Universität Bonn

Physikalisches Institut

The Belle II DEPFET Pixel Vertex Detector: Development of a Full-Scale Module Prototype

Mikhail Lemarenko

The Belle II experiment, which will start after 2015 at the SuperKEKB accelerator in Japan, will focus on the precision measurement of the CP-violation mechanism and on the search for physics beyond the Standard Model. A new detection system with an excellent spatial resolution and capable of coping with considerably increased background is required. To address this challenge, a pixel detector based on DEPFET technology has been proposed.
A new all silicon integrated circuit, called Data Handling Processor (DHP), is implemented in 65 nm CMOS technology. It is designed to steer the detector and preprocess the generated data. The scope of this thesis covers DHP tests and optimization as well the development of its test environment, which is the first Full-Scale Module Prototype of the DEPFET Pixel Vertex detector.

Physikalisches Institut der
Universität Bonn
Nussallee 12
D-53115 Bonn

BONN-IR-2013-20
November 2013
ISSN-0172-8741

Universität Bonn

Physikalisches Institut

The Belle II DEPFET Pixel Vertex Detector: Development of a Full-Scale Module Prototype

Mikhail Lemarenko
aus
Okoulovka

Dieser Forschungsbericht wurde als Dissertation von der Mathematisch-Naturwissenschaftlichen Fakultät der Universität Bonn angenommen und ist 2013 auf dem Hochschulschriftenserver der ULB Bonn http://hss.ulb.uni-bonn.de/diss_online elektronisch publiziert.

1. Gutachter:	Prof. Dr. Norbert Wermes
2. Gutachter:	Prof. Dr. Klaus Desch
Angenommen am:	31.10.2013
Tag der Promotion:	18.11.2013

Contents

Chapter 1

Introduction

> Consider the series 0, 1, 2, 3. . . . What is the next term? A good guess is 4. But the formula
>
> $$n + \frac{1}{24}n(n-1)(n-2)(n-3)$$
>
> also generates a series that begins 0, 1, 2, 3... In this case the series continues, not 4, 5, 6... but 5, 10, 21...
>
> Martin Gardner
> New Mathematical Diversions [1].

Since ancient times people have attempted to understand nature in order to discover the fundamental laws that govern our universe. The hope that all observations could fit into one simple model has not yet been justified: the more we learn about this world, the theories needed to describe it become increasingly sophisticated.

Modern science has experienced a true revolution over the last 200 years. Until the beginning of the 19^{th} century, very little was known about the basic structure of matter and its fundamental forces. Ancient Greek philosophers suggested the existence of indivisible pieces of matter, so-called *atoms* (from Greek $\alpha\tau o\mu o\varsigma$= atomos), but for many centuries this theory remained speculative. In 1897 J. J. Thomson discovered the electron by observing the *cathode rays* emitted from a hot filament. He demonstrated that these rays can be bent by a magnetic field and suggested that this was a stream of charged particles that he called *corpuscules*. This was soon followed by the discovery of the proton, neutron, positron and so on. By the mid 1960s, more than a hundred *elementary* particles were already known. Their embarrassingly large amount suggested that a more fundamental classification was necessary.

At present, the established theory of elementary particles and their interactions is called The Standard Model (SM). According to the SM, all matter consists of just three kinds of elementary particles (and their anti-partners): leptons, quarks and force mediators.

Symmetries in Nature

One of the fundamentals of modern physics is the notion of conserved quantities. Our intuition suggests that nature respects a certain class of symmetries. For example, we know from daily life that the outcome of identical experiments should be the same regardless of place, time or orientation in space. In fact, to each of these symmetries a corresponding conservation law can be assigned. In 1918 the German mathematician Emmy Noether published her famous theorem [2] in which she formalized the link between conservation laws and continuous symmetries.

Furthermore, we intuitively agree that a mirror image of a valid physical process should also be a valid physical process. This is known as the parity symmetry (**P**). Other examples of discrete symmetries are the time reversal symmetry (**T**) and the charge conjugation symmetry (**C**). These were accepted to be true symmetries for all processes in nature. There was such a firm belief that physics laws ought to be invariant under these transformations, that the validity of this assumption was never actually verified.

CP–Violation

A paradox, known as the "Theta-Tau Puzzle", arose in in the early 1950s of the past century [3]. Two newly discovered mesons, called θ^+ and τ^+, were identical in every respect (mass, charge, spin and so on) except their decays (θ^+ decays into two pions and τ^+ into three pions), this suggested that they had different parities.

Yang and Lee proposed that θ^+ and τ^+ are in reality the same particle (nowadays known as K^+), and that **P** is simply not conserved in one of the decays. This suggestion was confirmed by the experiment conducted by Chien-Shiung Wu on radioactive Cobalt 60, in which **P** was shown to be violated in weak interactions [4]. These results were published in 1957.

In the article published in the same year Lev Landau put forward the argument [5] that if $\theta - \tau$ were not the same, the lifetime difference in neutrino involved decays should be observed. This was not experimentally confirmed and he concluded that it was indeed a unique particle. Therefore, he indirectly demonstrated that **P** was violated in weak interactions. To restore the broken parity symmetry, he introduced a new Charge-Parity (**CP**) transformation (a combination of **C** and **P**). For a number of years this transformation was believed to hold in all interactions.

However, in 1963 Cronin and Fitch reported a weak **CP**-violation observed in neutral kaon (K^0) decays [6]. Later it was also shown for the first time that **CP** makes a distinction between matter and antimatter. For a long time this phenomenon had remained an unresolved issue until M. Kobayashi and T. Maskawa proposed a mechanism describing **CP**-violation in 1973 [7].

In fact, it is currently believed that **CP**-violation is a factor responsible for the existence of the universe as we know it. It was stated in 1967 by Andrei Sakharov [8] as one of the three necessary conditions for baryogenesis[1]. However, the amount of **CP**-violation discovered thus far is too small to describe the matter-antimatter asymmetry observed in the universe and more **CP**-violation sources are required.

B-Factories

For many years, neutral kaons was the only particle system for **CP**-violation studies. After the bottom quark discovery it was pointed out that neutral B-mesons can also be used for research into **CP**-violation. To enable the study of these systems, so-called 'B-factories' were constructed: the BaBar [9] experiment in California and the Belle in Japan [10][2].

Both experiments are carried out using a similar premise. For example, the Belle detector is located at the interaction region of an e^-/e^+ collider, called KEKB. The collider is tuned to have a center of mass energy of $\sqrt{s}$=10.59 GeV, which corresponds to the $\Upsilon(4S)$ resonance (a $b\bar{b}$ bound state). This energy is only slightly more than twice the mass of $B_{0,\pm}$-meson, such that $\Upsilon(4S)$ decays almost exclusively to a pair of $B\bar{B}$ (with a branching ratio higher than 96 %). The use of e^-/e^+ provides a clean measurement environment since the energies of colliding particles are well-known and a high signal to noise

[1] There conditions are: (1) Existence of at least one barion number violating process. (2) Existence of **C**- and **CP**-violating processes. (3) Interactions outside of thermal equilibrium.

[2] Further in the text we address only the Belle experiment, since this is the topic of the present thesis.

ratio could be achieved[1]. The beam energies are asymmetric giving a non-zero total momentum to the resulting $B\bar{B}$ states (Lorenz boost); this allows for a better spatial separation of different B-meson decay modes.

B^0 and $\bar{B}^0$ can decay to a common **CP**-eigenstate f_{CP}, where the transition is dominated by the $b \to c\bar{c}s$ process. If **CP**-violation does take place in this case, it can be characterized by the decay rate asymmetry:

$$A(t) = \frac{\Gamma(\bar{B}^0 \to f_{CP}) - \Gamma(B^0 \to f_{CP})}{\Gamma(\bar{B}^0 \to f_{CP}) + \Gamma(B^0 \to f_{CP})}$$

where $\Gamma(\bar{B}^0, B^0 \to f_{CP})$ is the rate for B^0 or $\bar{B}^0$ to f_{CP} at a proper time t after production.

For the case, then the $f_{CP} = J/\psi K_S$ or $f_{CP} = J/\psi K_L$, which are by construction both **CP**-eigenstates, the results of the **CP**-violation observation measured by the Belle detector are presented in Figure 1.1.

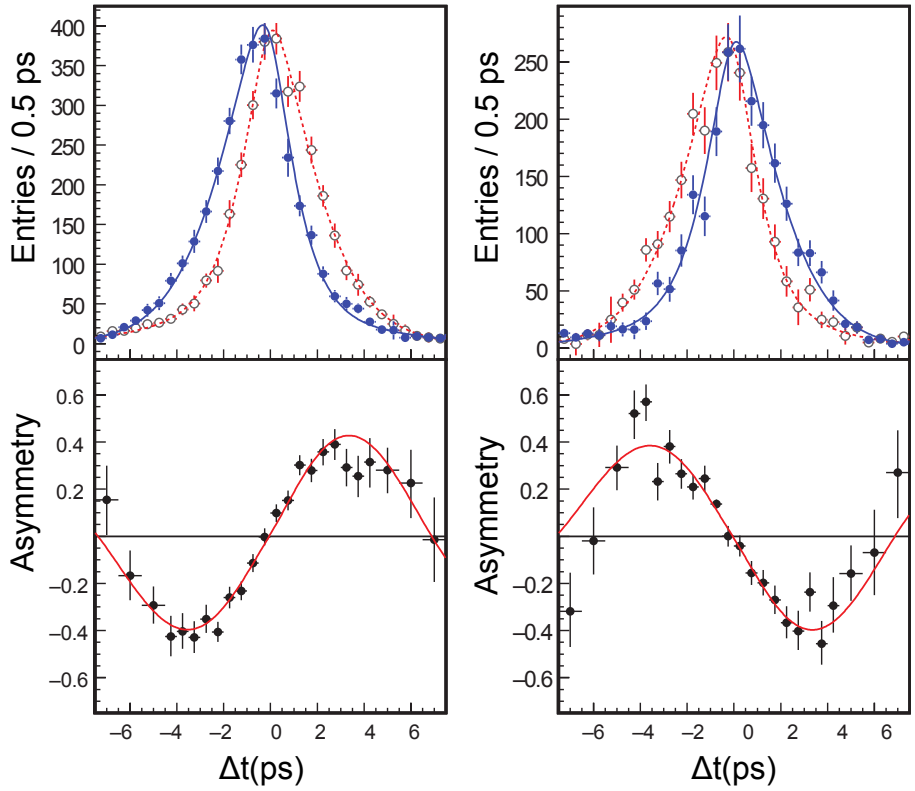


Figure 1.1: **CP**-violation observation in Belle for the case then one of the $B\bar{B}$ mesons decays into the following **CP**-eigenstates: $J/\psi K_S$ (left) and $J/\psi K_L$ (right). Information about which one of the B-mesons decayed into the f_{CP} is extracted from the complimentary B-meson (tag) decay, which is marked in red for the B_0 tags and in blue for the $\bar{B}_0$ tags [11].

During the operation of the detector until its final shutdown in 2010, the total integrated luminosity reached 1 ab^{-1}. Belle accomplished its main mission, which was the experimental verification of the Kobayashi and Maskawa proposal explaining the **CP**-violation mechanism. This experimental result was explicitly recognized by the Nobel Prize in Physics in 2008. Along with **CP**-violation observations,

[1] Other options, like to use protons, are more complicated: a proton, being a composite particle, has its kinetic energy shared by three quarks. Therefore, only an unknown fraction of the proton energy is available for a quark–quark collision, which effectively takes place.

Belle also made important discoveries in charm physics, τ-lepton physics, hadron spectroscopy and so on [11].

Belle Upgrade

For precise measurements of CKM[1]elements and for testing physics beyond the SM more statistical data is needed than has been thus far collected. Therefore, the Belle experiment is currently shut down for an upgrade.

From an accelerator point of view, this translates to the need for a luminosity increase. The planned upgrade of the KEKB (called SuperKEKB) entails two main strategies to achieve this goal: current increase and beam focusing. In total the expected luminosity will be increased by a factor of 40.

To cope with the increased background and the increased event rate a new detector, called Belle II, is being built. The upgraded central Silicon Vertex Detector (VXD) is among the main features of Belle II. For better spatial resolution, two additional layers of pixel detectors are planned together with the existing four strip layers of the VXD. For the implementation of the Pixel Detector (PXD) DEPFET technology was chosen, which has an excellent hit resolution and low material budget. Further details of this upgrade will be given in the next chapter.

The Presented Work

This thesis is devoted to the development of the Front-End readout of the PXD for Belle II; namely the Data Handling Processor (DHP) and its test environment. The DHP is a sophisticated All-Silicon Integrated Circuit that controls the PXD readout. Its second purpose is to reduce the data rates produced by the detector, which is achieved by discarding all data that does not contain a signal (zero-suppression). To do this efficiently, data have to be corrected for common mode noise and residual fixed pattern noise. To further reduce the quantity of data, an external trigger is applied. Thus, the generated ~3 Tbps of data can be reduced to 60 Gbps, which is limited by the bandwidth of all output links of the detector.

The chip is designed to meet the Belle II design specifications, i.e. to be radiation hard, to support up to 3 % of the data occupancy[2]and to have a high speed output link capable of transmitting the data over about 15 meters.

To meet the design goal, the chip underwent several steps of parameter space optimization and evaluation of radiation hardness. Furthermore, the chip was tested using a new specially designed test environment: a PXD module prototype, including all elements expected to be present in the detector, whilst being scaled down in channel count, i.e. a “full scale module prototype”. Finally, the prototype operation was tested during the DESY test beam campaign in Hamburg, confirming that the initial goals have been achieved.

The thesis is structured as follows:

- **Chapter 2** introduces the Belle II detector and its main elements. It also discusses upgrade related challenges, such as increased background and its influence on the detector implementation.
- **Chapter 3** provides a detailed description of the new PXD and DEPFET technology. The PXD concept, geometry, readout technique and each steering element are presented.
- **Chapter 4** introduces the DHP chip and discusses the related conceptual challenges. Furthermore, the proposed solutions are discussed.

[1] Cabibbo-Kobayashi-Maskawa matrix.

[2] That means that in the worst case scenario one can expect up to 3 % of pixels will contain signal information. However, this signal is mainly due to background and further offline processing to extract relevant physics events is necessary.

- **Chapter 5** discusses the chip optimization, which is needed in order to satisfy the Belle II requirements.
- **Chapter 6** presents the full scale module prototype of the PXD containing the full readout chain. It is followed by its performance evaluation and first test results.
- **Chapter 7**: discusses some aspects of the technology radiation tolerance. The evaluation of the potential risks and the implemented mitigation techniques are presented.
- **Chapter 8** summarizes the presented work and gives an outlook on the steps still needed to build the PXD.

Parts of the work were previously published [12, 13].

Chapter 2

Belle Upgrade

The operation of the Belle experiment has stopped since 2010 for the upgrade of the KEKB accelerator and the Belle detector. A schematic representation of SuperKEKB including the Belle II position is sketched in Figure 2.1. In this chapter, a general description of the new experiment is given.

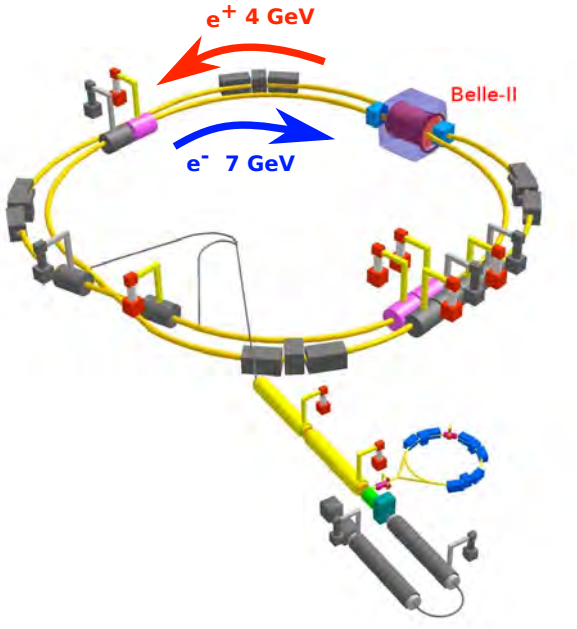


Figure 2.1: SuperKEKB collider and the position of the Belle II.

2.1 SuperKEKB

Within the scope of the current upgrade, the KEKB accelerator will be replaced by its upgraded version, SuperKEKB, whose planned luminosity will be about 40 times higher than the initial one [1]. This will be mainly achieved by an increase of the beam currents (3.6/2.6 A in LER[2]/HER[3] against initial 1.64/1.2 A) and by using the so-called Nano-beam scheme, in which the beam width is squeezed in one direction, thus increasing the interaction probability.

The beam-beam asymmetry was decreased from 3.5/8 GeV to 4/7 GeV [14] to solve the problem of an emittance growth due to high intra-beam scattering (see Section 2.3), which appears in the new

[1] 2.11×10^{34} $cm^{-2}s^{-1}$ reached in June 2009
[2] Low Energy Ring
[3] High Energy Ring

Nano-Beam scheme. The smaller beam asymmetry will induce about 30 % poorer vertex resolution. However, this is planned to be compensated by the new detection system.

2.2 Belle II

Figure 2.2 presents the Belle II detector. Typically used for collider experiments, Belle II follows an onion-like structure, covering almost the whole solid angle. As previously mentioned, Belle II is an upgraded version of the Belle detector adapted to support higher background event rate and having higher radiation hardness. A superconducting solenoid located in the inner volume of the detector provides a magnetic field of 1.5 T to bend the tracks of the charged particles.

A short description of its main subdetectors is presented below:

- **Silicon Vertex Detector (VXD)**. To enhance B and K_S vertices spatial resolution, which are produced near the PXD volume, the four layers of the Double-Sided Silicon Strip Detector (SVD) existing in the Belle design are completed with two layers of the Pixel Detector (PXD). A detailed description of the new PXD is given in Chapter 3.

- **Central Drift Chamber (CDC)**. The CDC starts just after the SVD and extends to a larger radius than Belle previously had. CDC is the central tracking device, measuring the charged particles' track position and their momenta. The measured dE/dx is also used for particle identification. The CDC also provides a fast trigger for charged particles.

- **Particle Identification System (PID).** The PID consists of the Time-Of-Propagation Detector (TOP) and Aerogel Ring Cherenkov Detector (ARICH). Both detectors use the Cherenkov effect for particle identification, especially for separation of kaons from pions.

  The Cherenkov effect can be resumed as follows: a particle flying though a medium faster than the speed of light in this medium induces a light cone along its track. The opening angle θ of this cone depends on the medium refraction index and the particle velocity according to the following equation:

  $$cos\,\theta = \frac{1}{n\beta}$$

  Cherenkov photons are then collected by UV–sensitive photomultipliers; their positions and arrival times are then registered for further evaluation.

- **Electromagnetic Calorimeter (ECL).** The ECL encloses the PID and consists of an array of CsI scintillators. It is used to detect and measure the precise energy of electrons and photons.

- K_L **and muon detector (KLM)** is the outermost Belle II subdetector. The KLM has a sandwich structure of alternating resistive plate chambers (RPC) and iron plates. The KLM distinguishes between muons and K_L using their respective signal signatures: a muon, having a low interaction cross section, creates a clear localized signature, whereas a K_L produces hadronic showers, leaving larger clusters.

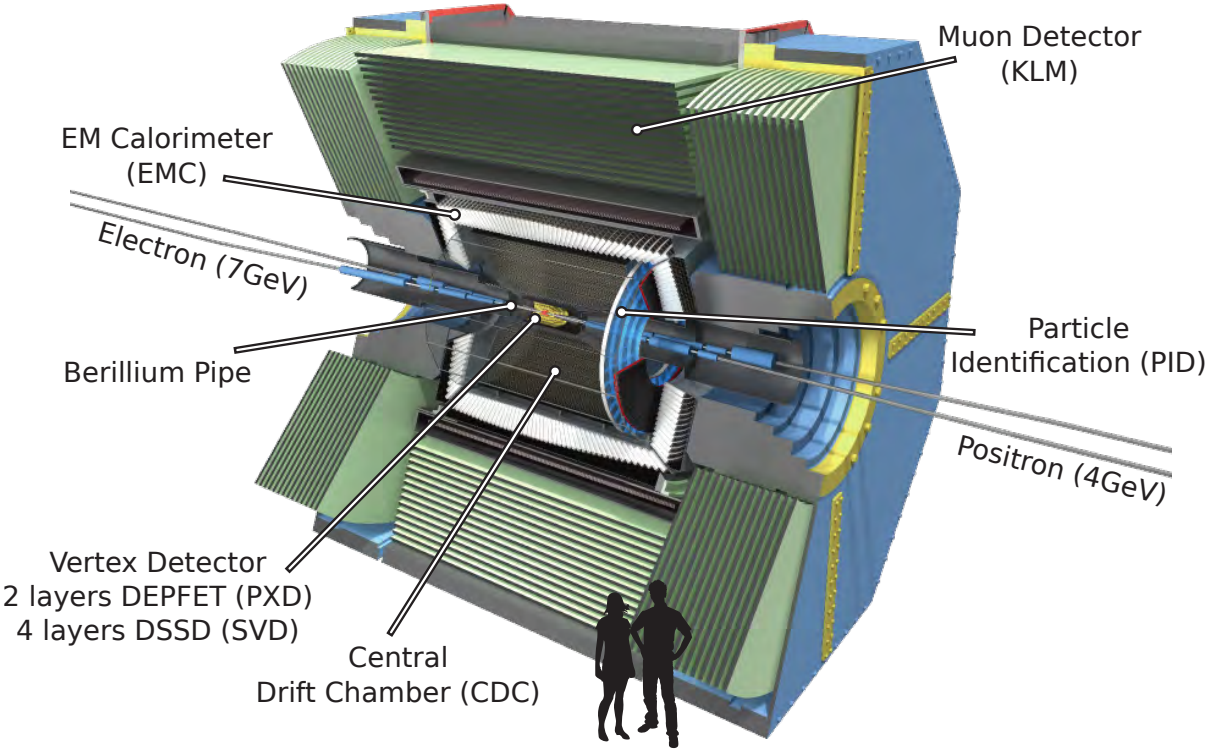


Figure 2.2: Belle II detector 3D representation, with indication of its main elements. This picture is designed by the Belle II Collaboration.

2.3 Belle II Background

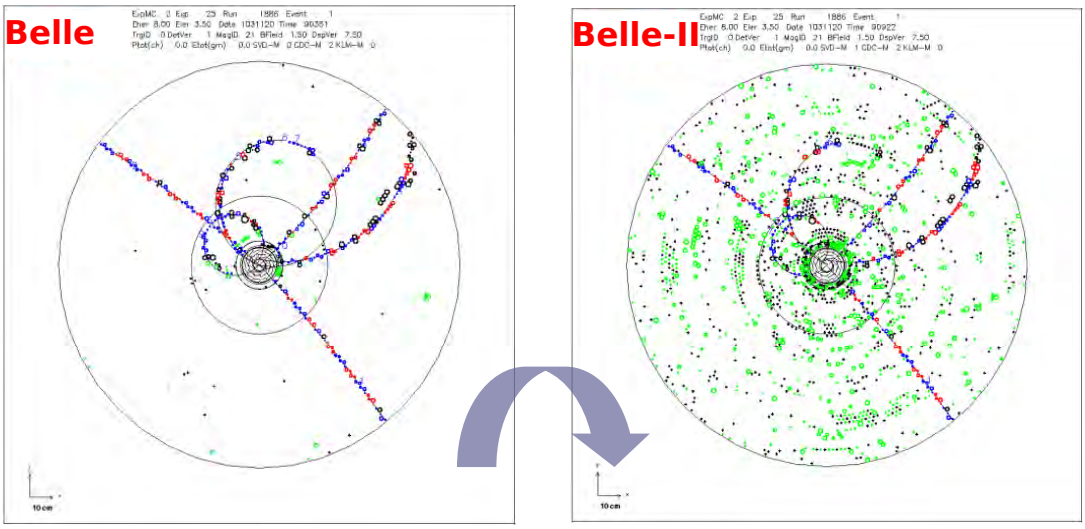


Figure 2.3: Comparative example of the background situation in Belle and Belle II experiments [14]. In the Belle II scenario, the increase of the background will make the zero suppression and the region-of-interest search non-trivial.

As mentioned above, the luminosity increase will be accompanied by a higher background level. Figure 2.3 represents an example of the background increase relative to Belle for a typical physics event. As one can see, the background will be by far the dominant data source of Belle II. It is very important to evaluate this effect for stable detector operation. In particular, the PXD is the most affected element, since it is the closest detector relative to the interaction point.

The expected background depends greatly on the implemented beam optics. In this section the main background sources are presented, and their latest simulated estimations are given. This information is relevant for the data processing logic optimization, which is implemented in the Data Handling Processor of the PXD. A detailed discussion about this optimization is presented in Chapter 5.

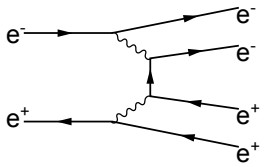
One can classify the background sources in the following way [15]:

- *Beam-Gas scattering*, or the scattering of the beam on residual gas. This effect is proportional to the product $I \times P$ (beam current $\times$ gas pressure). It results in Bremsstrahlung (energy loss due to photon emission) and Coulomb scattering (particle direction change).

- *Touschek effect* is an intra-bunch scattering. This effect is inversely proportional to the beam size and scales with E^{-3}. Therefore, the contribution from the LER dominates. Large angle Coulomb scattering causes an exchange of energy between the longitudinal and the transverse motion of the particles. Scattered particles leave the nominal beam orbit and hit the vacuum chamber and magnet walls. The resulting secondary shower particles can reach the PXD.

- *Radiative Bhabha scattering*, or e^+e^- scattering. In this case most of the scattered e^+e^- are lost very far downstream from the IP and are unlikely to produce particles that back-scatter into the detector. However, the emitted photons hit the downstream beam pipe and magnets, generating neutrons. This is the largest neutron background for the PXD (see Chapter 7).

- *Two-photon Radiation (QED)*: low momentum electron-positron pairs are produced via the two-photon process $e^+e^- \rightarrow e^+e^-e^+e^-$; it is one of the largest contributors to the background.



- *Synchrotron Radiation (SR)*. The SR is proportional to $I \cdot E^2 \cdot B^2$ (the beam current multiplied by the beam energy squared multiplied by the magnetic field squared). This background is mainly localized in the plane of the e^+e^- ring.

All background sources have a non-uniform contribution for both angular (ϕ) and longitudinal (z) directions; furthermore, the external layer of the PXD will see lower background as it is further away from the interaction point. A summary of their radial distributions and total occupancies per layer are presented in Figure 2.4 (SR excluded) and Table 2.1 [16]).

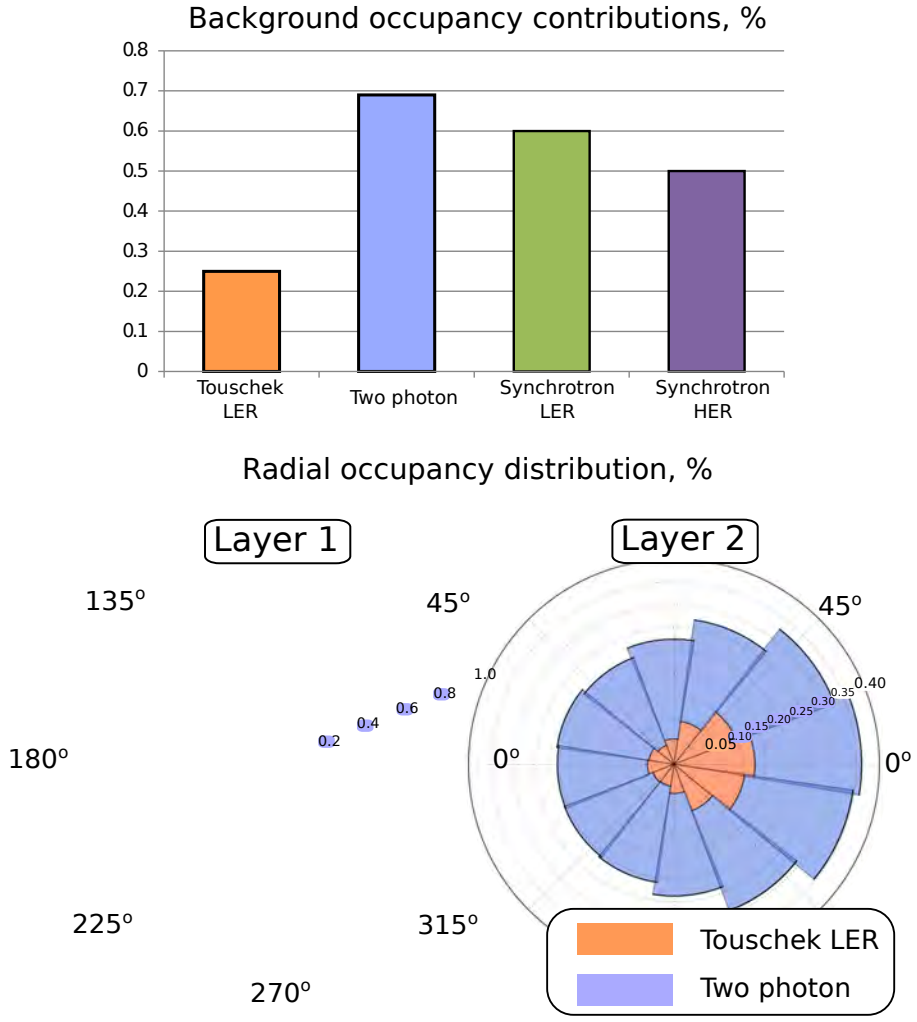


Figure 2.4: (Top) Background occupancy main contributors. The influence of the SR was not thoroughly investigated by the Belle II Collaboration at the time of writing this chapter. However, it is expected that the SR will be relevant only for the first layer of the detector for the plane situated in ϕ=0. (Bottom) The background occupancy (in %) radial distribution (SR excluded) for the internal (Layer 1) and the external (Layer 2) PXD layers. [16]

Background (in %)	Source	Layer 1	Layer 2
Two-Photon QED		0.69	0.24
Synchrotron Radiation[1]	LER	(0.6±0.15)[2]	0.0
Synchrotron Radiation[3]	HER	(0.5±0.3) [2]	0.0
Touschek	LER	0.25	0.18
Radiative Bhabha	HER	10^{-3}	10^{-3}
Radiative Bhabha	LER	10^{-4}	10^{-4}
Touschek	HER	0.0	0.0
Beam-Gas Coulomb	LER	0.0	0.0
Beam-Gas Coulomb	HER	0.0	0.0
Total		2.05±0.35	0.42

Table 2.1: Latest PXD background occupancy summary presented by the Belle II Collaboration [16, 17].

[1] LER: for the half ladder at $\phi = 0$.The occupancy in the other ladders can be neglected. Source: Y. Soloviev [17].
[2] Preliminary results with high uncertainty due to low statistics.
[3] Available data shows that SR radiation background from HER is distributed roughly uniformly over all PXD ladders (mostly scattered photons). Y. Soloviev [17].

Chapter 3

DEPFET Pixel Vertex Detector

This chapter is about the Pixel Detector (PXD), which consists of two layers of radially distributed sensors based on DEPFET technology. The chapter starts with general considerations as well as the detector concept, followed by a more detailed description of the main parts.

3.1 Detector Resolution

The key point of the PXD is precise reconstruction of the secondary decay vertices in order to distinguish them from the primary ones [18]. This distinction can be made in the following way: a reconstructed particle track is extrapolated back to its origin. If the resulting distance to the interaction point is significantly larger than the detector resolution, it is likely that this particle does not originate from the primary vertex. To satisfy the required precision, which is in the order of several tens of microns, it is important to optimize the total spatial resolution σ_{tot} of the detection system.

σ_{tot} is limited by the detector resolution error σ_{det} (geometry dependent) and the multiple scattering error σ_{MS}. σ_{tot} can be approximated by [18]:

$$\sigma_{tot}^2 = \sigma_{det}^2 + \sigma_{MS}^2 \tag{3.1}$$

with

$$\sigma_{MS}^2 = \sum (R_j \Delta\theta_j)^2$$

where $\Delta\theta_j$ is the scattering angle, while traversing detector layer j with radius R_j[1]; $\Delta\theta_j$ is approximately given by:

$$\Delta\theta_j = \frac{0.0136}{p[\text{GeV/c}]} \sqrt{\frac{\Delta X_j}{X_0}} \left[1 + 0.038\, ln\left(\frac{\Delta X_j}{X_0}\right)\right] \tag{3.2}$$

for the material of thickness ΔX_j with a radiation length X_0.

Figure 3.1 depicts a situation with two detector layers, as is envisaged in PXD[2]. In this case, the detector resolution error σ_{det} is a quadratic sum of two layers' resolutions projected on the beam axis σ_{One} and σ_{Two}:

$$\sigma_{det}^2 = \sigma_{One}^2 + \sigma_{Two}^2$$

As depicted, σ_{One} is obtained from the intrinsic resolution error of the first detector layer σ_1 by projecting it to the beam axis and assuming σ_2=0. Correspondingly, σ_{Two} is a projection of the σ_2 assuming σ_1=0.

[1] In case of a single detector, the outermost layer contribution would have been not relevant.
[2] Two dimensional simplification.

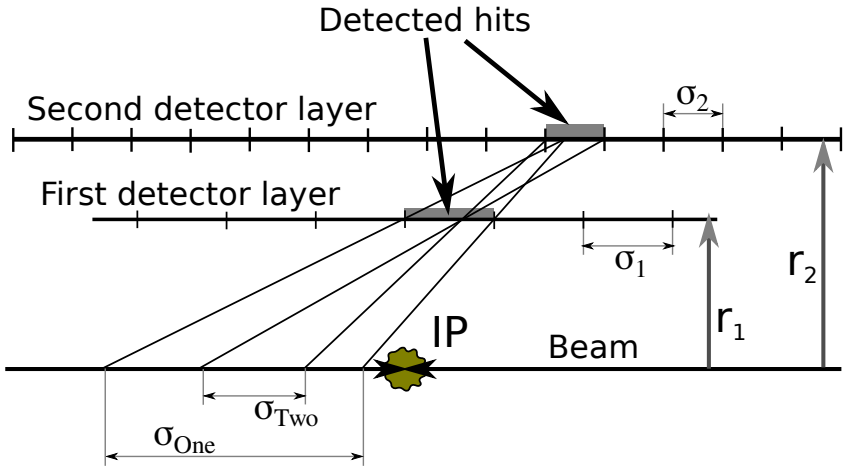


Figure 3.1: The total detector resolution is a quadratic sum of σ_{One} and σ_{Two}, which are respective projections on the beam axis of the intrinsic layer resolutions σ_1 and σ_2, as is presented in the figure.

Furthermore, from simple geometrical considerations it follows that:

$$\begin{aligned} \sigma_{One} &= \sigma_1 \left(\frac{r_2}{r_2 - r_1}\right) \\ \sigma_{Two} &= \sigma_2 \left(\frac{r_1}{r_2 - r_1}\right) \end{aligned} \tag{3.3}$$

Here r_1 and r_2 are the inner and outer layers' radii.

Several conclusions can be drawn from Equations (3.1) to (3.3) to minimize the σ_{tot}:

1. The position r_1 of the first detector layer should be as close to the interaction point as possible.
2. The distance between layers $r_2 - r_1$ should be maximized.
3. Intrinsic spatial resolutions σ_1 and σ_2 should be small, especially σ_1.
4. The material budget of the detector, especially for its first layer, should be minimized.

Using these criteria, the detector was optimized. Simulation results show [14] that for the proposed geometry (see Section 3.2) the total PXD resolution is estimated to be $\sigma_{tot} \sim 20\,\mu$m.

3.2 Overall Structure of the Detector

In Figure 3.2 a 3D model of the DEPFET Pixel Vertex Detector (PXD) is portrayed. It consists of two layers of independently controlled modules arranged around the beam pipe (with 10 mm radius). The inner layer is situated 14 mm away from the Interaction Point (IP) and has 8 modules with a Sensitive Area (SA) of 90 mm×12.5 mm; the outer layer is 22 mm away from the IP and has 12 modules with a SA of 124 mm×12.5 mm. Figure 3.3 presents the PXD mechanical prototype.

Each PXD module is an all-silicon sensor with its control and readout ASICs[1] directly bump-bonded on the module edges around the sensitive area. To shorten the readout period, modules are split in the middle; each half is separately steered by two independent sets of ASICs, thus each module represents two separate half-modules mechanically attached to each other. An example of this kind of half-module is shown in Figure 3.4; the related ASICs are situated outside the SA. Per half-module there are six Switcher chips responsible for the matrix control; they are situated on the 2 mm wide side-balcony.

[1] ASIC is the standard abbreviation for Application Specific Integrated Circuit.

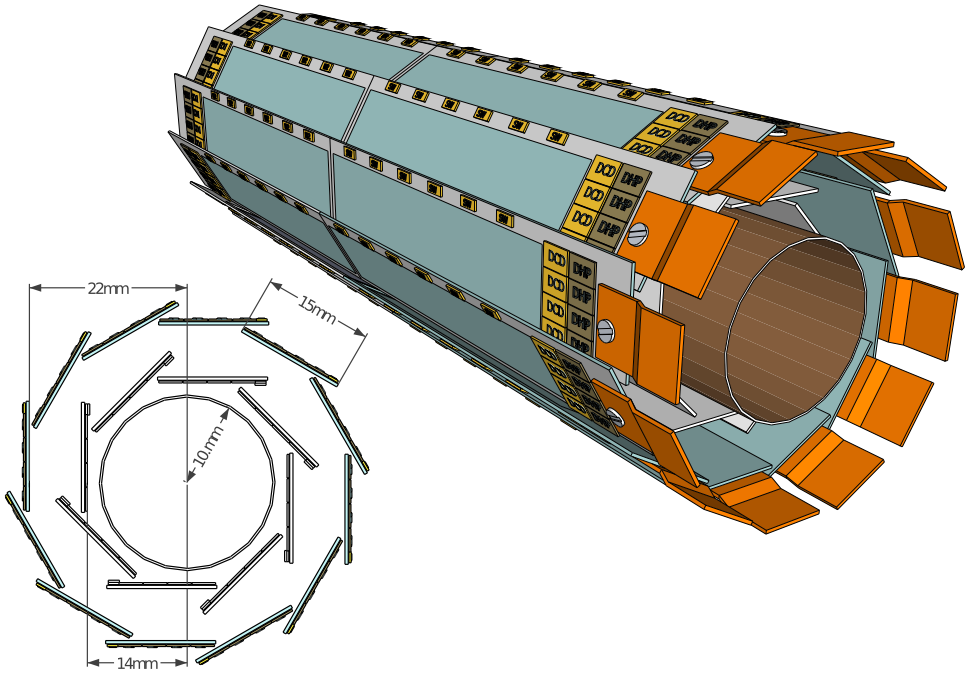


Figure 3.2: Isometric and front view projections of a 3D model of the detector (courtesy of Hans Krüger).

Figure 3.3: The mechanical mock-up with dummy modules of the DEPFET PXD (prepared by the MPI Munich).

Outside the SA an additional 25 mm wide balcony is foreseen to host other sets of ASIC: four Drain Current Digitizer (DCD) chips and four Data Handling Processor (DHP) chips.

The SA is reduced to 75 μm thickness using an anisotropic etching technique [19] to minimize track deviation of crossing particles due to multiple scattering. The balcony and end-of-stave regions have an original silicon wafer thickness (about 400 μm) to provide better mechanical stability.

The SA is an array of 1000×192 pixels implemented using DEPFET technology (Section 3.3). All drains of pixels belonging to a single column are connected together. The readout of the SA is done row-wise by Switcher chips. The DCD chips digitize drain currents coming from the pixels and the DHP chips pre-process the data and reduce their volume by suppressing pixels with a signal below a

predefined threshold. The produced zero suppressed data is sent to the detector's back-end electronics through a flexible Kapton®[1] cable attached to each module.

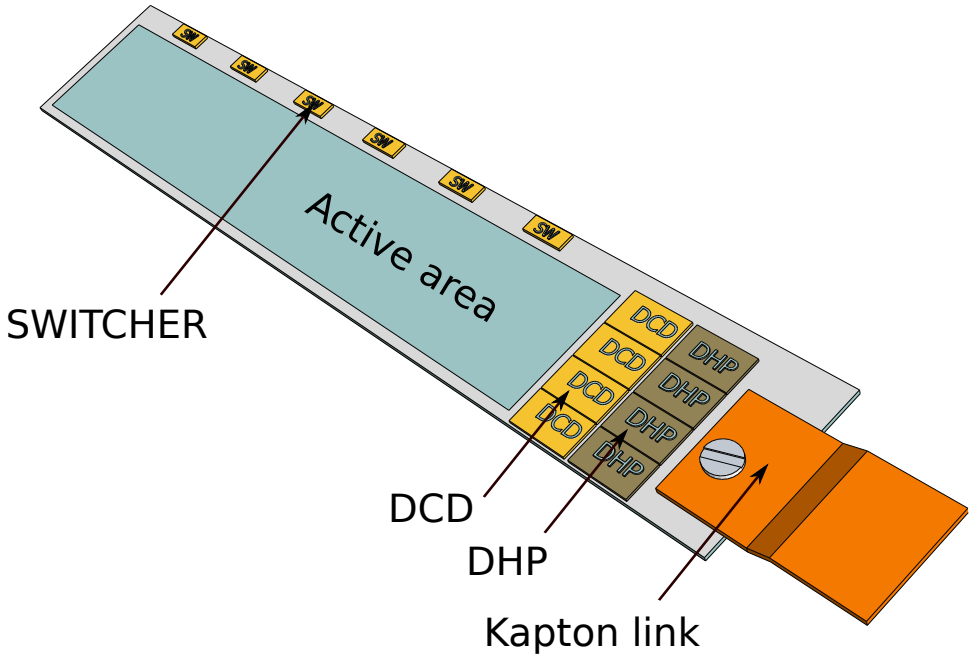


Figure 3.4: A sketch of the half-module structure with bump-bonded ASICs (Switcher, DCD and DHP chips) and their arrangements. The slow control, fast synchronization signals and the serial links of the DHP chips are connected to the back-end electronics via a flexible Kapton® cable, which is mechanically attached to each module.

