TECHNISCHE UNIVERSITÄT WIEN

DISSERTATION

Development and Testing of Depleted Monolithic Active Pixel Sensors in 180 nm CMOS technology for the Belle II VTX Upgrade

zur Erlangung des akademischen Grades

Doktor der technischen Wissenschaften

im Rahmen des Studiums

Doktoratsstudium Technische Physik

eingereicht von

Dipl.-Ing. Maximilian Babeluk
Matrikelnummer 01609369

ausgeführt am Institut für Hochenergiephysik
der Österreichischen Akademie der Wissenschaften
und am Atominstitut
der Technischen Universität Wien

Betreuung
Betreuer/in: Privatdoz. Dipl.-Ing. Dr.techn. Christoph Schwanda
Mitwirkung: Univ.Lektor Dipl.-Ing. Dr.techn. Thomas Bergauer

Wien, August 1, 2025

(Unterschrift Verfasser/in) (Unterschrift Betreuer/in)

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

Abstract

Depleted Monolithic Active Pixel Sensors (DMAPS) have been the subject of increasing interest in detector research for particle physics in recent years. New high-resistivity, mixed-signal silicon manufacturing processes enable pixel sensors with not only analog signal amplification but powerful digital data processing. This is a major simplification compared to the conventional hybrid sensor approach. With DMAPS, a single piece of silicon is the active sensing element, and only a few data- and power connections per chip to the data acquisition system are required.

The Belle II experiment is located at the SuperKEKB collider in Japan. Electrons and positrons collide with a center of mass energy of 10.58 GeV at extreme luminosity. A key objective is high-precision measurement of the standard model with a special focus on the b-sector. To allow the desired precision, a minimal material budget is required for the detector to reduce multiple scattering as much as possible. Therefore, DMAPS are a perfect candidate for upgrading the Belle II Vertex detector.

The Optimized BeLle Ii piXel sensor (OBELIX) is a DMAPS chip currently under development, intended for the upgrade of the Belle II Vertex detector scheduled for 2032. The OBELIX chip is based on the TJ-Monopix2 small-electrode DMAPS chip and uses a modification of the Tower 180 nm imaging CMOS technology.

While the pixel matrix of OBELIX is inherited from TJ-Monopix2 with only small adaptations, the periphery is thoroughly reworked. The OBELIX chip contains voltage regulators and a monitoring system with multiple temperature sensors and taps to measure voltage drop on chip. On the digital side of the periphery, the complete data processing is newly implemented. This includes a trigger memory to match the requirements for Belle II. In addition, a precision timing module allows for increased time resolution. The chip is also capable of sending low-granularity hit-data to the first-level trigger system of Belle II in parallel to normal operation.

This thesis focuses on the the development of the digital periphery of the OBELIX chip while also giving insight into the whole upgrade project and the other features of the OBELIX chip. The design is explained in technical detail, and simulations are conducted for performance evaluation at nominal- and over-load conditions. In particular, the author has implemented all of the new features of the digital periphery. This includes the trigger memory with performance simulations using realistic data patterns. Furthermore, the author conceptualized and implemented the precision timing module as well as the module for contribution to the Belle II first level trigger.

The performance of the TJ-Monopix2 chip has been evaluated using laboratory measurements and test-beams by the author. The timing measurements and analyses have been performed by the author, and significant contributions have been made to the efficiency measurements at the test-beams, which were a group effort. The resulting data is used to refine the simulations and the design of OBELIX.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

Zusammenfassung

Depleted Monolithic Active Pixel Sensors (DMAPS) haben in den letzten Jahren in der Hochenegiephysik an Bedeutung gewonnen. Neue Herstellungsverfahren für Microchips mit hoher Reinheit ermöglichen Pixelsensoren, die nicht nur analoge-, sondern auch digitale Datenverarbeitung bieten. Dies ist eine erhebliche Vereinfachung gegenüber dem herkömmlichen Hybridsensoransatz. Bei DMAPS ist ein einzelner Chip das Sensorelement, und nur wenige Daten- und Versorgungsverbindungen zum Datenerfassungssystem sind erforderlich.

Das Belle II Experiment befindet sich am SuperKEKB-Collider in Japan. Elektronen und Positronen kollidieren mit einer Schwerpunktsenergie von 10.58 GeV bei extremer Luminosität. Ein Hauptiel sind Präzisionsmessungen des Standardmodells mit Focus auf dem b-Sektor. Um die gewünschte Genauigkeit zu erreichen, ist ein minimales Materialbudget für den Detektor erforderlich. Daher sind DMAPS ein idealer Kandidat für ein Upgrade des Belle II Vertex-Detektors.

