsat}$. It holds

$$V_A(V_{DSsat}) = \frac{I_{DSsat}}{\frac{\partial I_{DSsat}}{\partial L}\left(\frac{\partial L}{\partial V_{DS}}\right)_{V_{DSsat}}} \ . \tag{B.87}$$

If we assume that the first derivative of the drain-source current function $I_{DS}(V_{DS})$ is continuous in $V_{DS} = V_{DSsat}$, the derivative $\frac{\partial I_{DSsat}}{\partial L}\left(\frac{\partial L}{\partial V_{DS}}\right)_{V_{DSsat}}$ must be equal to the derivative $\left(\frac{\partial I_{DS}}{\partial V_{DS}}\right)_{V_{DS}=V_{DSsat}}$, where $I_{DS}(V_{DS})$ is given by eq. B.73. Therefore

$$V_A(V_{DSsat}) = \frac{I_{DSsat}}{\left(\frac{\partial I_{DS}}{\partial V_{DS}}\right)_{V_{DSsat}}} \ . \tag{B.88}$$

If we calculate the derivative $\left(\frac{\partial I_{DS}}{\partial V_{DS}}\right)_{V_{DS}=V_{DSsat}}$, we obtain a very simple result

$$V_A(V_{DSsat}) = LE_{sat} + V_{DSsat} \ . \tag{B.89}$$

Finally, we have [52]

$$V_A = LE_{sat} + V_{DSsat} + L\left(1 + \frac{V_{GST}}{n^* LE_{sat}}\right)\frac{V_{DS} - V_{DSsat}}{x_j^*/2} \ . \tag{B.90}$$

Typical x_j^* value is $1.5 \cdot 10^{-7}\ m$ [52].

Fig. B.16 shows the drain-source current versus the drain-source voltage calculated by using the BSIM equations. In the region $V_{DS} < V_{DSsat}$ eq. B.73 has been used and for $V_{DS} \geq V_{DSsat}$ eq. B.82 and eq. B.90 have been used. In comparison, the variable depletion layer equations have been also plotted, eq. B.56. The difference between models is mainly the result of the velocity saturation effect which is taken into account only in the BSIM equations.

Overview

At the end of the chapter dedicated to the stationary transistor current equations, we will rewrite the most important formulas.

The drain-source current in weak inversion is given by

$$I_{DS} = U_T^2(n-1)\mu_n C'_{ox}\frac{W}{L}\exp\left(\frac{\Delta_\Phi}{U_T}\right)\exp\left(\frac{V_{GS}-V_{th}}{nU_T}\right)\left(1-\exp\left(\frac{-V_{DS}}{U_T}\right)\right) \ . \tag{B.91}$$

The slope factor n is defined as

$$n \equiv \frac{C'_{ox} + C'_{dep} + C'_{it}}{C'_{ox}} = 1 + \frac{C'_{dep} + C'_{it}}{C'_{ox}} \ , \tag{B.92}$$

where

$$C'_{dep} \equiv C'_{dep}(2\Phi + V_{SB}) = \left(\frac{dQ'_{dep}}{d\psi_S}\right)_{2\Phi+V_{SB}} \ ; \quad Q'_{dep}(\psi_S) = \sqrt{2e\epsilon_{Si}N_a\psi_S} \equiv C'_{ox}\gamma\sqrt{\psi_S} \ . \tag{B.93}$$

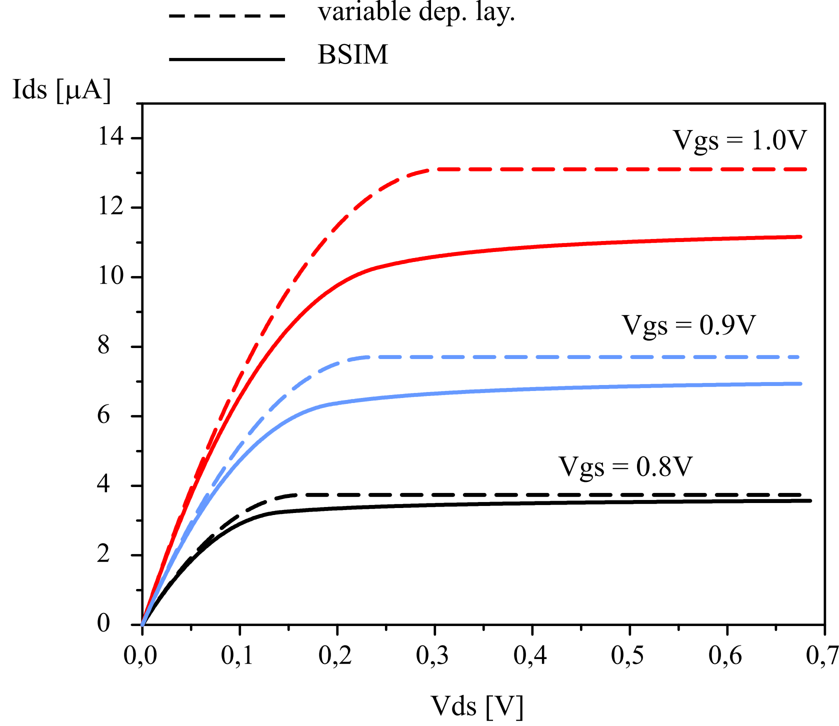


Figure B.16: Drain-source current versus V_{DS} in strong inversion. Solid line represent the BSIM equations eq. B.95 and eq. B.99. Gate length is 0.5 μm. Dashed lines represent the variable depleted layer model (eq. B.56). Difference between models is the result of velocity saturation. Standard parameters were used.