3.3 DEPFET Principle

The original paper published by Kemmer and Lutz in 1987 [21] proposed using either MOSFET or JFET technology for a new detector concept, which will be described below. These concepts were known under the names DEPMOS and DEPJFET respectively. This convention was later abandoned and both implementations are now known under the generic name DEPFET (DEpleted PFET or DEPleted FET depending on publication). Historically, JFET was the preferred option. However, for the Belle II design JFET technology was abandoned in favor of MOSFETs due to feature size issues [22].

Typically, a DEPFET pixel is a PMOS transistor situated on a high-ohmic bulk. Using the sidewards-depletion technique [23] (see Section 3.3.1 for details) the bulk is depleted in such a way that a potential minimum for electrons can be created in the region close to the transistor's channel (Figure 3.5). An additional n^+ implantation deposited right beneath the gate of the transistor generates an extra positive space charge, whilst being depleted, additionally confining the minimum in the x-axis. This implantation together with the sidewards-depletion creates a global spatial potential minimum for electrons in the pixel's surrounding neighborhood. The deep–n implantation is called the pixel's internal gate (for reasons described below).

[1] Kapton® is a polyamide film developed by DuPont that can remain stable in a wide range of temperatures, used (among others) in flexible printed circuits designs [20].

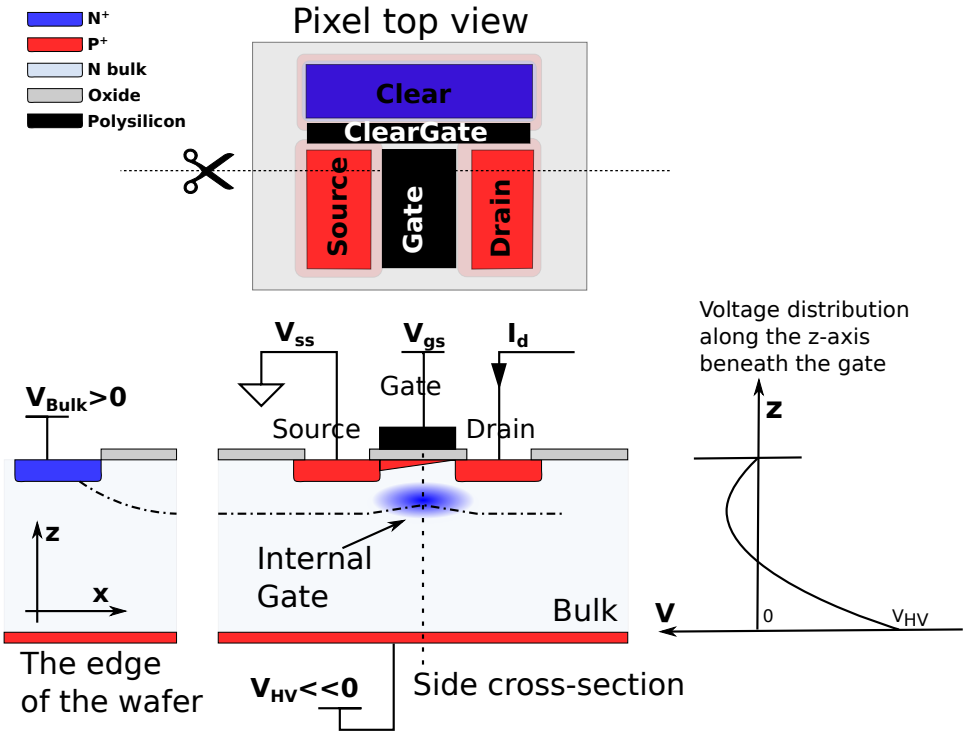


Figure 3.5: The DEPFET principle. The top view is presented to show how all the steering contacts of a pixel are arranged relative to each other. Furthermore, a cross-section in the direction of the source-gate-drain is presented; one sees the internal gate of a DEPFET relative to the external transistor gate. On the right, the voltage distribution inside bulk below the gate is drawn; the voltage maximum (the potential minimum for electrons) is situated in the internal gate region. Free electrons will drift and stay there.

Charged particles and absorbed photons leave a certain number of e–h pairs in the depleted bulk. The electric field present in the depleted bulk prevents their recombination: holes are attracted to the bulk back-side that is connected to a negative potential. Electrons in turn are collected in the potential minimum, i.e. in the internal gate.

The presence of trapped electrons near the P-channel influences its configuration and, consequently, the drain current. This explains the name 'internal gate': similar to a regular FET gate, it also controls the transistor's drain current. The stronger the coupling is (i.e. the closer the gate and the channel are), the stronger the influence is.

For the internal gate, the figure of merit is its charge referred transconductance g_q, which describes the sensitivity of the drain current I_{ds} to the change of the charge collected in the internal gate Q_i:

$$g_q = \frac{\partial I_{ds}}{\partial Q_i} = \frac{g_m}{C_{eff}} \qquad \left[\frac{nA}{e}\right] \tag{3.4}$$

It is equal to the V_{gs} referred transistor transconductance $g_m = \partial I_{ds}/\partial V_{gs}$ divided by the effective oxide capacitance C_{eff} [21].

In the transistor saturation region, which is the normal working regime of DEPFET devices, the g_q can be expressed as a function of L and W (the length and the width of the channel):

$$g_q \propto \sqrt{\frac{I_{ds} t_{ox}}{L^3 W}} \tag{3.5}$$

The parameters t_{ox} (gate oxide thickness), L and W (length and width of the gate) are purely geometrical and purely technology dependent. By increasing I_{ds}, g_q can be varied to some extent during the operation. In the saturation regime, the drain current of a MOSFET transistor is virtually independent of V_{ds} and equal to:

$$I_{ds} = 1/2 \frac{W}{L} \mu C'_{ox} \left(V_{gs} - V_{th} \right)^2 \tag{3.6}$$

with C'_{ox} and V_{th} being the oxide capacitance per unit area and the threshold voltage respectively. Thus, the gate voltage V_{gs} change is needed to influence the drain current during the operation.

DEPFET is an excellent concept for building elementary particle detectors, thanks to its large Signal-to-Noise Ratio (SNR).[1] The FET transistor of a DEPFET pixel provides an in-pixel pre-amplification of the charge collected in the internal gate, thus avoiding any additional parasitic capacitance that can be found in hybrid pixel detector front-ends[21]. An excellent SNR=96±4 has been reported for a measurement of the 6 keV k_α line of the Fe^{55} at room temperature with a shaping time of 10 μs [24].

Furthermore, the readout of DEPFET pixel is non-destructive: the same collected charge, stored in the internal gate, can be read several times.

3.3.1 Sidewards-Depletion Technique

Sidewards-depletion was invented by E. Gatti and P. Rehak [23] as a technique to run semiconductor drift chambers.

Using this technique, the depletion zone grows laterally into the bulk, as shown in Figure 3.6. The voltage needed to deplete the whole volume is four times smaller than that needed for regular depletion. This is explained by the fact that sidewards-depletion grows from both sides, therefore, the effective depletion depth is twice as small.

As in the case of p^+n junction, the depletion zone starts to grow on the border between the p^+ and n zones, parallel to that border. When two depletion zones are thick enough, they merge. With a small further increase, the total depletion situation is reached.

Another important detail that makes this method so special is the voltage profile, which is orthogonal to the wafer surface. In the previous picture, the upper and bottom planes are connected to the same potential. It is clear that even in this case, the potential in the bulk in between cannot be constant: it is forbidden by Gauss' theorem as the depleted bulk is charged. To understand the potential profile shape, it is sufficient to solve the Poisson equation with corresponding Dirichlet boundary conditions:

$$\begin{aligned} \Delta\varphi &= \frac{\rho}{\varepsilon\varepsilon_0} \\ \varphi|_{side} &= 0 \\ \varphi|_{top} &= \varphi_{top} \\ \varphi|_{bottom} &= \varphi_{bottom} \end{aligned} \tag{3.7}$$

[1] The signal left in the bulk by a charged particle is typically around several thousand electrons. To be detected it has to be amplified. It is particularly important to have the first amplification stage as clean as possible, since it has the highest contribution to the SNR.

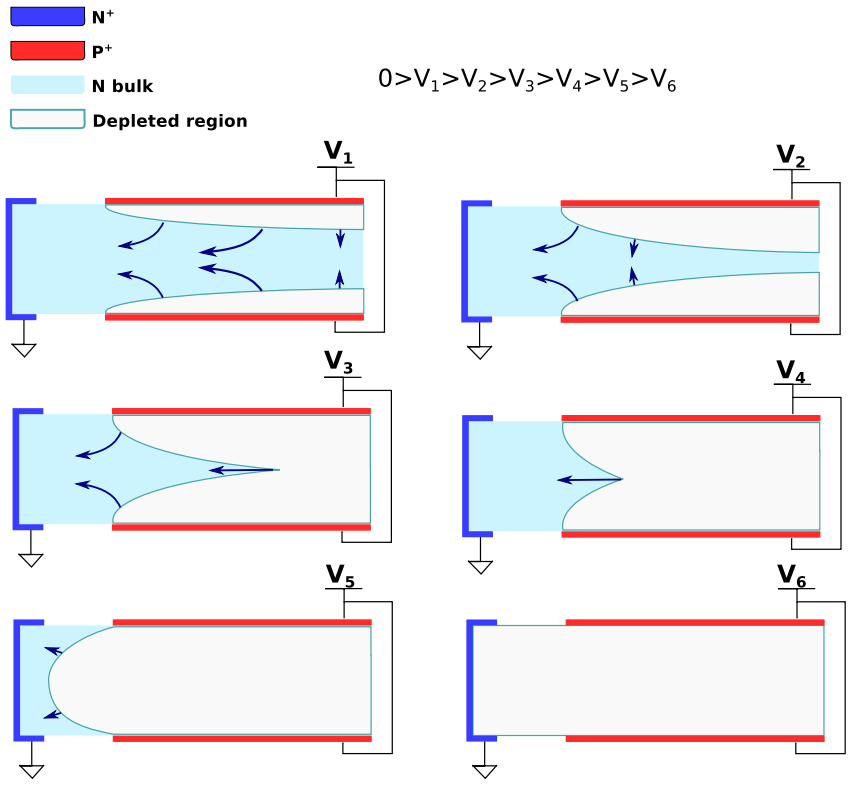


Figure 3.6: Sidewards-depletion principle. The depletion zone increases similarly to the conditions in a planar situation. However, the growth goes in a lateral direction, parallel to the p^+ contacts. With a further voltage increase, the two depletion zones merge and then, finally, the bulk is totally depleted.

In the regions far from the sensor edge, where side effects are negligible, this equation reduces to one dimension:

$$\begin{aligned} \varphi &= \varphi(z) \\ \Delta &= \frac{d^2}{dz^2} \end{aligned} \tag{3.8}$$

In this case, Equation (3.7) can be easily solved analytically and the solution takes the following form:

$$\varphi(z) = \frac{eN_D}{2\varepsilon\varepsilon_0} z(d-z) + \frac{z}{d}(\varphi_{bottom} - \varphi_{top}) + \varphi_{top} \tag{3.9}$$

It is an inverted parabola reaching its maximum in the interval (0,d), whose depth depends on the doping concentration N_D. The maximum position depends on the difference between the top and the bottom potentials; it is equal to:

$$z_{min} = \frac{d}{2} + \frac{\varepsilon\varepsilon_0}{eN_D d}(\varphi_{bottom} - \varphi_{top}) \tag{3.10}$$

By setting the appropriate voltage difference, the maximum position can be fixed at any point between two planes. This is the case of the DEPFET sensor: the voltage applied to the bottom of the substrate is very different relative to the top, such that this maximum is situated very close to the PMOS channel. This situation is presented in Figure 3.7

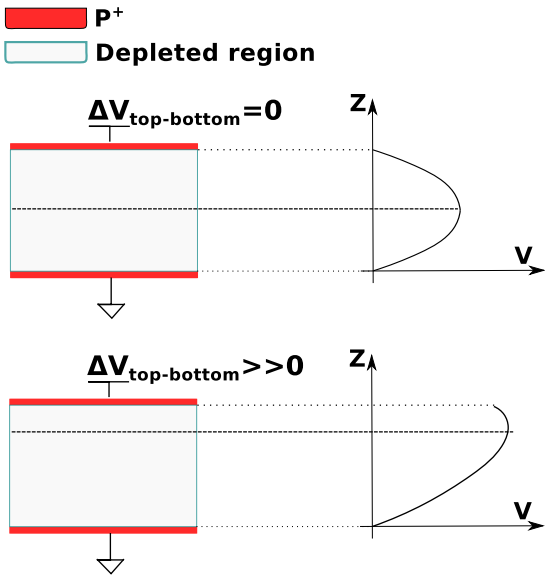


Figure 3.7: Potential position depending on the applied top and bottom voltages.

The Clear Contact

A DEPFET pixel is a dynamical system. The internal gate, while having a limited capacity, continuously gets filled with thermally generated electrons. Once totally filled, it stops attracting electrons.

A regular refreshment procedure should be executed to maintain the device in its working state. This is the so-called 'Clear' process. In order to clear, an additional n^+ implantation[1], called 'Clear', is situated close to the pixel (Figure 3.8) perpendicular to the drain-source direction.

To avoid competition in detection collection between the internal gate and the Clear contact during device operation, the Clear is embedded in an additional deep p-well. While depleted, the p-well gets negatively charged; this negative charge halo serves as a potential barrier repelling electrons during the charge collection period.

To remove electrons from the internal gate a short and high positive voltage pulse is applied to the clear contact. Via the punch-through mechanism [25] an electrically favorable path is created between the Clear and the internal gate. The electrons are then removed by drift as presented in Figure 3.8.

As reported by Sandow et al. [26], the clear operation can be as short as 10–20 ns, however the clear voltage should be 14 V high or more. The dependency between the clear voltage and the clear time is depicted in Figure 3.9.

The Clear-gate

The Clear process has for a long time been an issue for the DEPFET technology, being a timing bottleneck. To speed up the Clear process an additional gate (called clear-gate) between the Clear implantation and the internal gate has been added as depicted in Figure 3.8. It can lower the potential barrier created by the P-implantation surrounding the Clear contact; making the electrons' removal easier. First intro-

[1] the n^+ is necessary to provide the ohmic contact between the metal connection and the semiconductor

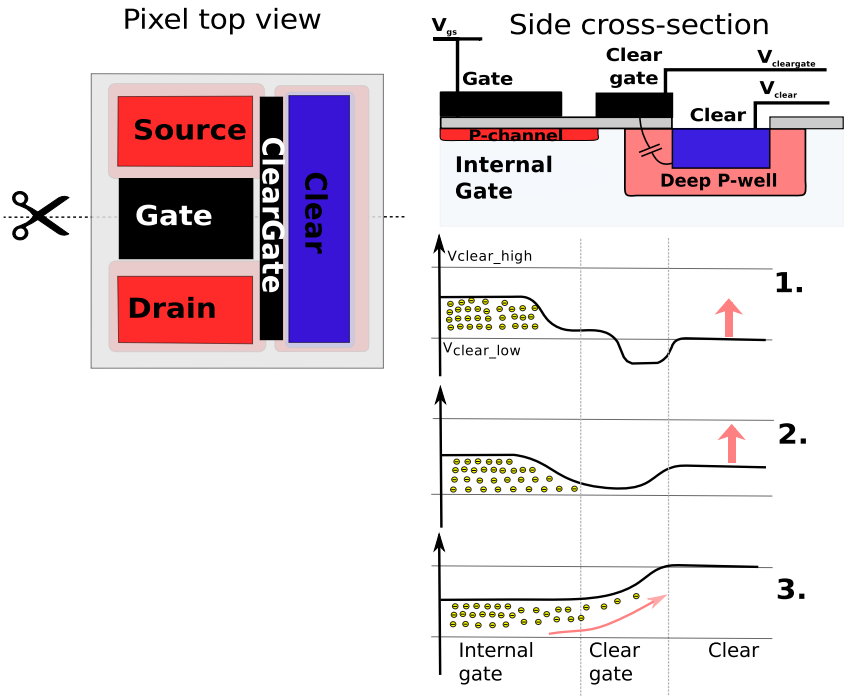


Figure 3.8: The DEPFET clear mechanism. During the clear operation, the potential of the clear contact is raised, as shown in plots (1)-(3). When the V_{clear} is high enough, the potential barrier between the internal gate and the Clear implantation disappears and the collected electrons can be removed from the internal gate by drift.

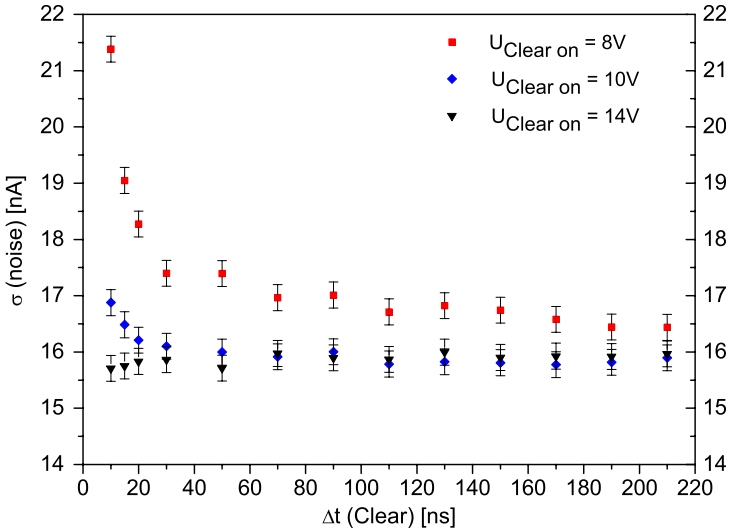


Figure 3.9: Pedestal reset noise depending on the clear pulse's duration. The measurement was based on the fact that after a complete clear the drain current is equal to its pedestal value regardless if the signal was detected before or not; hence the spread should decrease down to its plateau value while the clear pulse gets longer [26]. For high enough *clear_on* voltage, the clear pulse can be shorter than 20 ns.

duced in 2004 in the PXD4 DEPFET design [1], it was initially foreseen to activate this contact during the clear phase. However, this appeared to be not fast enough and added more complexity to the readout.

This contact is now controlled using a similar but more advanced technique, where the contact is not actively controlled by the steering logic but activated passively through the capacitive coupling with the Clear contact. This allows the fast charge removal without increasing the readout complexity. This technique is called the Capacitive Coupled Clear Gate (CCCG).

3.3.2 DEPFET Electrical Model

From the electrical operation point of view, a DEPFET pixel can be considered as a FET with two gate electrodes operating in parallel. Its electrical equivalent is sketched in Figure 3.10.

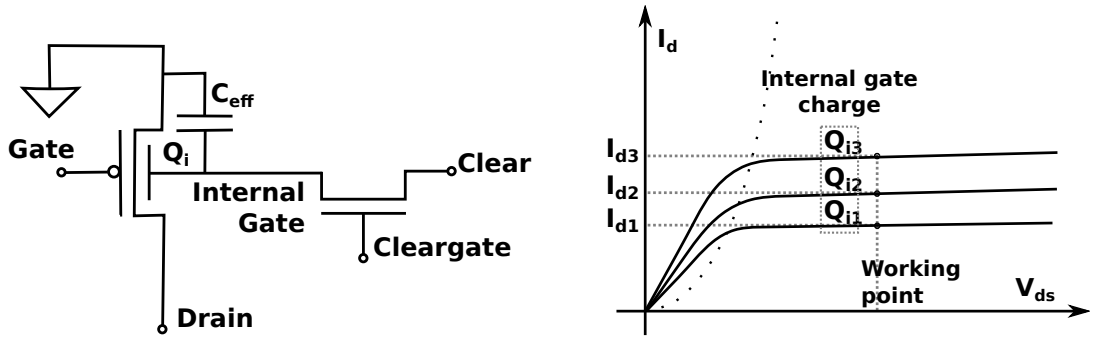


Figure 3.10: (left) DEPFET pixel equivalent circuit. A charge collected in the internal gate influences the drain current in the same way as the regular gate does. This charge can be removed by activating the Clear contact. (right) Drain currents for different charges collected in the internal gate. V_{gs} and V_{ds} are kept constant.

Using the small signal approximation a linear dependence between the transistor drain current I_{drain} and the charge Q_{int} collected in the internal gate can be written as:

$$I_{drain} = I_0(V_{gs}) + g_q \cdot Q_{int} \tag{3.11}$$

with g_q being the internal amplification defined before.
The pedestal current is defined as the drain current that flows when no charge is collected in the internal gate:

$$I_{ped} = I_{drain}(Q_{int} = 0) \tag{3.12}$$

In this case one obtains the collected charge using the following relation:

$$Q_{int} = \frac{1}{g_q}\left(I_{drain} - I_{ped}\right) \tag{3.13}$$

3.3.3 Readout Principles

It is possible to readout the DEPFET at least in two different ways, as is reviewed in [30, 31], i.e using the voltage or the current driven readouts.

[1] The PXD4 is the DEPFET matrix generation optimized for the ILC collider application. The ILC was the first big project where the DEPFET technology was considered to be used [27]. It still remains one of possible options [28, 29].

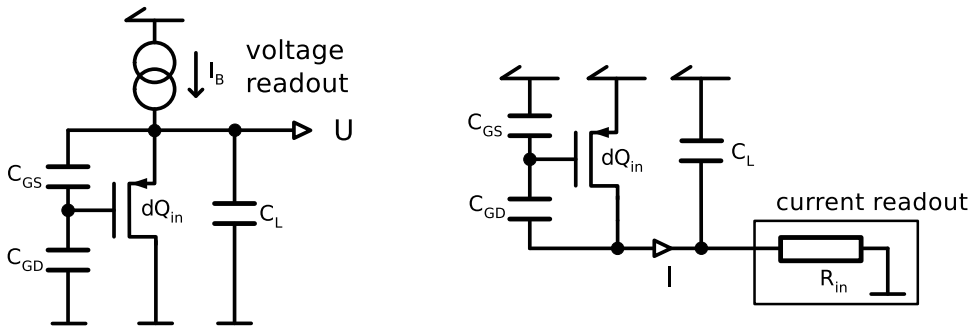


Figure 3.11: DEPFET readout options. Voltage readout (left) and current readout (right) [30, p.31].

Voltage Readout

The voltage driven or the source follower readout is presented in Figure 3.11(left). Here a constant bias current is applied to the transistor. The output voltage depends on the gates' configuration (external and internal ones). In this case the settling time for the output signal can be approximately given by [32, p. 42] or [31]:

$$\tau = 2.2\frac{C_L\left(1+\frac{C_{gs}}{C_{gd}}\right)+C_{GS}}{g_m} > \frac{C_L}{g_m} \tag{3.14}$$

For the long Belle II type matrix the C_L is estimated to be around 50 pF and the transconductance g_m to be 50 uS. This yields τ in order of microseconds, which is too slow for Belle II operation[1].

Current Readout

Presented by Figure 3.11 (right) the current driven readout has a fixed drain-source voltage. In this case the settling time does not longer depend on g_m but only on the load capacitance C_L[2] and the input low resistance of the acquisition electronics R_{in}

$$\tau = C_L R_{in} \tag{3.15}$$

Additionally, the gate switching time is limited by the driving power of the control electronics (see Section 3.5); according to estimations [33] it is expected that the gate switching time is ≈5 ns for a Belle II type matrix.

Using Equation (3.15) with the load capacitance of 50 pF it is enough to have an input resistance of 100 Ω in order to get the necessary value for τ. This is easily achievable by a transimpedance amplifier design, as described in Section 3.5.

For this reason the current driven readout is chosen for the PXD.

DEPFET Powering

In Figure 3.12 a typical powering necessary to run a DEPFET system is presented. Below the description of each bias potential is listed.

[1] As it will be presented in Section 3.4, the PXD readout is constrained by 100 ns per row.
[2] Assuming that $C_L \gg C_{GD}$ and C_{GS}

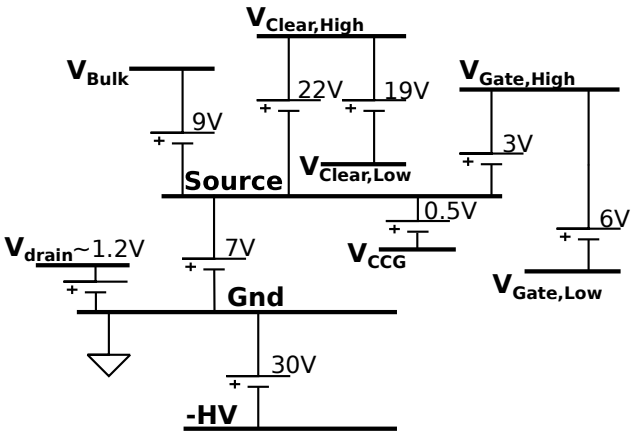


Figure 3.12: DEPFET powering with an example of possible voltages set for a DEPFET matrix (PXD6).

- **Source**. This voltage should be high enough to set the PMOS transistor to the saturation region, which is the working point of DEPFET devices. It is convenient to set most other voltages relative to the Source and not to the Ground.
- **Drain**. Denoted in Figure 3.12 as V_{drain}, this voltage is generated by the DCDB chip. It is kept constant, as required by the Current Readout (Section 3.3.3).
- **Bulk**. As described by the sidewards depletion technique (Section 3.3.1), to properly bias the parabolic potential (Figure 3.5), the bulk voltage V_{Bulk}, should be set higher than the Source voltage. The relation between HV, Source and V_{Bulk} sets the depth of the potential minimum.
- **-HV** (Depletion Voltage). This voltage is applied to the backside of the matrix relative to the ground potential. Together with **Bulk** potential, its purpose is to deplete the sensor.
- **Gate voltages**. There are two voltage levels responsible for the gate operation. The $V_{Gate,High}$ and the $V_{Gate,Low}$. The latter is used to activate and read the DEPFET pixel. The drain current (and the amplification) is defined by the $V_{gs} = V_{Gate,Low} - V_{Source}$ according to Equations 3.6 and 3.5. The $V_{Gate,High}$ voltage is applied to the gate during the collection period. It sets the PMOS into the off-state. Additionally, by applying higher voltage relative to source, one locally influences the potential distribution beneath the gate contact, where the internal gate is situated. This can make the internal gate more attractive for electrons; in this manner the collection efficiency can be increased.
- **Clear voltages**. Similar to the gate voltages, there are two independent clear voltage levels. The $V_{Clear,High}$ voltage defines how fast and efficient the clear process is executed. The Clear is off if it is set to $V_{Clear,Low}$. The value of the $V_{Clear,Low}$ is constrained: if it is set too low it can generate the electrons' back emission from the clear contact to the internal gate; if the $V_{Clear,Low}$ is chosen too high the clear contact may become attractive for electrons during the collection period, so that the charge loss would occur [34].
- **Cleargate voltage**. This voltage can alter the height of the barrier between the internal gate and the clear contact as sketched in Figure 3.8. A good setting of this potential fasten the clear procedure without producing charge losses.

This relatively large set of voltages belongs only to the active area of the sensor. Additionally, there are also ~7 digital and analog supply voltages needed to run the control ASICs.

Readout Sequences

In this chapter the readout concept and a résumé of the two most important techniques is presented. A recent summary of different readout techniques can be found in [35]. The standard technique to readout the signal of a DEPFET pixel is called Correlated Double Sampling (CDS), which is sketched in Figure 3.13. It consists of the following steps:

1. The gate of the pixel is switched to the low voltage $V_{gate,on}$ (PMOS is conducting). After the current has been settled, the value I_{sig} is sampled, its value is stored in a current memory cell[1].
2. The clear contact is connected to the $V_{clear,high}$ potential. Electrons from the internal gate drift towards the clear contact. The collected charge is cleared.
3. The drain current is sampled again but it is now free from the contribution of the internal gate. This is the pedestal current I_{ped}.
4. The value $\Delta I = I_{sig} - I_{ped}$ is acquired by subtracting these two currents and then digitized.

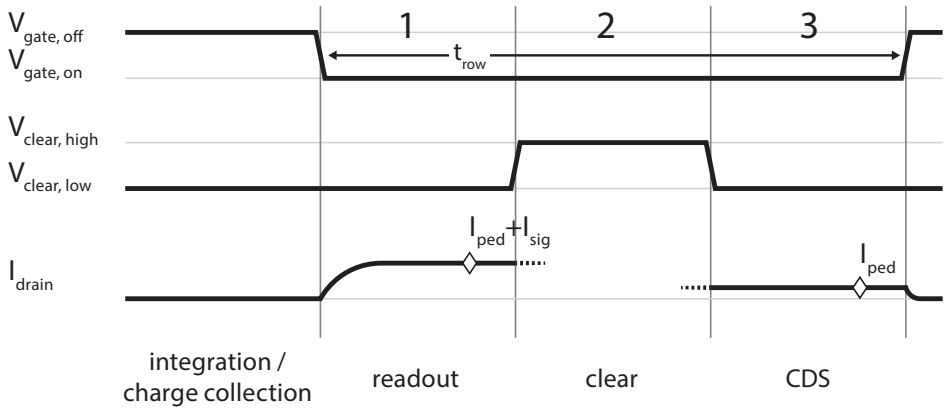


Figure 3.13: Correlated double sampling readout (CDS). Regions 1-2-3 correspond to sample-clear-sample readout steps [37, p.16].

Another way to acquire the signal is called Single Sampling Technique (Figure 3.14). The first two steps are the same as the ones from the CDS: transistor is switched on, the current is sampled and then the collected charge is cleared. In this case the drain current is digitized immediately. The pedestal subtraction happens in the digital logic. In this case it is assumed that the pedestal has been already digitized and its value is stable.

The CDS had been for a long time used as the preferred technique because of the immediate pedestal subtraction: it effectively acts as a high-pass filter and the low frequency noise is suppressed. Additionally, the dynamic range of the digitizer increases: one digitizes only the signal, since the pedestal current is subtracted in the analog domain before that. However, subtraction may not be perfect, residual pedestals still have to be corrected.

Moreover, if the readout speed is of concern, this method is quite constraining: an additional sampling of the I_{ped} to be subtracted from the I_{sig} considerably increases the best achievable readout interval. Assuming the clear operation to be much shorter than the sampling, the minimum acquisition period increases almost by a factor of two. As shown in [37], to fit the necessary 100 ns readout time, this method should be abandoned to the advantage of the Single Sampling technique.

[1] For timing reasons it is preferable to work with currents, while reading the DEPFET sensor, thus the subtraction happens in timing domain using the current memory cell as intermediate value storage element, whose concept can be found in the paper published by Hughes et al. [36].

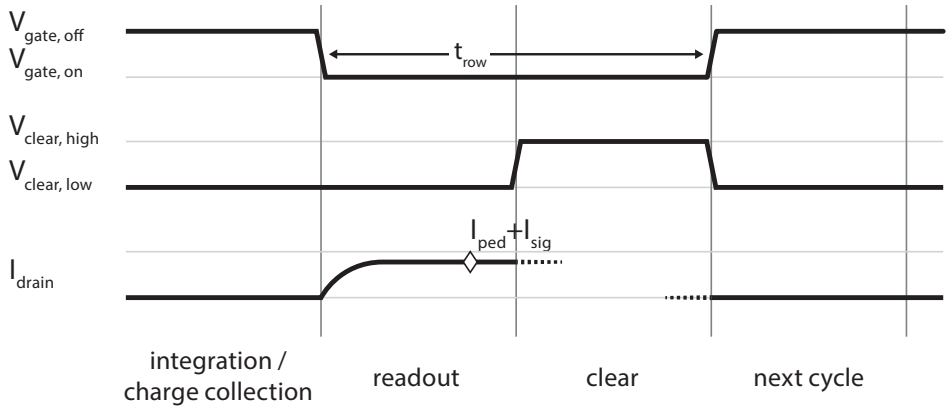


Figure 3.14: Single sampling readout scheme [37, p.16].

3.4 The DEPFET Matrix

To build a DEPFET pixel detector, the pixels are organized in a matrix. The current structure of this matrix is presented in Figure 3.15: Drains are connected by columns. Gate and clear contacts are connected by rows. Rows are steered by Switcher chips, which control the readout.

A typical readout is done in the following steps: at every moment Switcher chip activates one row; once its drains are digitized, the next row is activated and so on (for details, check Section 3.5). This row-wise sequencing is called rolling-shutter mode.

The main advantage of this technique is its scalability: it is easy to readout a large matrix with relatively little electronics. Further, the power consumption is kept low: it scales with the number of drains and not with the number of rows. However, the main drawback is timing: the longer the matrix, the slower the readout.

The timing issue was of concern during the design of the Belle II devices. In 20 µs of frame readout time [14] 1600 row should be digitized. This would result in unmanageable 12.5 ns per row. To relax the design, the matrix is split in two independently read parts of 2×768 pixels[1]. To further speedup the readout, rows are controlled by a set of four in parallel (see Figure 3.15). This results in approximately 100 ns per row. This timing is still tough for Belle II operation but manageable if the Single Sampling readout technique is used.

3.5 DEPFET Readout Electronics

The Switcher-B Chip

In the context of Belle II, the Switcher chip is called Switcher-B (B for Belle). In the PXD architecture it is conceptually the simplest ASIC. Designed in 180 nm HV CMOS technology [38], the Switcher-B chip aims to control the DEPFET matrix by means of providing high-voltage pulses to the clear and gate contacts for each matrix row.

As already written in the first section of this chapter (Figure 3.4), this chip will be mounted on the module's long-side balcony. Each chip is capable to control 32 rows of the matrix. 6 chips are necessary to steer 192 rows of a PXD module. To do so, the Switcher-B is designed to be used in a daisy chained way.

[1] Slightly decreasing the total number of pixel from 1600 down to 1536.

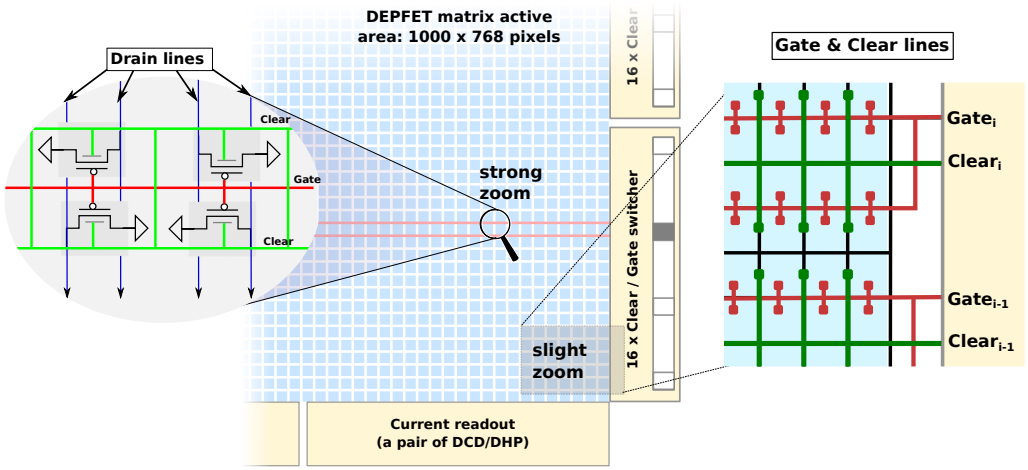


Figure 3.15: DEPFET matrix organization. (Strong zoom) Pixels drains are connected column-wise. At every moment only one row is active, whose pixel drain currents are digitized. (Slight zoom) One logic row corresponds to four physical rows: every gate and clear line is connected to four matrix rows at the same time in order to speed-up the readout.

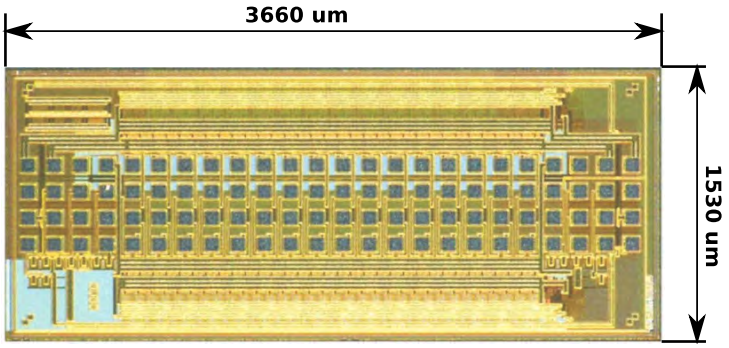


Figure 3.16: Switcher-B layout, photo from the Switcher-B manual [38].

The block diagram of the Switcher-B is sketched in Figure 3.17. It consists of two parts: the low voltage (logic block), where the control signals are generated and the high-voltage one, which is responsible for providing the necessary output signals.

The central element of the logic block is a simple 32-bit long shift register (SR): the *serin* strobe is sampled and propagated through the SR on every positive edge of the *clk*. The propagated bits activate the corresponding gate/clear controlling driver in the high-voltage block. With additionally provided *strG* and *strC* strobes the necessary control sequence (as sketched in 3.13 and 3.14) is generated in the logic block and serves to control the activated row driver.

The high voltage driver converts the generated logic signal into the high voltage constrained between two rails: $V_{gate,low}$ and $V_{gate,high}$ for the gate driver and $V_{clear,low}$ and $V_{clear,high}$ for the clear one.

In spite of the apparent simplicity, the Switcher-B has quite a challenging design, which is needed to fulfill the following constraints:

- As reported by [26], the output channels has to support up to 20 V. Therefore. a high voltage CMOS technology has to be used. The logic low voltage part of the chip must be electrically

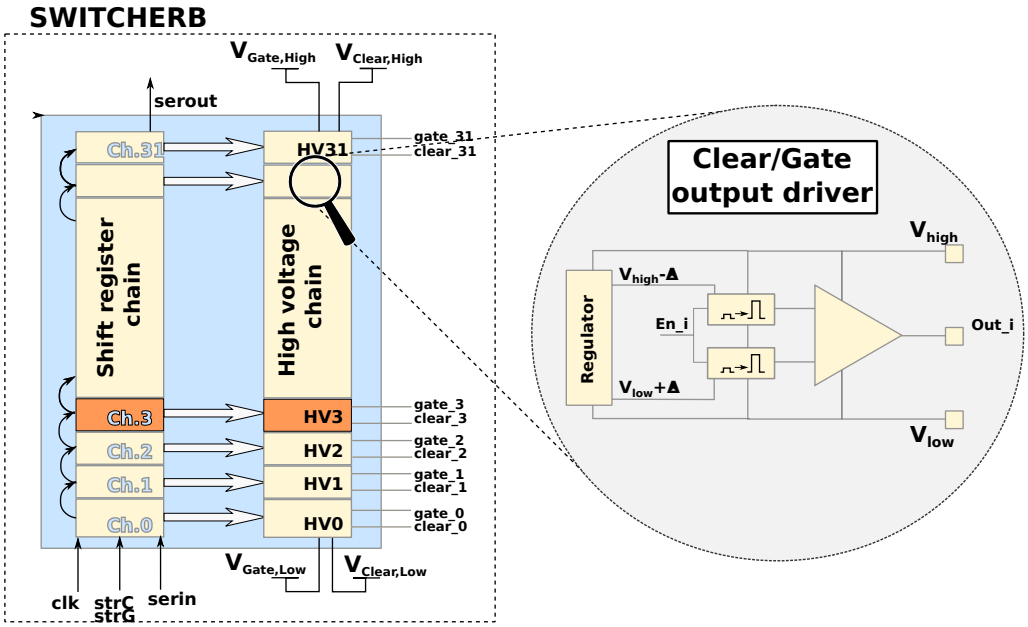


Figure 3.17: Switcher-B block diagram. The shift register controls the high voltage drivers for each channel. The serial output of the shift register can be connected to the serial input of the next chip.

decoupled from the output driver blocks. It is implemented by means of fast and low-power level shifters between these blocks.

- Analog electronics is particularly sensitive to radiation damage. To support 20 MRad radiation, as estimated by the Belle II Collaboration, special design techniques are used. According to radiation tests, the chip is expected to sustain at least 37 Mrad [14].
- Fast rise time: the output driver should provide enough current to provide about 20 ns pulse width for the output load of 50 pF [14].

The Switcher-B has a serial slow control JTAG interface compatible to the IEEE 1149 standard [39]. It provides testability features while integrating the chip to the PXD module. It can additionally alter default settings for in-chip current sources if necessary [33].

The Drain Current Digitizer

Overview

In the context of Belle II the Drain Current Digitizer (DCD) is called DCDB (again, B stands for Belle). It is a complex chip aiming to digitize the analog information coming from the detector. Its main block consists of an array of 512 ADCs[1]. To each of the 256 input channels belong two ADCs working alternatively. Each ADC is designed to run at 40 MHz clock frequency and needs 8 clock cycles for one conversion. This results in 2×40/8=10 Msps per channel or in 20.48 Gbps of the data-rate per chip[2]. The output of the DCDB is only 64-bit wide. To cope with this rate it is designed to run at the base clock frequency of 320 MHz[3].

[1] ADC stands for the Analog-to-Digital Converter
[2] This amount equals roughly to 1 DVD/sec/chip.
[3] In reality, it will run in PXD at 305 MHz. This clock will be introduced later in Chapter 4.1.

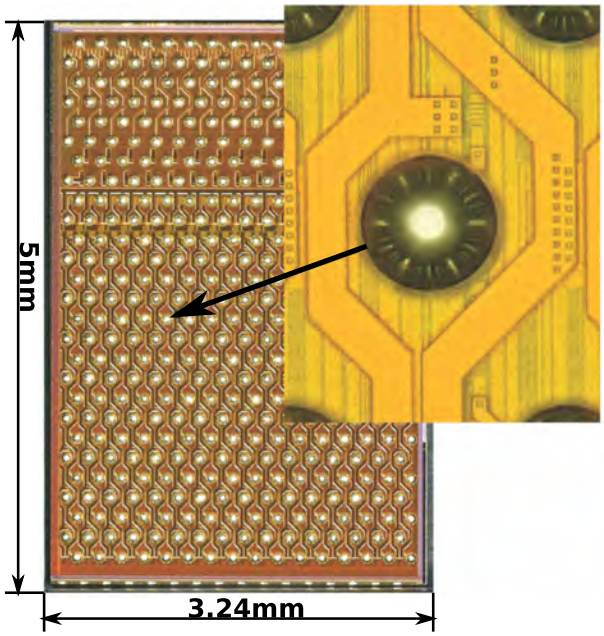


Figure 3.18: DCDB photomicrograph [38]. The chip dimensions and pads structure are shown. The DCDB is designed to be mounted using the bump-bonding to the DEPFET module. The chip has 256 current inputs (bottom part), 64 digital outputs (upper part) and others (power IOs, slow control etc).

The DCDB is produced using the 180 nm CMOS technology, with dimensions of 5×3.24 mm^2 as presented in Figure 3.18. It has 432 pads in total: 256 analog inputs, 64 digital outputs and others (power, slow control, sync etc.).

Input Stage

As discussed in Section 3.3.3, in order to fit the Belle II specified timing, a voltage based readout is too slow; therefore, the current based readout option has been chosen for the design of the DCDB chip.

The main block of the input stage of the chip is a transimpedance amplifier (TIA) shown in the centre of Figure 3.19, whose detailed description can be found in [40, 41]. Its main purpose is to keep the voltage on the input constant (i.e. keep the effective input impedance as low a possible) to maintain a fast settling time in presence of a large input capacitance; herewith the fast acquisition can be achieved. The output of the TIA is then converted to the current with the help of resistor R_s, as shown in Figure 3.19.

Next, this current is digitized by two 8-bit cyclic current-mode ADCs, working alternatively to speed up the acquisition. The nominal ADC dynamic range is 16 µA, which is less than the estimated pedestal current spread for the Belle II-type matrix. To cope with this spread the current can be adjusted in two ways:

- **The global adjustment**. The current offset can be globally tuned with two current sources: *NSubIn* (before the TIA) and *NSubOut* (after the TIA). The gain of the receiver can be globally adjusted being equal to one, two, three and four by setting the feedback resistance R_F as defined by:

$$G_{TIA} = \frac{R_f}{R_s}$$

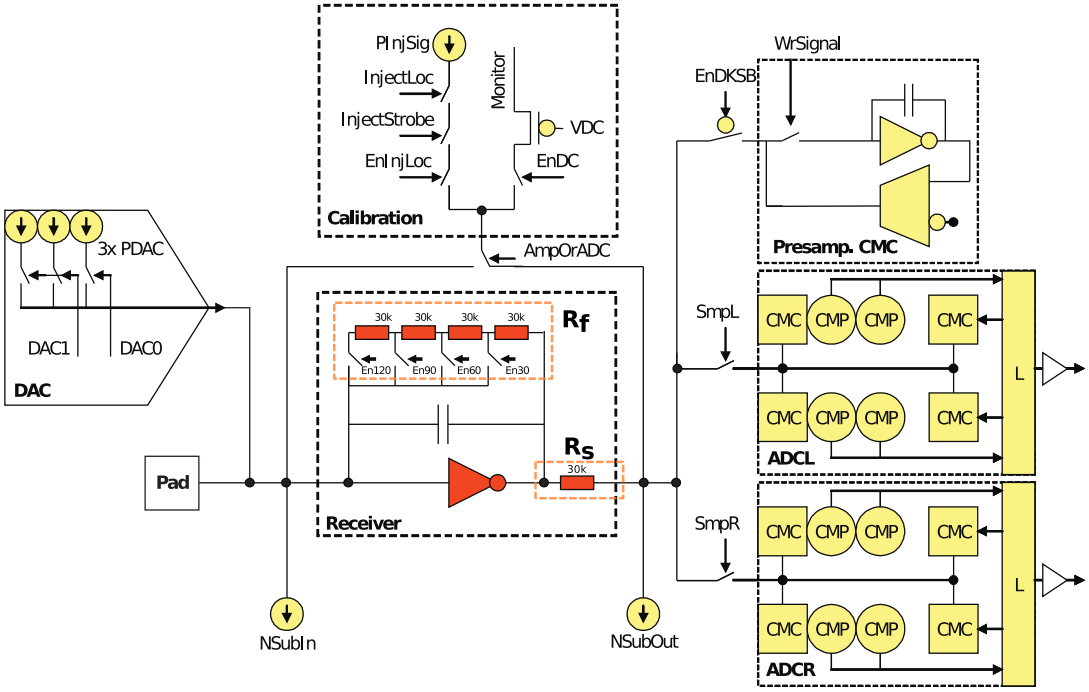


Figure 3.19: DCDB input stage. The main element is a transimpedance amplifier labeled as **Receiver**; its gain can be varied by adjusting the resistor R_f. Current sources *NSubIn* and *NSubOut* can be set to fit the input into the ADC's dynamic range. Additionally the offset of each input channel can be varied locally by using a dynamically adjustable DAC (in figure on the left side). Two cyclic current ADC's alternatively sample the input signal; they can be activated by switches *SmpR* and *SmpL*. Source: DCDB reference manual [40] .

- **Per channel adjustment**. The input can be dynamically adjusted by subtracting the current produced by two-bit current DACs. These DACs are externally controlled by the signals sent from the DHP chip.

Readout Modes

The DCDB chip is designed to support both acquisition modes discussed in Section 3.3.3, the CDS and the Single Sampling techniques.

The default mode is Single Sampling: the current is directly digitized and corrected in the digital domain by the DHP chip. However it is also possible to switch to the Correlated Double Sampling mode by applying the switch *EnDKSB*. In this case, the sampling happens twice: first before the clear: the value is stored in the current memory cell [41, p. 43], marked in the block diagram as *'Presamp.CMC'*. After that, the sampling is taken for the second time after the clear pulse; the currents are subtracted and converted by the corresponding ADC.

DHP

Different from the two previously presented ASICs, the Data Handling Processor (DHP) is an almost entirely digital chip. It is primarily used to reduce data rates produced by the DCD: in its absence the total rate of 160×20.48 Gbps≈3 Tbps produced by the whole detector would be very difficult to handle. Moreover, it is actually not feasible to transmit this amount of data to the back-end electronics due to the serial link bandwidth constraint (Section 3.6). To do the data reduction, the DHP does the

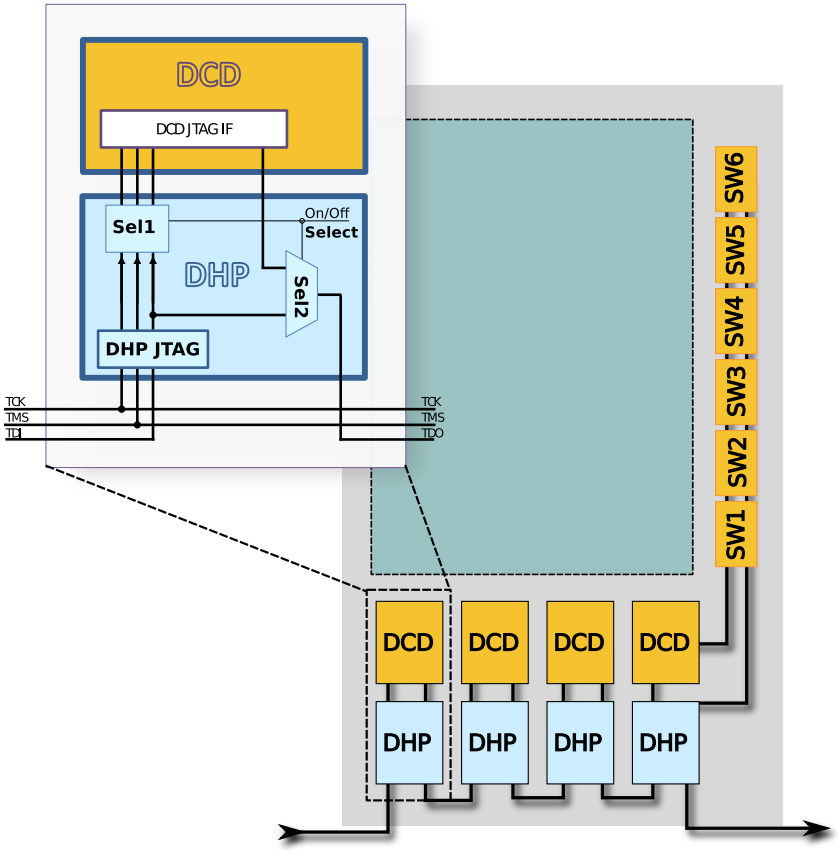


Figure 3.20: The slow control chain of a PXD half module. As shown by the left-top selection, Switcher-B and DCDB chips can be excluded from the configuration chain in order to have a faster access to the DHPT 1.0 chips.

data pre-processing consisting of the following steps: triggered frame selection, pedestal subtraction, common mode correction and zero suppression. Serial formatting is done in order to send the data to the DHH (Section 3.6) through a specially designed current mode logic serial link.

To correct for the high pixel-to-pixel pedestal spread each DHP sends to the corresponding DCDB the so-called offset sequence. Furthermore, one of the DHPs situated on the module is responsible for the control sequence generation for the Switcher-B chips chain, thus steering the matrix.

The DHP design development and optimization is in the focus of this thesis, more details about the DHP can be found in Chapter 4.1.

Module Slow Control

All ASICs of the PXD have a JTAG-standard [39] slow control configuration interface. In Figure 3.20, the global chain is presented, where all of all 14 ASICs are included.

Except of the initial boot-up situation, one would rarely need to access the configuration registers of the Switcher-B and the DCDB chips; for that reason the paths including these chips can be excluded from the chain by the DHPT 1.0. As is seen in the zoomed area in Figure 3.20 a special switch inside the DHPT 1.0 is implemented to bypass the chain.

3.6 Back-end Electronics

Under the generic term *Back-end electronics* we understand all the electronic blocks belonging to the PXD, but situated outside of the silicon sensor barrel. These include the Data Handling Hybrid (see below), data and power link, ATCA OnSen etc. (see Figure 3.21).

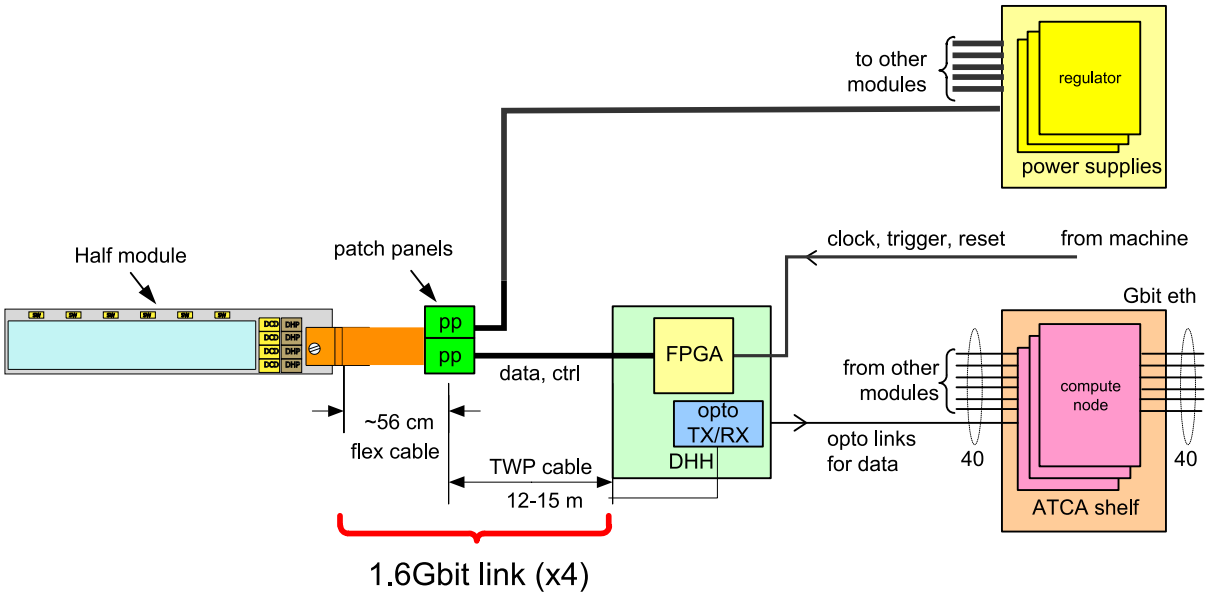


Figure 3.21: PXD link. Source: [42, DEPFET Workshop report]

System Clocks

All system clocks are derived from the main RF frequency of 508.887 MHz of the accelerator in order to remain synchronized. For example, the DHH receives the reduced RF frequency of 127.22 MHz. The main DHPT 1.0 clock of 76 MHz is obtained by its further multiplication by 6/10.

Data and Power Links

It is estimated that about 15–20 m of cabling is needed to interconnect the PXD detector with the first stage of the back-end electronics (the DHH, see Section 3.6).

Due to fragility and lack of space near the IP, the PXD modules are connected using a polyamide (Kapton®) cable, which is a flexible four-layer PCB[1], specially designed for the PXD. It transmits the power supply, slow control, synchronization and data signals; the length is estimated to be about 56 cm. On the opposite side it is connected to two patch panels where it is split in two parts as shown in Figure 3.21.

The first part serves for the digital data transmission; at this point it will be connected to the twisted pair (TWP) cable, which will be about 15 m long. The DHPT 1.0 serial output links are limited to 1.53 Gbps, which is 20×76 MHz base clock frequency provided to the PXD.

The second patch panel is connected to the detector power supply.

[1] PCB stands for Printed Circuit Board

Figure 3.22: DHH prototype. Based on Virtex-6 FPGA, version insertable in ATCA Module.

DHH

The Data Handling Hybrid or DHH is designed as a master module for the four DHP chips mounted on each half-module. Here is the list of its tasks [43]:

- **The galvanic isolation of the DEPFET modules from the external electronics.** Due to low radiation resistance of the commercially available optical modules, each PXD module is connected electrically to the DHH, which will be situated much further away from the IP than the PXD modules. Hence, the use of optical transmitter is possible on the DHH. In such a way the DHH provides the electrical ↔ optical signal conversion between the external electronics and the PXD.
- **Clock and trigger distribution.** The DHH is directly connected to the FTSW[1] board developed by KEK. The internal logic of the DHH decodes trigger and synchronization signals and forwards them to the DHP.
- **Remapping and clustering**. Due to complex matrix wiring, the zero-suppressed hit data received from the DHP chips will differ from the real hit position in the DEPFET matrix. A remapping procedure can be done on the DHH to correct this. Additionally, it is planned to forward the data-steam to the DCE3 [44] ASICs installed on DHH to clusterize the data before sending it to the compute node.
- **Module slow control**. Each PXD module is steered using the JTAG interface. The DHH receives JTAG commands from the DEPFET Slow Control Server, reinterprets and executes them using JTAG player core, which runs on the DHH FPGA.

ATCA System

The DHH sends the data to the ATCA[2] system, which consists of several Compute Nodes (CN), placed in one ATCA shelf (Figure 3.23). At this stage, the PXD data is combined with the strip detector data.

[1] FTSW stands for the Frontend Timing SWitch
[2] ATCA stands for the Advanced Telecom Computing Architecture

Figure 3.23: The ATCA shelf with two Compute Nodes installed [45].

DATCON[1] and the high level trigger are used to search for the Regions of Interest[2] for the further data reduction. The data is then sent to the data storage server.

[1] Data Concentrator

[2] The area where the track of interest is expected to be found is called Region of Interest. They are searched with the help of information coming from other detectors.

Chapter 4

The Data Handling Processor

This chapter is dedicated to the Data Handling Processor chip development. First, an overview of the chip is given and it is explained in which context it is used in the PXD. Then I proceed with a detailed description of its elements. Further, implementation aspects are discussed.

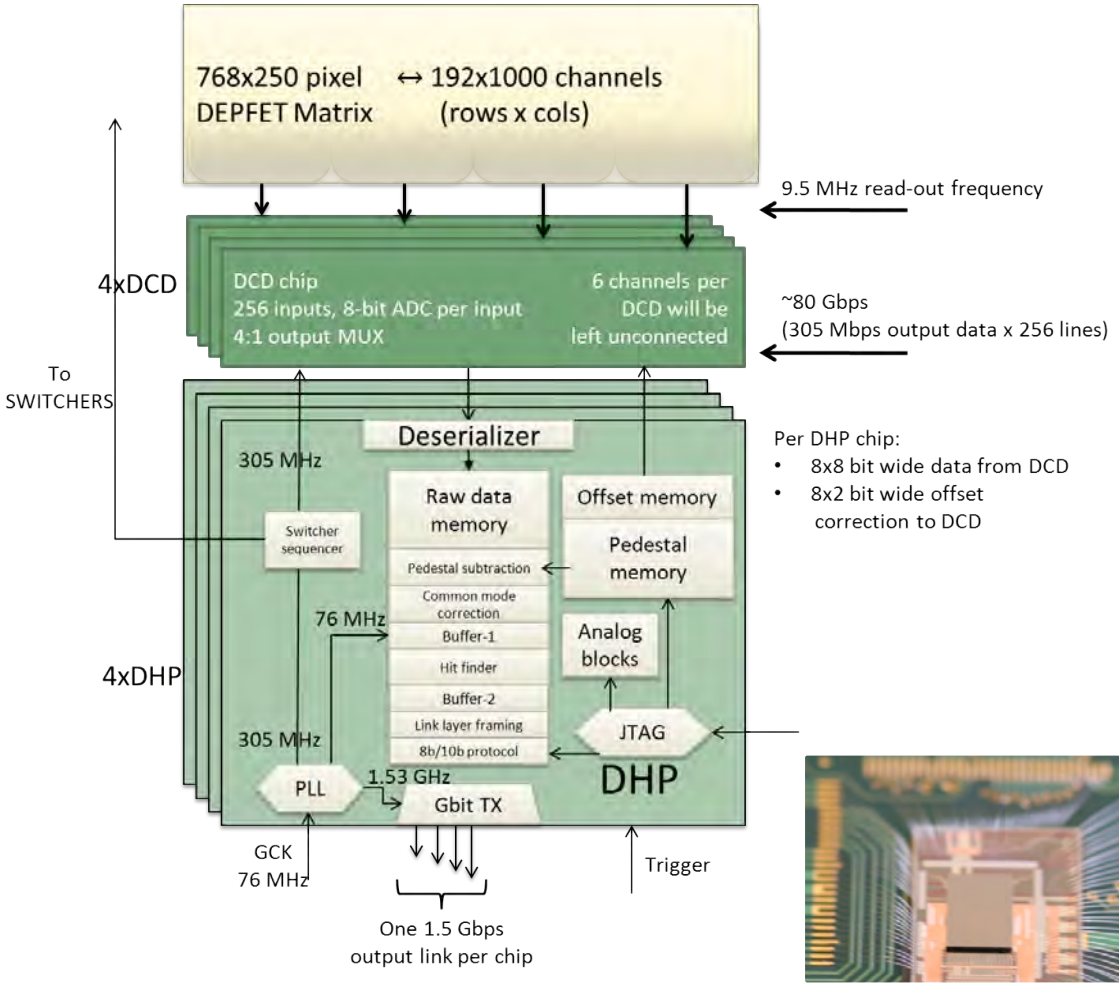


Figure 4.1: DHP block diagram. Four DHP chips interact with four DCD chips; these are necessary to readout one DEPFET matrix of a PXD half-module.

Figure 4.2: DHP 0.2 on the wirebond adapter.

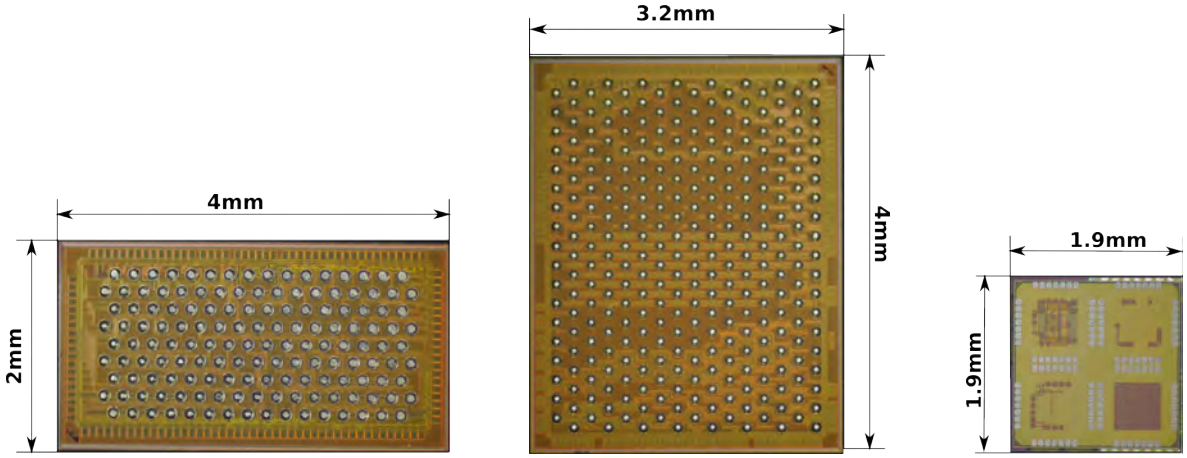
4.1 Overview

The DEPFET matrix consists of 768×250 pixels, corresponding to the logic geometry of 1000 drain columns × 192 gate rows, as mentioned in Section 3.4. The row sampling frequency is ≈9.5 MHz with a digital resolution of 8 bits. This corresponds to ≈80 Gbps data rate per half ladder. As described in Chapter 3.5, the data reduction is necessary on this stage.

The Data Handling Processor (DHP), whose block diagram is presented in Figure 4.1, reduces by at least a factor of 15 the data rate produced by the DCDB chip. This is achieved in two ways:

1. **Trigger selection**. The external trigger defines the time regions of interest, whose data is transmitted to the back-end electronics.
2. **Zero Suppression**. The raw data is preprocessed in the DHP and only the information about non-empty pixels is selected. To do so, pedestal subtraction, common mode correction and hit-search are needed.

The DHP chip also steers the readout process on the PXD module, i.e. it generates timing signals and switching sequences for the DCDB and Switcher-B chips.



(a) DHP 0.1 chip (b) DHP 0.2 chip (c) DHPT 0.1 chip

Figure 4.3: Three test chips submitted for the DHP project. The DHP 0.1 and DHP 0.2 are implemented in IBM CMOS 90 nm technology. The DHPT 0.1 is implemented using TSMC CMOS 65 nm technolgy.

Currently, three development chip versions have been produced and one production version is submitted.

- **DHP 0.1** (Figure 4.3a). The first test chip, submitted in March 2010, uses the IBM 90 nm CMOS technology. It measures half of the planned size: 2 mm×4 mm, with only half of the inputs. It has the basic data processing elements, serial interface and all prototypes of analog blocks, such as PLL[1], the output link CML driver[2], ADCs and DACs[3].
- **DHP 0.2** (Figure 4.3b) is the second chip iteration made after first results obtained from DHPT 0.1 tests. Submitted in October 2011, this full size chip has an improved and complete data processing

[1] PLL stands for Phase Lock Loop, see Section 4.3.1 for details
[2] CML stands for Current Mode Logic, see Section 4.3.2 for details
[3] ADC stands for Analog to Digital Converter DAC stands for Digital to Analog Converter

chain. The design is targeted to be very close to the final version. This chip is used in the Full-Scale Module Prototype (Chapter 6.2).

- **DHPT 0.1** [1](Figure 4.3c). First prototype in TSMC 65 nm. We switched to the CMOS TSMC 65 nm technology since the CMOS 90 nm IBM was not available any more for prototyping and it would be too expensive to continue using it. This chip contained several independent circuits such as one PLL, one CML, and one digital test block. In the scope of this thesis, this chip was used to investigate the radiation sensitivity of the 65 nm technology in order to estimate the digital error rate during the run of the detector (see Chapter 7).

- **DHPT 1.0** (not submitted at the moment of writing this chapter). With some important enhancements and minor corrections, it will have similar data processing as the DHP 0.2 chip. This chip is planned as the first production chip to be used in the PXD.

The following notation is valid in the upcoming sections: when speaking about the conceptual solution the simple abbreviation "DHP" is used; if speaking about one of the test chips listed above, its full name, for example DHP 0.2, is used.

4.2 Data Processing Blocks

The data reduction is the main task of the chip, which is split in several steps, as shown in Figure 4.1. Below the step-by-step description is presented.

4.2.1 Deserialiser

The 256 input channels of the DCDB chip are organized in 8 double-columns with 32 ADC outputs each. Each column has one 8-bit wide output link. This makes in total 64 outputs connected to DHP inputs. The data transmission runs at 305 MHz clock rate.

First, the data is deserialized: the received data is reorganized and written row-by-row into the raw memory data buffer. For each row 32 DCD clock cycles are necessary. The buffer has a depth of one frame and is designed as a ring buffer: data that is older than one frame period is overwritten.

This ring buffer serves as a programmable delay element: further data processing starts upon trigger arrival coming from the outside. The estimated trigger latency is expected to be about 5 µs [14]. This delay corresponds to ≈50 rows, since each row has 100 ns sampling time. Therefore, this buffer should be at least 50 rows deep. The current ring buffer can have a maximal depth of 256; it is more than enough to adjust for any foreseeable latency.

4.2.2 Raw Data

The input values to the DHP chip consist of four components:

$$I_{cr,t} = S_{cr,t} + P_{cr} + CM_t + N_{cr,t} \tag{4.1}$$

$I_{cr,t}$ denotes the DHP raw data amplitude. Indexes c and r stand for logic coordinates of the digitized pixel, i.e. the DCD column and Switcher row. t stands for time.
$S_{cr,t}$ is the signal we want to detect.
P_{cr} is the static (or slowly changing) offset, which is different from pixel to pixel due to technology

[1] Here T is for TSMC

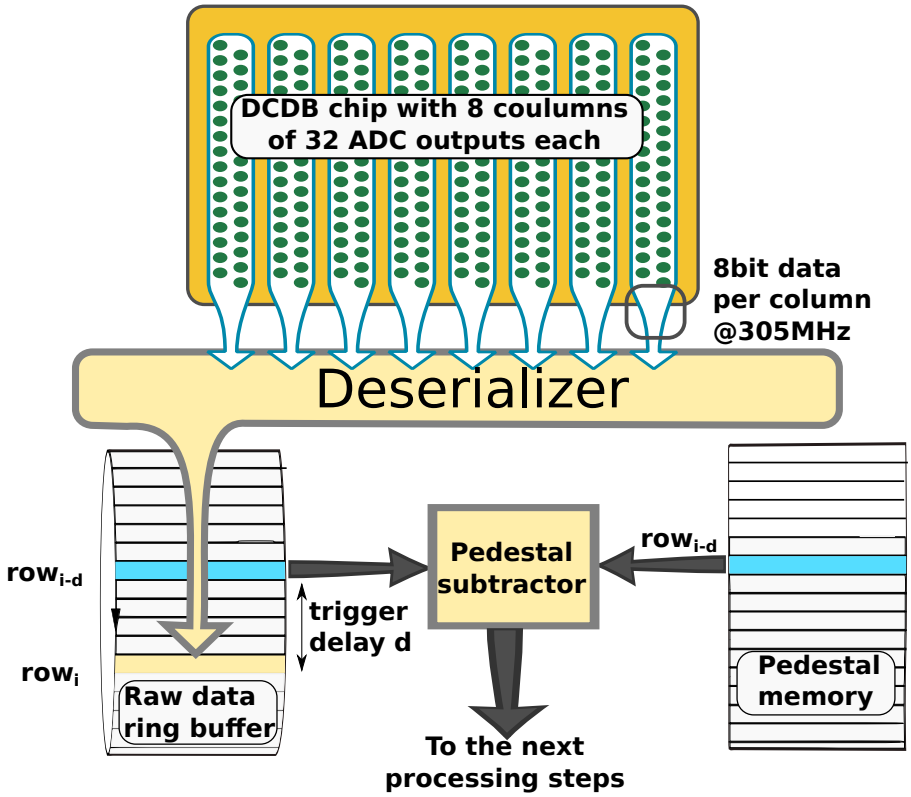


Figure 4.4: Deserialization and pedestal subtraction block diagram. The total input rate of the DHP is 305 MHz $\times$ 8 $\frac{\text{bits}}{\text{column}}$ $\times$ 8 columns = 19.5 Gbps. The deserialized data is continuously written into the Raw Data Ring Buffer. Upon the trigger arrival the data and pedestals are sampled with a certain programmable latency and sent to the Pedestal Subtraction block.

variations or power supply distribution. This offset is called pedestal. It is also sometimes denoted as the fixed pattern noise.
CM_t is the pick-up noise. Being common to all pixels that are sampled at the same time, it is possible to estimate and correct for it.
$N_{cr,t}$ is random noise of other nature, which is independent per channel (thermal noise etc).

Since the sampling is executed in a row-wise manner, the time dependency can be replaced by the row dependency, hence the number of indices in this equation can be reduced and rewritten as:

$$I_{cr} = S_{cr} + P_{cr} + CM_r \tag{4.2}$$

The remaining noise $N_{cr,t}$ is not included.

4.2.3 Pedestal Subtraction

Let i be the current row number written into the memory and $\delta_{trigger}$ the trigger latency expressed in the number of rows the DHP would digitize during the corresponding delay. Upon the trigger arrival the row number $r = i - \delta_{trigger}$ is read from the ring buffer. Together with the raw-data buffer the DHP chip contains the pedestal memory storage of equal size, which is read at the same time at the same row

position number r. Both data vectors are sent to the Pedestal Subtraction block, which calculates the output values:

$$R_{out,cr} = I_{cr} + \Delta_{CM} - P_{cr} = S_{cr} + CM_r + \Delta_{CM} \tag{4.3}$$

where R_{out} is the pedestal-free output vector and Δ_{CM} is the offset we need to add, which should be at least as big as the maximum expected Common Mode noise (see Section 4.2.5). This is needed to avoid working with signed numbers, so one adds this offset to the final result to be sure that the answer will always be positive. All these processing steps are presented in Figure 4.4.

4.2.4 Pedestal Update

In principle, the pedestal current can be subtracted using the Correlated Double Sampling (CDS) method [46]. However, the CDS reduces the readout speed by a factor of two in comparison to the single sampling mode. The readout speed limitations thus make the usage of the CDS impossible. That is why the DHP stores all the pedestal information to subtract it from the raw data. These pedestals have to be continuously refreshed as they are very sensitive to a number of external factors such as temperature, radiation etc.

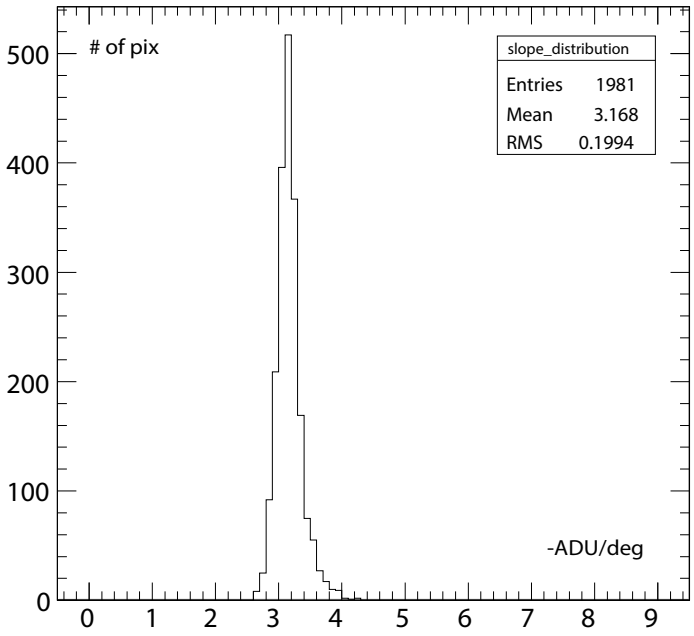


Figure 4.5: Histogram of pedestal temperature sensitivities [-ADU/deg] for a PXD6 type matrix with typical voltage settings, as described in Section 3.3.3. Courtesy of T. Kishishita

Figure 4.5 presents a typical pedestal sensitivity to temperature. The observed spread was measured for standard voltage settings that are used for the PXD6 generation of DEPFET matrices. As seen from the histogram, a variation from -2.7 ADU/°C to -4 ADU/°C can be observed from pixel to pixel. Taking as an example the ambient temperature change of 2 °C, the pedestal value of a pixel with the minimum

sensitivity would decrease by 3.4 ADU and that of a pixel with the maximum would decrease by 8 ADU, completely distorting the resulting pedestal map.

If even a slight variation of external conditions leads to important distortions, it is mandatory to have a continuous pedestal monitoring and update procedure. The update procedure can be either controlled by the chip (online update) or externally (offline update).

From the hardware development point of view, the simplest option is to keep the static pedestal memory inside the chip and do all necessary calculations offline and use the slow control interface to upload the pedestals values. This procedure is however time consuming. In the worst case scenario, it may take up to 15–30 mins [47]. That means if one deals with pedestals variations faster than this period, it will not be possible to compensate for them during the PXD operation. To solve this problem the integrated on-chip solution was considered to be used, which was called *dynamic pedestals* (see Appendix B). This method was tested in the DHP-emulator described in Section 6.1. However, the dynamic pedestals solution was too resource demanding. Moreover, laboratory tests showed that the pedestals are rather stable, if keeping power supply and temperature constant. Consequently, the offline update option was chosen for the final design.

4.2.5 Common Mode Correction

From Equations 4.3 and 4.2 the data coming from the output of the Pedestal Subtraction block looks like:

$$R_{cr} = S_{cr} + CM_r \tag{4.4}$$

(The trivial summand Δ_{CM} is not included).

In this processing step the necessary corrections for the Common Mode (CM) noise are done. The CM is the noise component, whose amplitude is the same for all values sampled at the same time. It is often a result of a pick-up fluctuation, affecting the analog electronic circuits. It is corrected before the zero suppression, where the threshold cut is applied.

Processing Constrains

For the processing chain to be sustainable, processing steps cannot last longer than the time delay until new data comes. As was previously said, the new DCD row arrives on every 32 DCD clock cycles (305 MHz) or with a total delay of ~100 ns. The digital processing runs using the GCK clock, which is 4 times slower. Consequently all operations should fit within 32/4=8 GCK clock cycles (76.3 MHz). To relax the design, the data is processed in four chunks of 64 bytes, arriving every 2 clock cycles instead of one 256 byte chunk of data arriving each 8 clock cycles.

The complexity of the task consists in finding the solution that fits in the given time delay.

Simple Average

The simplest way to estimate the common mode value $\widetilde{CM}_t$ is to take the average of the signal. As follows from Equation 4.4 this is equal to:

$$\widetilde{CM_r} = \frac{\sum_{c=0}^{N-1} R_{cr}}{N} = CM_r + \overline{S_r}$$
$$\text{with } \overline{S_r} = \frac{\sum_{c=0}^{N-1} S_{cr}}{N} \tag{4.5}$$

From Equations 4.4 and 4.5, the extracted signal is then equal to:

$$\widetilde{S_{cr}} = R_{cr} - \widetilde{CM_r} = S_{cr} - \overline{S_r} \tag{4.6}$$

where $\widetilde{S_{cr}}$ is the estimated signal value. This bias can be corrected offline (see Appendix A for details).

The bigger problem is that by systematically underestimating $\widetilde{S_{cr}}$ hits with small amplitudes are lost when they are close to the threshold value. This depends on the per-row hit occupancy.

Median

A better way to estimate the CM is to take a widely used statistical method, the median value (further in text it is just called the median). The median is less affected by outliers and skewed data, which is the case if the detected signals are rather scarce.

Let m be the median value of a set $\mathcal{S}$ containing N values. The elements of $\mathcal{S}$ are indexed in increasing order:

$$\begin{aligned} &\mathcal{S} = \{x_1, x_2, , , , x_N\} \\ &\text{such that:} \\ &x_i \leq x_{i+1} \ \forall i \end{aligned} \tag{4.7}$$

In this case the median will be defined as follows [48]:

$$m = \begin{cases} x_{\frac{N+1}{2}} & \text{if N is odd} \\ \frac{1}{2}\left(x_{\frac{N}{2}} + x_{\frac{N+1}{2}}\right) & \text{if N is even} \end{cases} \tag{4.8}$$

From this definition 4.8 it follows that to find the median value of an unarranged set one 'simply' needs to sort it. However, for typically used sorting algorithms such as *binary tree sort* [49] the complexity is estimated to be $O(n \cdot ln(n))$ or worse. That is, for an array of 256 elements O(2000) operations would be needed. For a software implementation on a high performance CPU this would result in several microseconds of delay, which is more than ten times longer than acceptable. In ASIC design one has rather limited resources (area, memory), however one is free to choose the hardware architecture giving a large degree of flexibility. In this case the pipelineability and parallelizability of an algorithm play an important role.

Alternative and more performing algorithms can be found, such as described in [50], which is based on cumulative histogramming. The algorithm can be summarized as follows:

1. The baseline is subtracted from the input data set to fit the output into the predefined limits. An assumption is made that the data do not vary too much and can be binned in a small number of cells: 16, 24 (32 in the initial paper).[1]

[1] The higher the binning the slower is data processing. Eventually, 32 bins proposed initially were estimated to be too slow relative to the required delay.

2. A histogram of the input data is generated.
3. A cumulative histogram (CH) is generated. For a set with N members, the first CH element whose frequency is higher or equal to the $N/2$ is the median, as presented in Figure 4.6.

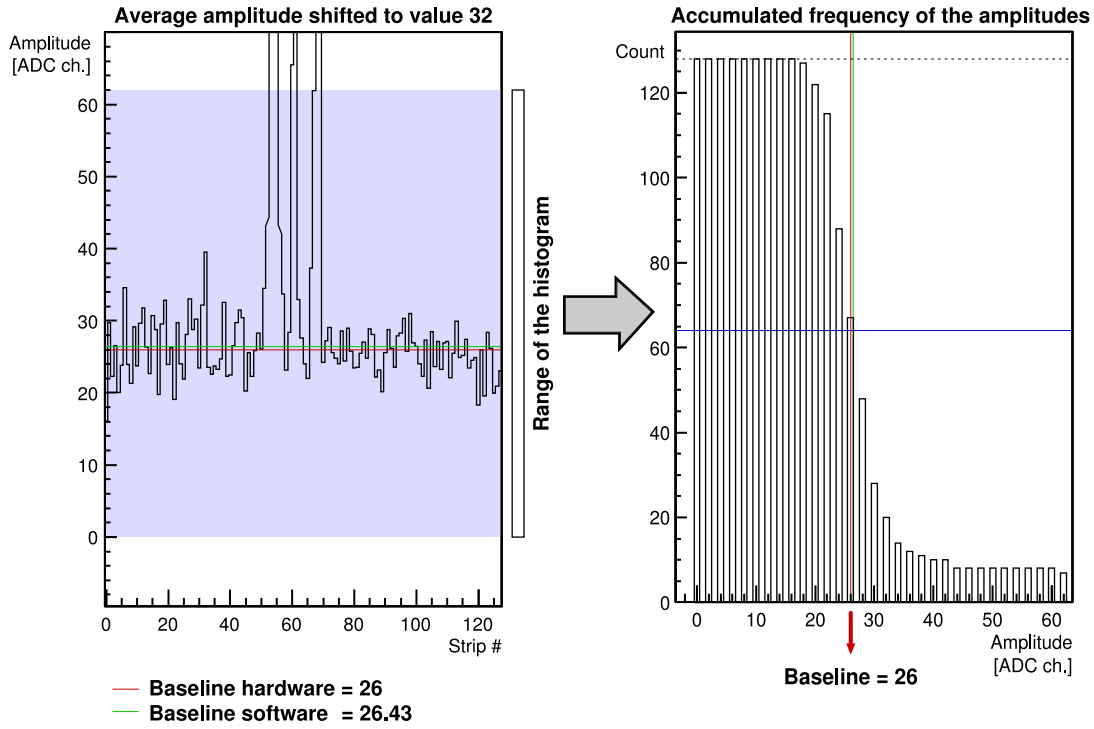


Figure 4.6: Principle of the median algorithm. After baseline subtraction the cumulative histogram is generated. In this example the input vector containes 128 members. The first element of the CH with value higher than 64 is taken for the new baseline to be subtracted [50].