The threshold voltage V_{th} is

$$V_{th} \equiv V_{fb} + 2\Phi^* + \frac{Q'_{dep}(2\Phi^* + V_{SB}) + Q'_{it}(2\Phi^* + V_{SB})}{C'_{ox}} \ . \tag{B.94}$$

$\Delta_\Phi = 2\Phi^* - 2\Phi = 0 - 6U_T$; $2\Phi = 2U_T \ln\left(N_a/n_i\right)$. The drain-source current in strong inversion for $V_{DS} < V_{DSsat}$ is given by the equation

$$I_{DS} = \frac{\mu_{eff} C'_{ox} W}{L\left(1 + \frac{V_{DS}}{LE_{sat}}\right)} \left(V_{GST} V_{DS} - n^* \frac{V_{DS}^2}{2}\right) \ . \tag{B.95}$$

The effective mobility μ_{eff} is given by eq. B.61. The n-factor in strong inversion is defined as

$$n^* \equiv \frac{C'_{ox} + {C'_{dep}}^* + {C'_{it}}^*}{C'_{ox}} = 1 + \frac{{C'_{dep}}^* + {C'_{it}}^*}{C'_{ox}} \ , \tag{B.96}$$

where

$${C'_{dep}}^* \equiv C'_{dep}(2\Phi^* + V_{SB}) = \left(\frac{dQ'_{dep}}{d\psi_S}\right)_{2\Phi^* + V_{SB}} \ . \tag{B.97}$$

The saturation voltage V_{DSsat} is equal to

$$V_{DSsat} = \frac{V_{GST}}{n^*\left(1 + \frac{V_{GST}}{n^* LE_{sat}}\right)} \ . \tag{B.98}$$

The drain-source current for $V_{DS} \geq V_{DSsat}$ can be expressed as

$$I_{DS} = \frac{\mu_{eff} C'_{ox} W}{L\left(1 + \frac{V_{GST}}{n^* L E_{sat}}\right)} \frac{V_{GST}^2}{2n^*} \left(1 + \frac{V_{DS} - V_{DSsat}}{V_A}\right) \tag{B.99}$$

and the Early Voltage V_A is equal to

$$V_A = LE_{sat} + V_{DSsat} + L\left(1 + \frac{V_{GST}}{n^* L E_{sat}}\right) \frac{V_{DS} - V_{DSsat}}{x_j^*/2} \ . \tag{B.100}$$

The parameters used in the previous equations are explained in the following table. Also, the standard parameter values that we used many times so far are shown.

symbol	name	expression	typical value
U_T	thermal voltage	kT/e	$26mV$ $(300K)$
ϵ_{Si}	dielectric constant (Si)	$11\epsilon_0$	$11 \cdot 8.854 \cdot 10^{-12} F/m$
e	elementary charge		$1.602 \cdot 10^{-19} C$
N_a	acceptor density in bulk		$1.7 \cdot 10^{17} cm^{-3}$
n_i	intrinsic carrier density (Si)		$1.45 \cdot 10^{10} cm^{-3}$ $(300K)$
2Φ		$2U_T \ln(N_a/n_i)$	$0.85V$
$2\Phi^*$		$2U_T \ln(N_a/n_i) + 6U_T$	$1.01V$
C'_{ox}	oxide capacitance/unit area	ϵ_{SiO2}/t_{ox}	$2.951 fF/\mu\text{m}^2$
ϵ_{SiO2}	dielectric constant ($Si0_2$)	$3.9\epsilon_0$	$3.9 \cdot 8.854 \cdot 10^{-12} F/m$
t_{ox}	oxide thickness		$11.7nm$
γ	body effect coefficient	$\sqrt{2e\epsilon_{Si}N_a}/C'_{ox}$	$\gamma = 0.78$
μ_n	electron mobility		$0.067m^2/Vs$
μ_p	hole mobility		$0.025m^2/Vs$
W	gate width		
L	gate length		
E_{sat}	electron saturation field		$2.4V/\mu\text{m}$
E_{sat}	hole saturation field		$6.4V/\mu\text{m}$
V_{GST}		$V_{GS} - V_{th}$	
V_{th}	threshold voltage	see eq. B.35	$0.57V$ for $2\Phi^* = 2\Phi$
V_{fb}	flat-band voltage		$-1V$
x_j^*	junction depth		$150nm$
n	slope factor	see eq. B.92	1.42
n^*	n-factor	see eq. B.92	1.42 for $2\Phi^* = 2\Phi$

B.1.6 Short-Channel Effects

If the channel length is small, the use of the gradual channel approximation becomes suspect. The gradual channel approximation assumes negligible longitudinal electric field. As the channel length decreases, the widths of the source and drain depleted areas become comparable with channel dimensions and the electric field becomes effectively two-dimensional. The short channel effects are very pronounced in the present MOSFET technologies, since the transistor lengths are near $0.1\,\mu$m.[13] The mobility reduction, velocity

[13] The junction depth is, typically, $0.15\,\mu$m .

saturation, channel length modulation and the drain induced barrier-lowering (DIBL) are the most important short channel effects. We have already explained the first three effects when we derived the transistor equations that are part of the BSIM model.

The DIBL effect occurs when the drain-source distance becomes so small that the positively ionized acceptors in the n^+ regions can induce additional channel charge acting as a second gate. This is illustrated in fig. B.17. The DIBL effect causes a decrease of the threshold voltage by a value that is proportional to the drain-source voltage V_{DS}. The drain-source current in saturation increases as function of V_{DS}, which degrades the output current characteristic. The extreme case of DIBL is the punch through effect, which occurs when the drain and source depleted zones touch. The DIBL and punch-through effects can the diminished by increasing the substrate doping density.

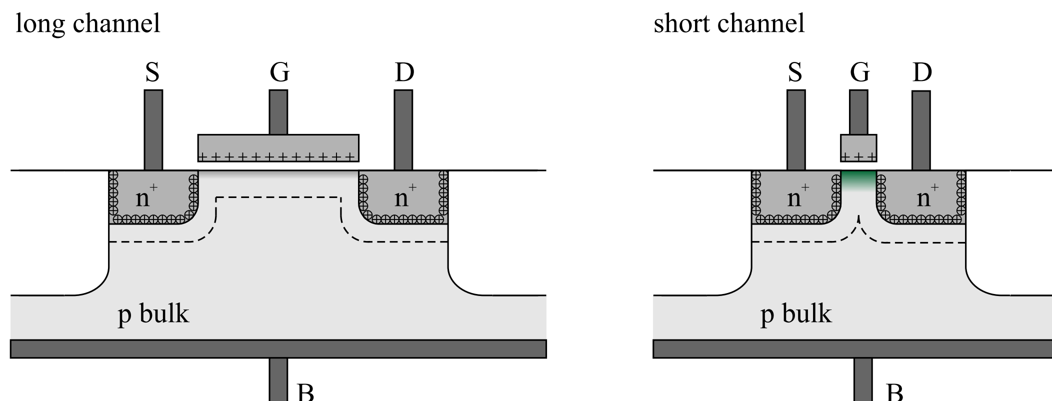


Figure B.17: Illustration of barrier-lowering effect.