This algorithm was implemented for test purposes as an FPGA module. The method has the following advantages/disadvantages:

+ No bias due to the signal presence. In this method only the most frequent values are counted, outliers are ignored.
+ The histogram is a highly parallelizable process with a short processing time[1].
− Limited dynamic range. The option with 16 bins seemed reasonable but too restrictive for the CM assumptions. The 24 bins version went against a high power consumption and large area (~30 mW and 6 mm^2). The 32 bins version is too slow and expensive in terms of size and power.
− Only integer result precision.

Two Parse Average

The median search algorithm being too resource prone, an alternative solution was found, combining advantages of implementation simplicity and result precision. We called it “Two parse average algorithm“ (TPA), as reported in [12].

[1] relative to the element sorting option

The TPA can be resumed as: to get the CM value, the averaging procedure is executed twice. First a rough estimation of the CM is taken as a simple average ($\widetilde{CM}$). Then the first signal detection step takes place: if a signal is detected, it is replaced by the $\widetilde{CM}$ estimation. In this way the sample is cleaned from the signal presence. Then the average can be taken again and, within the digitization limit, this gives the unbiased CM.

$$\widetilde{I}_j = \begin{cases} I_j & \text{if } [I_i < \widetilde{CM} + \Delta_{Thres}] \quad \Leftrightarrow \text{if no signal is detected} \\ \widetilde{CM} & \text{if } [I_i \geqslant \widetilde{CM} + \Delta_{Thres}] \quad \Leftrightarrow \text{if a signal is detected} \end{cases} \tag{4.9}$$

$$CM = \frac{\sum_j \widetilde{I}_j}{n} \tag{4.10}$$

Δ_{Thres} denotes a certain threshold value of the minimum signal to be detected. Finally, after the CM subtraction, the threshold is applied again[1] to detect eventual signals and proceed to the Zero Suppression step.

Algorithm Comparison

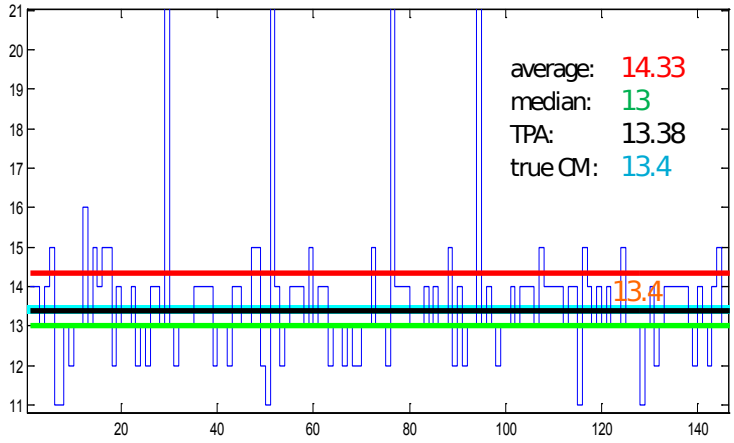


Figure 4.7: Comparison of Common Mode search algorithms. A test vector, containing some outliers (signals) is taken. Comparison of the discussed three algorithms is presented. The TPA achieves the best estimate of the true CM. The median gives only the integer precision, the simple average result is biased.

Figure 4.7 presents a comparison of the discussed three algorithms. Both TPA and median algorithm are equally good within integer precision. The simple average estimator is biased. However, TPA has a potential to give an even better precision. Additionally, being simpler to implement and less resource prone, the TPA is chosen to be used in the DHP design. A comparative summary of discussed algorithms is presented in Table 4.1.

Algorithm	Simple Average	Median	TPA
Bias	yes	no	no
Effort to implement	Low	High	Medium
Resources	Few	Many	Relatively few
Possible precision	$1/\sqrt{N}$	integer	$1/\sqrt{N}$

Table 4.1: Common Mode estimation algorithms comparison

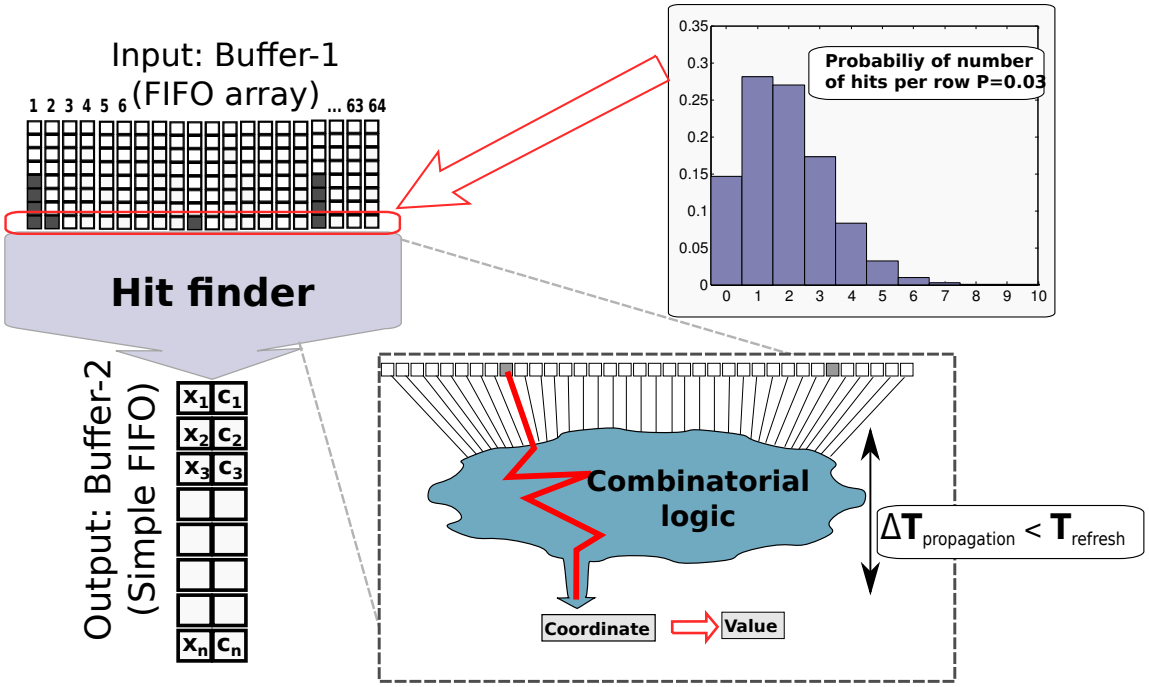


Figure 4.8: Hit finder structure. (Right-top) Number of hits per input vector probabilities, assuming 3% Poissonian input. (Left) To take care of input and output data-rate fluctuations, the input and the output buffers are necessary. (Right-bottom) Hit-finder representation: logic propagation time should be smaller than the interval between incoming data refresh.

4.2.6 Hit-finder Module

After the CM correction stage the data vector to be processed contains only possible hits we want to process S_{cr}:

$$S_{cr} = \begin{cases} \text{Hit amplitude} & \text{if detected} \\ 0 & \text{if not detected} \end{cases} \tag{4.11}$$

As described in Section 4.2.1, a 256 byte-long vector arrives every eight GCK clock cycles. For design reasons this data is split in four chunks of 64 bytes. That is, one 64 byte-long chunk arrives with every second clock cycle.

At this stage one needs to find a pair $[x_i, c(x_i)]$ (hit position and its corresponding amplitude) and write it to the output buffer before sending it out, as schematically drawn in Figure 4.8. This task is executed by the Hit Finder block.

[1] the second time one applies the threshold comparison to the unbiased vector, so possible small signals, which could be lost after first comparison, can be detected here

Finding hits in such a vector is not a trivial task. For example, using a sequential linear search, the complete 64-elements long input vector should be scanned, i.e. 64 comparisons in total. If this is done clock-synchronously, 64 clocks would be needed and this would result in extremely poor hit-finding efficiency (long dead times). It is clear that an asynchronous solution is needed. In order to function correctly, as sketched in Figure 4.8, the total propagation time of such logic should be less than the period between two consecutive data arrivals. An elegant solution is a binary tree logic structure [30, 51]. This search algorithm was used in an FPGA DHP emulation (see Section 6.1). The solution implemented in the DHP 0.2 chip follows a similar logic of combinatorial search. The current performance of the implemented Hit Finder allows us to detect one hit per clock cycle. Since the data arrives once per two clock cycles, that means that two hits per row is its maximum sustainable processing capacity.

However, the hit arrival rate is not constant[1]. Figure 4.8 (right-top) shows the probability distribution for a number of hits per input vector to arrive, if one assumes 3% of data occupancy[2]. This natural dispersion of numbers of hits per row can lead to eventual data losses: if, for example, three hits are present in one row, the Hit Finder will not have enough time to detect all of them, as in two clock cycles new data arrive. To take care of those variations additional buffers in front and behind the Hit Finder were introduced, see Figure 4.8 (left).

These buffers are done in the following way:

- The front Buffer-1 is implemented as an array of 64 individual FIFOs[3], one for each input vector position. Upon the CM suppression, all non-zero values are pushed to the corresponding positions of this FIFO array (FIFO 1).
- The Hit Finder, if running continuously, produces the data faster than the Serializer can send them. That is why upon data processing, the Hit Finder puts the results into the intermediate Buffer-2, which takes care of statistical variations (see Section 4.2.7).

Both these buffers, the Buffer-1 and Buffer-2 should ideally be very deep to take care of any possible data fluctuation. Due to limited resources, sizes of these buffers and the data processing architecture should be optimized (see Section 5).

4.2.7 Serializer

After the Hit Finder stage, the zero suppressed data is put into the output Buffer-2 as presented in Figure 4.9.

The output data is split into blocks of 16-bit words and organized in frames. Each frame starts with a header containing the information about the type of data in the frame. To parse the data, each word has a flag signalling what kind of information is packaged. The data format is described in detail in the DHP 0.2 manual [52].

Each generated zero-suppressed data it encapsulated into Aurora frames using a special protocol developed by Xilinx [53], where the data is additionally 8b/10b encoded [54]. This ensures that on average the number of transmitted 'ones' is equal to the number of transmitted 'zeros' regardless of the data pattern. This is necessary to keep the transmission link DC-balanced[4].

[1] For first estimations we use assumption of Poissonian hit distribution.
[2] we target 3% as the maximum sustained occupancy, see Section 2.3
[3] FIFO stands for First In First Out queue
[4] This provides many advantages for a wire link. Transmitter, receiver and equalizer design can be simplified [54]. Additionally it provides means to have a reliable clock recovery, which is especially important, when transmitter and receiver clock are not exactly equal.

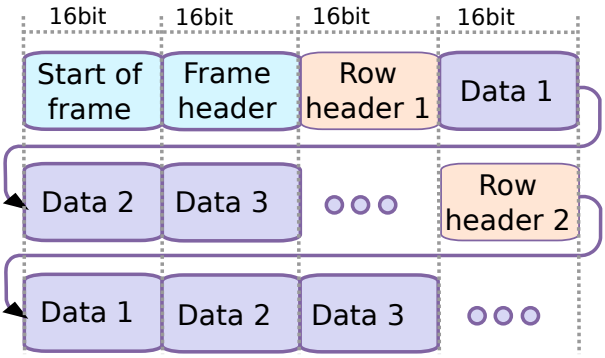


Figure 4.9: DHP Frame structure

Finally, the data is sent frame-wise with a rate of 1.53 Gbps using the CML transmitter, described in Section 4.3.2.

4.2.8 Switcher Sequencer

One of four DHP chips is connected to the Switcher-B chips chain and generates signals necessary to steer the DEPFET matrix. These are three cyclic signals: *SW_CLK, SW_STR_G and SW_STR_C*. By adjusting relative phases of these signals one can achieve any necessary periodic steering pattern, as described in the Switcher-B manual [33]. These signals run with a F_{ser}=305 MHz clock. In the DHP 0.2 version it is periodic every 32 clock cycles. More complex patterns will be programmable in the production DHPT 1.0 chip.

A frame strobe signal *SERIN* is sent once per frame in order to propagate the control sequence through all matrix rows. Test results of this module are presented in Section 6.3.

4.3 Custom Blocks

4.3.1 Clock Generation with PLL

The readout of the PXD is synchronized with the operation of the SuperKEKB. To do so, all operation clocks are derived from the RF frequency of 508.79 MHz and the beam revolution cycle of 10 µs.

To each PXD half module the base clock of 76.32 MHz is provided by the corresponding DHH[1]. Two other clocks that are necessary to run the system (serializer F_{ser} and deserializer F_{des} clocks) are generated on-chip using the Phase Lock Loop (PLL) block, whose circuit is presented in Figure 4.10

The PLL topologies for both CMOS 90 nm[2] and 65 nm[3] technologies are similar and originally inherited from the pixel front-end chip (FE-I4) for the upgraded Atlas pixel detector [56].

The main element of the PLL is a Voltage Controlled Oscillator (VCO), consisting of three inverters connected in a loop. The resulting oscillating frequency of the VCO can be adjusted by the control voltage V_{ctrl} around the target clock frequency of F_{ser} =1.53 GHz. The deserializer clock F_{des} is obtained from the F_{ser} by its division by five.

[1] this clock is derived from the RF frequency by multiplying it by $\frac{3}{20}$
[2] for DHP 0.1 and DHP 0.2
[3] for the DHPT 0.1 and DHPT 1.0

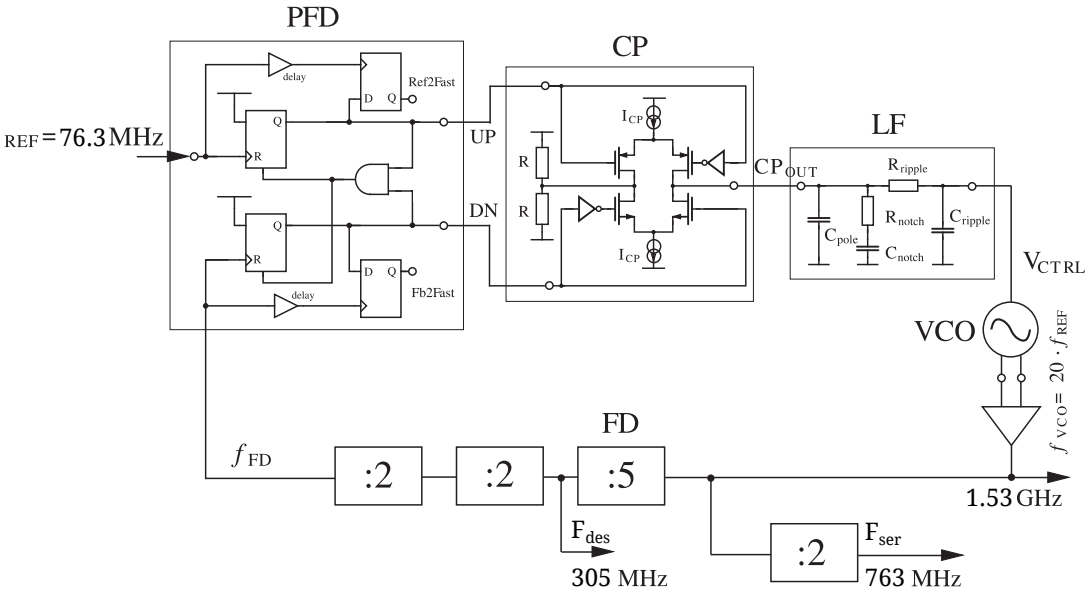


Figure 4.10: PLL block diagram. Clocks F_{des}=305 MHz, and F_{ser}=1.53 GHz are generated from the input reference clock of 76.3 MHz. Source: [55]

To lock the output on the desired frequency, the VCO is stabilized via the negative feedback (FB) controlled by the base clock F_{ref}. The FB consists of three elements: the Phase Frequency Detector (PFD), the Charge Pump and the Loop Filter (LF), as sketched in Figure 4.10.

The PFD has two outputs, UP and DOWN, that can be activated depending on the relative phase between F_{ref} and F_{FD}. These two outputs control the Charge Pump, which in turn can charge or discharge the capacitor C_{pole} (the first in the LF circuit). The charge on the C_{pole} defines V_{ctrl}, which steers VCO. More information about this implementation can be found in paper written by Kishishita et al. [55].

All three clocks are used in the PXD:

- F_{ref}=76.3 MHz, also noted as GCK, is used in DHP as the clock for the digital processing blocks.
- F_{des}=305.3 MHz or the deserializer clock is primarily used as the clock to run DCDB chip. Consecutively, the DHP deserializer, which reshuffles the input data in the necessary order, uses the same clock.
- F_{ser}=1.53 GHz is the serializer clock that is used by the output of the DHP chip to send the zero-suppressed data out via a serial high-speed link (see Section 4.3.2)

4.3.2 Output Link: Data Transmission with Current Mode Logic Link

In the Serializer (Section 4.2.7) the incoming 20 bit wide data becomes 1 bit wide. Therefore, the Serializer clock F_{ser} is 20 times faster than the reference clock F_{ref} and its frequency is equal to F_{ser}=F_{ref}×20=1.53 GHz.

For such a high data rate a so-called Current Mode Logic (CML) output driver is used in the chip, whose schematic is sketched in Figure 4.11. The resulting differential pair of signals TX_P and TX_N

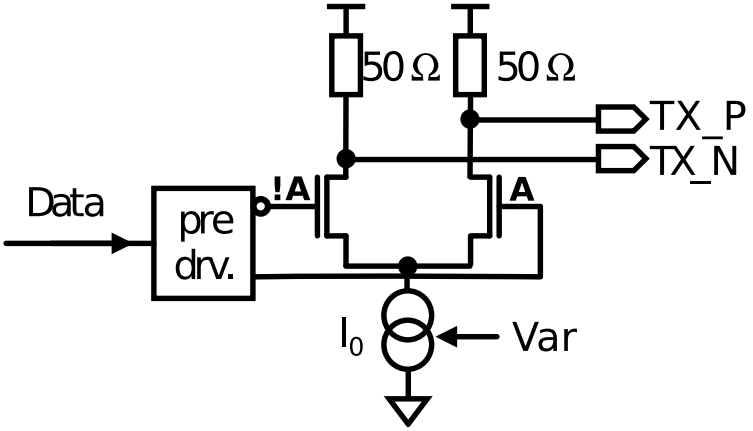


Figure 4.11: Simple Current Mode Logic driver design used in DHP 0.1. The nominal current I_0=20 mA can be varied according to the configurable register value [57].

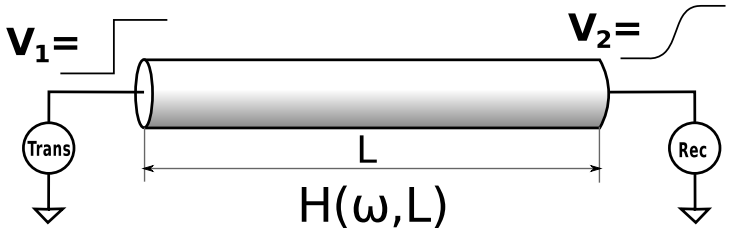


Figure 4.12: An example of a step pulse transmission. The cable acts as a low-pass filter with a corresponding transfer characteristic H(ω,L). This can result in long rise time of the output signal limiting the maximum bandwidth.

is driven by the serialized output **A** and its inverted version **!A**, which are applied to the gates of two transistors working as switches.

Pre-emphasis

As discussed in Section 3.6 the DHP output signal is expected to cross about 15 m of cable. As will be shown in Section 6.3.2, the signal is strongly attenuated on such high frequencies, as schematically presented in Figure 4.12.

This is explained by the fact that the cable acts as a low-pass filter with a certain characteristic H(ω,L). For the example presented in Figure 4.12, a step signal after crossing the cable gets a finite raise time, limiting the maximum cable bandwidth.

To increase the cable performance one uses the so-called pre-emphasis technique. It adds a distortion to the initial signal, steepening the rising (or the falling) edge of the output signal. To better understand what it is, the following question can be asked: what kind of signal should one send so that the output signal would have the steepest slope possible (Figure 4.13)?

This can be written as follows: the output voltage in the frequency domain is the input voltage multiplied by the cable's transfer characteristic:

$$V_2(\omega) = V_1(\omega) \cdot H(\omega, L) \tag{4.12}$$

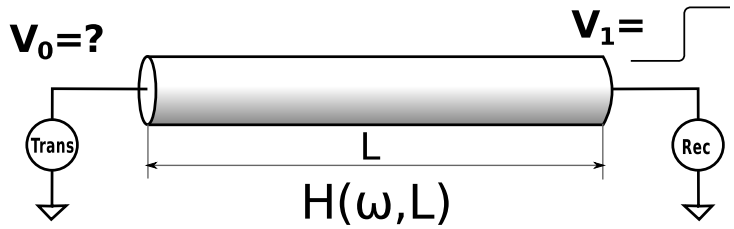


Figure 4.13: What kind of pulse should one send to get the steepest raise on the output?

The same relation is true for V_0 and V_1, presented in Figure 4.13. One can write it in the following way, multiplying both sides by $H^{-1}(\omega, L)$:

$$V_0(\omega) = V_1(\omega) \cdot H^{-1}(\omega, L) \tag{4.13}$$

After this transformation, high harmonics of V_1 are boosted in V_0 to compensate for the low-pass characteristic of the transmission line transfer function.

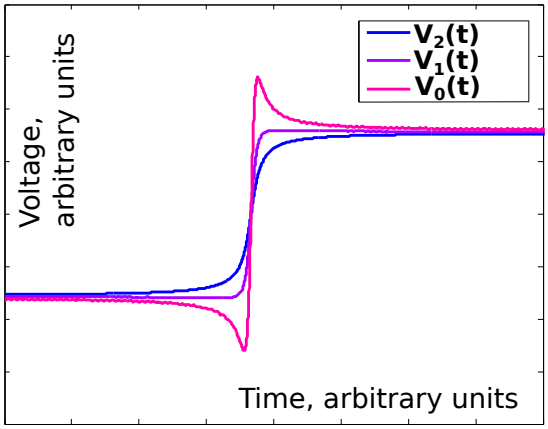


Figure 4.14: Transmission line as a filter: V_2 is a filter response to the initial step function V_1. V_0 is obtained from V_1 by applying the inverse transfer function, in other words the filter response of V_0 will be V_1.

For visualization, the filter behavior has been tested using Matlab simulation (Figure 4.14): a low pass filter[1]has been applied to a step-like function $V_1(t)$ to simulate the cable response $V_2(t)$, then an inverse filtering was applied to get the prototype $V_0(t)$ of the initial function.

From the result one can see that the inverse filtering creates overshoot effects during transition periods. This is used in the pre-emphasis technique: one artificially creates overshoot effects during 1→0 or 0→1 transitions to compensate for the cable high frequency attenuation.

[1] The transfer characteristic H of the filter was chosen so that the resulting shape V_2 would look similar to what was observed by our system. It was done to illustrate the phenomenon and not to quantify it.

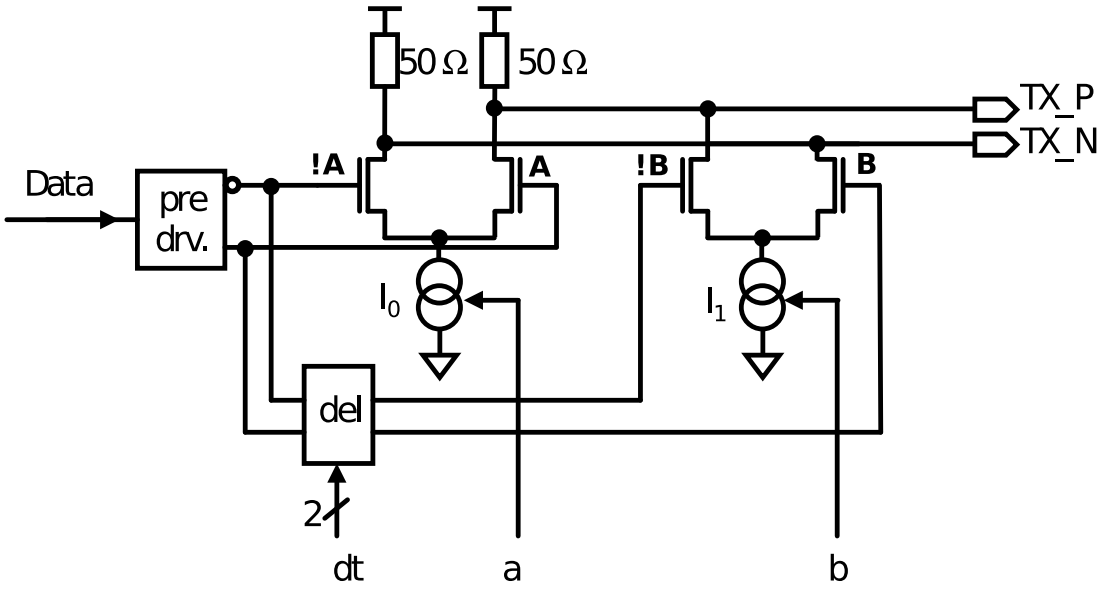


Figure 4.15: CML driver with pre-emphasis design. An additional smaller current source steered by the inverted and delayed version of the same output signal creates the preemphasis effect. **a**, **b** and **dt** steer the shape of the output signal [57].

Digital Implementation of Pre-emphasis

To implement the pre-emphasis technique, the CML driver, presented in Figure 4.11, is improved as shown in Figure 4.15. An additional current source I_1, smaller than I_0, is added. It is switched in opposite phase relative to the current source I_0 but with a certain delay Δt:

$$B(t) = -A(t - \Delta t)$$

In such a way, the output voltage is proportional to:

$$V_{out}(t) \propto I_0 A(t) - I_1 A(t - \Delta t)$$

Currents I_0, I_1 and the time delay Δt can be varied within the DHP chip by applying corresponding settings. An example of their tuning is presented in Figure 4.16. This pre-emphasis implementation is not perfect, since it does only one of two necessary overshoots during the state transition but this considerably increases the data line transmission characteristics, as it will be presented in Chapter 6.3.2.

4.3.3 Other Custom Modules: LVDS, DACs and ADC

With the exception of the serial link that uses CML transmitter, all other signals between the DHP and DHH module are driven using LVDS [58] transmitters and receivers.

Programmable current sources exist on the chip exist on the chip to properly bias other custom blocks. To measure the external temperature on-board using the intrinsic diode temperature dependence, the DHP 0.2 has a built-in ADC. All these blocks were custom designed by our group in collaboration with the University of Barcelona.

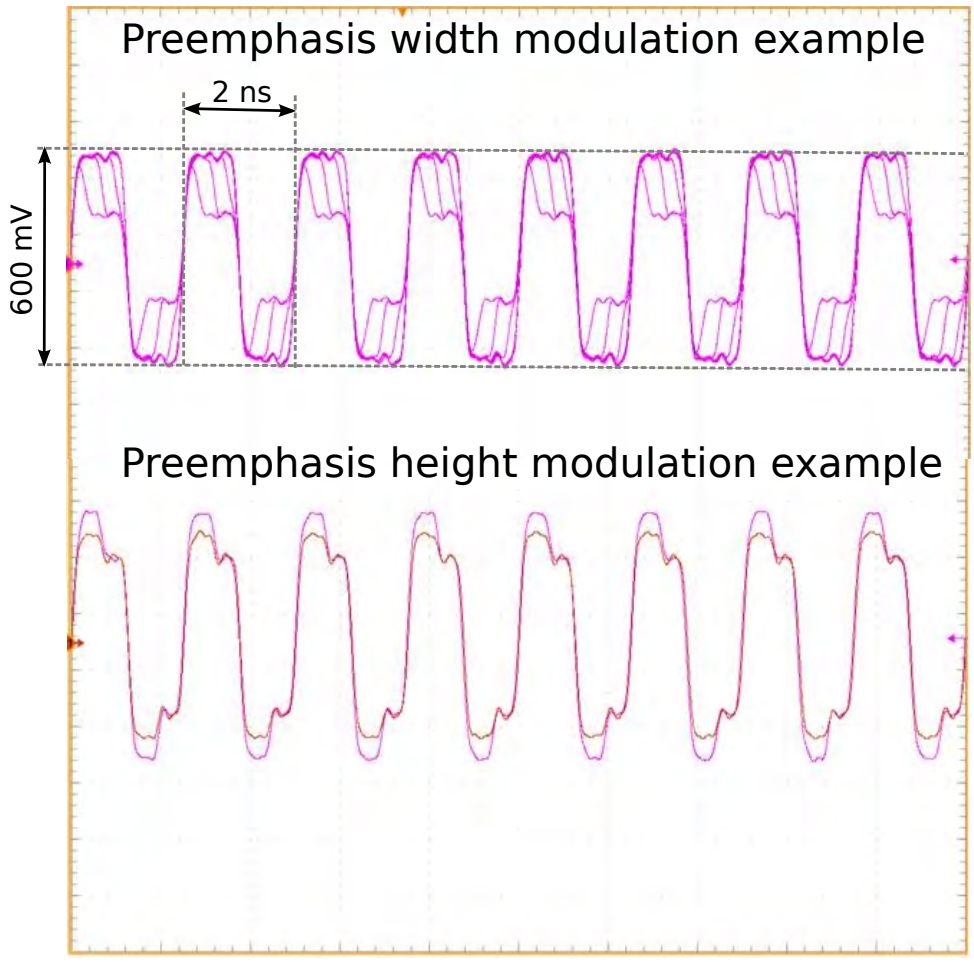


Figure 4.16: Example of preemphasis settings. Top: the width of the overshoot is tuned. Bottom: the height of the overshoot is tuned [57].

Chapter 5

DHP Architecture Optimization

5.1 Design Considerations and Constraints

In this section the data processing efficiency and its optimization of the DHP chip will be discussed. Higher efficiency means higher sustainable data occupancy that DHP chip can process with no (or very few) losses.

There are several processing bottlenecks in the DHP design that limit its performance, namely:

1. **Hit Finder**. As mentioned above, the current Hit Finder implementation can process up to 2 hits per row. This corresponds to a maximum sustained occupancy of $P_{HF} = 2/64 \cdot 100\% = 3.125\%$.

2. **Output link**. After the Hit Finder the data is stored in the output buffer before being sent through the Gigabit link with a bandwidth of 1.53 Gbps. This further reduces the maximum supported occupancy, down to $P_{out} = 2.5\%$, as shown in Figure 5.1.

3. **Buffers**. To derandomize the data, two additional buffers are needed to take care of occupancy fluctuations during the Zero Suppression procedure. The first buffer (Buffer-1) is put in front of the Hit Finder and consists of a FIFO array with one FIFO per column. The second buffer (Buffer-2) is put behind the Hit Finder, it stores the zero suppressed output data. The total memory we can use for these buffers is constrained by the chip area.

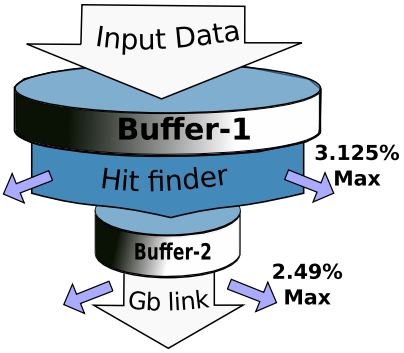


Figure 5.1: The DHP occupancy tolerance diagram. The maximum occupancy supported by the Hit Finder is 3.125%, the one supported by the output transmitter is 2.5%

5.1.1 Triggering

To increase the effective bandwidth, the DHP is operated using a so-called triggered mode rather than continuously, only processing the data from the time intervals of interest upon the trigger arrival. Typically a trigger lasts one frame but if the next trigger appears before the end on the previous one, the trigger execution is prolonged. For these conditions the proportion of the processing time with regard to the total time (the Busy Factor, BF) can be calculated as[1]:

$$BF = 1 - e^{-F_{tr}/F_{fr}} \tag{5.1}$$

[1] For the demonstration of Equation 5.1, the assumption of the Poissonian trigger rate is taken, see Appendix C.

Where F_{tr} is the trigger rate and F_{fr} is the frame rate. For the Belle II scenario the frame rate is equal to F_{fr}=50 kHz. The highest estimated trigger rate is F_{tr}=30 kHz [14]. This results in $BF \approx 0.45$. Later in this chapter two different data processing modes will be discussed: the continuous mode (the trigger is always on) and the triggered mode if one triggers only occasionally.

Knowing how the data is packed, one can estimate the data rate, which follows Equation 5.3, and determine the maximum data rates:

Continuous acquisition mode is limited by the output link capacity of 1.53 Gbps, which is achieved at the occupancy of 2.5%[1].

Triggered acquisition mode at 30 kHz: The triggering effectively decreases the data rate by the facor BF. Using the current data format (see Section 5.3) this yields 6.3% of the maximum supported occupancy.

5.2 Ideal DHP Model

The DHP chip can be seen to function "ideally" if the sizes of its buffers are big enough to take care of any data occupancy fluctuations. The maximum supported occupancy L_0 is then defined only by its output link bandwidth (Figure 5.2). An ideal chip has zero losses if the averaged occupancy is less than the maximum value L_0. The losses in absolute values and in percentage of the incoming rate are presented in Figure 5.2.

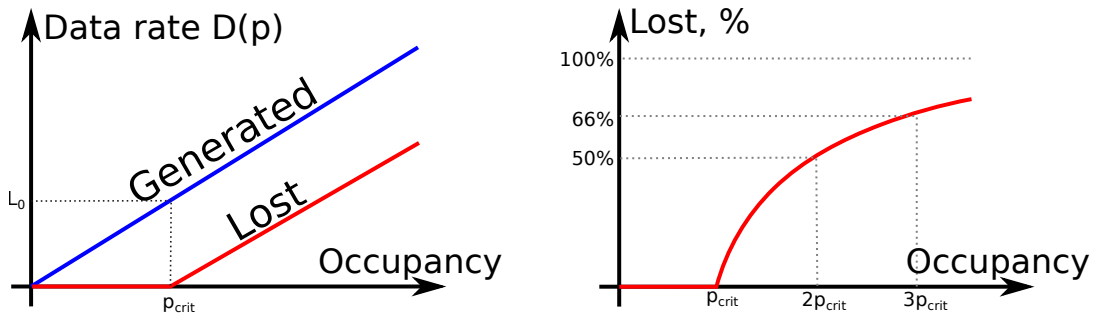


Figure 5.2: (Left) An ideal DHP does not lose any data unless the average data rate exceeds the output link capacity L_0. The data losses are equal to the difference between the generated data $D(p)$ and L_0. (Right) Relative data losses have a hyperbolic shape (see Equation 5.2).

The generated data is linearly proportional[2] to the occupancy (p):

$$D(p) \approx \alpha p$$

We have zero losses for this ideal model if the generated data rate is less than $L_0 = D(p_{crit})$ and $D(p) - L_0$ otherwise. Hence the loss curve of the ideal DHP processor can be represented by a simple hyperbolic function, as sketched in Figure 5.2.

$$L(p) = \begin{cases} 0 & \text{if } D(p) < L_0 \\ 1 - \frac{L_0}{\alpha p} & \text{if } D(p) \geq L_0 \end{cases} \tag{5.2}$$

[1] this is variable depending on the output data format. The number is given for the data format that will be used in DHPT 1.0

[2] with a good approximation especially at high occupancies. At low occupancies (where DHP data loss is not an issue anyway) it slightly differs from linear behavior due to data format reasons.

5.3 Output Formatting

The final amount of data produced by the zero suppressed DHP output depends on output format definition. In particular, to store one zero suppressed data point one should save the information, which is summarized in Table 5.1:

	Hit value	Row	Column	Common Mode
range	8-bit ADC	192 rows	256 columns	10-15 ADU
# bits needed	8	8	8	6-5

Table 5.1: Zero suppressed pixel information to be transmitted by the output link. Option: instead of using geometry row-column notation, one can also use electrical Switcher–column and drain–row indexing, which makes the effective size of the matrix equal to 768×64.

Figure 5.3 shows the data rate as a function of occupancy for the initial and the optimized data formats. As one can see, a well structured zero suppressed frame format can save up to 40% of data traffic. The details of the two data formats are given below:

A **Simple 24-bit per pixel format**: <10-bit row, 6-bit column, 8-bit ADC>. One Common Mode is additionally sent per each switcher row. Per frame one 32-bit header is sent.

B **16-bit row header each 256 pixels. + 16-bit per hit**: <1-bit row-flag+8-bit row+6-bit CM> + n*<1-bit hit-flag+7-bit column+8-bit ADC>. Per frame one 32-bit header is sent. Row header is not sent if no hits are detected.

For the data format currently in use (B), the generated data amount can be estimated according to Equation 5.3:

$$D(p) \approx 10/8 \cdot F_{row}(256 \cdot 16 \cdot p + 2 \cdot 16) \cdot BF = 40F_{row}(128 \cdot p + 1) \cdot BF \tag{5.3}$$

8b/10b encoding
switcher row frequency
pixels per row
bits per pixel
occupancy
Busy Factor
two info headers per row

Here, $F_{row} = GCK/8 = 9.5$ MHz is the frequency at which the matrix rows are read. The factor 10/8 is due to the 8b/10b encoding. $256 \cdot p$ is the average amount of pixels processed per row[1]; it is then multiplied by 16, since 16 bits are needed to encode one pixel in a row. $2 \cdot 16$ are two row headers per one row. The details of the format can be found in the DHP 0.2 Reference Manual [52].

5.4 Chip Optimization Goals, Buffer Sizes

By design the first buffer (Buffer-1) consists of 64 independent FIFOs, each processing 20-bit wide words (Section 4.2.6). All FIFOs have the same depth D_1. Buffer-2 is a single FIFO with depth D_2 accepting 32-bit wide words. Both depths D_1 and D_2 are constrained by the area S_{max} that can be used for their implementation. The goal of our design is to minimize the DHP losses L for the maximum given occupancy $x = x_{max}$.

[1] per row we have 256 elements. However, this number was recently reduced to 250 due to geometry constraints.

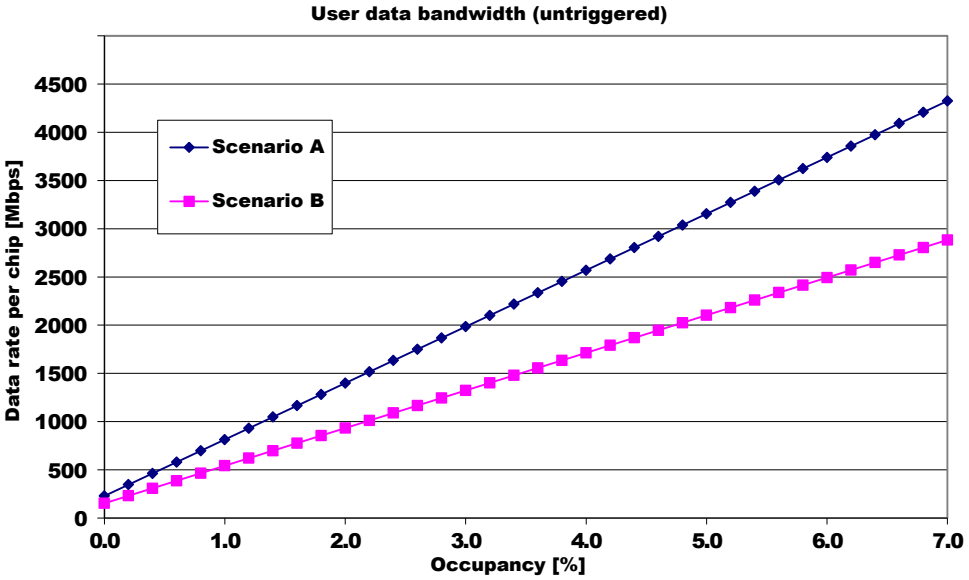


Figure 5.3: DHPT 1.0 data rates for data format scenarions A and B, as described in Section 5.3, for the untriggered readout case (equivalent to BF=1 in Equation 5.3). Source: [59]. Triggering effectively results in data rate decrease by a BF factor.

This problem can be seen as a three dimensional constrained optimization problem (occupancy and buffer sizes). This problem can be defined in two alternative ways:

1. **Constrained Loss Minimization.** Assuming that one unit of depth of Buffer-1 would cost S_1 mm^2 of the chip area and S_2 is the cost of Buffer-2 per unit of depth, then these constraints can be writen as:

$$\begin{cases} S_1 D_1 + S_2 D_2 & \leq S_{max} \\ L(p, D_1, D_2)_{p=p_{max}} & = min \end{cases} \tag{5.4}$$

   S_1 and S_2 are proportional to their respective sizes in bits[1]: If S_0 is the average surface necessary to implement one-bit register, then:

$$\begin{aligned} S_1 = 64 \cdot 20 \cdot S_0 &= 1280\, S_0 \\ S_2 &= 32\, S_0 \end{aligned}$$

   Hence, Equation 5.4 can be rewritten as:

$$\begin{cases} 1280 D_1 + 32 D_2 & \leq S_{max}/S_0 \\ L(p, D_1, D_2)\Big|_{p=p_{max}} & = min \end{cases} \tag{5.5}$$

2. **Constrained Best Supported Occupancy Optimization.** An alternative way to express the optimization problem is: the DHP chip has a limited area S_{max} and the maximum tolerated losses

[1] As the first approximation. It also differs depending on the implementation: how it is routed, if the SRAM of register array was used and so on

are L_{max} (typically smaller than one percent). What is the highest allowed hit occupancy? This can be written as:

$$\begin{cases} S_1 D_1 + S_2 D_2 \leq S_{max} & \leftarrow \text{limited space condition} \\ L(p, D_1, D_2) \leq L_{max} & \leftarrow \text{limited losses condition} \\ p(S \leq S_{max}, L \leq L_{max}) = max & \leftarrow \text{the best occupancy condition} \end{cases} \tag{5.6}$$

In Section 5.9 it will be shown how it is possible to optimize the chip according to these two definitions.

A full chip simulation is necessary for that purpose. A simpler approach is to analyze several curves $L(x)$ for different $\{D_1, D_2\}$ scenarios in use. Then one choses an acceptable scenario and checks whether its parameters would fit into the chip implementation.

5.5 C++ Chip Model

The model based on Hardware Description Language (HDL), which is used to create the chip, is tested using a specially designed verification environment (Section 5.6.1). However, it is rather difficult to use the same environment for the algorithm and the parameter optimization. Many parameters are hard written in the code and cannot be easily varied, which is especially important for parameter scan tests. To test a module, time consuming test-bench writing is necessary. Therefore a C++ model has been developed in parallel to the HDL code, being simpler but still sufficiently precise. Owing to the Object-Oriented nature of the language, C++ offers nice modular and parameter flexibilities, which are so important in the chip characterization.

For the chip development the following strategy was chosen: the parameter optimization was done using the C++ chip model. After having selected the desired parameters, they were used for the HDL chip code. Finally, to double check that the developed C++ model is correct, it is then compared against the HDL code with the same parameters. The structure of the C++ model is presented in Figure 5.4.

5.6 HDL Chip Verification

The DHP chip[1] is written using SystemVerilog HDL. It is extremely important to be sure that there are no design mistakes before the chip is produced: any major bug found during chip testing means a chip resubmission with corresponding costs [2] and at least 2-5 months of delay. In order to verify if the design is correct it is checked using the verification environment specially designed for that purpose. This is described in the following section.

5.6.1 UVM Methodology and the Test Environment

The Universal Verification Methodology (UVM)[3] is a special verification library based on the System Verilog language supported by industry standard EDA[4] tool vendors: Aldec, Cadence Design Systems, Mentor Graphics and Synopsis.

[1] for all of its versions: DHP 0.1, DHP 0.2 and DHPT 1.0
[2] 60-80k€ for DHP 0.2 or DHPT 1.0 submissions
[3] originally known as Open Verification Methodology
[4] Electronic design automation

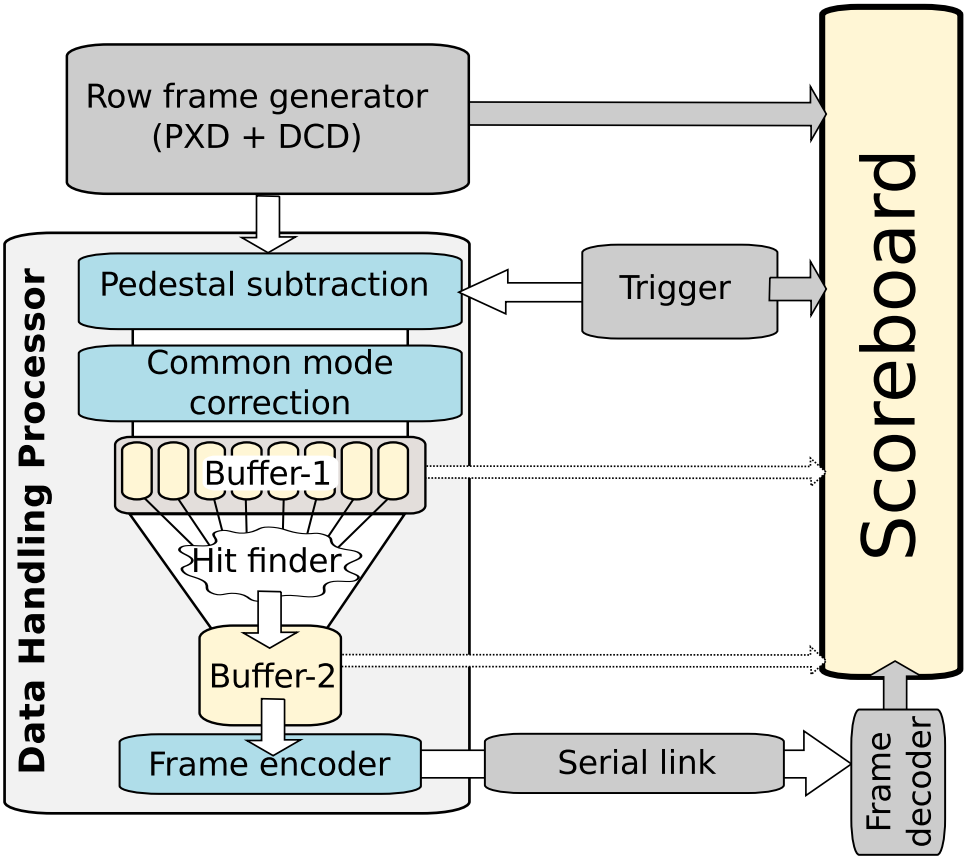


Figure 5.4: C++ DHP testing environment structure. The constructor of the environment allows to vary any possible parameter and to flexibly replace modules. The Frame generator imitates the behavior of the PXD+DCDB, generating the raw data with variable occupancies and distributions. Upon trigger arrival the DHP starts to process the data. The input frames, triggers, output frames and other information is monitored and collected by the Score Board, where the final analysis is done.

The purpose of the UVM is to provide standardized methodology to create a verification environments of complex digital systems, in particular for ASICs.

The development of a full covering verification environment is very time consuming and takes a major part of the ASIC design. This is, however, the only way to correctly test the circuitry before it actually exists. In our particular case the DHPT 1.0 chip has to work in a complex digital environment:

- It has 80 high speed digital data lines connected to DCDB.
- It has a serial JTAG configuration interface
- It has a high speed serial link with 8b/10b encoding.
- It controls Switcher-B
- It receives and interprets synchronization signals coming from the DHH.

To test the whole system, the verification environment containing all interfaces has been created, as presented in Figure 5.5.

The UVM allows to define a set of tests using the created environment to make a complete verification. It is important to underline that the UVM test environment is very different from the C++ model. The former is needed for detailed verification, the latter is for parameter optimization.

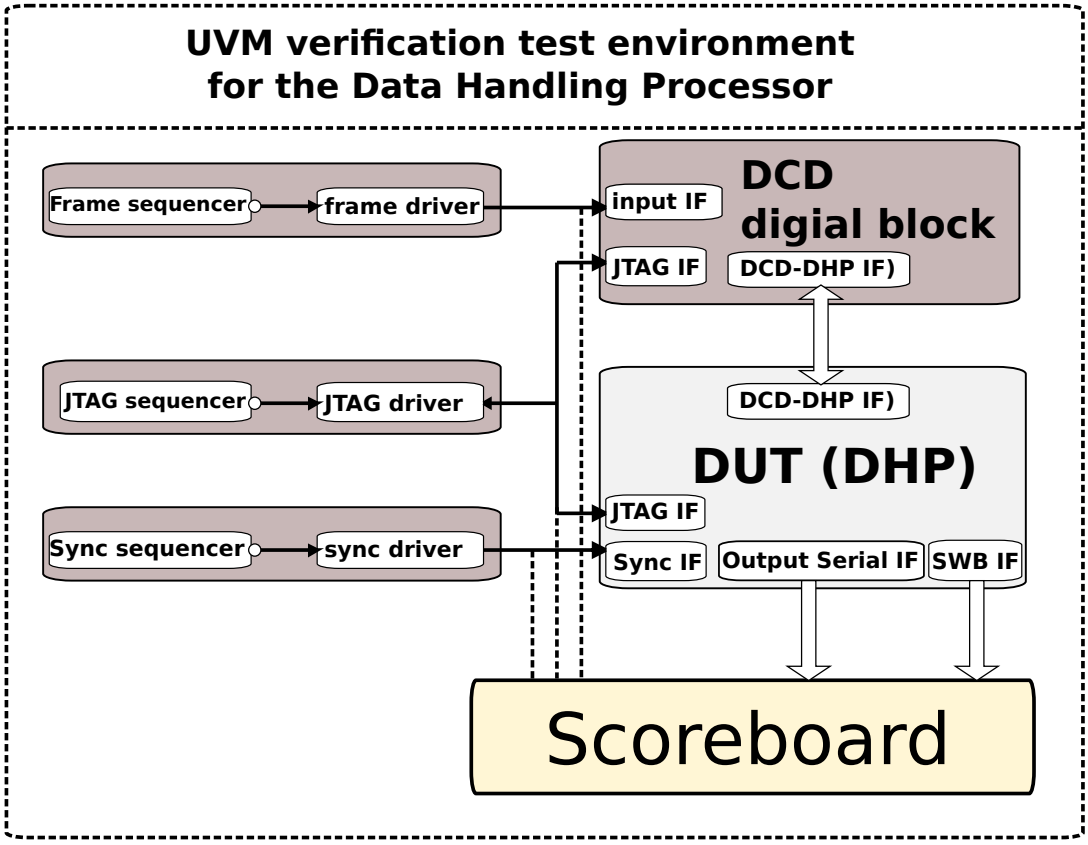


Figure 5.5: Simplified representation of the DHP verification environment based on UVM verification methodology. All three chip interfaces are present: the slow control (JTAG), the synchronization interface and the PXD frame interface. The input data is driven by the digital block used in DCDB. The Scoreboard analyzes and evaluates the chip behavior and compares it to expectations. Upon that a decision is taken if the test succeeds. For completeness a set of tests is defined in this environment.

5.7 DHP 0.2 Optimization

For the DHP 0.2 optimization several samples of the simulated background were used. Additionally, a simple Poisson distributed data was used to compare if and how much the processing efficiency depends on the nature of the background. An example of samples, which were used, is presented in Figure 5.6. The background contributions are listed in Table 2.1 on page 12.

For the same occupancy but different hit distribution the data processing efficiency results are slightly different. Distributions with long low-p_t tracks for Touschek background, as presented in Figure 5.6, result in a lower data processing efficiency, especially if Buffer-1 is shallow (8-16 words deep). There is a simple explanation for that: these horizontal tracks can easier jam one of the FIFOs of Buffer-1. Although the Touschek data bring a relatively small contribution in the total background, it was the main test pattern we used for the DHP 0.2 chip model optimization to be sure that even worst case conditions are met.

Figure 5.7 presents the data processing efficiency results for different DHP 0.2 design options and using the Touschek background as input data. In this simulation the trigger rate of 30 kHz was used, which is the expected rate for the worst case scenario [14]. In this case the effective best-supported occupancy considerably increases (presented in Figure 5.7 by the right-most curve) but it is resource

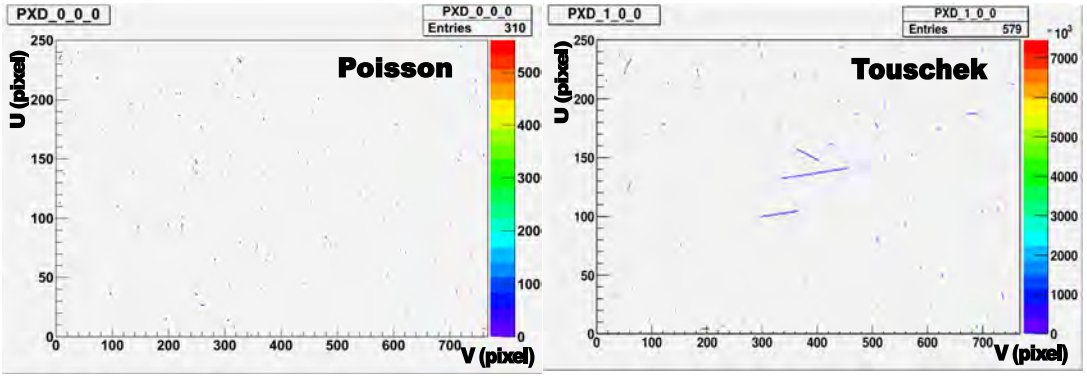


Figure 5.6: DHP 0.2 events examples. Poisson-like background (left). The worst case: Touschek background (right), its long clusters are likely to jam FIFOs of the Buffer-1. The U and V correspond to the column and the row coordinates of the sensitive area.

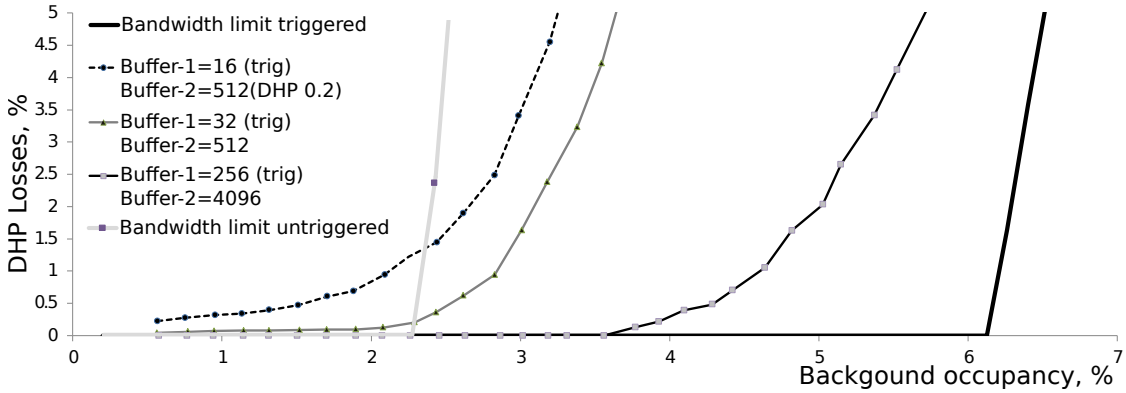


Figure 5.7: Simulation results for the Touschek background. For continuous (un-triggered) readout all three options are very close to the bandwidth limit represented by the gray curve. While randomly triggering at 30 kHz, the bandwidth increases but deep buffers are necessary to approach the best possible performance.

prone to approach this best possible performance. The DHP 0.2 chip was designed with Buffer-1=16 and Buffer-2=512. As it follows from Table 2.1 on page 12, the background occupancy is not expected to be higher than 2.5%. At this rate the estimated DHP 0.2 losses are about 1%.

5.8 Chip Tests and Comparison to its Model

The obtained optimized buffer parameters were used for the HDL model of the chip. To double check that the chip model correctly represents real implementation, the HDL code was tested on the Poisson distributed random generated data and compared against the C++ model. The results are presented in Figure 5.8. As the HDL chip model is very slow, it was not possible to process enough data to eliminate the statistical noise, as it was done in the C++ model. Nevertheless, general trends are the same.

The DHP 0.2 chip was submitted for fabrication in July, 2011 and received at the end of the same year. As will be shown in Chapter 6, it successfully passed verification tests, confirming declared functionality and data processing capabilities.

The next sections of this chapter will present possible optimization options for further chip iterations, such as DHPT 1.0.

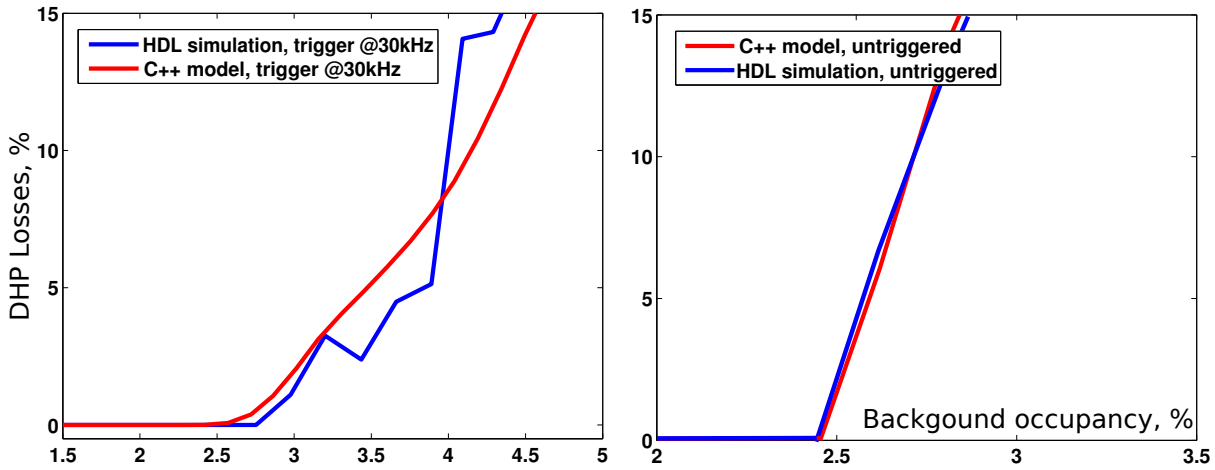


Figure 5.8: Efficiency comparison of the C++ chip model and the HDL implementation for the parameters selected for DHP 0.2 implementation (Buffer-1 =16, Buffer-2 =512). Note: for triggered readout a very high statistics is needed for results to converge: the HDL model is much slower than the C++ model and has lower statistics. Therefore, a high deviation is observed.

5.9 From DHP 0.2 to DHPT 1.0. Further Optimization

The TSMC[1] CMOS 65 nm technology that is used for the DHPT 1.0 production has a higher logic density compared to the IBM CMOS 90 nm technology used for the DHP 0.2 chip production. Keeping the same chip geometry, this offers a possibility to increase the Buffer-1 and Buffer-2 sizes if that profits to the processing efficiency.

For this technology the logic density is about 0.6-2 μm^2 per one bit register depending on the memory type and routing used. We estimate that it is possible to use up to 2 mm^2 out of the total chip area of 12 mm^2 for this purpose. This gives for the most conservative estimations at least 1 Mbit of the available memory for buffer implementations.

To further improve the chip performance for the next chip iteration, the optimization problem was approached in a systematic way, as described in Section 5.4.

Constrained loss optimization

Using the first approach described by Equation 5.4, one can rewrite the constraints as follows (for the worst case scenario with 2 μm^2/bit):

$$\begin{cases} 1280D_1 + 32D_2 & \leq 10^6 \\ L(p, D_1, D_2)\Big|_{p=p_{max}} & = min \end{cases} \tag{5.7}$$

To proceed with this optimization problem, the efficiency scan for different buffer sizes and for several maximum supported occupancies was done.

Simulation shows that these additional resources offer the possibility to have a decent performance for 4.5% data occupancies and even more. The result of this simulation is presented in Figure 5.9.

The straight line represents the maximum memory limit, as described by the first line of Equation 5.7.

[1] Taiwan Semiconductor Manufacturing Company

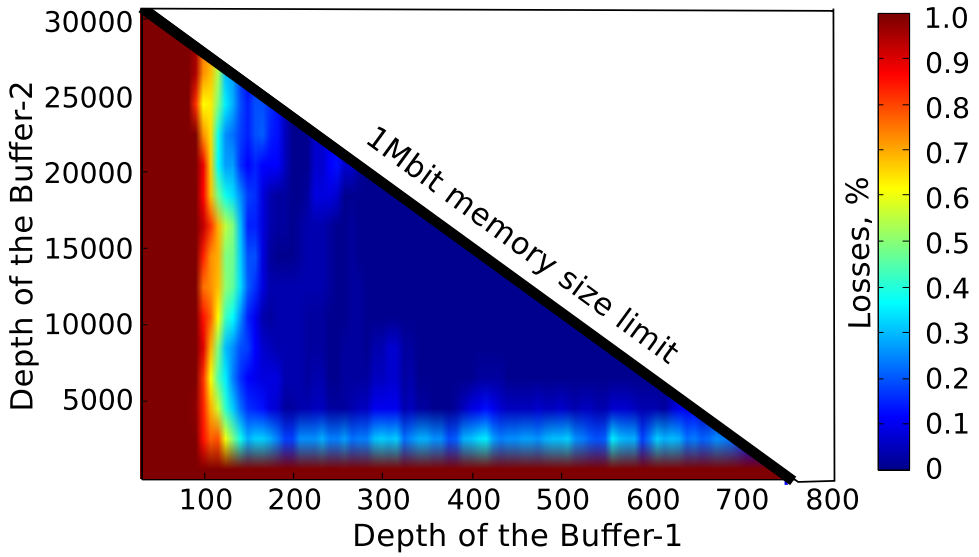


Figure 5.9: The chip efficiency scan for different Buffer-1 and Buffer-2 sizes and for 4.5% occupancy Poisson distributed data. By scanning through different Buffer-1 and Buffer-2 parameters (represented by horizontal and vertical axes respectively), one minimize the data losses, as it is color-coded in the figure. The straight line represents the limit of the maximum implementation area. The trigger is randomly distributed with a rate of 30 kHz.

Constrained Best Supported Occupancy Optimization

After inserting numbers in Equation 5.6 it results in:

$$\begin{cases} 1280D_1 + 32D_2 \leq 10^6 & \leftarrow \text{limited space condition} \\ L(p, D_1, D_2) \leq L_{max} & \leftarrow \text{limited losses condition, for example } L_{max}=1\% \\ p(S \leq 2\text{mm}^2, L \leq L_{max}) = max & \leftarrow \text{the best occupancy condition} \end{cases}$$

In this case the scan parameters are chosen to be Buffer-1 depth D_1 and the data occupancy p. Buffer-2 is defined by the condition that the total memory size is equal to 1 Mbit, thus replacing the "≤" sign by "=" in the first condition of the equation.

The outcome of the resulting simulation is presented in Figure 5.10. We see that the maximum sustained occupancy can be as high as 5.5 % by setting the Buffer-1 depth between 400 and 600 (between 15000 and 7000 for Buffer-2 correspondingly).

These two optimizations will be used as guidelines for the upcoming DHP chip submissions.

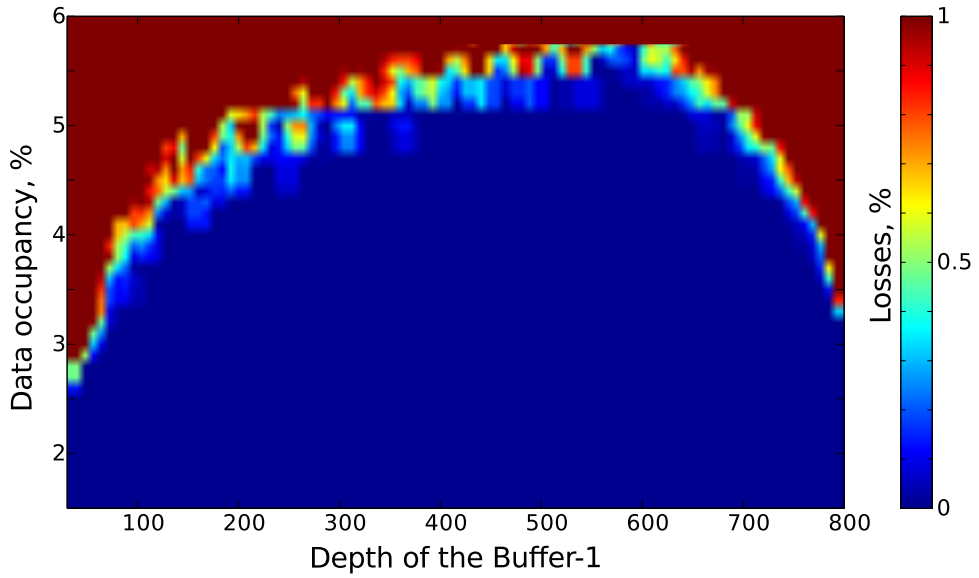


Figure 5.10: Memory constrained chip efficiency scan for different Buffer-1. Buffer-2 compliments Buffer-1 such that their total memory size remains constant and equal to 1 Mbit, as defined by Equation 5.7. By scanning through different data occupancies (vertical axis) and different memory distribution (horizontal axis), it is possible to optimize for the highest supported occupancy. As it is seen from the figure, the optimal configuration is achieved, then the Buffer-1 depth is about 400-600 words. In this case the best sustainable occupancy can be as high as 5.5 %.

Chapter 6

System Tests

This chapter presents the test systems that have been created during the DHP development. First, a short overview about the DHP-emulator system is given, where the DHP was implemented in an FPGA as an additional virtual entity. Then, the main test system is introduced, which was originally designed for the DHP 0.2 chip characterization; later it was used to create the PXD Full-Scale Module Prototype. Finally, the test results will be presented and discussed.

6.1 FPGA DHP Emulation

6.1.1 System

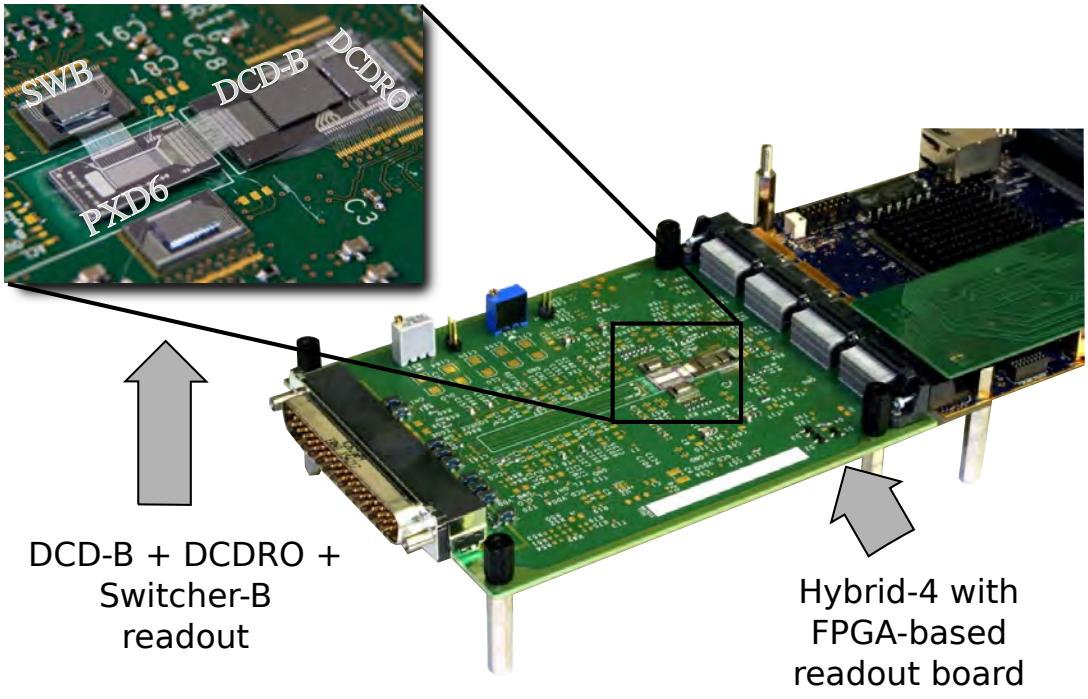


Figure 6.1: The predecessor of Hybrid-5, the Hybrid-4 test setup has been used for a long time as the main PXD prototype. The first zero suppression readout has been here introduced as a virtual entity in the FPGA firmware.

The first DHP test system was implemented on the basis of the Hybrid-4 test readout system presented in Figure 6.1. In this prototype a single DCDB together with the Switcher-B chip are steered by the

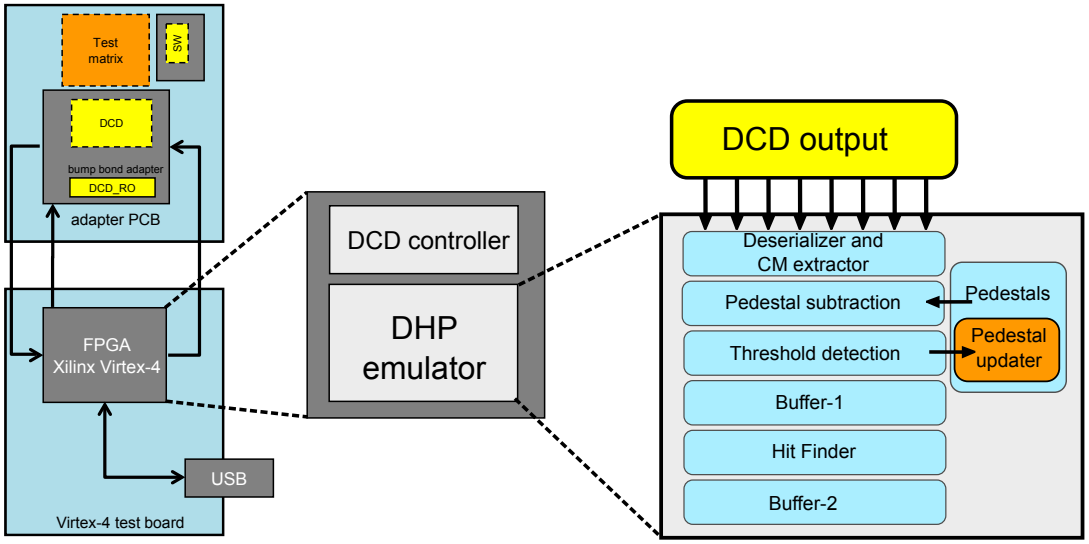


Figure 6.2: Hybrid-4 with a DHP-emulator block diagram. The DHP-emulator is added at the end of the main digital block that controls the DCDB and processing its data. The data are intercepted and zero-suppressed before being transmitted to the acquisition system.

multi-purpose Virtex-4 FPGA based board[1]. Initially, this system has a non-zero-suppressed readout and all data post-processing is done offline.

To test the DHP functionality in real conditions an additional module emulating the DHP behavior was written in Verilog and added to the FPGA firmware. Since the logic resources are limited, this prototype is simpler than the full-size DHP; it is designed to process only small matrices. However, all processing blocks necessary for the zero-suppression were present. Moreover, the dynamic pedestals correction (Appendix B) option was added in the code. This simplified the pedestals handling: dynamic pedestal algorithm, being intrinsically similar to the sliding average, can automatically acquire pedestals after processing several frames.

Internally, the DHP-emulator module was added to the existing DCD readout block, intercepting raw data before they are sent to a computer as shown in Figure 6.2.

6.1.2 Results

This chip emulation was tested in November 2010 during a Test Beam campaign at the CERN-SPS using 120 GeV pions (see Figure 6.3). Figure 6.4 presents several rare nuclear events registered by the system and processed by the DHP-emulator. Apart from the verification of the DHP concept, the zero suppressed readout enabled to considerably increase the data acquisition rate and allowed higher statistics.

Although many system features of the real DHP chip were not present (JTAG interface, CML gigabit link, PLL etc.), the proof of concept for the data processing part was successfully demonstrated. The experience acquired during the development and tests of this system was used to refine the conceptual solutions for the final DHP design.

[1] This board was designed by Manuel Koch for the previous generation of the Drain Current Digitizer chip, the DCD2 [37]. Later this environment was adapted to host the DCDB in the scope of J. Knopf PhD thesis [41].

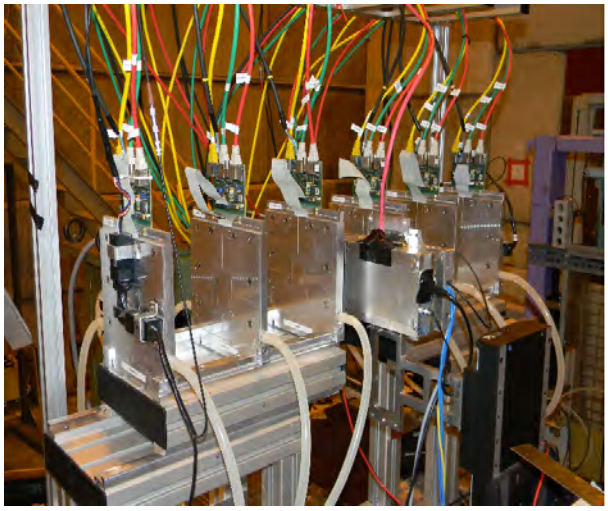


Figure 6.3: Test Beam 2010 in CERN-SRS. EUDET Telescope + Hybrid 4.1.01 (shown in the center). The Zero Suppression was added using the FPGA DHP emulation. This allowed to test the DHP processing chain with real data and to considerably increase the data acquisition rate.

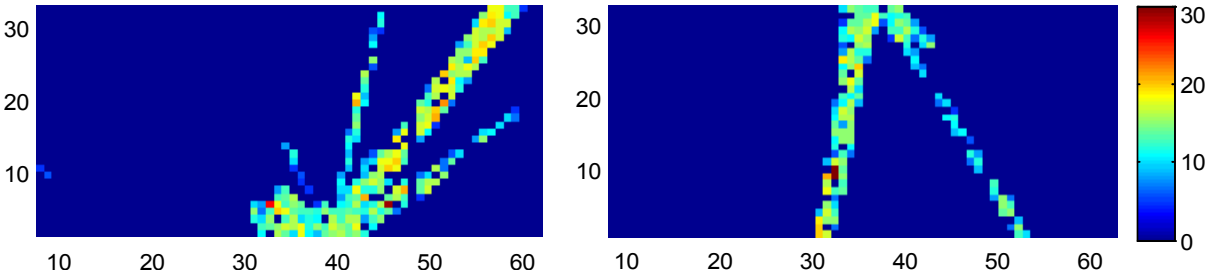


Figure 6.4: DHP emulator results. Rare events: nuclear events registered by the DHP-emulator.

6.2 The Full-Scale Module Prototype of the PXD

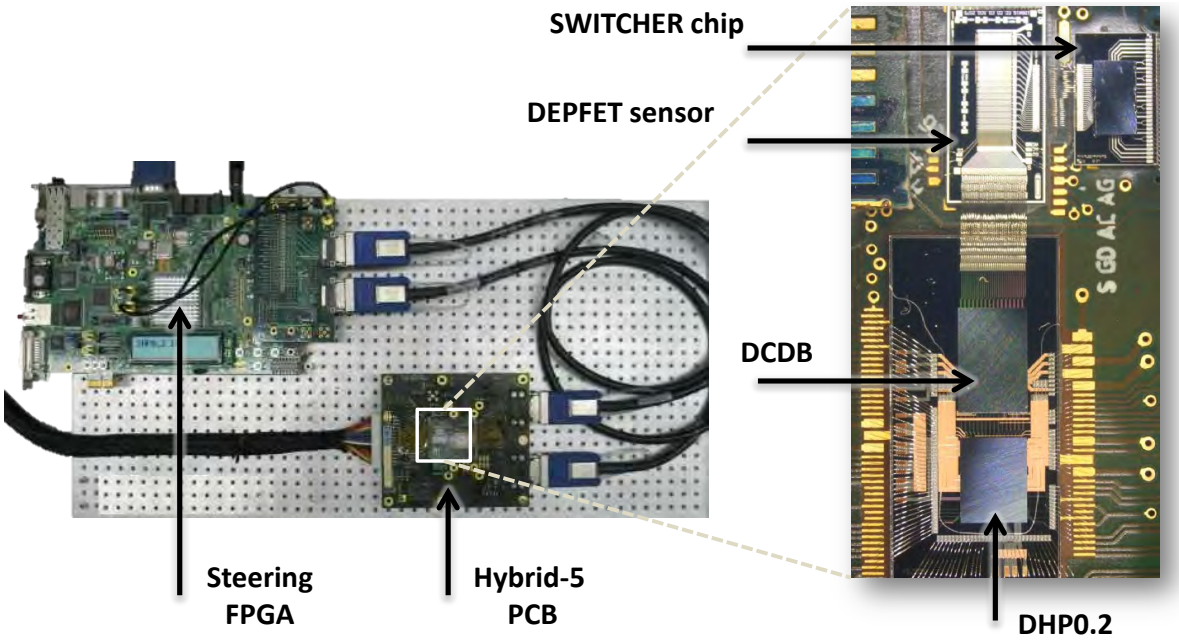


Figure 6.5: The Full-Scale Module Prototype photograph. The FPGA board is controlled from a computer via the TCP/IP protocol. The steering signals and one serial link between the FPGA board and the Hybrid-5 were sent via a pair of InfiniBand®[1] cables. Hybrid-5 hosts a small DEPFET matrix and all necessary ASICs to steer it.

The Full-Scale Module Prototype (FSMP) depicted in Figure 6.5 has been designed as the first prototype of a PXD module (Figure 3.4) containing the minimum amount of all elements necessary for the PXD to function. That is: one small DEPFET matrix[2], one Switcher-B, one DCDB and one DHP 0.2 controlled by one DHH. The idea behind the FSMP is to test the complete steering and processing chain however scaled down to the essentials. If the FSMP is tested with success, it is in principle a matter of scaling to build the full DEPFET module. The FSMP consists of the Hybrid-5 board (Section 6.2.1) and its control system, presented by the DHH emulator (Section 6.2.2).

This prototype has been used to evaluate the DHP 0.2 and test the first version of the full processing chain of the PXD. The results of these tests will be presented in this Section.

6.2.1 Hybrid PCB

A custom made PCB had to be designed to host ASICs and interconnect their inputs and outputs with the readout system. Due to the fact that several kinds of ASICs are present on it at the same time, this is called a hybrid PCB. It was not possible to directly mount ASICs on the PCB, therefore a so-called Wirebond Adapter was additionally designed for these purposes (Figure 6.6).

The initial version of the hybrid PCB has four metal layers and is 120×120 mm^2 large. It is designed to host DHP 0.2 and DCDB chips only. In the next design iteration, the PCB is extended to additionally host one Switcher-B chip and a small DEPFET matrix. This board is called Hybrid-5 and is depicted in Figure 6.5.

[1] InfiniBand® is an industry-standard specification that defines an input/output architecture used to interconnect servers, communications infrastructure equipment, storage and embedded systems [60].

[2] We used a standard test matrix produced in 2010-2011 with 50 μm thickness of the active area (PXD6).

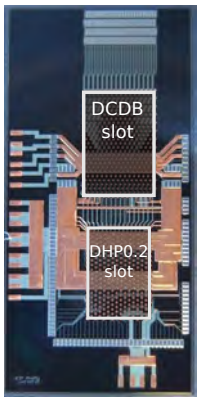


Figure 6.6: The Wirebond Adapter for DCDB and DHP 0.2 chips. The ASIC is placed in the center of the adapter using the bump-bonding technique [62].

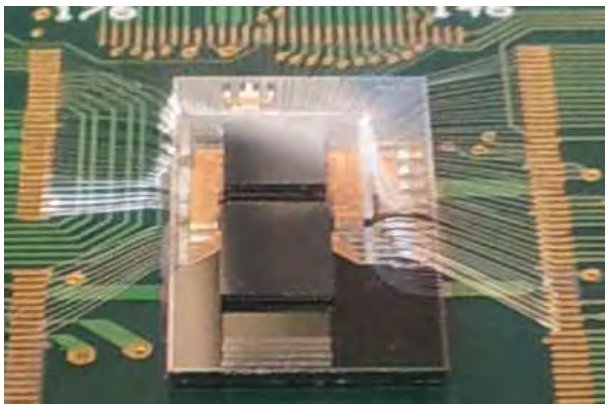


Figure 6.7: A photo of the Wirebond Adapter connected to the hybrid PCB.The main purpose of the Hybrid PCB is to interconnect the Wirebond Adapter shown in Figure 6.6 with the control interface and the power supply.

The design of Hybrid-5 was inspired from previous existing hybrid PCBs, such as Hybrid-4 [37] and S3B system [61]. However, this design was simpler: thanks to the DHP 0.2 presence, only one high-speed link (the DHP serial output) was necessary to be implemented.

6.2.2 FPGA Readout System

As described in Section 3.6, in the final Belle II design of the detector each half module of the PXD is controlled by one dedicated DHH. In the moment of the FSMP development the DHH hardware was not available. Therefore, a temporary replacement emulating the DHH functionality has been created in order to control the system. There is currently quite a good choice of the general-purpose development boards, which could be used for the creation of such kind of systems. After some market research, the decision to use the XUPV5 evaluation board [63] for this purpose was taken. This choice was defined by a rich amount of on-board industry standard interfaces and a good price[1].The experience we got during this development brought a valuable contribution to speed-up the DHH firmware development.

To some extent[2], it was a full replacement of the DHH, mimicking its functionalities. For this reason it is called DHH Emulator in this text.

As depicted in Figure 6.8, the DHH Emulator consists of two parts: the XUPV5 board and the Expansion PCB that was additionally designed by our group in order to interface the Hybrid-5. All relevant blocks are marked by numbers in orange circles:

1. **VIRTEX-5 FPGA**. The main element of the DHH Emulator hosting the control firmware, based on Mircoblaze controller.
2. **RocketIO® GTP transceiver**[3] is used to interface the high-speed DHP 0.2 serial link with the FPGA.
3. **The expansion pinout array** is used for all other signals necessary to control the system: JTAG, synchronization, Trigger Logic Unit, etc.

[1] This system is subsidized by Xilinx for academic purposes, making this choice very attractive.
[2] The main limitation is that using this system we can steer only one single DHP.
[3] The RocketIO® GTP Transceiver is a multi-gigabit highly configurable transceiver designed by Xilinx. It integrates a variety of features, among which are: CML buffer with configurable termination, RX equalization; it supports rates up to 3.75 Gbps.

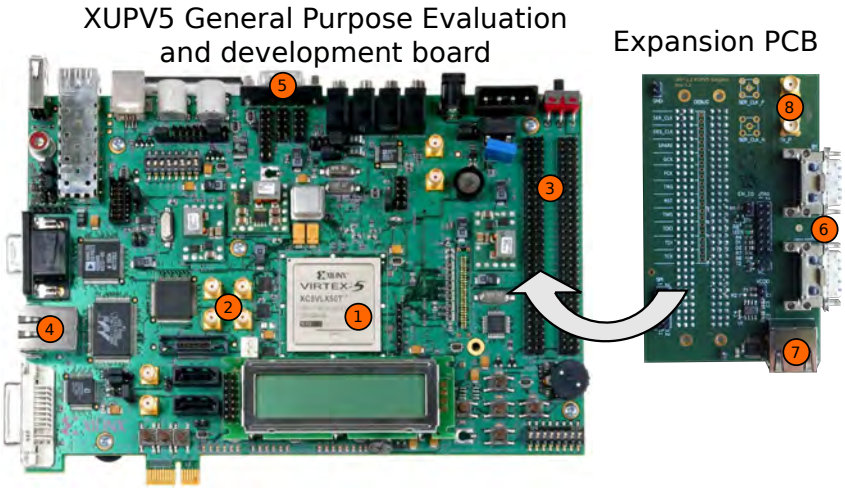


Figure 6.8: XUPV5 general purpose development platform with expansion adapter.