B.2 Transient-Current Equations

In the previous sections we have derived the formulas for DC currents. These equations have the following form[14]

$$\begin{aligned} I_D &\approx I_{DS}(V_{GS}, V_{DS}, V_{BS}) \\ I_G &= 0 \\ I_B &\approx 0\ . \end{aligned} \tag{B.101}$$

After any change of terminal potentials, transistor needs some time to settle down. Physically it can be explained in the following way. Each stationary current is related to a specific arrangement of space charge regions inside MOS structure. In order to make a transition between stationary states, the charge in transistor has to be rearranged, which cannot occur instantaneously. In order to quantify the transient effects, we will calculate the amount of charge related to each transistor terminal.

The gate charge Q_G is the mirror charge of the channel and depleted zone charge, Q_{ch} and Q_{dep}. Q_{dep} is associated with the bulk contact, while the channel charge comes from

[14]We constrain us to the case $V_{DS}, V_{BS} > 0$.

source and drain.[15] Therefore we have

$$\begin{aligned} Q_G &= -(Q_{ch} + Q_{dep}) \\ Q_B &= Q_{dep} \\ Q_S &= Q_{ch} \ . \end{aligned} \tag{B.102}$$

In this chapter we define all charges to have the physically correct sign, for example $Q_{ch}, Q_{dep} < 0$ in inversion. The total charge Q_{ch} and Q_{dep} can be calculated by integration of the charge distributions $Q'_{ch}(x)$ and $Q'_{dep}(x)$ along the channel.

Strong Inversion

For the moment, we will constrain us to the case of strong inversion. In the previous section we derived the function $Q'_{ch}(V_x)$. The exact formula is given by eq. B.36 and the approximative by eq. B.72. Let us rewrite the approximative expression[16]

$$-Q'_{ch}(V_x) = C'_{ox}(V_{GST} - nV_x) \ . \tag{B.103}$$

The depleted zone charge is given by the following equation

$$\begin{aligned} -Q'_{dep}(V_x) &= \gamma\sqrt{2\Phi + V_{SB} + V_x} = -Q'_{dep}(V_x = 0) - \left(\frac{dQ'_{dep}}{dV_x}\right)_{V_x=0} V_x \\ &= \gamma\sqrt{2\Phi + V_{SB}} + C'_{ox}(n-1)V_x \ , \end{aligned} \tag{B.104}$$

where we used eq. B.93 and eq. B.92. We have

$$\begin{aligned} Q_{ch} &= -WC'_{ox}\int_0^L (V_{GST} - nV_x)dx = \\ &= -WLC'_{ox}V_{GST} + WC'_{ox}n\int_0^L V_x dx \end{aligned} \tag{B.105}$$

and

$$\begin{aligned} Q_{dep} &= -WC'_{ox}\int_0^L \left[\gamma\sqrt{2\Phi + V_{SB}} + C'_{ox}(n-1)V_x\right] dx = \\ &= -WLC'_{ox}\gamma\sqrt{2\Phi + V_{SB}} - WC'_{ox}(n-1)\int_0^L V_x dx \ . \end{aligned} \tag{B.106}$$

The gate charge is simply

$$Q_G = -(Q_{ch} + Q_{dep}) \ . \tag{B.107}$$

The integral $\int_0^L V_x dx$ can be solved by substitution $dx \rightarrow \frac{dx}{dV_x}dV_x$. The term $\frac{dx}{dV_x}$ can be calculated by using the formula for the drift current (we neglect the velocity saturation effect)

$$I_{DS} = W\mu_0|Q'_{ch}|\frac{dV_x}{dx} = W\mu_0 C'_{ox}(V_{GST} - nV_x)\frac{dV_x}{dx} \ . \tag{B.108}$$

[15] We neglect the charge in interface traps.

[16] To simplify the discussion we omit the stars in n^* and $2\Phi^*$.

Therefore

$$\begin{aligned}\int_0^L V_x dx &= \int_0^{V_{DS}} \frac{W\mu_0 C'_{ox}(V_{GST} - nV_x)}{I_{DS}} V_x dV_x = \\ &= \int_0^{V_{DS}} \frac{W\mu_0 C'_{ox}(V_{GST} - nV_x)}{\mu_0 C'_{ox}\frac{W}{L}\left(V_{GST}V_{DS} - n\frac{V_{DS}^2}{2}\right)} V_x dV_x \ . \end{aligned} \quad \text{(B.109)}$$

In order to derive the last expression, we used the equation for the drain-source current eq. B.95, where we neglected the terms that describe velocity saturation and mobility reduction. Integration of eq. B.109 leads to

$$\int_0^L V_x dx = L\left(\frac{V_{DS}}{2} - \frac{nV_{DS}^2}{12\left(V_{GST} - \frac{n}{2}V_{DS}\right)}\right) \ . \quad \text{(B.110)}$$

So far we can write the following formulas for the terminal charges in strong inversion [52]:

$$Q_S + Q_D = Q_{ch} = -WLC'_{ox}V_{GST} + nQ_0 \ , \quad \text{(B.111)}$$

$$Q_B = Q_{dep} = -WLC'_{ox}\gamma\sqrt{2\Phi + V_{SB}} - (n-1)Q_0 \ . \quad \text{(B.112)}$$

and

$$Q_G = -(Q_{ch} + Q_{dep}) = WLC'_{ox}V_{GST} + WLC'_{ox}\gamma\sqrt{2\Phi + V_{SB}} - Q_0 \ , \quad \text{(B.113)}$$

where

$$Q_0 \equiv WLC'_{ox}\left(\frac{V_{DS}}{2} - \frac{nV_{DS}^2}{12\left(V_{GST} - \frac{n}{2}V_{DS}\right)}\right) \ . \quad \text{(B.114)}$$

The formulas for the terminal charges are valid also in the case of current saturation. V_{DS} has to be then substituted by the drain-source saturation voltage:

$$V_{DS} = V_{DSsat} \approx \frac{V_{GST}}{n} \ . \quad \text{(B.115)}$$

It is generally not easy to calculate which part of the channel charge is delivered from source and which from the drain terminal. There are a few charge partitioning schemes all implemented within the scope of the BSIM transistor model. Physically most realistic model is based on following equations [52]:

$$\begin{aligned} Q_S &= W\int_0^L Q'_{ch}(x)\left(1 - \frac{x}{L}\right) dx \\ Q_D &= W\int_0^L Q'_{ch}(x)\frac{x}{L} dx \ . \end{aligned} \quad \text{(B.116)}$$