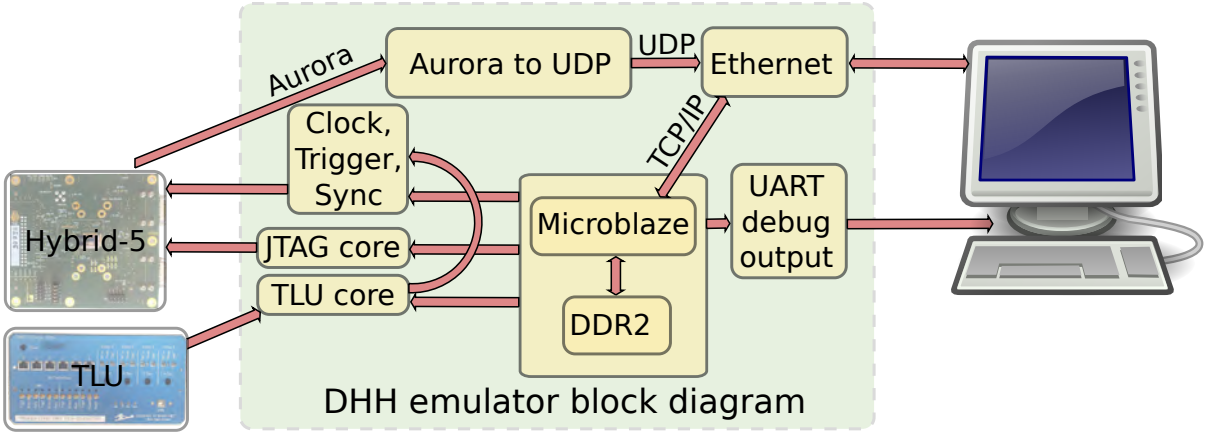


Figure 6.9: DHH Emulator firmware's block diagram. Each rectangle represents an independent HDL module. The main block is a soft core Microblaze [64] processor that controls other modules using a data bus.

4. **Ethernet cable plug**: the system was controlled from PC via TCP/IP and UDP protocols.
5. **RS-232 Debug interface**.
6. **InfiniBand® cable plugs**.
7. **A plug to interface a TLU**[1]
8. **DHP gigabit link output** is connected to the GTP transceiver via a pair of SMA cables.

The DHH Emulator is designed to support one DHP-based system. It is controlled from a user's PC via an Ethernet cable. A JTAG core is used to configure Hybrid-5. For a user to receive the Hybrid-5 output data, Aurora frames are converted into UDP[2] packets and sent to the control PC. The control firmware is written using HDL language with a use of the Microblaze soft-core processor [64], allowing executing C/C++ programs. This hybrid solution combines the flexibility of a software written code and the speed performance of the hardware modules. The block diagram of the system is depicted in Figure 6.9.

[1] The TLU stands for Trigger Logic Unit, which is designed to trigger and transmi t trigger numbers. It is needed if the Hybrid-5 is run synchronously with other detectors, for example during a Test Beam.
[2] User Datagram Protocol, a simple transmission protocol belonging to the standard Internet protocol family.

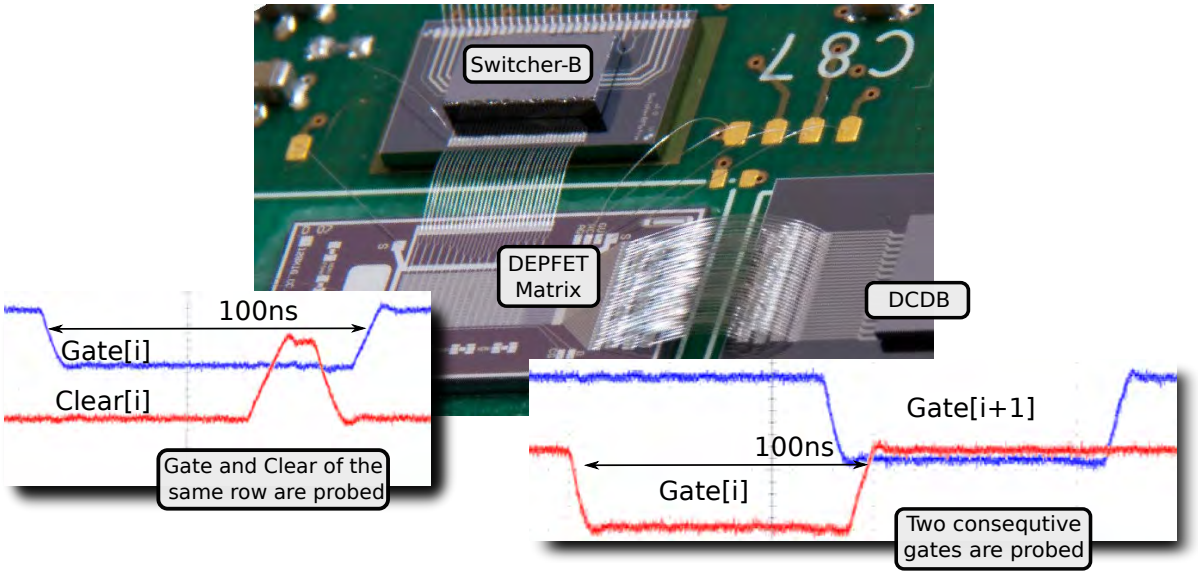


Figure 6.10: The DEPFET matrix wirebonded to the Switcher-B. After configuring the DHP 0.2 we were able to generate the steering sequence necessary to run the system. The presented signals were directly probed from the Switcher-B outputs.

6.3 DHP 0.2 Tests

6.3.1 System Start

The first step in DHP 0.2 testing was to see if we can run the system, control the chip and if the data processing functions correctly.

As a result, the PLL, the LVDS outputs and the CML link showed the expected behavior. In particular we were able to control the Switcher-B chip. The observed control sequence on the outputs of the Switcher-B is shown in Figure 6.10.

To test the data processing, a test frame and the corresponding pedestals are loaded into the chip memory. With no surprise, the data processing logic behaves exactly as in simulation. This justified the big effort spent on the creation of the HDL verification environment described in Section 5.6.

6.3.2 Serial Link

To analyze a transmitted signal, the *eye diagram* [65] is a very common tool in use, giving a first idea about the signal quality in high speed digital transmissions. The eye diagram is an overlayed version of the received waveforms, triggered at the master clock. After acquiring many transitions, a so-called eye is obtained.

The sampling point is usually chosen to be as far as possible form the transition regions to maximize the difference between the low and high logic states. By measuring how much the eye is opened one can get an idea about the quality of the signal: the larger are the eye width and eye height, the more margin has a receiver to correctly sample and evaluate the result. An eye diagram received from Hybrid-5 after crossing 1 m InfiniBand® cable is presented in Figure 6.11.

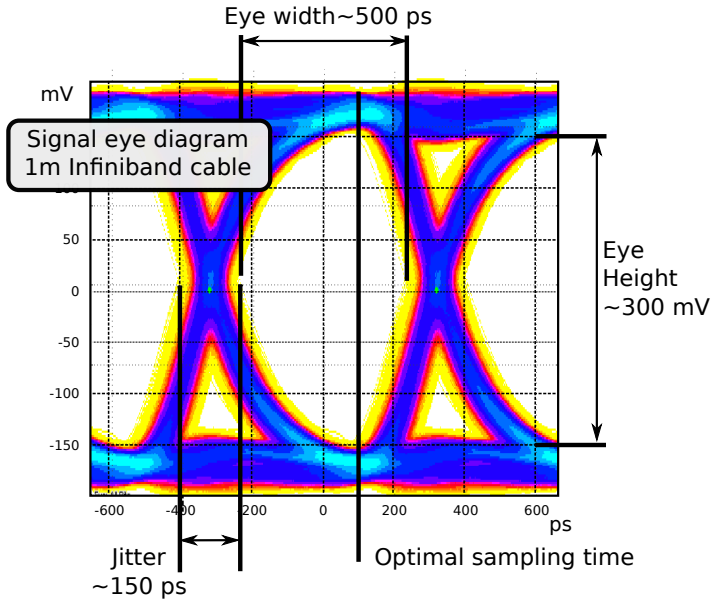


Figure 6.11: An overlay of a large amount of received waveforms, sampled with a data clock, forms a so-called eye diagram used in data transmission quality analysis. The more the eye is "opened", the lower is the error rate. In this example the signal crossed 1 m of InfiniBand® cable.

As described in Chapter 3.6, a cable length of ~15 m is expected to be used for each PXD module to interconnect it with the corresponding DHH[1]. In such a long cable, the signal undergoes a strong attenuation at high frequencies.

We tested if the transmission still works with 15 m cable between the Hybrid-5 and the DHH Emulator. Unfortunately, the signal quality was so poor that the Aurora link [2] could not be synchronized. The corresponding eye diagram is presented in Figure 6.12(left). The total signal amplitude remained almost the same, however the eye height considerably decreased (from 300 mV for 1 m to 80 mV for 15 m cable length), resulting in a bad signal to noise ratio.

Thanks to the preemphasis of the DHP 0.2 CML driver (described in Section 4.3.2) it was possible to improve the eye diagram quality of the received signal up to the acceptable level, so that the data link between the transmitter and the receiver could be established. The eye diagram for the best achieved preemphasis setting is depicted in Figure 6.12 (right).

Bit Error Rate

To quantify the quality of the high speed serial link, an eight-bit pseudo-random pattern is used for the data transmission test lasting one day, which corresponds to the $N_{transmitted}$=10^{14} bits. No errors were registered. For the Poisson distributed errors no more than N_{errors}=6 errors with a confidence level (CL) of 99.7% can occur during this period of time. Hence, the total Bit Error Rate (BER) is equal to:

$$\mathrm{BER}(@\mathrm{CL}=0.997) = \frac{N_{errors}}{N_{transmitted}} = 6 \cdot 10^{-14}$$

[1] Additionally a 40-60 cm Kapton® cable will be connected in series. Since it is considerably shorter than the main cable, we do not expect high attenuation in it.

[2] see Chapter 4.2.7

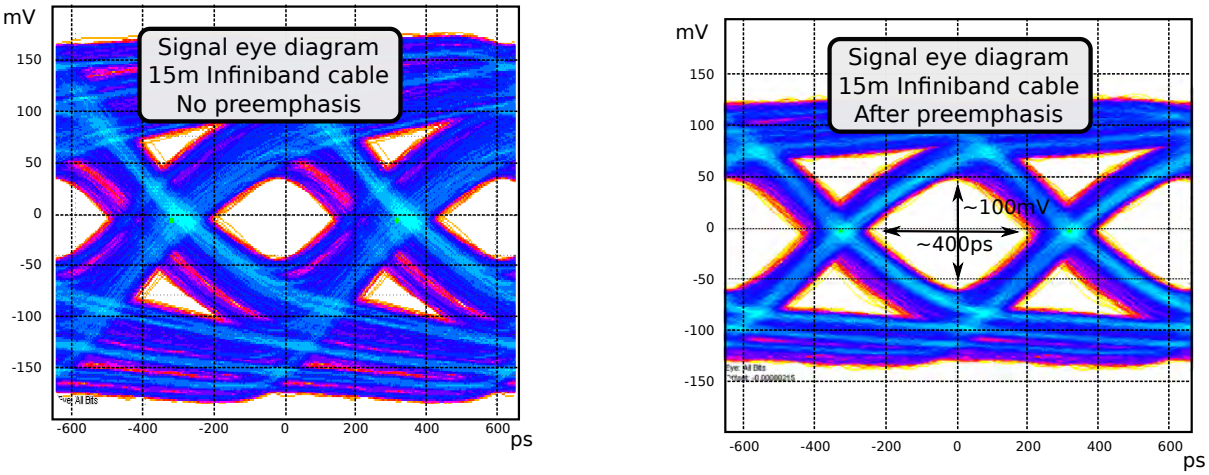


Figure 6.12: 15m cable tests. With no preemphasis (left). With the best achieved preemphasis using the DHP 0.2 CML driver (right).

6.3.3 DHP 0.2 + DCDB Tests

A detailed analysis of the DCDB performance on the basis of the Hybrid-4 system has been reported in [41]. Using these results and the optimal DCDB settings found previously, is was possible to establish a link between two chips and control the DCDB chip on the Hybrid-5 setup. Three steps were necessary to control DCDB:

1. **Chip configuration**. The DCDB chip is configurable from the common JTAG chain described in Section 3.5.

2. **Digital data transmission**. A built-in test pattern to verify the digital communication has been implemented in DCDB. By adjusting the input reference clock latency, the data transmission correctly works up to the nominal speed of 305 MHz.

3. **DCDB analog performance**. To test if all ADC channels of the chip are working correctly, current injection tests via a debug input were executed[1]. A correct behavior for all channels has been observed up to 160 MHz input clock rate. At the nominal speed of 305 MHz too many (more than 15 % – 20 %) channels have shown a noisy behavior, making Hybrid-5 unusable. These two situations are presented in Figure 6.13[2]. This problem was previously observed while running standalone DCDB chips. It can be remedied by output lines' delays fine-tuning. This option is implemented in the next DHP submission, the DHP 0.1 chip.

A detailed analysis of the DCDB analog channels, such as INL, DNL characteristics, the Gain and the Offset spreads are omitted, being a topic of another thesis; previous results can be found in [37, 41].

[1] A precision current source Keithly 2400 was used for these purposes.

[2] In Hybrid-4 system it has been reported that there is no problems to run the DCDB chip at the nominal speed. We suspect that this observed in Hybrid-5 problem can be due to the limited DHP 0.2 delay adjustment capabilities and Deserializer clock duty cycle. This issue is scheduled to be corrected in the next DHP chip iteration, the DHPT 1.0.

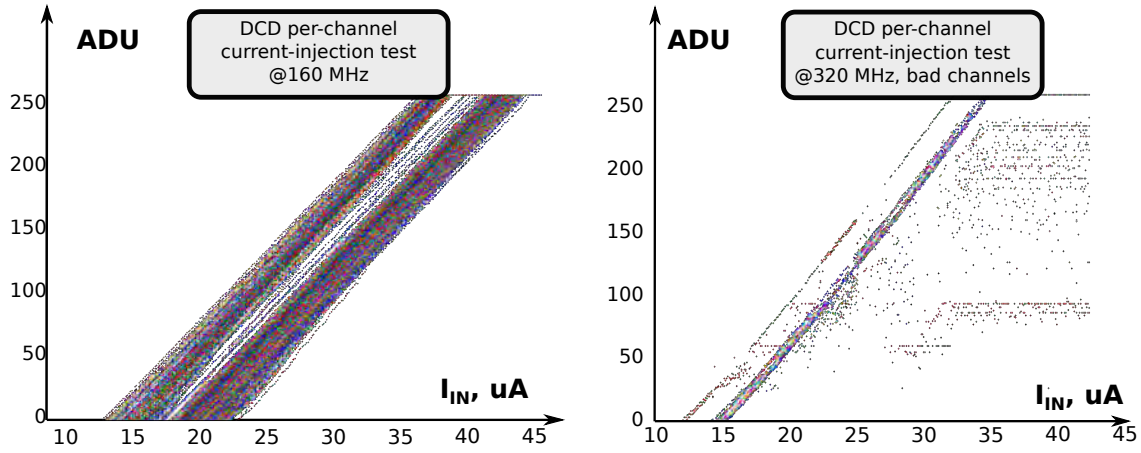


Figure 6.13: DCD curves. The DCDB can correctly run with 160 MHz nominal speed, left figure represents ADC scans for all channels. At full speed (305 MHz) more than 10% of the channels showed a noisy behavior. These bad channels are shown in the right-hand figure.

6.4 FSMP Tests

6.4.1 Hybrid-5 Limitations

The Hybrid-5 PXD prototype runs with the following limitations:

- **Half of the nominal speed**. As reported in Section 6.3.3, we can only run the system at 152 MHz, which is approximately half of the nominal speed.
- **Partial Common Mode correction**. A relatively small DEPFET matrix with 16 switcher rows and 128 drain channels is used in Hybrid-5. This means that only half of the 256 inputs of the DCDB are connected. The Common Mode correction block of the DHP 0.2 has been designed assuming that all channels will contain common mode information. Therefore, the zero suppressed output data is only partially CM corrected, small residuals will be still there. It is possible to remove these residuals offline, since the information about the CM is sent together with the data. However, this partial CM correction forces to apply a relatively high threshold for the signal detection; this limits the detector energy and the position resolution; additionally it causes a partial data loss for low energy signals.

6.4.2 Matrix Laser Scan

The small DEPFET matrix connected to the DCDB is a 32×64 pixels structure[1]. Each pixel has a size of 50×75 μm^2, this corresponds to the total matrix size of 1.6×4.8 mm^2.

Due to complex wiring the received coordinates of zero-suppressed data does not correspond to physical pixel positions in the matrix. Therefore, a laser scan is done to determine the correspondence: in this automatic procedure every pixel of the DEPFET matrix was hit by a laser pulse and the zero suppressed data position was registered; knowing the coordinates of the laser pulse and the output coordinates, the desired transformation was found. In Figure 6.14 the resulting matrix image is presented (left): thanks to the sub-pixel precision of the scan, the high and the low sensitivity areas of the matrix can be observed.

[1] 32 columns×64 row corresponds to the electrical topology of 16 Switcher-B rows×128 drain columns.

Moreover, since the matrix was illuminated from the back-side, one also sees the non-transparent metallization area. In the zoomed area a precision scan (with a step size of 10 µm in both directions) of the 4×6 pixels region is presented. One can observe a ~10% position dependent variation of the cluster signal, which contributes to the total energy resolution limitations.

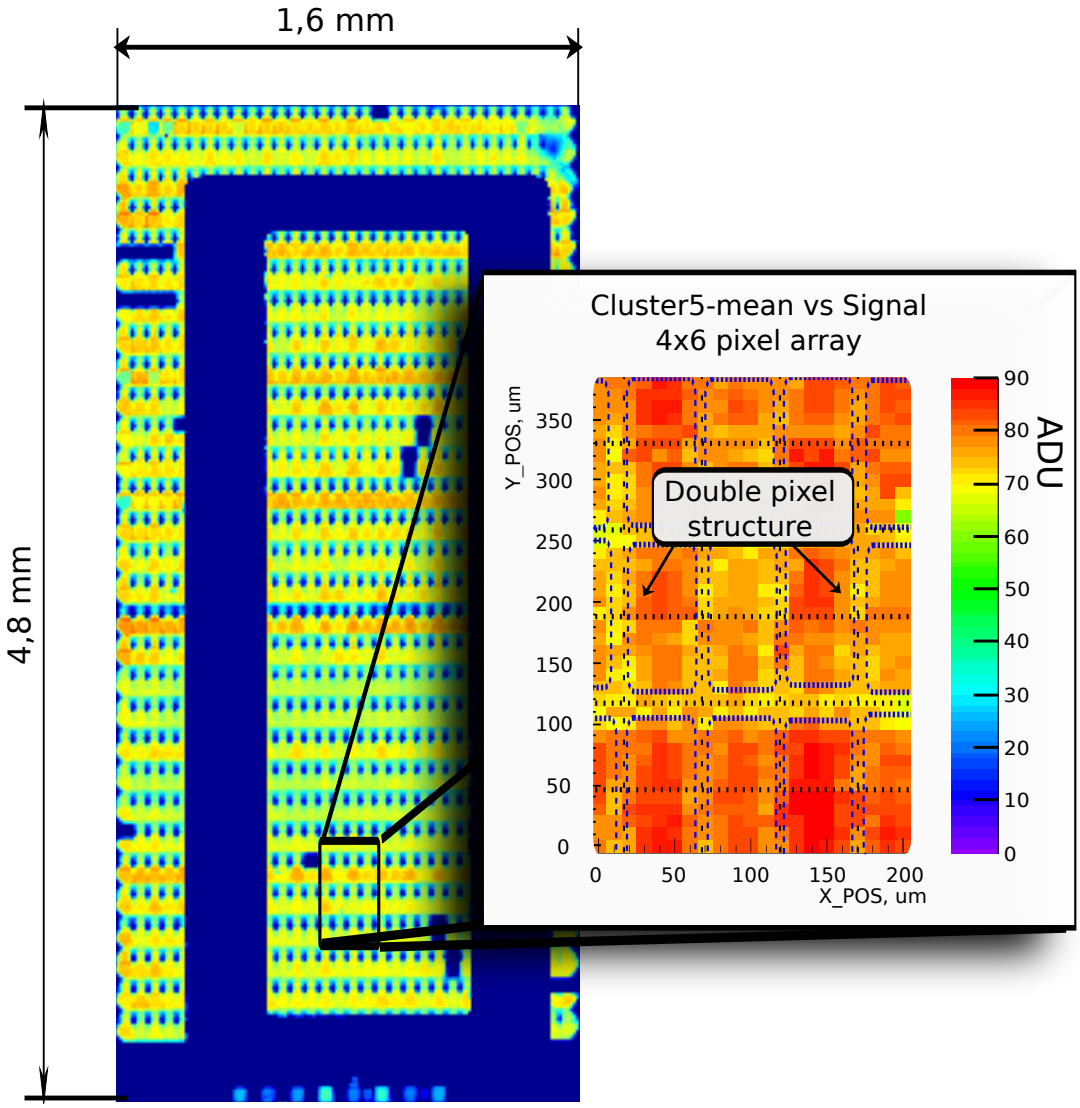


Figure 6.14: (Left) The laser scan of the whole matrix (seed signal) to test if pixels are correctly mapped. The back side metallization is seen as a blue rectangle, since it absorbs photons and no light reaches the sensitive area. In the zoomed area a precise homogeneity scan of the 4×6 pixel area with a step of 10 μm in both directions is presented (cluster signal, optimized voltages). Picture and measurement are taken from [66]

6.4.3 Source Tests

A 74 MBq Am^{241} source was used for the calibration of Hybrid-5. It is shielded by a 0.25 mm beryllium foil to suppress the α-particles emission. From the Am^{241} reference table, one can find the following γ-lines with the highest intensity in the spectrum [67]:

1. **59.5 keV** is one of the brightest lines in the spectrum with a relative intensity of $I_{59.5keV}$=35.9 %.
2. **33.2 keV** is a line with a relative intensity of $I_{33.2keV}$=0.13 %.
3. **26.3 keV** is a line with a relative intensity of $I_{26.3keV}$=2.4 %.
4. **13.9 keV** a very bright line with a relative intensity of $I_{13.9keV}$=42 %. However, it is situated close to the system's detection threshold, so it cannot be used for the spectrum calibration[1].

To acquire the Am^{241} spectrum, Hybrid-5 was irradiated for 10 hours with 10^7 resulting registered events. The system was set in the triggerless mode[2] and it was able to capture all absorption events. This allowed having a high event statistics, even though the absorption probability is extremely low. This gives a great advantage for the new system: the previous prototype is only able to capture $\sim 0.1\%$ of total events [41] and more than a year would be necessary to gather the same amount of events. Furthermore, because the data is already sent in zero-suppressed format, no additional data post processing is needed while running Hybrid-5.

Due to the negligible probability to obtain a signal, the calculated cluster energy was taken as a simple sum over all received hit values. Indeed, for a hit rate of 300 events per 50.000 frames, this give an average rate of λ=0.006 events per frame. Assuming Poisson statistics of the events arrival, this gives the following probability to acquire more than one event per frame:

$$P(N \geq 1) = 1 - P(N = 0) - P(N = 1) = 1 - (1 + \lambda)e^{-\lambda} = 1.79 \cdot 10^{-5}$$

In other words, roughly 180 out of 10^7 registered events contained double hits.

The results of this acquisition are presented in Figure 6.15.

This histogram was calibrated using 59.5 keV and 26.8 keV lines. By using the Gaussian fit of the 59.5 keV line, the detector noise can be estimated to be 860 e^- (or 5.2 % of the energy resolution).

Most of the signal is contained in one pixel cluster, a certain amount of two and three pixel clusters was also observed. Clusters of other sizes were present in negligible amounts (less than 0.5 %). The contribution of each type of clusters is presented in the stacked histogram in Figure 6.15.

Spectrum analysis

1. **Interaction probability**. The photon beam, while crossing the matter, attenuates exponentially (known as the Bouguer-Lambert–Beer law in optics):

   $$N = N_0 e^{-\mu L}$$

   Here, μ is the attenuation coefficient, which depends on the photon energy and the material in use.

   As depicted in Figure 6.16, in the energy range of 10–100 keV two processes play a major role in the photon beam attenuation in Silicon: photo effect [69] and Compton scattering [70].

[1] This situation can be correcting if using the upcoming DHPT 1.0 chip, which has a more advanced common mode correction module.

[2] i.e. trigger is set to be always on

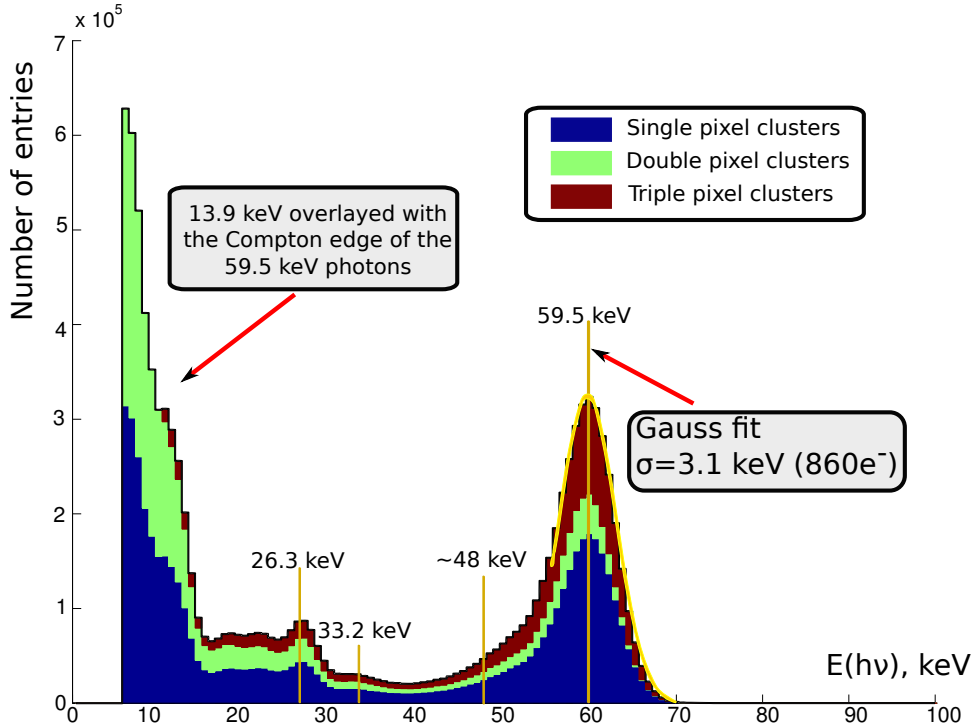


Figure 6.15: Am^{241} spectrum. 10^7 is the total number of events (error bars are smaller than the line width). The scaling was done using two lines: the 59.5 keV and the 26.3 keV. One can additionally recognize a weak 33.2 keV line also expected in the spectrum. The 59.5 keV Compton Edge (11.8 keV) mixed with the 13.9 keV line is very pronounced. Left-right asymmetry of the 59.5 keV line can be explained by back-scattered photons with energies ≥48 keV.

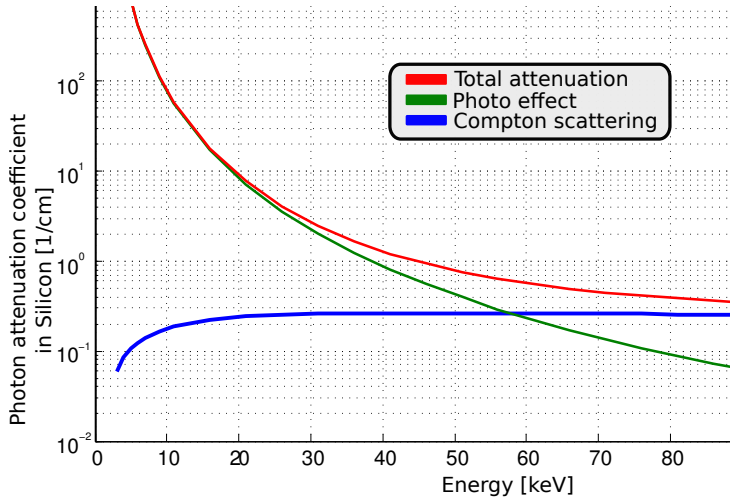


Figure 6.16: Photon attenuation coefficient in Silicon. In the range of 10–100 keV Compton scattering and Photo-effect bring the main contribution in the total attenuation [68]. At 60 keV Compton scattering is slightly dominant, which explains the large Compton Edge observed in Figure 6.15.

$$\mu = \mu_{Compton} + \mu_{Photo}$$

In the example of the 59.5 keV photons of the Am^{241} spectrum, the total interaction probability for a thin sensor of 50 µm can be calculated as:

$$p = (1 - e^{-\mu_{tot}\Delta L}) = 1 - \exp(\text{-}0.74\ \text{cm}^{-1} \times 0.005\ \text{cm}) \approx 0.0037 \tag{6.1}$$

2. **Compton edge**. From the cross section we see that Compton scattering should be slightly dominant. The energy transferred to the substrate varies with the scattering angle θ and has a broad energy spectrum with a sharp edge, called Compton edge. This transferred energy can be calculated as:

$$E_T = E_\gamma - E_\gamma^{'} = E_\gamma \left(1 - \frac{1}{1 + \frac{E_\gamma}{m_e c^2}(1 - cos\theta)} \right)$$

The Compton edge corresponds to the maximum energy transfer E_T, which is reached at θ=180°:

$$E_{Tmax} = E_T(\theta = 180^\circ) = \frac{E_\gamma^2}{E_\gamma + \frac{m_e c^2}{2}}$$

For the photon energy E_γ = 59.5 keV, the Compton edge is at 11.2 keV. This is present in the spectrum (Figure 6.15). However, it is overlayed with the expected 13.9 keV peak. The problem is that this signal is partially lost being too close to the detector's threshold. For a better analysis at these energies a lower threshold would be necessary.

3. **Back scattering**. The 59.5 keV photons that indirectly reach the detector (for example back scattered from the support table below the sensor) would have a lower energy. Assuming only one scattering, their energy cannot be lower than the initial energy minus the highest energy they can transfer:

$$E_{\gamma min}^{'} = E_\gamma - E_{Tmax} = 59.5\ \text{keV-}11.2\ \text{keV}=48.3\ \text{keV}$$

Indeed the 59.5 keV peak is asymmetric and visibly thicker from the left side. These back scattered photons can be the reason for that.

4. g_q **measurement**. Using the DCDB calibration data presented in Section 6.3.3 the calculated charge referred transimpedance is estimated to be g_q=420±19 pA/e^-. However, the g_q being proportional to the regular transistor's transimpedance g_m can be varied as a function of V_{gs}.

6.4.4 Test Beam Results

The Test Beam (TB) is the closest to reality way to test the system. It allows reaching conditions that are similar to what the PXD will experience in Belle II. It is also an excellent test for some other properties of the system, such as the detector efficiency and its spatial resolution. Therefore, to be kept up to date on the system performance and learn more how to handle it, TBs are scheduled on a regular basis, once or twice per year.

The main program of the TB that took place in DESY (Hamburg) in May 2013 was to test the performance of Hybrid-5 together with its full acquisition chain.

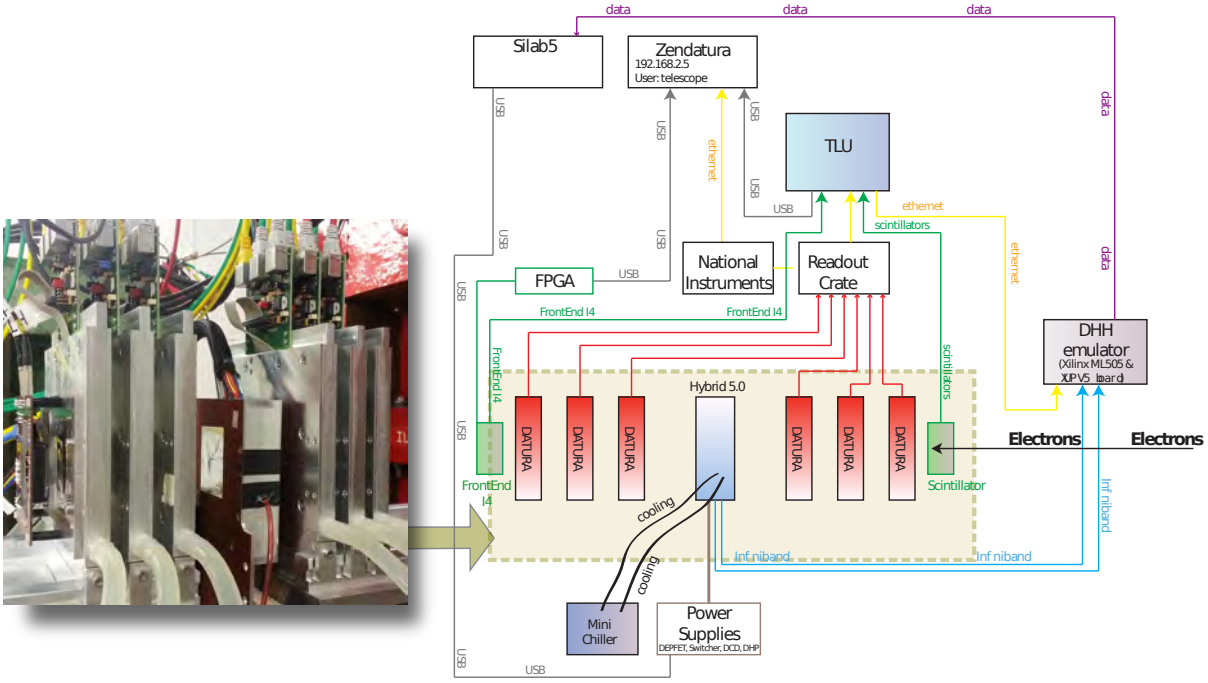


Figure 6.17: The Test Beam setup in DESY Experimental Hall-2. The Hybrid-5 is installed between six reference planes of the telescope, whose block diagram is presented in the right. The photo and diagram are taken from the Test Beam log book.

The test setup was mounted using the DATURA EUDAQ telescope [71] with six reference planes for tracking reconstruction. The Hybrid-5 was the device under test and was situated in the center of the telescope. Two scintillators were installed in front and behind the telescope. Upon simultaneous event detection by all reference planes and scintillators, the Trigger Logic Unit (TLU) issues a trigger for the device under test. This allows for a precise track reconstruction to measure the detector efficiency and spatial resolution. Hybrid-5 is steered by the DHH Emulator and sends the data to the collecting computer. A Block Diagram of the test setup is sketched in Figure 6.17.

The TB was done at the DESY Beam Experimental Hall-2, where and electron beam of variable energy up to 6 GeV can be supplied. The beam intensity profile for different energies is presented in Figure 6.18.

The Hybrid-5 was thoroughly tested weeks before in laboratory conditions. Therefore, it was not expected to have any problems linked with the setup hardware. The main challenge of this TB was the full acquisition chain verification, including other system elements, not belonging to Hybrid-5, namely the EUDAQ telescope integration and the TLU synchronization, which is needed for the tracks' reconstruction.

From the system point of view, TB went very smoothly: Hybrid-5 was ready to take data on the second day after installation. Figure 6.19 shows the example of the first ten million events recorded by the system. From the expected energy deposition one could deduce the g_q=450 pA/e$^-$. Overall, the Hybrid-5 showed the excellent performance beyond expectations and all scheduled tests were successfully carried out.

A positive outcome of this TB is a milestone of a great importance in the whole PXD development, as now the full module production can start. Once the large matrices for the modules are ready, it is a matter of elements replication to build the full PXD.

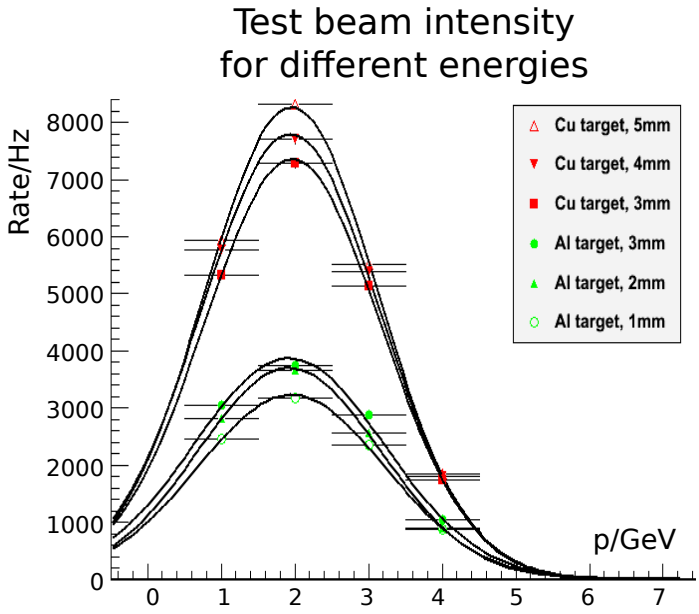


Figure 6.18: Electron beam rates. Using a collimator and a variable magnetic field to select the electrons, energies up to 6 GeV were accessible.

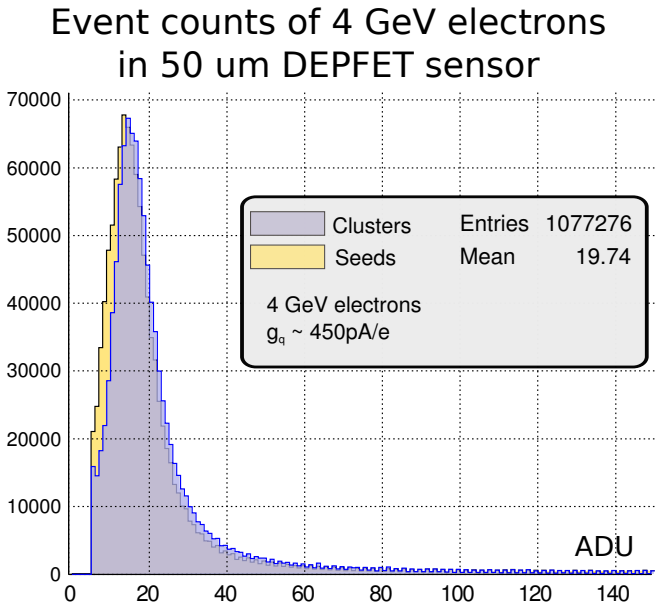


Figure 6.19: Histogram of the signals from 4 GeV electrons.

Chapter 7

Technology SEU sensitivity

7.1 Introduction

> Almost everyone who has had extended experience with electronic computers has witnessed unexplained events in which a single digit of a number appears to change spontaneously, or perhaps the computer itself suddenly stops, and no way can be found for it to repeat the failure. Within the computer industry these problems are known as "soft fails", which differentiates them from the "hard fail" of a bad electronic circuit that must be replaced. A soft fail in a computer memory may be defined as the spontaneous flipping of a single binary bit, which when later tested will prove to be operating correctly.
>
> The appearance of soft fails in computers has recently become prominent because of the α particle problem. This problem was suddenly recognized in 1978 after a new generation of electronics with very small circuit components was introduced. Alpha particles (helium nuclei) are the decay particles of radioactive chains of atoms which start with uranium or thorium atoms and have emission energies between 5 and 10 million electron volts. They are produced by traces of uranium or thorium in or near the electronic circuits. These α particles can produce up to 3 million electron-hole pairs (but not more) within the silicon crystal on which the electronic circuits are fabricated. Until 1978 electronic components in computers were apparently not sensitive to noise bursts of 3 million electrons, so the problem was not recognized earlier.
>
> J.F Ziegler and W.A. Lanford,
> "Effect of Cosmic Rays on Computer Memories"[72]

A very important issue of modern submicron electronic circuits is their tolerance to radiation. In a most generic way these effects can be classified in two categories:

1. Cumulative effects
2. Single Event Effects (SEE)

Cumulative ionization effects being a function of Total Ionizing Dose (TID) fall into the first category. Because of its discrete nature, digital electronics is insensitive to these kind of effects[1] up to a certain and rather high dose of radiation: these effects are not seen as long as the switching logic is properly biased and stays inside its working region .

To the second category belong localized effects induced by a single particle collision effect. A particular type of SEE is Single Event Upset (SEU). It is a stationary effect, meaning in simple words a sudden bit-flip of a logic memory cell, happening if a crossing particle generates enough charge to overwrite

[1] Affecting such parameters as threshold voltage V_{th} or leakage current I_{leak}.

the previous value, resulting in a binary error.

In this chapter a short overview of this phenomenon will be given, which we need to discuss for the characterization of the radiation tolerance of the CMOS 65nm, the technology used for the DHPT 1.0 fabrication.