When we substitute the formula for Q'_{ch} (eq. B.105), by using the substitution $dx \rightarrow \frac{dx}{dV_x}dV_x$, we can solve the integrals in eq. B.116 [52]:

$$Q_S = -WLC'_{ox}\frac{\left(V_{GST}^3 - \frac{4}{3}V_{GST}^2 nV_{DS} + \frac{2}{3}V_{GST}(nV_{DS})^2 - \frac{2}{15}(nV_{DS})^3\right)}{2\left(V_{GST} - \frac{n}{2}V_{DS}\right)^2}$$

$$Q_D = -WLC'_{ox}\frac{\left(V_{GST}^3 - \frac{5}{3}V_{GST}^2 nV_{DS} + V_{GST}(nV_{DS})^2 - \frac{1}{5}(nV_{DS})^3\right)}{2\left(V_{GST} - \frac{n}{2}V_{DS}\right)^2} \ . \quad (B.117)$$

Weak Inversion

Assuming weak inversion, the channel charge can be neglected with respect to the depletion zone charge. Since the energy bands and quasi Fermi-level for holes are constant along the channel, see fig. B.10, the depleted zone charge is also uniformly distributed. Therefore $Q_{dep} = LWQ'_{dep}$. The depleted zone charge is given by eq. B.13:

$$Q'_{dep} = -\sqrt{2eU_T\epsilon_{Si}N_a}\left[\exp\left(-\frac{\psi_S}{U_T}\right) + \frac{\psi_S}{U_T} - 1\right]^{1/2} \quad (B.118)$$

or simplified

$$Q'_{dep} = -\sqrt{2e\epsilon_{Si}N_a\psi_S} \equiv C'_{ox}\gamma\sqrt{\psi_S} \ . \quad (B.119)$$

The surface potential ψ_S can be calculated by using the expression eq. B.23. Let us rewrite it here:

$$V_{GB} = V_{fb} + \psi_S + \frac{|Q'_{dep}(\psi_S)|}{C'_{ox}} \ . \quad (B.120)$$

The terminal charges are given by the following equations

$$Q_S + Q_D = LWQ'_{ch} \approx 0; \quad Q_B = LWQ'_{dep}; \quad Q_G = -LWQ'_{dep} \ . \quad (B.121)$$

Capacitances

The equations eq. B.111, eq. B.112, eq. B.113, eq. B.114 and eq. B.117 express the terminal charges in strong inversion. The equations eq. B.118, eq. B.119, eq. B.120 and eq. B.121 are valid in the case of weak inversion. All these equations have the following general form

$$\begin{aligned} Q_D &= Q_D(V_{GS}, V_{DS}, V_{BS}) \\ Q_G &= Q_G(V_{GS}, V_{DS}, V_{BS}) \\ Q_B &= Q_B(V_{GS}, V_{DS}, V_{BS}) \\ Q_S &= Q_S(V_{GS}, V_{DS}, V_{BS}) \ . \end{aligned} \quad (B.122)$$

Note, V_{th} defined by eq. B.35 depend on V_{SB}. By using eq. B.122 and eq. B.101, we can write the equations for the transient (time dependent) transistor currents

$$\begin{aligned} i_D(t) &\approx I_{DS}(V_{GS}, V_{DS}, V_{BS}) + \frac{\partial Q_D}{\partial V_{GS}}\frac{dV_{GS}}{dt} + \frac{\partial Q_D}{\partial V_{DS}}\frac{dV_{DS}}{dt} + \frac{\partial Q_D}{\partial V_{BS}}\frac{dV_{BS}}{dt} \\ i_G(t) &\approx \frac{\partial Q_G}{\partial V_{GS}}\frac{dV_{GS}}{dt} + \frac{\partial Q_G}{\partial V_{DS}}\frac{dV_{DS}}{dt} + \frac{\partial Q_G}{\partial V_{BS}}\frac{dV_{BS}}{dt} \\ i_B(t) &\approx \frac{\partial Q_B}{\partial V_{GS}}\frac{dV_{GS}}{dt} + \frac{\partial Q_B}{\partial V_{DS}}\frac{dV_{DS}}{dt} + \frac{\partial Q_B}{\partial V_{BS}}\frac{dV_{BS}}{dt} \ . \end{aligned} \quad (B.123)$$

It is useful to define the capacitances $C_{ij} \equiv \frac{\partial Q_i}{\partial V_j}$ (for example $C_{gs} \equiv \frac{\partial Q_G}{\partial V_S}$). It holds $\frac{\partial Q_i}{\partial V_j} = \frac{\partial Q_i}{\partial V_{jS}}$; $\jmath \neq S$ (for example $C_{gb} \equiv \frac{\partial Q_G}{\partial V_B} = \frac{\partial Q_G}{\partial V_{BS}}$) since the voltages are referenced to

the source potential V_S. Due to charge conservation law, it holds $Q_G + Q_S + Q_D + Q_B = 0$ and therefore $\sum_i C_{ij} = 0$. If we increase all potentials by the same value ΔV, the charge Q_i does not change. Therefore $\Delta Q_i = \sum_j \frac{\partial Q_i}{\partial V_j} \Delta V = 0 \Rightarrow \sum_j C_{ij} = 0$. A property of the charge partitioning model based on eq. B.116 is non-reciprocity of the capacitances $C_{ij} \neq C_{ji}$.

Although eq. B.123 contain everything needed for calculation of the transient currents and voltages, it is useful to construct the equivalent circuit that behaves in agreement with derived formulas. Such an equivalent circuit is shown in fig. B.18 [1]. The capacitances

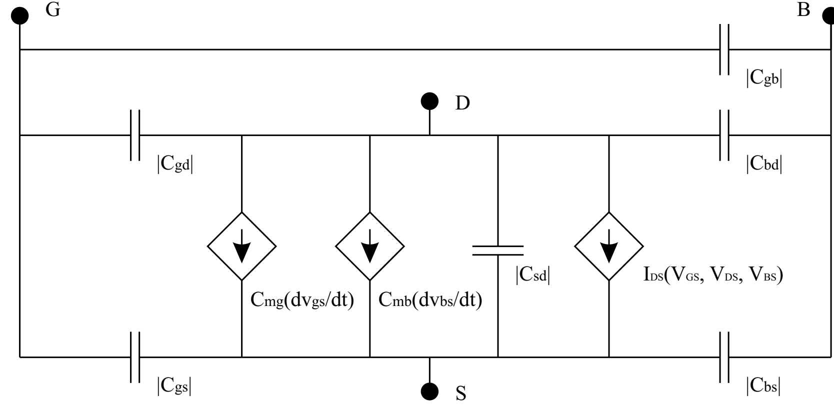


Figure B.18: Equivalent transistor circuit.

are defined as follows

$$
\begin{aligned}
|C_{gd}| &= -\frac{\partial Q_G}{\partial V_D}; \quad |C_{gs}| = -\frac{\partial Q_G}{\partial V_S}; \quad |C_{gb}| = -\frac{\partial Q_G}{\partial V_B} \\
|C_{bs}| &= -\frac{\partial Q_B}{\partial V_S}; \quad |C_{bd}| = -\frac{\partial Q_B}{\partial V_D}; \quad |C_{sd}| = -\frac{\partial Q_S}{\partial V_D} \ .
\end{aligned} \tag{B.124}
$$

C_{mg} and C_{mb} are transcapacitances that account for nonreciprocity of the capacitances. We have

$$
\begin{aligned}
C_{mg} &= \frac{\partial Q_D}{\partial V_G} - \frac{\partial Q_G}{\partial V_D} \\
C_{mb} &= \frac{\partial Q_D}{\partial V_B} - \frac{\partial Q_B}{\partial V_D} \ .
\end{aligned} \tag{B.125}
$$

The expression for the drain-source current can be linearized in the vicinity of the DC voltages between transistor terminals. The small-signal equivalent circuit is shown in fig. B.19. Beside the intrinsic capacitances, we took into account a few other capacitances that occur in real devices. C_{js} and C_{jd} are the capacitances of the inverse polarized source and drain $p-n$ junctions. C_{os} and C_{od} are the gate-source and gate-drain overlap capacitances. In the case of strong inversion and current saturation ($V_{DS} = V_{DSsat}$), the formulas eq. B.124 and eq. B.125 lead to the following results

$$
\begin{aligned}
|C_{gd}| &= 0; \quad |C_{gs}| = \frac{2}{3}C_{ox}; \quad |C_{gb}| = \frac{n-1}{3n}C_{ox} \\
|C_{bs}| &= (n-1)\frac{2}{3}C_{ox}; \quad |C_{bd}| = 0; \quad |C_{sd}| = 0 \ ,
\end{aligned} \tag{B.126}
$$

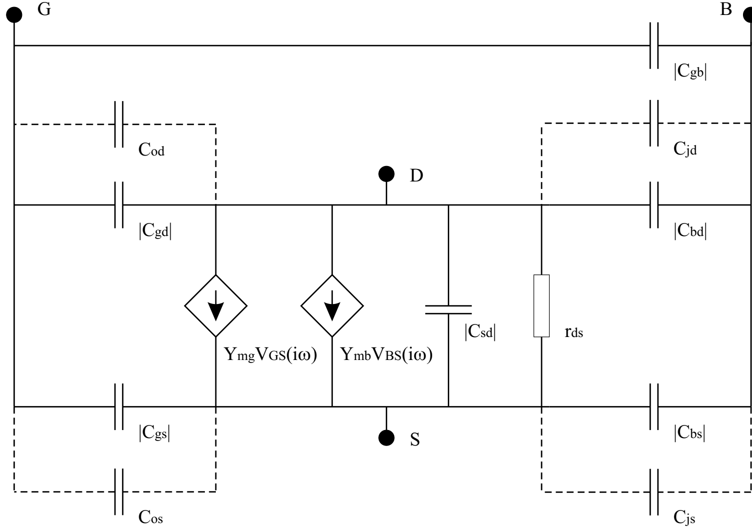


Figure B.19: Equivalent small-signal circuit.

$$Y_{mg} = g_{mg} + i\omega \frac{4}{15} C_{ox}; \quad Y_{mb} = g_{mb} + i\omega (n-1) \frac{4}{15} C_{ox} \; , \tag{B.127}$$

where $g_{mg} \equiv \frac{\partial I_{DSsat}}{\partial V_{GS}}$, $g_{mb} \equiv \frac{\partial I_{DSsat}}{\partial V_{BS}}$ and $r_{ds} \equiv \left(\frac{\partial I_{DSsat}}{\partial V_{DS}}\right)^{-1}$. The oxide capacitance is $C_{ox} = WLC'_{ox}$. The parameters Y_{mg} and Y_{mb} are called transadmittances. There is a phase shift in Y_{mg} and Y_{mb} to account for the delay between the change of the control voltage and the change in the drain-source current. The trans-capacitances can be interpreted as the first order approximation of non quasi-static effects. The parameters g_{mg} and g_{mb} are called transconductances. In the case of weak inversion we have

$$\begin{aligned} |C_{gd}| &= 0; \quad |C_{gs}| = 0; \quad |C_{gb}| = \frac{C_{ox} C_{dep}(\psi_S)}{C_{ox} + C_{dep}(\psi_S)} \\ |C_{bs}| &= 0; \quad |C_{bd}| = 0; \quad |C_{sd}| = 0 \; , \end{aligned} \tag{B.128}$$

$$Y_{mg} = g_{mg}; \quad Y_{mb} = g_{mb} \; . \tag{B.129}$$

$C_{dep}(\psi_S) = \frac{d|Q_{dep}|}{d\psi_S}$ is the depleted zone capacitance that can be calculated by using eq. B.118 or eq. B.119. The surface voltage can be expressed as a function of the gate potential by using eq. B.120. The formula for $|C_{gb}|$ can be easily verified. We have

$$|C_{gb}| = -\frac{\partial Q_G}{\partial V_B} = \frac{\partial (Q_{dep} + Q_{ch})}{\partial V_B} \approx \frac{\partial Q_{dep}}{\partial V_B} = \frac{dQ_{dep}}{d\psi_S} \frac{\partial \psi_S}{\partial V_B} \; . \tag{B.130}$$

By differentiating eq. B.120 (valid in weak inversion), we obtain

$$\frac{\partial \psi_S}{\partial V_B} = -1 + \frac{\partial Q_{dep}}{\partial V_B} \frac{1}{C_{ox}} \; . \tag{B.131}$$

By substituting of eq. B.131 into eq. B.130 we have

$$\frac{\partial Q_{dep}}{\partial V_B} = \frac{-\frac{dQ_{dep}}{d\psi_S}}{\left(1 - \frac{dQ_{dep}}{d\psi_S} \frac{1}{C_{ox}}\right)} = \frac{C_{ox} C_{dep}(\psi_S)}{C_{ox} + C_{dep}(\psi_S)} = |C_{gb}| \; . \tag{B.132}$$