7.2 Definitions

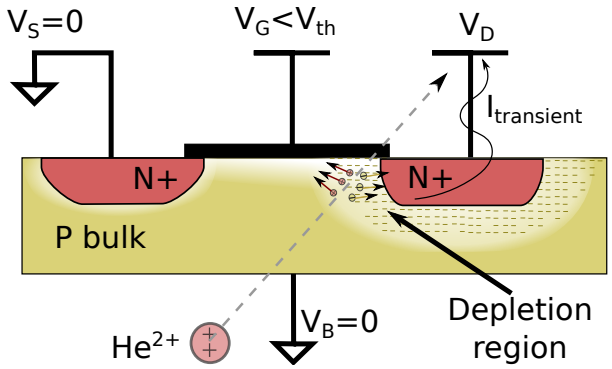


Figure 7.1: For a FET in the off-state the drain–bulk contact represents a reversely biased junction whose depleted region beneath the implantation represents an SEU sensitive region. Charged particles, upon crossing this region, may generate undesirable transient currents susceptible to produce an SEU.

To introduce the SEU phenomenon, an example of an SRAM memory cell will be used. The SRAM cell is one of very commonly used bricks of digital electronics.

Sensitive Region

The sensitive region (SR) is the depleted region of the reverse biased p–n junction (Figure 7.1). A charged particle with high Linear Energy Transfer, typically a charged ion, while crossing the SR can generate a large amount of e–h pairs, inducing unwanted transient currents in the circuit. Contrary to standard silicon detectors, where this property is desirable, the SR is a side effect.

SRAM Cell

An SRAM cell is represented in Figure 7.2, consisting of two inverters. By connecting the input of one inverter to the output of the second one, a feed-back loop is created, the logic state remains constant as long as no external steering signal is applied. Indeed, assuming for a moment that node **A** has the logic state zero, this will imply that the inverter **1** forces the node **B** to be in the logic state one. The inverter **2** in return maintains **A** in its value zero and the total construction is stable. These two equilibrium states are used to program two values of one bit: zero and one.

Particles crossing the bulk may create enough charge to flip the memory state. Figure 7.3 sketches the same SRAM on transistor level. Studies[73, 74] show that there are two points in an SRAM memory cell sensitive to the SEU – the two OFF transistors, more precisely, their drains (in Figure 7.3 marked as yellow). The sensitive region of the NMOS transistor is intentionally sketched larger than that of PMOS one. It is explained by the fact that PMOS are less sensitive to SEU. Indeed, in a standard CMOS technology the PMOS is situated in an additional N-well. The reverse-biased P-N junction of

the PMOS drain enters to the competition with the P-N junction of the bulk–N-well region. Due to charge sharing less signal is captured.

Because of the intrinsic symmetry of the SRAM cell, the transitions 0→1 and 1→0 are equiprobable.

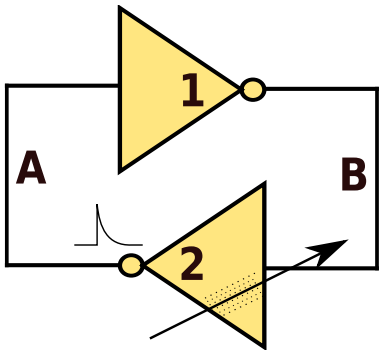


Figure 7.2: A schematic representation of an SRAM memory cell. The arrow crossing the inverter **2** represents a crossing charged particle, which deposits a certain amount of charge in the bulk.

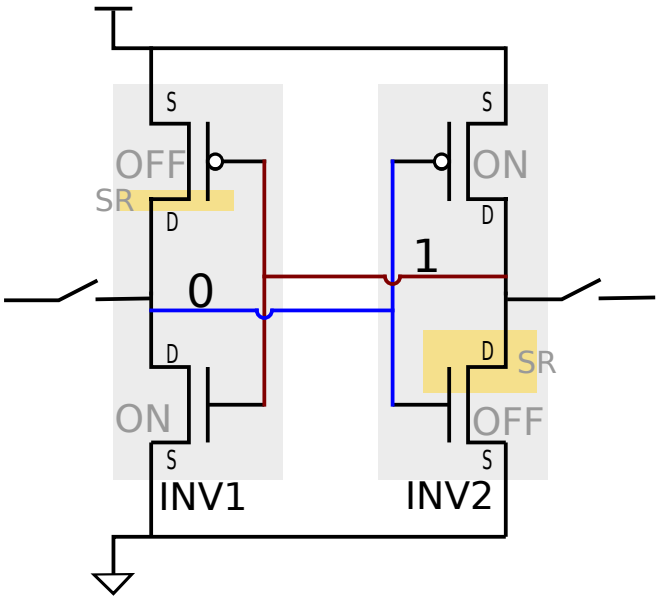


Figure 7.3: SRAM cell electrical diagram. Sensitive areas, which are drains of the "OFF" transistors, are marked with yellow. The NMOS has larger SR than PMOS.

Sensitive Volume

The volume of the depleted region of the sensitive node, where the ionisation has to take place, is called the sensitive volume (V_s). In general, it is proportional to the technology feature sizes and the thickness of the sensitive layer.

Critical Charge

Charges produced by a crossing particle in V_s is then evacuated through the open channel of the second transistor. If the induced voltage is large enough, the state of the feed back transistor can be inverted and the whole cell changes its logic state. The minimum charge Q_{crit} necessary to cause an SEU is called the critical charge. In case of an SRAM cell, having an active feed-back loop, Q_{crit} is a function of the feed-back time (T_{fb}). T_{fb} is defined as the time necessary for an output change of one inverter to propagate back to its input via the feedback loop. SRAM irreversibly switches and the SEU occurs[73]. Q_{crit}, T_{fb} and the current induced during the transient period (I_{drain}) are linked with a simple relation:

$$Q_{crit} = \int_0^{T_{fb}} I_{drain} \mathrm{d}t \tag{7.1}$$

The critical charge is directly dependent on the struck node capacitance and the supply voltage. As both generally decrease from one technology generation to the next one, Q_{crit} decreases correspondingly. Figure 7.4 shows an example of the critical charge simulated evaluation by scaling the technology or decreasing the supply voltage[75].

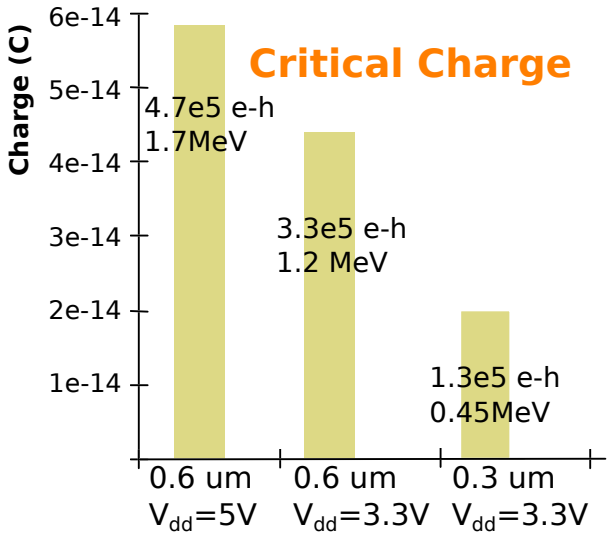


Figure 7.4: Simulation results giving the critical charge and the critical energy for 0.6 and 0.3 µm technologies. The critical energy is converted to a number of electron–hole pairs using the conversion factor of 3.6 eV per pair creation on average [73].

Energy Deposition ($\frac{dE}{dx}$) of a Particle

The stopping power, or the linear energy transfer (LET) of a charged particle is described by the Bethe-Bloch formula[76]. Figure 7.5 shows the dependency for several different types of particles in the MeV–GeV range. Units are scaled appropriate to the phenomenon, that is how many e–h pairs does a particle release by crossing one micrometre of silicon.

Which Particles Induce SEU?

By summarizing previously mentioned definitions, one can deduce the kind of particles capable to produce an SEU. For that let us look again to values of Figure 7.4. For a particle to be a direct cause of an SEU (i.e. for the particle itself and not its secondary products), two conditions should be fulfilled:

1. The energy of the particle should be high enough, so that it is capable to release Q_{crit}. That means it should be in the order of at least one MeV.
2. The stopping power of the particle should be high enough to release Q_{crit} within V_s.

In Figure 7.5 the first and the second conditions are represented by the yellow (vertical) and the blue (horizontal) regions respectively[1]. Their intersection region (green), where both of them are valid, belongs to heavy ions, starting from the α-particle and heavier.

[1] The limits for these two conditions vary with the technology in use

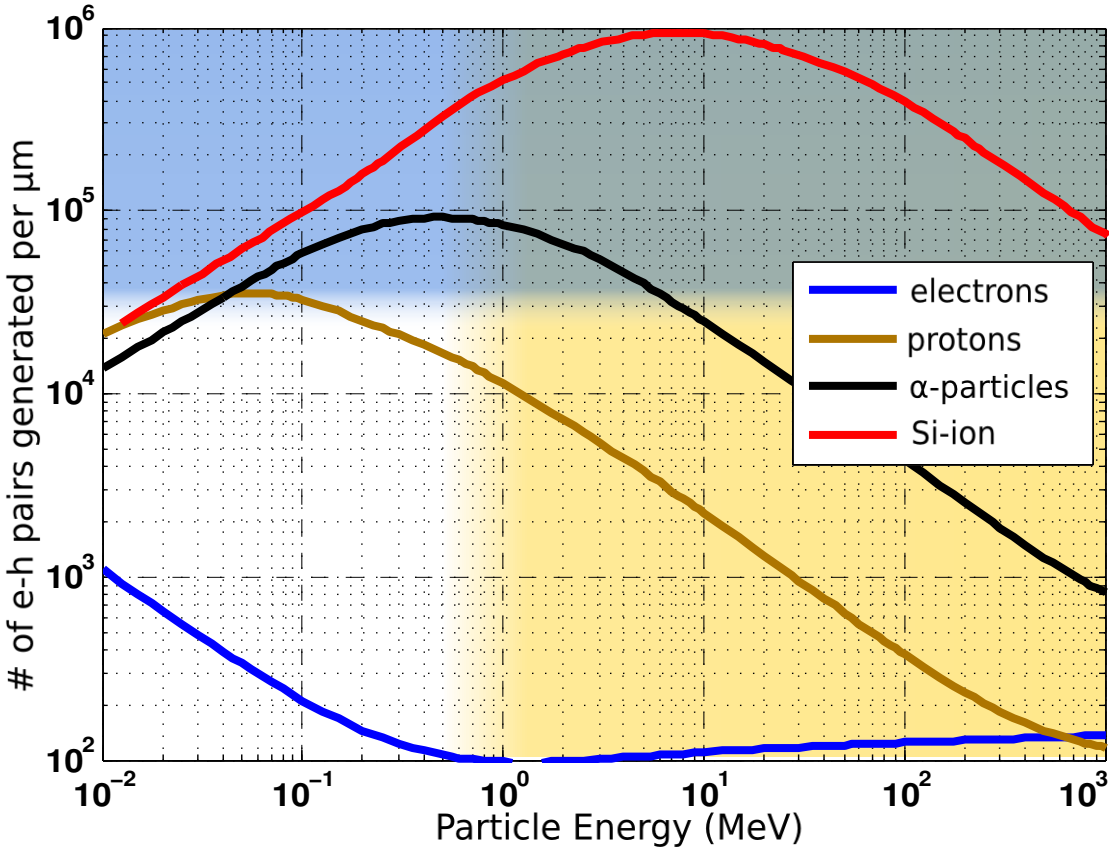


Figure 7.5: Ionization wake density of electron-hole pairs in silicon following the passage of various charge particles. An electron-hole pair is created for about every 3.6 eV of energy lost to target electrons by each particle. Yellow region: particles have enough energy to generate Q_{crit}. Blue region: the ionization densitiy of a particle is high enough to generate Q_{crit} within V_s [72].

Proton - Neutron Induced SEU

One can also see, that a proton does not directly induce an SEU[1]. The only energy region, where the proton LET is high enough to produce an SEU (to satisfy the second condition) , corresponds to the energy of about 100 keV. But with this energy a proton would produce only 30000 electron–hole pairs, too few for an SEU[2] (the first condition is not satisfied). A neutron, being a neutral particle, does not produce any ionizing radiation at all.

However, neutron and proton both represent an important source of SEUs due to the hadronic interaction between a nucleon (proton or neutron) and a nucleus. These interactions can be either elastic or inelastic. The former can be written as:

$$nucleon + target \longrightarrow nucleon + target \tag{7.2}$$

[1] The smaller the technology feature size is, the less true is this statement due to the decrease of the Q_{crit}. Due to the Landau tail of the energy deposition distribution, a non-zero probability always exists.
[2] for technology feature size ≥100 nm.

The latter can be written as:

$$nucleon + target \longrightarrow X_1 + X_2 + \ldots X_n + residual\ nucleus \tag{7.3}$$

where X_i is anything not heavier (or equal) than helium.

According to studies [77], elastic collisions bring relatively negligible contributions in SEU generation; inelastic ones are the main source of SEUs for protons and neutrons.

Proton and neutron can be considered as different isospin states of the same particle. This symmetry is exact if the nuclear interaction is the only one in operation. Indeed at high energies (for our application above 50 MeV [77]) both proton↔nucleus and neutron↔nucleus reactions are very similar. This is no longer true at low energies once the Coulomb repulsion starts to play an important role.

SEU Cross-section Definition and its Energy Dependency

The SEU is characterized by its cross-section, which is defined as the probability of a soft error per unit of irradiation fluence (Φ) per memory bit, in other words:

$$\sigma_{SEU} = \frac{N_{errors}}{\Phi \cdot N_{bits}} \qquad \left[cm^2\right] \tag{7.4}$$

The σ_{SEU} is energy dependent and it can be well fitted to a Weibull curve [78] having the following form:

$$\sigma_{SEU} = \sigma_0 \left(1 - e^{-\left(\frac{E_K - E_0}{W}\right)^s}\right) \tag{7.5}$$

where E_K is the kinetic energy of the projectile, W and s are shape parameters (s is typically around one), σ_0 is the plateau or saturation cross-section value and E_0 is the threshold (onset) energy of the projectile.

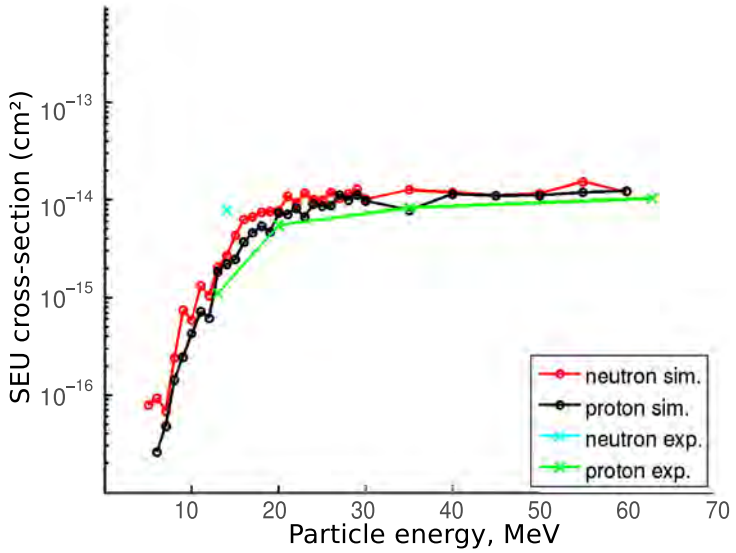


Figure 7.6: Experimental and simulation SEU neutron and proton cross-sections for a 250 nm technology node[79].

Proton and neutron cross-sections behave very similarly, especially in the saturation region; however, the neutron has a lower threshold energy E_0. A comparison between neutron and proton SEU for the 250 nm technology[79] is presented in Figure 7.6.

Soft Error Rate (SER)

It is easy to deduce the SER for the mono-energetic beam from Equation 7.4. It can be estimated as a product of the flux ($\dot{\Phi}$), the memory size and the SEU cross section:

$$SER = \sigma_{SEU}\dot{\Phi}N_{bit} \tag{7.6}$$

For a broad particle energy spectrum this formula is generalized as:

$$SER = N_{bit}\int \sigma_{SEU}(E)\frac{d\dot{\Phi}(E)}{dE}dE \tag{7.7}$$

To simplify Equation (7.7) one approximates the shape of the σ_{SEU} curve by a Heaviside function, meaning that particles below the threshold energy do not produce SEU, those with the energy higher than the threshold have an equal SEU probability. In this case Equation (7.7) can be simplified to:

$$SER = N_{bit}\sigma_0\dot{\Phi}(E > E_0) \tag{7.8}$$

7.3 PXD Related Background

The PXD will be installed very close to the Interaction Point of Belle II, r=14 mm for the internal layer and r=22 mm for the external one [14]. According to estimations of the Belle II Collaboration [80], summarized in Table 7.1, an average flux of 10^4 cm^{-2} neutrons is expected through the surface of each DHPT 1.0.

	Switcher-B		DCD		DHP	
	-Z	+Z	-Z	+Z	-Z	+Z
Touschek	1302	3197	2976	2380	2605	2306
Beam-Gas Coulomb	2133	473	744	0	1786	297
Radiative Bhabha	1420	6398	3869	2822	5060	5805
4-fermion final state QED	1049	1023	1543	1730	1859	2100
Total # neutrons per cm^{2}s	5904	11091	9132	6932	**11310**	**10508**

Table 7.1: Neutron background, total contribution from different sources given for the forward (+Z) and the backward (-Z) directions. Number of particles per cm^2·s. Data presented by A. Moll on Vienna Belle II SVD PXD meeting, February, 2012[80].

As will be discussed in the next paragraph, this kind of background is an SEU source whose risks have to be evaluated for the production chip DHPT 1.0 and, if necessary, mitigated.

Several results have been published concerning the 65 nm technology's SEU sensitivity evaluation [81, 82]. However, the result strongly depends on a particular implementation.

7.4 The DHPT 0.1 Chip

To measure the SEU cross-section (σ_{SEU}) of the 65 nm technology, a test chip DHPT 1.0 has been designed; its layout is depicted in Figure 7.7. The purpose of this test chip was to test all critical full-custom design elements to be included in the DHPT 1.0. As shown in Figure 7.7, it consists of four independent areas, three of them containing test structures for the future DHPT 1.0 chip.

For the SEU tests one of these reticles contained standard memory blocks that will be used for the digital design (bottom left part of the chip).

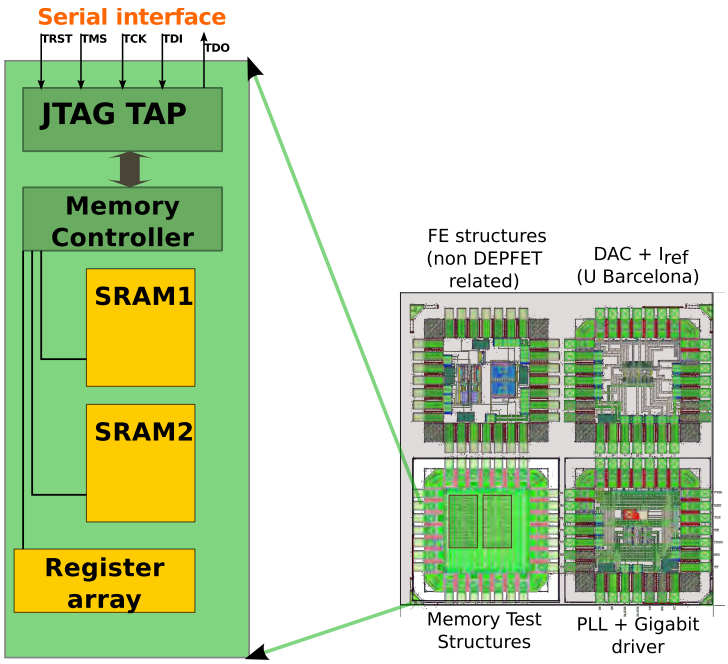


Figure 7.7: DHPT 0.1 test structures. One of four reticles in the chips is used for memory tests. The access to these memories is implemented using the standard JTAG [39] interface.

These structures are summarized in Table 7.2 on page 92. These are three different memory types: two different SRAM blocks and one generated register array.

7.4.1 Radiation Facility

The CERN irradiation facility IRRAD-I at the CERN PS East Hall[84] was used to test the σ_{SEU} of the digital blocks of DHPT 0.1. The DUT[1] was exposed to a 24 GeV/c proton beam (area 2x2 cm^2). Protons were arriving per bunches with variable fluencies from 2×10^9 p/cm^2 up to 9×10^9 p/cm^2. The full dosimetry information was provided by the facility.

Figure 7.8 demonstrates that for this energy the particle's stopping power is close to the stopping power of the Minimum Ionizing Particle. Furthermore, the σ_{SEU} for these energies reaches its saturation value as is shown in Figure 7.6.

[1] DUT stands for the Device Under Test

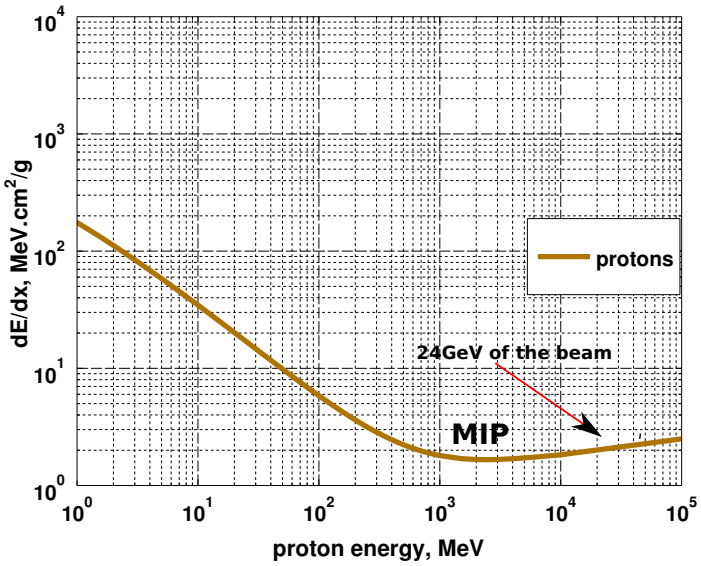


Figure 7.8: Proton stopping power for the Si. Data was taken from the PSTAR database [83]. The beam contained 24 GeV protons; at this energy σ_{SEU} reaches its saturation value.

7.4.2 Test Setup

The readout system is based on the same FPGA platform that is used to steer Hybrid-5 (Section 6.2). A small PCB was designed to mount and interconnect the DHPT 0.1 with the readout system. To drive the chip serial interface across 25 m distance the CMOS signals are converted to the LVDS using a dedicated radiation tolerant chip that was previously designed in our group. This chip was additionally mounted on the same PCB.

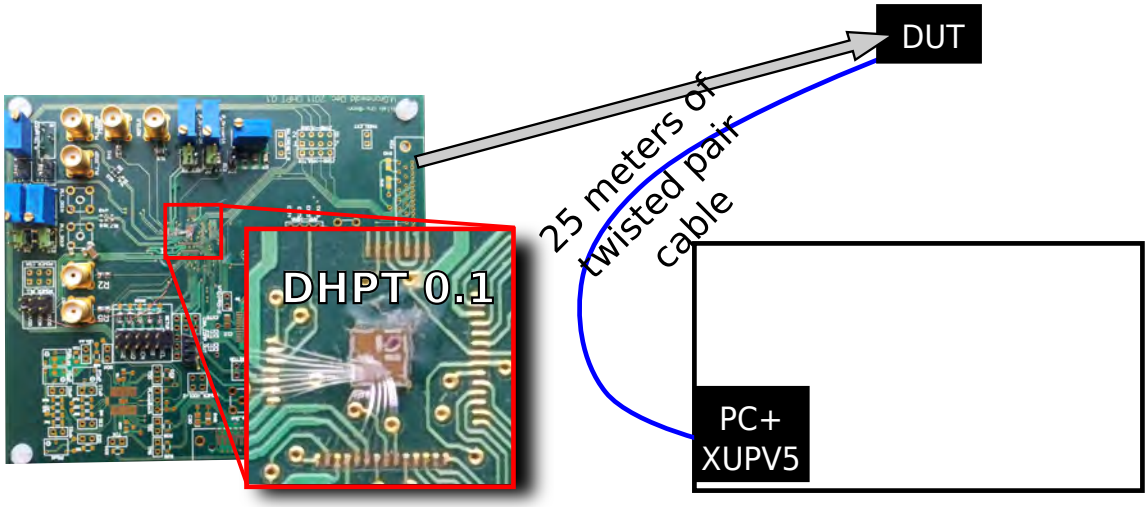


Figure 7.9: The system installed in CERN to measure the SEU cross-sections of the 65 nm CMOS technology that will be used to produce the DHPT 1.0 chip.

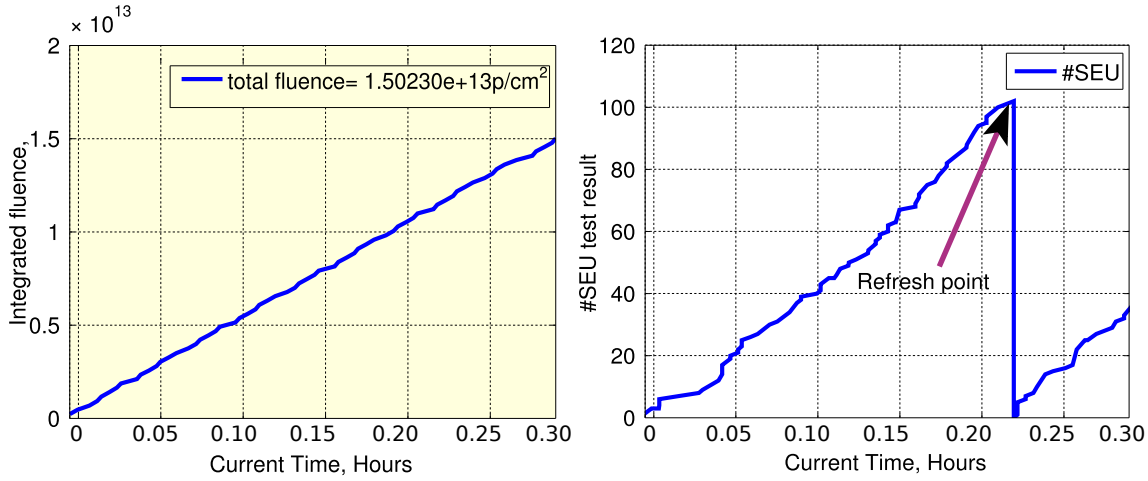


Figure 7.10: An example of the SEU counting procedure. The left-hand plot shows the fluence record for the same period of time. The right-hand plot shows the number of SEU events as a function of time. σ_{SEU} can be directly obtained by dividing the number of SEU events by the fluence and by the memory size. The resulting measurement error is assumed to be $^{\Delta\sigma_{SEU}}/_{\sigma_{SEU}} = {}^{1}/\sqrt{N}$.

The test setup was organized in a way drawn in Figure 7.9. The DUT was installed in the irradiation area. The power supply voltages and the LVDS signals were sent across 25 meters of twisted-pair cable. The FPGA board and the control PC were installed in the control room nearby. After the begin of the tests, the system was remotely steered from Bonn via SSH tunnel during 3 weeks of the irradiation campaign.

7.4.3 Measurements

Unless explicitely stated, in the tests presented below the alternating 101010-pattern was used. No pattern dependency of the σ_{SEU} was studied in the scope of these experiments.

Generated Register Array

First, the σ_{SEU} of the generated register array with a size of 32x72 bits was measured. It was relatively easy to do, since the direct measurement was possible. An example of such a measurement is presented in Figure 7.10. The measurement was a simple counting of the cumulated number of SEU events. In order to keep the double hit probability negligible, it was necessary to refresh the memory from time to time, a soon as it was reaching 100 events, which is 4% of the memory. This way the probability of double hits was kept below 0.2%. The end result is summarized in Table 7.2.

SRAM Memories

It was not possible to directly measure σ_{SEU} for the two SRAM memories due to the following reasons:

- The total σ_{SEU} of the SRAM was measured to be approximately ten times higher than the σ_{SEU} of the register array. Hence the memory refresh rate for this test has to be correspondingly higher; one beam spill is enough to create many SEU events, even too many for certain regions of the SRAM memory, particularly sensitive to the SEUs. The memory has to be refreshed after each spill.

- The readout of the DUT is slow, it takes more than one minute to refresh the memory and transmit all data to the PC. Since the readout time and the time between spills are more or less the same, one cannot be sure that after each memory dump there is only one spill and not more. It can be even possible that one part of the memory is affected only by one spill while another part, due to the long dump time, can be affected by the next coming beam spill.
- σ_{SEU} appeared to be inhomogeneous, as presented in Figure 7.11. In high sensitivity regions the probability of the multiple hits was not negligible, this fact had to be taken into account and corrected.

Because of these reasons it was impossible to get directly the SRAM σ_{SEU} by simply measuring the number of errors and dividing it by the measured fluence.

Instead, the high statistics was gathered of how many errors per each memory cell was registered. Dividing by total number of spills, the probability for each particular bit to register an SEU per single spill was calculated as presented in Figure 7.11. To correct for multiple hits, Equation (7.9) was used. The proof of this Equation is presented in Appendix D.

$$E_r = -\frac{1}{2}N \cdot ln\left(1 - 2 \cdot \frac{E_m}{N}\right) \tag{7.9}$$

Here N is the memory size, E_m is the measured number of soft errors and E_r is the expected number of soft errors that actually happened.

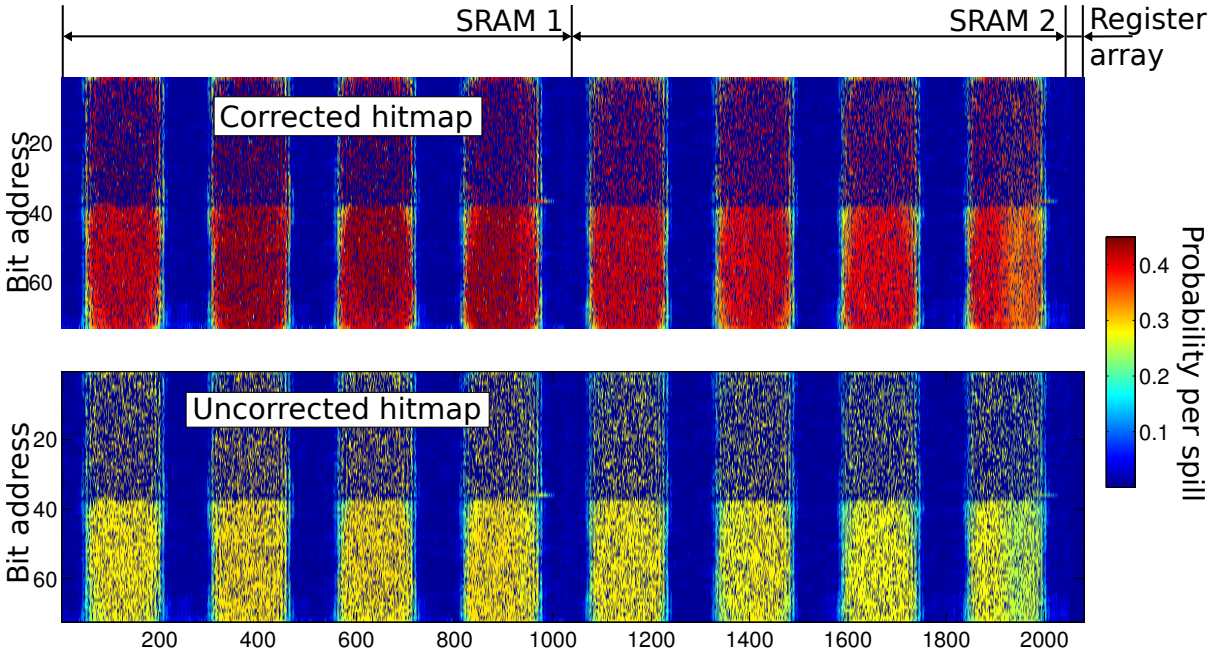


Figure 7.11: DUT color-coded hitmap, representing the SEU probability per spill. The DUT consisted of three memories, each of them being an array of 72-bit words. The word address is plotted on the horizontal axis, the bit address in the word is plotted vertically. Addresses 0-1023 correspond to the SRAM 1, 1024-2047 to the SRAM 2 and 2048-2080 to the generated register array. The lower image represents the apparent hit probability, the upper one represents the estimated real hit probability, corrected for multiple hit events using Equation 7.9

Further, average probabilities for an SEU per cell for these three types of memories were calculated.

Finally, from the definition of the σ_{SEU} follows that these probabilities are proportional to their respective cross-sections:

$$\frac{p_1}{p_2} = \frac{\sigma_1}{\sigma_2} \tag{7.10}$$

Knowing the value of the SEU cross-section for the generated register array, it is possible to calculate the other two cross-sections, their values are summarized in Table 7.2.

The observed σ_{SEU} regular pattern shows about ~10 factor of the difference in sensitivities between corresponding maximum and minimum values. Since we do not possess the physical layouts of these memories due to legal reasons, we can not evaluate this. Instead a simple average for all values was taken.

	Type	**Size, bits**	**Cell area,um^2**	**SEU cross section for irradiation doses below 20 Mrad**
1.	SRAM 1, TS1N65LPA	1024 x 72	1.05×0.5	σ_{SRAM_1}=(0.89 ± 0.1) · 10^{-13} cm^2
2.	SRAM 2, TS1N65LPLL	1024 x 72	1.05×0.5	σ_{SRAM_2}=(0.97 ± 0.1) · 10^{-13} cm^2
3.	Generated register array	32 x 72	5.2×1.8	σ_{FF}=(0.64 ± 0.06) · 10^{-14} cm^2

Table 7.2: Summary results of the memories under tests. The values for the presented sections are valid for total ionizing doses below 20 Mrad

7.5 Further Results

The above presented results were obtained from the data acquired during the first day of the SEU test campaign, corresponding to 20 Mrad of irradiation. The total experiment duration was 21 days with the total acquired dose of more than 600 Mrad.

The results beyond 20 Mrad, though not being directly relevant for Belle II, are still interesting, here is their summary:

- With this radiation campaign we confirm that the TSMC 65 nm technology is well suited for high radiation dose, indeed, the DUT has shown a stable behavior up to 300 Mrad (Beyond that dose an effect of "sticky bits" is observed, i.e. some bits started to hold their values regardless attempts to overwrite them). This ionization dose is more than enough for Belle II application.

- When the experiment was repeated one week later after the start, a non negligible drift of the σ_{SEU} was observed. The high-contrast structure, presented in Figure 7.11 has evolved and became more homogeneous. See Figure 7.14.

- While describing the SRAM cell in the paragraph 7.2 it was mentioned that the cell design is symmetric, hence the cross-sections $\sigma_{SEU}(1 \rightarrow 0)$ and $\sigma_{SEU}(0 \rightarrow 1)$ should be equal. In the beginning of the irradiation campaign an additional test to verify this statement was done.

  It was based on the principle that if both cross-sections are exactly the same, then, by exposing the memory of size N (where N is large) long enough to the radiation, one would obtain equal number of ones and zeros randomly distributed. In Figure 7.13 the result of the described test is graphically presented. As it turns out, a slight asymmetry in favor of 0->1 transitions was observed.

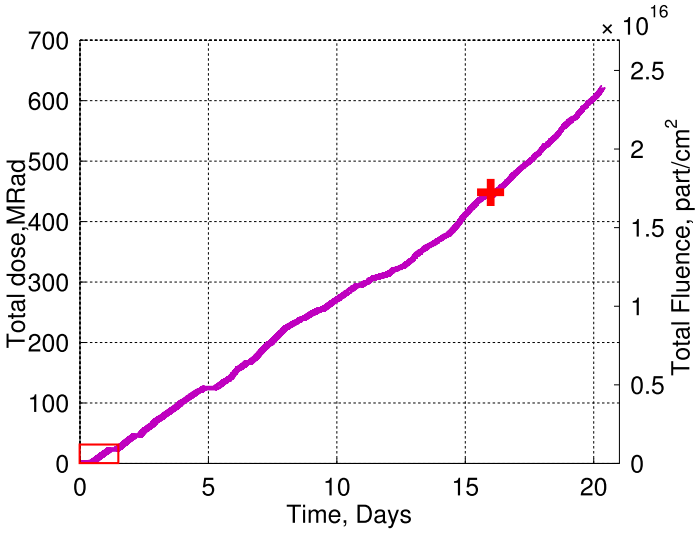


Figure 7.12: The radiation dose log for the whole duration of the experiment. The DUT was irradiated with 600 Mrad during three weeks. After ten days, or approximately 300 Mrad, the DUT showed unstable behaviour. After 16 days, or more than 400 Rad, the connection to the DUT was lost; this is indicated by the cross in the drawing. The Belle II specified dose of 20 Mrad was acquired during the first day and represented by a red rectangle in the figure.

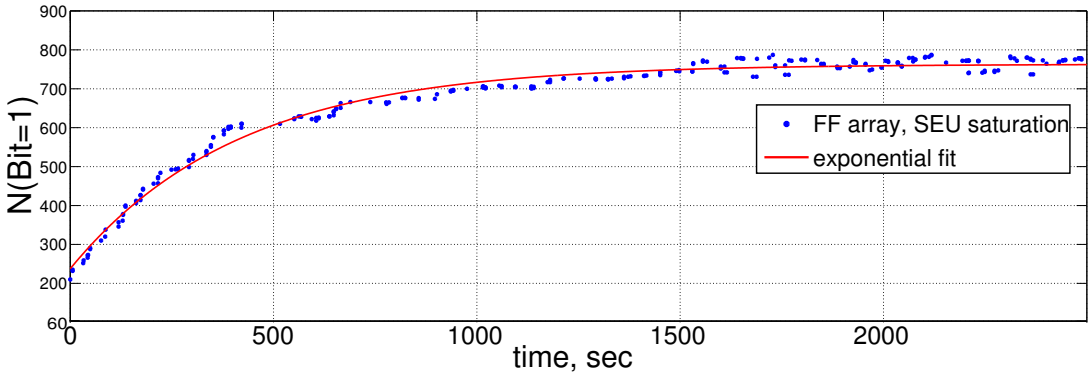


Figure 7.13: SEU symmetry test. The generated register array (type 3 memory, see Table 7.2) of 1440 bits were initially filled with zeros. If one waits long enough, an equilibrium between 1->0 and 0->1 transitions can be observed (here after ≈25 mins). Equilibrium number of ones is estimated to be 764±4 (expected value is 720).

7.6 DHP Sensitivity to SEU

To estimate the SER of the DHPT 1.0, the above mentioned estimated fluence and measured cross-sectioned are used. However, some further assumptions are necessary:

- The exact structure and the total amount of the memory of the DHPT 1.0 are not known yet. However, we do not foresee drastic changes in the chip design in comparison with the current test chip DHP 0.2, therefore, it is a good choice to take the same memory sizes.
- As previously discussed, a non-negligible difference between the SEU cross-section of neutrons and the SEU cross-section of protons can be observed only in their low-energy tails due to the absence of the Coulomb barrier. Even though the neutron energy spectrum was simulated by the

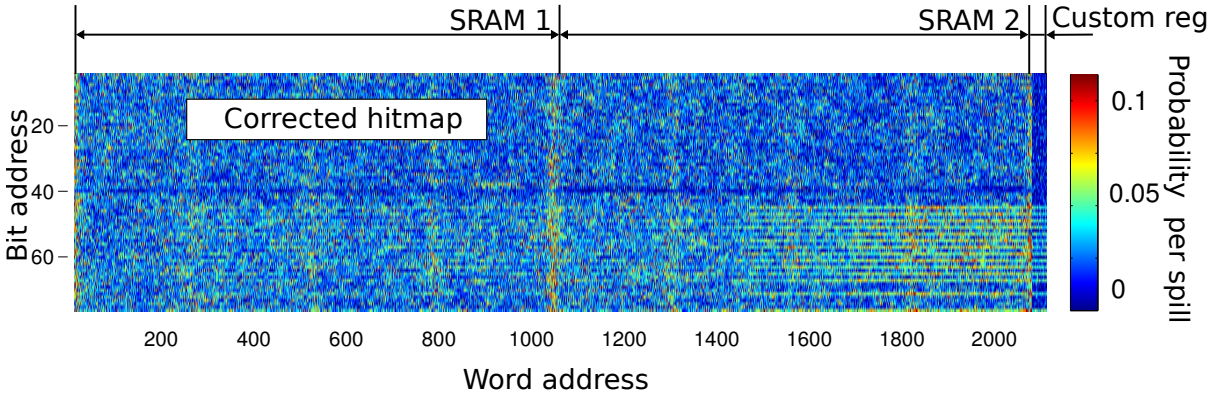


Figure 7.14: The SEU probability per spill map after about 300 Mrad of irradiation. The SRAM σ_{SEU} increased by 30 % and became more homogeneous; the initial wavy structure of the σ_{SEU} shown in Figure 7.11 is now hardly seen. The σ_{SEU} of the custom register array has increased by a factor of ten and comparable to the σ_{SEU} of the SRAM memories.

Belle II Collaboration, the σ_{SEU} threshold energy is still unknown. Hence only an upper bound estimation of the SER value can be done. Here we assume that all particles having the same σ_{SEU} equal to its saturation value.

- σ_{SEU} of the block memories to be used in the chip are taken to be equal to σ_{SRAM_2}, which is the worst case scenario; all data processing logic cross-sections are taken equal to σ_{FF} (FF stands for Flip-Flop).