Fig. B.20 shows the capacitance C defined as $C \equiv \frac{C_{ox} C_{tot}(\psi_S)}{C_{ox}+C_{tot}(\psi_S)}$, where $C_{tot}(\psi_S) = \frac{d(|Q_{dep}|+|Q_{ch}|)}{d\psi_S}$. The capacitance C is equal to the total gate capacitance $C_{gg} \equiv \frac{\partial Q_G}{\partial V_G} = C_{gb} + C_{gs} + C_{gd}$. Note, in weak inversion $C_{gg} \approx C_{gb}$. The exact formula for Q'_{tot} (eq. B.10) has been used. $V_{SB} = V_{DB} = 0$. It is interesting that the depleted zone capacitance, defined as $\frac{d|Q_{dep}|}{d\psi_S}$, has a finite value for $V_{GB} = V_{fb} = -1.0\ V$, although the depleted zone then disappears. This effect can be modelled only by using the correct formula for Q'_{tot}. The use of the approximative formula for Q'_{ch}, eq. B.119, leads to $C = C_{ox}$ for $V_{GB} = V_{fb}$, which is wrong. The measurement result is also shown.

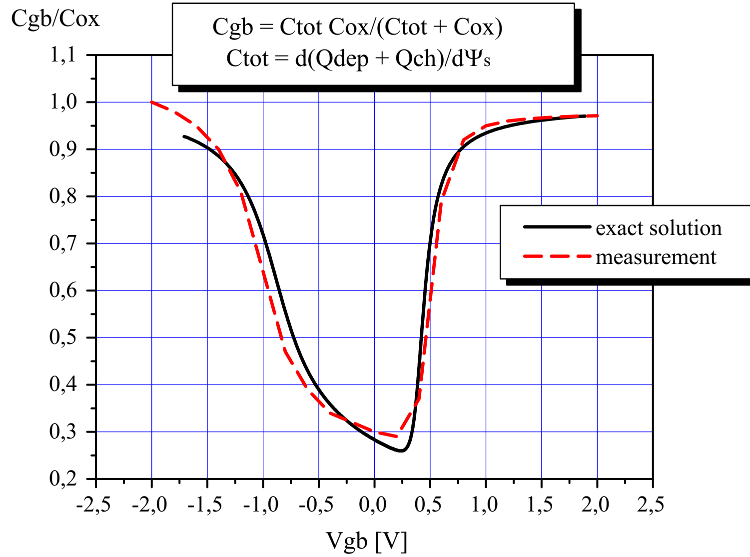


Figure B.20: Total gate capacitance C_{gg}, $V_{SB} = V_{DB} = 0$. A PMOS transistor is implemented in a 0.25 μm technology. The bulk dopant concentration is $N_d = 4.06 \cdot 10^{-17}\ cm^3$, oxide thickness $t_{ox} = 5.65\ nm$ (technology data sheet). Flat-band voltage assumed is $V_{fb} = -1.0\ V$.

Bibliography

[1] Yannis Tsividis, *Operation and Modeling of the MOS Transistor.* New York: McGraw-Hill, 1987

[2] Kevin F. Brennan, *The Physics of Semiconductors: With Applications to Optoelectronic Devices.* Cambridge: University Press, 1999

[3] W. R. Leo, *Techniques for Nuclear and Particle Physics Experiments.* Springer Verlag, 1994

[4] Jacob Millman, Arvin Graber, *Microelectronics.* McGraw-Hill, 1989

[5] Gerhard Lutz, *Semiconductor Radiation Detectors: Device Physics.* Springer Verlag, 1999

[6] L. Blanquart, J. Richardson, P. Denes, K. Einsweiler, E. Mandelli, G. Meddeler, P. Fischer, M. Ackers, G. Comes and I. Peric, "Analog Front-End Cell Designed in a Commercial 0.25 μm Process for the ATLAS Pixel Detector at LHC," *IEEE Trans. Nucl. Sci.* vol. 49, no. 4, pp. 1778-1782, August 2002.

[7] E. Mandelli, L. Blanquart, P. Denes, K. Einsweiler, R. Marchesini, G. Meddeler, M. Ackers, P. Fischer, G. Comes and I. Peric, "Digital Column Readout Architecture for the ATLAS Pixel 0.25 μm Front End IC," *IEEE Trans. Nucl. Sci.* vol. 49, no. 4, pp. 1774-1777, August 2002.

[8] T. Stockmanns, P. Fischer, F. Hugging, I. Peric, O. Runolfsson, N. Wermes, "Serial powering of pixel modules," *Nucl. Inst. Meth.* Band A 511, pp. 174-179, 2003.

[9] L. Blanquart "Overview of front-end Design," Proc. Pixel front-end final design review, CERN 2002 (http://agenda.cern.ch/fullAgenda.php?ida=a021405)

[10] L. Blanquart "Analog design improvements for FE-I2, " Proc. Pixel front-end final design review, CERN 2002 (http://agenda.cern.ch/fullAgenda.php?ida=a021405)

[11] L. Blanquart "Front-End Analog Cell Designed in a Commercial 0.25 μm Process for the ATLAS Pixel Detector at LHC," Proc. 5th International Meeting on Front End Electronics, Fermilab 2003 (http://www-ppd.fnal.gov/EEDOffice-w/conferences/FEE_2003/Agenda.html)

[12] L. Blanquart, A. Mekkaoui, V. Bonzom, P. Delpierre, "Pixel analog cell prototypes for ATLAS in DMILL technology" *Nucl. Inst. Meth.* Band A 395, pp. 313-317, December 1997.

[13] K. Einsweiler "The Front-end ASIC for the ATLAS Pixel Detector," Proc. 5th International Meeting on Front End Electronics, Fermilab 2003 (http://www-ppd.fnal.gov/EEDOffice-w/conferences/FEE_2003/Agenda.html)

[14] K. Einsweiler, "Overview of design and specifications," Proc. Pixel front-end final design review, CERN 2002 (http://agenda.cern.ch/fullAgenda.php?ida=a021405)

[15] K. Einsweiler, "Design improvements for FE-I2," Proc. Pixel front-end final design review, CERN 2002 (http://agenda.cern.ch/fullAgenda.php?ida=a021405)

[16] K. Einsweiler, "ATLAS Pixel FE-I1 Chip: Design Description," Proc. Pixel front-end final design review, CERN 2002 (http://agenda.cern.ch/askArchive.php?base=agenda&categ=a021405&id =a021405s1t1/documents)

[17] N. Wermes "Design and Prototype Performance of the ATLAS Pixel Detector," *Nucl. Inst. Meth.* Band A 477, pp. 121-128, 2002.

[18] N. Wermes "Pixel Detectors for Particle Detection and Imaging Applications," *Nucl. Inst. Meth.* Band A 512, pp. 277-288, 2003.

[19] N. Wermes "Trends in Pixel Detectors: Tracking and Imaging," submitted to *IEEE Trans. Nucl. Sci.,* 2004.

[20] P. Fischer "The ATLAS pixel front-end chip FE-I in $0.25\,\mu$m technology" Proc. 8th workshop on electronics for LHC, September/2002, Colmar, France

[21] C. Berg, L. Blanquart, V. Bonzom, P. A. Delpierre, K. Desch, P. Fischer, M. Keil, S. Meuser, B. A. Raith, N. Wermes "Bier and Pastis, a pixel readout prototype chip for LHC, " *Nucl. Inst. Meth.* Band A 439, pp. 80-90, 2000.

[22] L. Blanquart, V. Bonzom, G. Comes, P. Delpierre, P. Fischer, J. Hausmann, M. Keil, M. Lindner, S. Meuser, N. Wermes "Pixel readout electronics for LHC and biomedical applications" *Nucl. Inst. Meth.* Band A 439, pp. 403-412, 2000.

[23] P. Fischer "Design considerations for pixel readout chips, " *Nucl. Inst. Meth.* Band A 501, pp. 175-182, 2003.

[24] P. Fischer "Pixel electronics for the ATLAS experiment," *Nucl. Inst. Meth.* Band A 465, pp. 153-158, 2001.

[25] F. Krummenacher "Pixel detectors with local intelligence: an IC designer point of view," *Nucl. Inst. Meth.* Band A 305, pp. 527-532, 1991.

[26] G. Anelli et al. ”Radiation tolerant VLSI circuits in standard deep submicron CMOS technologies for the LHC experiments : practical design aspects,” *IEEE Trans. Nucl. Sci.* vol. 46, no. 6, pp. 1690-1696, December 1999.

[27] A. Giraldo ”Evaluation of deep-submicron technologies with radiation tollerant layout for electronic in LHC enviroment,” PhD thesis, CERN, December 2000.

[28] G. Anelli ”Design and characterization of radiation tolerant integrated circuits in deep submicron CMOS technologies for the LHC experiments” PhD thesis at the university of Padova, Italy, December 1998 (http://rd49.web.cern.ch/RD49/RD49Docs/anelli/these.html)

[29] W. Snoeys et al. ”Integrated circuits for particle physics experiments,” *IEEE J. Solid State Circuits* vol. 35, no. 12, pp. 2018-1030, December 2000.

[30] W. Snoeys et al. ”Layout techniques to enhance the radiation tolerance of standard CMOS technologies demonstrated on a pixel detector readout chip,” *Nucl. Inst. Meth.* Band A 439, pp. 349-360, January 2000.

[31] W. Snoeys et al. ”A new NMOS layout structure for radiation tolerance” *IEEE Trans. Nucl. Sci.* vol. 49, no. 4, pp. 1829-1833, August 2002.

[32] T. Calin et al. ”Upset hardened memory design for submicron CMOS technology” *IEEE Trans. Nucl. Sci.* vol. 43, no. 6, pp. 2874-2878, December 1996.

[33] X. Aragones et al. ”Substrate coupling trends in future CMOS technologies,” paper presented at PATMOS’97, pp. 235-244, Louvain-La-Neuve, Belgium, 1997

[34] J. Bastos, A. M. Marques, M. S. J. Steyaert, W. Sansen ”12-Bit Intrinsic Accuracy High-Speed CMOS DAC,” *IEEE J. Solid State Circuits* vol. 33, no. 12, pp. 1959-1969, December 1998.

[35] J. C. Chen, D. Sylvester, C. Hu ”An on-chip interconnect capacitance characterization method with sub-femto-farad resolution,” *IEEE Trans. Semiconduct. Manufact.* vol. 11, issue. 2, pp. 204-210, May 1998.

[36] N. Wermes et al. ”New results on DEPFET pixel detectors for radiation imaging and high energy particle detection,” submited to *IEEE Trans. Nucl. Sci.*, 2003.

[37] R. H. Richter, L. Andricek, P. Fischer, K. Heinzinger, P. Lechner, G. Lutz, I. Peric, M. Reiche, G. Schaller, M. Schnecke F. Schopper, H. Soltau, L. Struder, J. Treis, M. Trimpl, J. Ulrici, N. Wermes ”Design and technology of DEPFET pixel sensors for linear collider applications,” *Nucl. Inst. Meth.* Band A 511, pp. 250-256, September, 2003.

[38] M. Trimpl et al. ”A fast readout by using switched current techniques for a DEPFET pixel vertex detector at TESLA,” *Nucl. Inst. Meth.* Band A 511, pp. 257-264, September, 2003.

[39] J. Ulrici et al. ”Spectroscopic and imaging performance of DEPFET pixel sensors,” *Nucl. Inst. Meth.* Band A 465, pp. 247-252, June, 2001.

[40] W. Neeser et al. ”The DEPFET-Pixel Bioscope,” *IEEE Trans. Nucl. Sci.* vol. 47, no. 3, pp. 1246-1250, 2000.

[41] P. Fischer, J. Hausmann, M. Overdick, B. Raith, N. Wermes, L. Blanquart, V. Bonzom and P. Delpierre ”A counting pixel readout chip for imaging applications,” *Nucl. Inst. Meth.* Band A 405, pp. 53-59, March, 1998.

[42] P. Fischer et al. ”A counting pixel chip and sensor system for x-ray imaging,” *IEEE Trans. Nucl. Sci.* vol. 46, no. 4, pp. 1070-1074, 1999.

[43] P. Fischer et al. ”A photon counting pixel chip with energy windowing,” *IEEE Trans. Nucl. Sci.* vol. 47, no. 3, pp. 881-884, 2000.

[44] M. Lindner et al. ”Medical X-ray imaging with energy windowing,” *Nucl. Inst. Meth.* Band A 465, pp. 229-234, June, 2001.

[45] P. Fischer et al. ”Single Photon Counting X-ray Imaging with Si and CdTe Single Chip Pixel Detectors and Multichip Pixel Modules ,” submitted to *IEEE Trans. Nucl. Sci.*

[46] P. Fischer, E. Kraft ”Low swing differential logic for mixed signal applications,” *Nucl. Inst. Meth.* Band A 518, pp. 511-514, February, 2004.