The summary of the SER estimations are presented in Table 7.3:

Memory type	**Size**	**SEU cross section**	**Mean time between SEU**	**Refresh rate**	**Mitigation measures**
Pedestals	0.5 Mbit	σ_{SRAM_2}	30 min	15 min[1]	Hamming code protection
Raw data buffer	0.5 Mbit	σ_{SRAM_2}	30 min	20 us[2]	
Data processing logic	45 kbit	σ_{FF}	4 days	20 us	
Configuration Register	368 bit	σ_{FF}	490 days	1 day	Triple redundancy

Table 7.3: DHP SER assuming the measured cross sections. SER was calculated using Equation 7.7. The mean time between SEU is the inverse of the SER.

There are two relatively large memories in the DHPT 1.0: the pedestals and the data memory, both of the same size of 0.5 Mbit. However, an SEU event does not have the same importance depending where it takes place: if an SEU corrupts the data memory, this will be simply interpreted as an additional background event[1], later filtered out by the track recognition system.

[1] This estimation was done by Slow Control group of the Belle II Collaboration [47]
[2] 20 us is the time necessary to process one PXD frame, running at 50 kHz.
[3] it can be also generate a data loss, but the probability of this is even smaller taking into account that the estimated hit occupancy is less than 3%

An SEU in the pedestal memory is more disturbing: in case of the pedestal value decrease it is interpreted as a hot pixel. On the other hand, if an SEU event increases the pedestal value, this results in a possible data loss: there is a chance of not detecting an event if its value is too small and one subtracts the wrong pedestal. Fortunately these errors will not accumulate, since the memory update period is estimated to be in the order of 15 min, as summarized in Table 7.3. Nevertheless, the impact of these events should be minimized. For that reason an automatic error correction module based on the Hamming code algorithm has been implemented on-chip, correcting SEU on each memory access, i.e. each 20 us. This is a standard technique in electronics industry, for example used in ECC RAMs for servers or super computers, where long-term stability and data integrity are important. Its advantage is that it needs quite little overhead for implementation (additional 8 parity bits for 64 bit words resulting in total length of 72 bits). It is able to correct for single SEU and detect double SEU event. To track them an SEU counter is also implemented on-chip. This algorithm is presented in presented in Appendix E.

Along with large data buffers, there is also the data processing logic in the chip (row 3 in Table 7.3). Due to small sizes and small cross-section the SER is low. Moreover, as in case of the data memory, these errors do not cumulate and are not so critical. Hence no error correction techniques have been implemented.

The last and quite important part of the chip is its configuration registers (the last row of Table 7.3). Though having extremely low SER, it is still very important to protect the chip from them. A possible SEU can stop the acquisition and the system reboot would be necessary. To maximally protect the configuration register from this risk, each of its bits has been protected by the triple redundancy technique[5], which further reduces the effective σ_{SEU}.

These studies were compared with other publications [81, 85] where the SEU cross-sections for CMOS 65 nm technology were estimated, the results were found to be coherent, as presented in Table 7.4. The work was presented in the TWEPP-2012, Oxford, UK, and summarized in its proceedings[13].

Comparison of CMOS 65 nm SEU cross-sections from different sources (cm/bit)

Xilinx Reliability Report [81]	Commercial SRAM [85]	**My Results**
Configuration Memory:		Generated Register Array
σ_{SEU} =0.67 × 10^{-14}		σ_{SEU} =0.64 × 10^{-14}
SRAM:	SRAM (all DUT confused):	SRAM:
σ_{SEU} =0.4 × 10^{-13}	σ_{SEU} =0.2 – 0.9 × 10^{-13}	σ_{SEU} =0.97 × 10^{-13}

Table 7.4: SEU comparisons with other publications. As one sees, depending on implementation the σ_{SEU} can vary; however, these results look coherent.

[5] This means that each bit is repeated tree times. In case of a singe SEU affecting only one bit, it is immediately corrected by the majority vote feed-back logic. The triple redundancy correction is a particular case of the Hamming code (3, 1). This does not unfortunately correct for double SEU but this can be minimized by setting a large spatial distance between each of triplicated bits, thus excluding double SEU event having a single particle as a source.

Chapter 8

Conclusions

8.1 Summary

The Belle II experiment, which will start after 2015 at the SuperKEKB accelerator in Japan, will focus on the precision measurement of the **CP**-violation mechanism and on the search for physics beyond the Standard Model. In order to gain significantly more statistics than have been gathered to date, the newly upgraded accelerator will provide an unprecedented luminosity of $8{\times}10^{35}$ cm^{-2}s^{-1}, which is about 40 times greater than that of its predecessor.

A new detection system capable of coping with considerably increased background is required. Moreover, one of the main challenges of the present upgrade is reconstruction of B_0 and K_S vertices with a precision in the order of 10 μm. To address this challenge, a pixel detector based on DEPFET technology has been proposed. Its excellent spatial resolution (in the order of several microns) and low material budget was one of the decisive factors determining the choice of this technology for the first time.

The DEPFET Pixel Vertex Detector is a complex system consisting of 40 modules arranged across two layers. Each module has a sensitive area, which is thinned down to 75 μm and steered with three types of ASICs: Switcher, Drain Current Digitizer (DCD) and Data Handling Processor (DHP). Switcher chips are designed to steer the pixel matrix of the sensitive area. The DCD chips digitize the drain current coming from the pixels. All ASICs will be directly bump-bonded to the balcony of the all-silicon DEPFET module. The total amount of raw data generated by the detector is equal to ~3 Tbps, which is not transportable to the back-end electronics due to mechanical and electrical constraints. Therefore, in-PXD zero suppression is needed.

The third ASIC, the Data Handling Processor, has been developed using CMOS 90 nm technology. It is designed to steer the readout process by sending the control signal to DCD and Switcher chips. Secondly, it is responsible for data zero suppression, which is done by performing common mode correction, pedestal subtraction and signal detection [1]. The output is sent upon the trigger arrival for further data reduction. The zero suppressed data is transmitted by DHP to the back-end electronics over a 15 m long electrical output link with a rate of 1.6 Gbps using 8b/10b encoding. Several custom designed blocks, such as PLL, DAC, ADC, CML transmitter have been designed for this chip. The high performance constraints for digital data processing make the DHP implementation extremely challenging. Several conceptual solutions for the digital data processing blocks were proposed and implemented. Currently, three chip prototypes have been produced and one has been submitted. The second chip iteration, the DHP 0.2 submitted in 2011, is the first full-size chip version.

The scope of this thesis covers DHP tests and optimization as well the development of its test environment, which is the first Full-Scale Module Prototype of the DEPFET Pixel Vertex detector. The work consisted of several steps:

[1] By means of threshold comparison.

- **The proof of concept** was carried out by implementing the DHP data processing core as a virtual entity in the steering FPGA of the existing DEPFET test system. The findings obtained were used in the chip design that was launched in parallel.
- **The design optimization**. To meet the design requirements, several chip models were used to optimize the design. Moreover, the chip design verification entailed a significant amount of work, which preceded the chip submission.
- **The test system development**. A new test system, called the Full-Scale Module Prototype (FSMP), was developed. One of its main elements is the presence of the DHP 0.2 chip. The FSMP is the very first PXD prototype containing all the elements necessary to run the DEPFET detector.
- **System tests.** The results presented in this thesis showed that the FSMP performance exceeded the expectations, producing very good results in first laboratory tests, as well as in the latest Test Beam campaign at DESY 2013.
- **SEU technology evaluation**. The PXD will be installed in a harsh radiation environment. In these kinds of conditions the digital electronics are prone to soft errors, mainly known as Single Event Upsets. To verify that the DHP can function in this environment, this problem has been evaluated and its risks have been mitigated.

8.2 Outlook

The latest feasibility proof of the PXD concept was shown during the DESY Test Beam campaign in May 2013. The FSMP has demonstrated that it works correctly on the system level. The full acquisition chain, including all ASICs designed to run the detector, has been tested.

The next step is to scale the existing FSMP by increasing the number of ASICS simultaneously present in the test system. This will reproduce the situation expected in the final design.

The so-called Electrical Multi-Chip Module (EMCM) has been developed for this purpose and is currently being assembled (Figure 8.1). It is anticipated that the existing readout chain will be reused for EMCM tests.

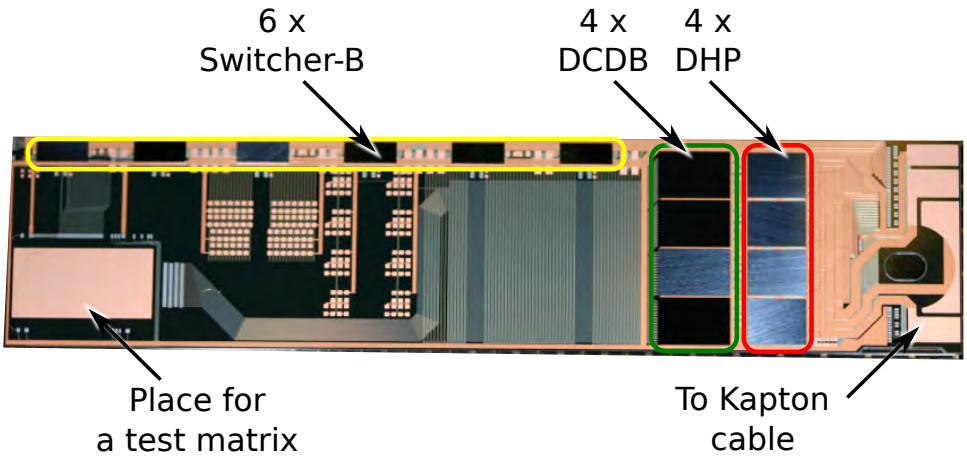


Figure 8.1: One of the EMCM modules, which is being prepared for further tests. All ASICs have already been mounted (6 x Switcher-B, 4 x DCDB, 4 x DHP 0.2). EMCM is electrically equivalent to a PXD module, however without a sensitive area.

The first production version of the DHP chip, called the DHPT 1.0 has been submitted and is expected to be delivered by the end of 2013. This chip will be an improved and corrected version of the DHP 0.2, implemented in new CMOS 65 nm technology. It is expected to be the production chip version, suitable for assembly on PXD modules.

The test systems are becoming increasingly complex and scaling problems now present the main challenges. The DEPFET collaboration has been doing its utmost to deliver the PXD on time.

Appendix A

Offline Correction for the Simple Average Estimation of the Common Mode

The $\overline{S_r}$ defined in Equation 4.5 can be rewritten as:

$$\overline{S_r} = \frac{\sum_{c=0}^{N-1} S_{cr}}{N} = \frac{\sum_{c=0}^{N-1} S_{cr}}{k_r} \times \frac{k_r}{N} = S_r p_r \tag{A.1}$$

where k_r is the number of non-zero signals in the row r, the $p_r = k_r/N$ is the corresponding local occupancy and the S_r is the average of all non-zero signals present in this row r.

Then Equation 4.6 can be rewritten as:

$$\widetilde{S_{cr}} = S_{cr} - \overline{S_r} = S_{cr} - S_r p_r \tag{A.2}$$

Taking the average of this relation over all non-zero signals, we get:

$$\widetilde{S_r} = S_r - p_r S_r = S_r(1 - p_r)$$

Hence, using Equations A.1 and A.2 we get the final result how to correct for the bias introduced by the Simple Average CM estimator:

$$S_{cr} = \widetilde{S_{cr}} + \frac{p_r}{1 - p_r}\widetilde{S_r}$$

Appendix B

Online Pedestals Update

The pedestal monitoring can be done on-chip to speed-up the pedestal update. However, it is preferable to avoid complex solutions due to the limited amount of intelligence that one can implement in a small ASIC. A way to do it is a simple recursive algorithm, which does not need many processing resources. It proceeds as follows: if a signal inside the pixel is not found, then its new pedestal value P_{t+1} should be updated using the current input P_t and the previous pedestal P_{t-1} according to the Equation:

$$P_{t+1} = \frac{(2^n - 1)P_{t-1} + P_t}{2^n}, \tag{B.1}$$

where n is a memory factor, i.e. it characterizes how long the memory would keep traces of old values (higher n makes the output result smoother, similar to the sliding average operation).

After rewriting Equation B.1 as $P_{t+1} = P_{t-1} + 2^{-n}(P_t - P_{t-1})$ it is clear that one of the simplest hardware implementation can be achieved in a straightforward manner using basic bit operations:

$$P_{t+1} = P_{t-1} + [(P_t - P_{t-1}) >> n], \tag{B.2}$$

where the >> stands for the bit-wise right shift operator.

This kind of operational mode has been successfully implemented in the FPGA emulation of the DHP (Chapter 6.1). However, the following arguments have been raised against this option during the digital design development.

1. **Memory factor**. Equation B.2 illustrates that to correctly perform the operation, one would additionally need n bits per pixel[1]. For example, designing the update mode $P_{t+1} = \frac{15P_{t-1}+P_t}{16}$ or having n=4 with nominal storage of 8 bits per pixel, this translates into 50% memory increase; having limited area resources this would imply an excessive need of extra area.
2. **Pedestal monitoring**. The update of the pedestal map on the hardware level implies no knowledge about the exact pedestal values by the DAQ. That means it is not possible to reconstruct the raw data values, which can be very useful to do some additional post-processing, such as corrections for the digitization non-linearities[2].
3. **Slow variations**. According to the latest prototype tests, the initial fear of having relatively fast pedestal variations does not confirm. One expects rather stable values on the time scale of hours.

For these reasons, it is planned to keep the chip's design simple and to update the pedestals values offline.

[1] Otherwise all variations smaller than 2^n would be ignored.
[2] One of solutions would be to periodically dump the memory content to keep track of pedestals' time variations.

Appendix C

DHP Random Triggering

During the PXD readout with estimated frame rate if F_{fr}=50 kHz, the random triggering with a maximum rate of F_{tr}=30 kHz will be used, if the next triggers starts before the end of the previous one, they are considered as one long trigger. For efficiency estimation purposes, it is important to know, how much time in percent does such scenario result.

To solve this problem, it is easier to work with time variables:

- Trigger length $T = 1/F_{fr}$
- Average distance between triggers $\tau = 1/F_{tr}$

The probability that two consecutive triggers intersect is:

$$P_i = \int_0^T \frac{1}{\tau} e^{-t/\tau} dt = 1 - e^{-T/\tau} \tag{C.1}$$

accordingly, the probability that two triggers do not intersect is the complementary case:

$$P_n = \bar{P}_i = e^{-T/\tau} \tag{C.2}$$

For the scenario, where two triggers intersect, the average length of the first trigger, before the next starts, is equal to:

$$\langle L_i \rangle = \frac{\int_0^T \frac{1}{\tau} t e^{-t/\tau} dt}{\int_0^T \frac{1}{\tau} e^{-t/\tau} dt} = \ldots = \tau \frac{1 - (1 + T/\tau) e^{-T/\tau}}{1 - e^{-T/\tau}} \tag{C.3}$$

otherwise, if triggers did not intersect, the first trigger lasts the time T.

Let us suppose we triggered N times (N is a large number). This lasts on average τN seconds. Among all these triggers $N \cdot P_i$ of them will intersect, $N \cdot P_n$ will not. The total average length of all triggers then equals to:

$$L_T = N \langle L_i \rangle P_i + NTP_n = \ldots = \tau N (1 - e^{-T/\tau}) \tag{C.4}$$

By dividing the result by τN, this gives the required busy factor:

$$BF = \frac{L_T}{N\tau} = 1 - e^{-T/\tau} = 1 - e^{-F_{tr}/F_{fr}} \tag{C.5}$$

Appendix D

SEU Multiple Hit Estimation

To measure the SEU rate one counts the number of loaded bit-pattern errors in the memory array under test (the DUT) while exposing it to a particle beam. If the exposure time of the DUT is long, some memory cells will undergo double or even triple SEU. However, in reality one observes only those errors whose positions have an uneven SEU multiplicity: meaning that they were hit 1, 3, 5 or more times(Figure D.1).

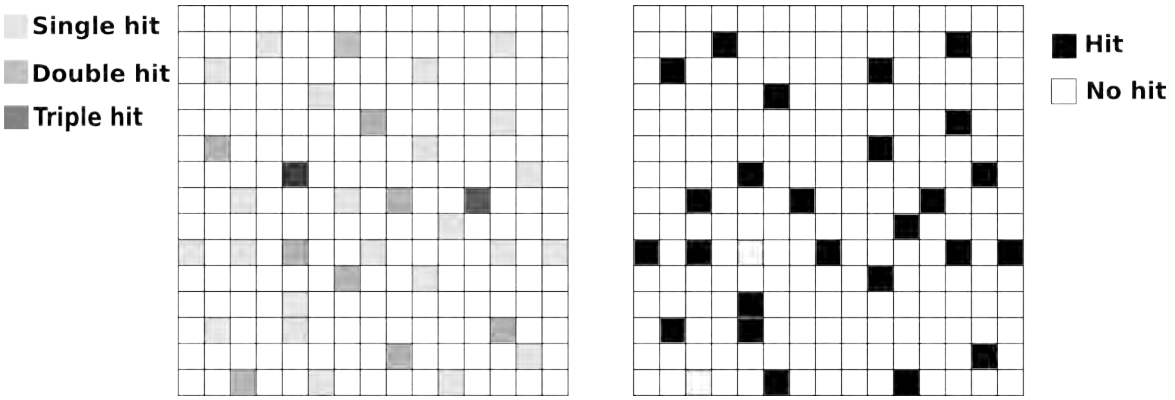


Figure D.1: Both grids represent the same memory array. Each memory cell can be hit by an incoming beam with a certain probability. If this probability is not small, some cells can be hit multiple times (left-hand figure). If each hit inverts a cell value, only the uneven hit multiplicity can be observed, i.e. 1, 3, 5 etc. times (right-hand figure).

In this case the measured SEU cross-section is smaller than the real one, since double, quadriple etc. SEU are unseen. In this appendix the necessary correction of how to estimate the real number of errors from the measured ones will be given.

Let us assume that each cell was hit λ times on average (for our case λ is a product of σ_{SEU} and the beam fluence $\dot{\Phi}$; it is, however, not important for the final result).

Knowing that each SEU event is independent and equiprobable in space and time, the probability that each memory cell will be hit k times follows the Poisson distribution[86].

$$p(\lambda, k) = \frac{\lambda^k e^{-\lambda}}{k!} \tag{D.1}$$

An even number of inversions does not change the initial value, an odd inversion multiplicity is seen as a single inversion.

$$\begin{aligned} p(No\ Error) &= p(\lambda,0) + p(\lambda,2) + p(\lambda,4) + ... \\ &= e^{-\lambda}\sum_{k=0}^{\infty}\frac{\lambda^{2k}}{(2k)!} = e^{-\lambda}\frac{e^{\lambda}+e^{-\lambda}}{2} = \frac{1}{2}\left(1+e^{-2\lambda}\right) \end{aligned} \tag{D.2}$$

Hence,

$$p(Error) = 1 - p(No\ Error) = \frac{1}{2}\left(1-e^{-2\lambda}\right) \tag{D.3}$$

Let N be the number of cells in the array (N is large), E_m be the measured number of events and E_r the real number of events. From the definition of λ and $p(Error)$:

$$\begin{aligned} \lambda &= \frac{E_r}{N} \\ p(Error) &= \frac{E_m}{N} \end{aligned} \tag{D.4}$$

From Equations D.2 and D.4 one gets the result how to obtain the real hit number from the measured one:

$$\frac{E_r}{N} = -\frac{1}{2}ln\left(1-2\frac{E_m}{N}\right) \tag{D.5}$$

For Equation (D.5) to make sense, the logarithm argument must belong to interval (0,1]. This is true if $E_m \in \left[0, \frac{N}{2}\right)$. Indeed, the number of measured errors is always less than $\frac{N}{2}$ within statistical fluctuation region. At very high exposures, then the initial information is overwritten, one expects on average exactly a half of bits to be wrong. However, for such extreme cases this equation is unreliable.

To visualize Equation (D.5), the plot of the $\frac{E_r}{N}$ as a function of the $\frac{E_m}{N}$ is drawn in Figure D.2.

From this figure one can see that if the hit occupancy (the ratio between the number of hits and the total cell number) is lower than 10%, there is only a small difference between measured and real event numbers, when at 40% there is a difference of a factor of 2. It is explained by the fact that at 10% occupancy there is a relatively small probability of hit overlay.

One can also observe that the ratio $\frac{E_m}{N}$ asymptotically tends to 50% while $\frac{E_r}{N}$ tends to infinity.

$$\lim_{\frac{E_r}{N}\to\infty}\frac{E_m}{N} = 0.5 \tag{D.6}$$

As previously said, at high exposures each cell gets a random value after many hits. As result, the average measured error number reaches its limit value of 50%.

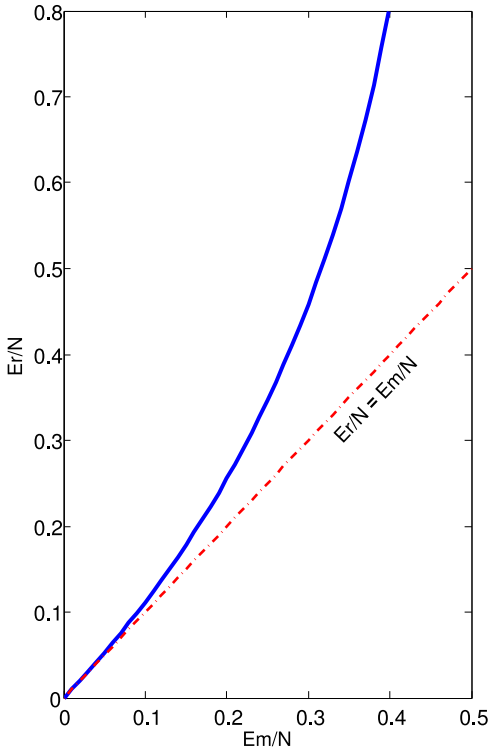


Figure D.2: Real number of hits as a function of the measured ones. Both quantities normalized to the total number of cells in the memory array.

Appendix E

SEU Error Correction Using Hamming Code

In this appendix an example of the extended (8,4)*-Hamming Error Correcting Code (ECC) will be explained. This method is short to describe and is easily extensible for longer word lengths. For example, the (72,64)-Hamming error correcting technique is used in the DHPT 1.0 chip to correct for single error and detect double error. In literature [87] this kind of ECC is called SECDED codes = Single Error Correction, Double Error Detection.

For this example, an 8-bit long Hamming encoded word consists of 4 data and 4 parity bits:

- Three parity bits p_0, p_1 and p_2. They are placed in positions 1, 2, 4†. (see Table E.1).
- One total parity bit p_t stays in position zero.
- All others are data bits

Bit position	Dec:	0	1	2	3	4	5	6	7
	Bin:	000	001	010	011	100	101	110	111
p_2:									
p_1:									
p_0:									
Total parity p_t:									

Table E.1: The (8,4)-Hamming error correction code. In the first row the positions of the parity bits are painted in blue. The data bits are painted in gray. In the second row is shown how to calculate p_2: it is a parity of data bits from positions 5, 6 and 7. In two next rows it is shown how the next two parity bits p_1 and p_0 are calculated. Finally, the parity p_t can be calculated as the total parity of all data and parity bits.

- p_2 is the parity of those bits, whose index binary representations have their third bit equal to one, i.e. 5, 6 and 7.
- p_1 is the parity of those bits, whose index binary representations have their second bit equal to one, i.e. 3, 6, 7.
- p_0 is the parity of those bits, whose index binary representations have their first bit equal to one, i.e. 3, 5, 7.

In general, each parity bit p_i calculates the parity of other bits with indexes j whose bitwise $AND(i, j) > 0$ (Equation E.1)

* Eight is the total message length, four is the number of data bits.

† In general for an N-bit long Hamming encoded word containing K parity bits $\{p_0, p_1, \ldots p_{k-1}\}$ their positions are: $\{2^0, 2^1, \ldots 2^{k-1}\}$.

$$p_i = \sum_{j>i} b_j \times [AND(i,j) > 0] \tag{E.1}$$

The total parity bit p_t is calculated at the end, then all other three bit p_1, p_2, and p_3 are known. p_t is equal to the total parity of all other bits in the word, including previously calculated p_1, p_2, etc. (Equation E.2)

$$p_t = \sum_{j\geq 1} b_j \tag{E.2}$$

The structure of this technique, it is summarized in Table E.1.

Let us consider an example of data transmission between two users. The sender (Alice) sends a message to the receiver (Bob). Upon the reception, Bob extracts the parities from the message:

$$P = \{p_t, p_0, p_1, p_2\}$$

To be sure that there is no error, he calculates his own parities based on the received data:

$$\widetilde{P} = \{\widetilde{p_t}, \widetilde{p_0}, \widetilde{p_1}, \widetilde{p_2}\}$$

Then, he constructs the following vectors that he calls the *syndrome*:

$$S = XOR(P, \widetilde{P})$$

It is evident that in absence of an error Bob will find the syndrome value equal to zero, since the exclusive or operation (XOR) of two same values is zero.

In case of one error Bob will see the syndrome looking like that:

$$S = \{1, x_1, x_2, x_3\}$$

The first bit is one because only one bit-flip changes the total parity. The vector $X = \{x_1, x_2, x_3\}$ will indicate the position, where the error took place. Indeed, let us for example suppose that on the bit position number three was an error. One sees from the Table E.1 that the syndrome will be equal to: $S = \{1, 011\}$, the 011 is the binary representation of three, which is the error's correct position.

In case of two errors the syndrome looks like this:

$$S = \{0, X\}, X \neq 0$$

The first bit of the syndrome equals zero since the total parity flips twice and returns to the initial value. One would be able to state, that two errors (or more) took place, however not knowing where, so Bob will have to either drop the message or ask Alice to resend it.

This algorithm is easily extensible for longer words. For example (16,11)-Hamming code with the word length of 16. Here one would need one additional parity bit p_4, with the total number of parity bits equal to 5 and 11 data bits.

Bibliography

[1] M. Gardner, *New mathematical diversions from Scientific American*, A Fireside book, Simon and Schuster, 1966.

[2] E. Noether, 'Invariante Variationsprobleme', *Nachr. d. König. Gesellsch. d. Wiss. zu Göttingen, Math-phys. Klasse, Seite 235-157* (1918), eprint: www.physics.ucla.edu/\~\cwp/articles/noether.trans/german/emmy235.html.

[3] D. Griffiths, *Introduction to Elementary Particles*, New York, USA: John Wiley & Sons, 1987.

[4] C. Wu et al., 'Experimental test of parity conservation in beta decay', *Phys.Rev.* 105 (1957) 1413–1414, DOI: 10.1103/PhysRev.105.1413.

[5] L. Landau, 'On the conservation laws for weak interactions', *Nuclear Physics* 3.1 (1957) 127–131, ISSN: 0029-5582, DOI: 10.1016/0029-5582(57)90061-5.

[6] J. H. Christenson et al., 'Regeneration of K_1^0 Mesons and the $K_1^0 - K_2^0$ Mass Difference', *Phys. Rev.* 140 (1B 1965) B74–B84, DOI: 10.1103/PhysRev.140.B74.

[7] M. Kobayashi and T. Maskawa, 'CP Violation in the Renormalizable Theory of Weak Interaction', *Prog.Theor.Phys.* 49 (1973) 652–657, DOI: 10.1143/PTP.49.652.

[8] A. D. Sakharov, 'Violation of CP invariance, C asymmetry, and baryon asymmetry of the universe', *JETP Lett.* 5 (1967) 24–27.

[9] B. Collaboration., 'The BaBar detector', *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 479.1 (2002), Detectors for Asymmetric B-factories 1–116, ISSN: 0168-9002, DOI: 10.1016/S0168-9002(01)02012-5.

[10] B. Collaboration, 'KEK B and BELLE experiment', *Nuovo Cim.* A109 (1996) 1055–1060, DOI: 10.1007/BF02823646.

[11] J. e. a. Brodzicka, 'Physics achievements from the Belle experiment', *Progress of Theoretical and Experimental Physics* 2012.1 (2012), DOI: 10.1093/ptep.

[12] M. Lemarenko et al., 'The data handling processor for the Belle II pixel vertex detector: Efficiency optimization', *JINST* 7 (2012) C01069, DOI: 10.1088/1748-0221.

[13] M. Lemarenko et al., 'Test Results of the Data Handling Processor for the DEPFET Pixel Vertex Detector', *JINST* (2013).

[14] T. Abe et al., *Belle II Technical Design Report*, 2010.

[15] I. Adachi et al., 'sBelle Design Study Report', *KEK-REPORT-2008-7* (2008), arXiv:0810.4084 [hep-ex].

[16] M. Andreas, 'PXD Background', tech. rep., Belle II collaboration, 2012.

[17] S. Yuri, 'Synchrotron Radiation Background to PXD', tech. rep., DESY, 2013.

[18] G. Lutz and A. S. Schwarz, 'Silicon devices for charged-particle track and vertex detection', *Ann.Rev.Nucl.Part.Sci.* 45 (1995) 295–335, DOI: 10.1146/annurev.ns.45.120195.001455.

[19] L. Andricek et al., 'Processing of ultra thin silicon sensors for future e+e- linear collider experiments', *IEEE Nuclear Science Symposium Conference Record*, vol. 3, Oct. 2003 1655–1658, DOI: 10.1109/NSSMIC.2003.1352196.

[20] DuPont, *Kapton polyimide film offical site*.

[21] J. Kemmer and G. Lutz, 'New detector concepts', *nimpra* 253 (1987) 356–377, DOI: 10.1016/0168-9002(87)90518-3.

[22] R. H. Richter et al., 'Design and technology of DEPFET pixel sensors for linear collider applications', *Nucl.Instrum.Meth.* A511 (2003) 250–256, DOI: 10.1016/S0168-9002(03)01802-3.

[23] E. Gatti and P. Rehak, 'Semiconductor drift chamber - an application of a novel charge transport scheme', *Nucl.Instrum.Meth.* A225 (1984) 608–614.

[24] J. Ulrici et al., 'Spectroscopic and imaging performance of DEPFET pixel sensors', *Nucl.Instrum.Meth.* A465 (2000) 247–252, DOI: 10.1016/S0168-9002(01)00401-6.

[25] L. Rossi et al., *Pixel Detectors: From Fundamentals to Applications*, Particle Acceleration and Detection, Springer, 2006, ISBN: 9783540283324.

[26] C. Sandow et al., 'Clear-performance of linear DEPFET devices', *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 568.1 (2006) 176–180, ISSN: 0168-9002.

[27] L. Andricek et al., 'The MOS-type DEPFET pixel sensor for the ILC environment', *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 565.1 (2006), Proceedings of the International Workshop on Semiconductor Pixel Detectors for Particles and Imaging PIXEL 2005. 165–171, ISSN: 0168-9002, DOI: 10.1016/j.nima.2006.05.045.

[28] O. Alonso et al., 'DEPFET active pixel detectors for a future linear e^+e^- collider', *arXiv* (2012).

[29] I. Collaboration, 'LC Reference Design Report Volume 3 - Accelerator', *arXiv* (2007).

[30] M. Trimpl, 'Design of a current based readout chip and development of a DEPFET pixel prototype system for the ILC vertex detector', PhD thesis, University of Bonn, 2005.

[31] P. Fischer et al., 'Readout concepts for DEPFET pixel arrays', *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 512.1–2 (2003), Proceedings of the 9th European Symposium on Semiconductor Detectors: New Developments on Radiation Detectors 318–325, ISSN: 0168-9002, DOI: 10.1016/S0168-9002(03)01909-0.

[32] A. S. Niculae, 'Development of a low noise analog readout for a DEPFET pixel detector', PhD thesis, University of Siegen, 2003.

[33] P. D. P. Fischer et al., *Switcher-B Reference Manual*, Document Revision 1.1, Nov. 2010.

[34] K. Gärtner and R. Richter, 'DEPFET sensor design using an experimental 3D device simulator', *nimpra* 568 (2006) 12–17.

[35] S. Rummel, 'Investigation of DEPFET as Vertex Detector at ILC — Intrinsic properties, radiation hardness and alternative readout schemes', PhD thesis, Technische Universität München, 2009.

[36] J. Hughes, N. Bird and I. Macbeth, 'Switched currents-a new technique for analog sampled-data signal processing', *Circuits and Systems, 1989., IEEE International Symposium*, vol. vol. 3, 1989 1584–1587, DOI: 10.1109/ISCAS.1989.100663.

[37] M. Koch, 'Development of a Test Environment for the Characterization of the Current Digitizer Chip DCD2 and the DEPFET Pixel System for the Belle II Experiment at SuperKEKB', Presented 05 Sep 2011, PhD thesis, Bonn U., 2011.

[38] I. Peric et al., 'DCDB and SWITCHERB, the readout ASICS for belle II DEPFET pixel detector', *Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC), 2011 IEEE*, 2011 1536–1539, DOI: 10.1109/NSSMIC.2011.6154365.

[39] JTAG, *IEEE Standard Test Access Port and Boundary Scan Architecture*.

[40] I. Perić et al., *DCD-B Reference Manual*, Revision 0.1, 2010.

[41] J. Knopf, 'Development, characterization and operation of the DCDB, the front-end readout chip for the pixel vertex detector of the future BELLE-II experiment', PhD thesis, Heidelberg University, 2011.

[42] H. Krüger, 'Module Link Report,4th International Workshop on DEPFET Detectors and Applications, Schloß Ringberg', tech. rep., DEPFET Collaboration, May, 2010.

[43] D. Levit, 'DEPFET Standalone Data Acquisition System for Test Beams and Test Setups', MA thesis, Technische Universität München, 2012.

[44] A. Wassatsch and R. Richter, 'DCE3 - An universal real-time clustering engine', *Circuits and Systems (ISCAS), 2012 IEEE International Symposium on*, May 2012 3242–3245, DOI: 10.1109/ISCAS.2012.6272015.

[45] M. Schnell, 'Data Concentrator for the Belle II Pixel Detector', MA thesis, University of Bonn, 2012.

[46] M. White et al., 'Characterization of surface channel CCD image arrays at low light levels', *Solid-State Circuits, IEEE Journal of* 9.1 (Feb.) 1–12, ISSN: 0018-9200.

[47] F. Sergey, 'PXD pedestals monitor and DQM', tech. rep., Belle II collaboration, 2012.

[48] W. Härdle and L. Simar, *Applied Multivariate Statistical Analysis*, Springer London, Limited, 2007, ISBN: 9783540722441.

[49] T. H. Cormen et al., *Introduction to Algorithms*, 2nd, McGraw-Hill Higher Education, 2001, ISBN: 0070131511.

[50] B. Grube, 'The Trigger Control System and the Common GEM and Silicon Readout for the COMPASS Experiment', MA thesis, Technische Universität München, 2001.

[51] P. Fischer, G. Comes and H. Krüger,
'First implementation of the MEPHISTO binary readout architecture for strip detectors',
Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment 461.1-3 (2001), 8th Pisa Meeting on Advanced Detectors 499–504, ISSN: 0168-9002, DOI: 10.1016/S0168-9002(00)01283-3.

[52] M. Lemarenko et al., *DHP 0.2 Reference Manual*, Revision 1.0, 2012.

[53] *Aurora 8B/10B Protocol Specification*.

[54] A. Widmer and P. Franaszek, 'A DC-Balanced, Partitioned-Block, 8B/10B Transmission Code', *IBM Journal of Research and Development* 27.5 (1983) 440–451, ISSN: 0018-8646, DOI: 10.1147/rd.275.0440.

[55] T. Kishishita et al.,
'Prototype of a gigabit data transmitter in 65nm CMOS for DEPFET pixel detectors at Belle-II',
Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment (2012), ISSN: 0168-9002,
DOI: 10.1016/j.nima.2012.11.013.

[56] A. Kruth et al., 'Charge Pump Clock Generation PLL for the Data Output Block of the Upgraded ATLAS Pixel Front-End in 130 nm CMOS', *Proceedings of TWEPP 09*, 2009.

[57] H. Krüger, 'Measurements on the first TSMC 65nm Test Chip (DHPT 0.1), Vienna', tech. rep., DEPFET Collaboration, 2012.

[58] F. Semiconductor, 'LVDS Fundamentals, AN-5017', tech. rep., Institution, 2005.

[59] H. Krüger, 'DHP 0.2 Status', tech. rep., Belle II collaboration, May, 2011.

[60] *InfiniBand Trade Assiciation*.

[61] S. Rummel, *Design review hybrid 3.0 for the S3B system*, 2007.

[62] Y. Bessho et al., 'A Stud-bump-bonding Technique For High Density Multi-chip-module', *Electronic Manufacturing Technology Symposium, 1993., Proceedings of 1993 Japan International*, 1993 362–365, DOI: 10.1109/IEMT.1993.639807.

[63] Xilinx, *UG347. ML505/506/507 Evaluation Platform User Guide*, Document Revision 3.1.2, 2011.

[64] Xilinx, *MicroBlaze Processor Reference Guide*, 4.0, Xilinx, 2004.

[65] Agilent, *Eye-Diagram Analysis User's Guide*.

[66] F. Lütticke, *private communication*.

[67] T. R. Society, *Gamma-ray spectrum catalog of isotopes*.

[68] M. J. Berger et al., 'XCOM: Photon Cross Sections Database',
NIST Standard Reference Database 8 (XGAM) (2010).

[69] A. Einstein, 'Über einen die Erzeugung und Verwandlung des Lichtes betreffenden heuristischen Gesichtspunkt', *Annalen der Physik* 322.6 (1905) 132–148, ISSN: 1521-3889,
DOI: 10.1002/andp.19053220607.

[70] A. H. Compton, 'A Quantum Theory of the Scattering of X-rays by Light Elements',
Phys. Rev. 21 (5 1923) 483–502, DOI: 10.1103/PhysRev.21.483.

[71] EUDET JRA1 Group, 'EUDET Pixel Telescope Data Taking Manual - Updated Version 2009',
EUDET-Memo-2009-03 (29th Apr. 2010).

[72] J. F. Ziegler and W. A. Lanford, 'Effect of Cosmic Rays on Computer Memories', *Science* 206.4420 (1979) 776–788, DOI: 10.1126/science.206.4420.776.

[73] C. Detcheverry et al., 'SEU critical charge and sensitive area in a submicron CMOS technology', *Nuclear Science, IEEE Transactions on* 44.6 (1997) 2266–2273, ISSN: 0018-9499, DOI: 10.1109/23.659045.

[74] P. Dodd, M. Shaneyfelt and F. Sexton, 'Charge collection and SEU from angled ion strikes', *Nuclear Science, IEEE Transactions on* 44.6 (1997) 2256–2265, ISSN: 0018-9499, DOI: 10.1109/23.659044.

[75] F Faccio, C Detcheverry and M. Huhtinen, 'First Evaluation of the Single Event Upset (SEU) Risk for Electronics in the CMS Experiment', tech. rep. CMS-NOTE-1998-054, Geneva: CERN, 1998.

[76] Particle Data Group, K. Nakamura et al., 'Review of Particle Physics', *J. Phys. G* 37 (2010) 075021.

[77] H. H. K. Tang, 'Nuclear physics of cosmic ray interaction with semiconductor materials: Particle-induced soft errors from a physicist's perspective', *IBM Journal of Research and Development* 40.1 (1996) 91–108, ISSN: 0018-8646, DOI: 10.1147/rd.401.0091.

[78] A. Papoulis and S. Pillai, *Probability, random variables, and stochastic processes*, McGraw-Hill electrical and electronic engineering series, McGraw-Hill, 2002, ISBN: 9780073660110.

[79] D. Lambert, F. Desnoyers and D. Thouvenot, 'Investigation of neutron and proton SEU cross-sections on SRAMs between a few MeV and 50 MeV', *Radiation and Its Effects on Components and Systems (RADECS), 2009 European Conference on*, 2009 148–154, DOI: 10.1109/RADECS.2009.5994571.

[80] M. Andreas and R. Martin, 'Update on Background Simulation in the PXD', tech. rep., Belle II collaboration, 2012.

[81] Xilinx, 'Device reliability report', *tech-report* (2007) 23.

[82] S. Bonacini et al., 'Characterization of a commercial 65 nm CMOS technology for SLHC applications', *JINST* 7 (2012) P01015, DOI: 10.1088/1748-0221.

[83] NIST, *PSTAR online database*.

[84] M. Glaser, F. Ravotti and M. Moll, 'Dosimetry Assessments in the Irradiation Facilities at the CERN-PS Accelerator', *Nuclear Science, IEEE Transactions on* 53.4 (2006) 2016–2022, ISSN: 0018-9499, DOI: 10.1109/TNS.2006.880569.

[85] P. Roche et al., 'A commercial 65nm CMOS technology for space applications: Heavy ion, proton and gamma test results and modeling', *Radiation and Its Effects on Components and Systems (RADECS), 2009 European Conference on*, 2009 456–464, DOI: 10.1109/RADECS.2009.5994696.

[86] F. Haight, *Handbook of the Poisson distribution*, Publications in operations research, Wiley, 1967.

[87] P. Sweeney, *Error Control Coding: From Theory to Practice*, Wiley, 2002, ISBN: 9780470843567.

List of Figures

List of Tables

Acknowledgements

- I would like to thank my wife Olga, who was the main reason why I decided to do a PhD and move to Bonn. In good and bad periods, she was always my main support and source of motivation.
- I would like to thank Prof. Dr. Norbert Wermes for accepting me in his group. All along my PhD studies I learned a lot about experimental particle physics and its methods. Moreover, this was a great opportunity to be a member of a big collaboration and to take part in the international project to develop a new concept pixel detector.
- I would like to thank Dr. Hans Krüger for supervising my work. During past four years, he has been sharing this experience in electronics, which was a great help for my professional development.
- Finally, I would like to thank my colleagues and friends from *Silizium Labor Bonn*, especially those who helped me with proofreading of this thesis.