[47] Boyd Fowler, Abbas El Gamal and David Yang, Information Systems Laboratory, Stanford University, California ”Techniques for Pixel Level Analog to Digital Conversion,”
(http://www-isl.stanford.edu/ abbas/group/papers_and_pub/aerosense98.pdf)

[48] C. J. Brunnett ”Method and apparatus for improved radiation detection in radiation scanning systems,” U. S. Patent, no. 4,052,620, October, 1977.

[49] M. Overdick, W. Rutten, T Zaengel ”Röntgendetektor mit grossem Dynamikbereich,” European Patent, no. EP 1 231 485 A2, August, 2002.

[50] Helmuth Spieler ”Radiation Detectors and Signal Processing, Lecture Notes,” VII. Heidelberger Graduate Lectures in Physics, University of Heidelberg, 2001
(http://www-physics.lbl.gov/ spieler/Heidelberg_Notes/index.html)

[51] Helmuth Spieler ”Introduction to Radiation Detectors and Electronics, Rate of Noise Pulses in Threshold Discriminator Systems, Lecture Notes,” University of California, Berkeley, 1998
(http://www-physics.lbl.gov/ spieler/physics_198_notes/)

[52] X. Xiu et al. *"BSIM4.3.0 User's manual,"* University of California, Berkeley, 2003

[53] Y. Cheng et al. *"BSIM3v3 Manual,"* University of California, Berkeley, 1996

[54] B. J. Sheu, D. L. Scharfetter, P. Ko, M. Jeng "BSIM: Berkeley short - channel IGFET model for MOS transistors ," *IEEE J. Solid State Circuits* vol. sc-22, no. 4, pp. 558-566, August 1987.

[55] C. Galup-Montoro M.C. Schneider, A. Arnaud and H. Klimach "Self-Consistent Models of DC, AC, Noise and Mismatch for the MOSFET," Technical Proceedings of the 2004 NSTI Nanotechnology Conference and Trade Show, Volume 2, pp 494 - 499, 2004 (http://www.nsti.org/publ/Nanotech2004v2/817.pdf)

[56] A. Arnaud, C. Galup-Montoro "A compact model for flicker noise in MOS transistors for analog circuit design," *IEEE Trans. Electron Devices* vol. 50, no. 8, pp. 1815-1817, August 2003.

[57] A. Arnaud, C. Galup-Montoro "Simple continuous and consistent physic based model for flicker noise in MOS transistors," (http://iie.fing.edu.uy/vlsi/papers/ib02_Flicker.pdf)

[58] C. Galup-Montoro et al. "Self-Consistent Models of DC, AC, Noise and Mismatch for the MOSFET," presentation (http://www.nanotech2004.com/presentations/WCM2004/WCM2004-CGalupMontoro.pdf)

[59] M. J. Pelgrom, et al. "Matching properties of MOS transistors ," *IEEE J. Solid State Circuits* vol. 24, no. 5, pp. 1433-1439, October 1989.

[60] Ana Isabela Araujo Cunha et al. "An MOS transistor model for analog circuit design ," *IEEE J. Solid State Circuits* vol. 33, no. 10, pp. 1510-1519, October 1998.

[61] Fabian Hügging "Der state-of-the-art Pixelsensor für teilchenphysikalische Anwendungen mit extrem hohen Strahlungsfeldern", PhD thesis, Universität Dortmund, Juni, 2001. (http://eldorado.uni-dortmund.de:8080/FB2/ls4/forschung/2001/Huegging; internal&action=buildframes.action)

[62] Markus Lindner "Enwicklung eines zählenden Pixeldetektors für "Digitales Röntgen"", PhD thesis, Universität Bonn, August, 2001.

[63] Jochen Cammin "Study of a Light Standard Model Higgs Boson in the $t\bar{t}H^0$ Channel with ATLAS at LHC and Decay-Mode Independent Searches for Neutral Higgs Bosons with OPAL at LEP", PhD thesis, Universität Bonn, March, 2004.

[64] P. Fischer "Noise calculations", internal document

[65] S. P. Karna et al. "Point defects in Si-Si0_2 systems: current understanding," publisched in G. Pacchioni et al. (eds.), *"Defects in* Si0_2 *and related dielectrics: science and technology,"* Kluwer Academic Publishers, pp. 599-615, (2000)

[66] R. Schrimpf "Radiation effects on MOS devices," radiation effects, short course, Vanderbilt university, Nashville, Tennessee, 2001. (http://www.mrchsv.com/docs/Vanderbilt/Radiation%20Effects %20on%20MOS%20Devices.pdf)

[67] R. Schrimpf "Component modeling and simulation," radiation effects, short course, Vanderbilt university, Nashville, Tennessee, 2001.

[68] T. Holman "Circuit and layout issues," radiation effects, short course, Vanderbilt university, Nashville, Tennessee, 2001.

[69] Technical design report of the ATLAS inner detector (http://atlas.web.cern.ch/Atlas/GROUPS/INNER_DETECTOR /TDR/tdr.html)

[70] ATLAS technical proposal (http://atlas.web.cern.ch/Atlas/TP/tp.html)

[71] ATLAS detector and physics performance (http://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/TDR/access.html)

[72] Pixel Detector Technical Design Report (http://atlas.web.cern.ch/Atlas/GROUPS/INNER_DETECTOR/ PIXELS/tdr.html)

[73] LHC design report (http://ab-div.web.cern.ch/ab-div/Publications/LHC-DesignReport.html)

Acknowledgments

First of all, I would like to thank my supervisor Prof. Dr. Norbert Wermes, who offered me the great opportunity to work on very exciting projects - on the design of the central part of one of the largest and most exciting experiments in high-energy physics ever and on the development of interesting novel radiation-detector concepts. I am very grateful for his support and kindness.

I am also very grateful to Prof. Dr. Peter Fischer, my second supervisor for his continuous support and guidance. While working with him I learned so much about design of complex electronic microchips.

I wish to thank all who worked on the three projects presented here; without the hard team-work of all these people I would not have been able to write this thesis.

I would like to thank all who worked in Bonn for the friendly atmosphere and much help I received during the four years of my PhD studies.

I would like to thank all in Berkeley, where I spent a few nice months. Especially, I wish to thank Dr. Kevin Einsweiler and Laurent Blanquart for providing me support and for interesting discussions.

I would like to thank my parents for giving me so much support.

Ivan Perić