Der Optimized BeLle Ii piXel Sensor (OBELIX) ist ein DMAPS Chip derzeit in Entwicklung, und für dieses Upgrade im Jahr 2032 vorgesehen ist. OBELIX basiert auf TJ-Monopix2 mit kleinen Elektroden und verwendet eine Modifikation der Tower 180 nm CMOS Imaging Technologie.

Während die OBELIX Pixelmatrix mit geringfügigen Anpassungen von TJ-Monopix2 übernommen wurde, wurde die Peripherie grundlegend überarbeitet. OBELIX enthält Spannungsregler und ein Monitoringsystem mit mehreren Temperatursensoren und Spannungsabfallmessung. Die gesamte digitale Datenverarbeitungskette wurde neu implementiert. Dazu gehört ein Triggerspeicher, der den Anforderungen von Belle II entspricht. Darüber hinaus ermöglicht ein Präzisions-Timing-Modul eine verbesserte Zeitauflösung. Der Chip erlaubt parallel zum normalen Betrieb reduzierte Hit-Daten an das Level 1 Triggersystem zu senden.

Diese Arbeit konzentriert sich auf die Entwicklung der digitalen Peripherie des OBELIX-Chips und gibt gleichzeitig einen Einblick in das gesamte Upgrade-Projekt und die anderen Funktionen des OBELIX Chips. Das Design wird mit technischen Details erläutert und es werden Simulationen zur Leistung unter Nenn- und Überlastbedingungen präsentiert. Der Autor hat alle neuen Funktionen der digitalen Peripherie im Zuge dieser Arbeit implementiert. Dazu gehört der Triggerspeicher mit Simulationen unter Verwendung realistischer Datenmuster. Darüber hinaus hat der Autor das Präzisions-Timing-Modul sowie das Modul für den Beitrag zum Belle II Level 1 Trigger konzipiert und implementiert.

Die Leistung des TJ-Monopix2-Chips wurde vom Autor anhand von Labormessungen und Testbeams evaluiert. Die Zeitmessungen und deren Analyse wurden vom Autor durchgeführt. Die Testbeams zur Effizienzmessung wurden mit wesentlichen Beiträgen des Autors in eier Gruppe durchgeführt. Die daraus resultierenden Daten werden zur Verfeinerung der Simulationen und des Designs des OBELIX Chips verwendet.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

If you hold a Linux shell to your ear, can you hear the C?

— *Unknown Author*

Acknowledgements

First and foremost, I would like to thank the main supervisor of my thesis, Priv.-Doz. Dipl.-Ing. Dr.techn. Christoph Schwanda, for his support and guidance throughout my time as a PhD student while giving me a great deal of freedom in choosing the specific topics to work on.

I am also truely grateful to my co-supervisor, Univ.-Lekt. Dipl.-Ing. Dr.techn. Thomas Bergauer, for supporting my scientific career and mentoring me at HEPHY over the past five years.

My deepest gratitude goes to Christian Irmler for the collaboration and countless discussions whenever I needed new ideas. It was always highly productive to discuss with you on specific topics.

I am thankful to Jérôme Baudot for always having an open ear for my comments. Thank you for inviting me to Strasbourg twice and allowing me to deepen my understanding of chip design at IPHC. A special mention goes to Giuliana Rizzo - I truly admire your scientific rigor and meticulousness and aspire to live up to that standard, even if that occasionally prolongs meetings to borderline durations.

I would also like to thank the entire OBELIX design team, and in particular Hung Pham, Luca Federici, Abdelkader Himmi, Alexander Walsemann, Hans Krüger, Xu Danwei and Roberto Russo for their excellent collaboration.

Many thanks to Alice Gabrielli, Marike Schwickardi, Yannik Buch, Benjamin Schwenker and Bernhard Pilsl - I have always enjoyed working with you and particularly the time of the test-beams in your company. Special thanks also goes to Christian Bespin and Lars Schall for providing large parts of the DAQ system in a working state without spoiling the fun of debugging for us.

I am also grateful to the students I had the pleasure of working with during my time as PhD student: Harald Handerkas, Juraj Habala, Pia Wenty, Paul Schellmann and Tobias Blesgen.

Lastly, I would like to thank my family and friends for their support throughout my PhD studies. Special thanks to my parents, and to Rita and Martin, for shielding me from many day-to-day affairs as much as possible, especially in recent months. I also want to thank Tillmann for supporting me and enduring my occasional mental absence, especially during this final phase.

The test-beam measurements leading to results in this thesis have been performed at the Test Beam Facility at DESY Hamburg (Germany), a member of the Helmholtz Association (HGF).

This work has received funding from the European Union's Horizon 2020 Research and Innovation Program under grant agreement no 101004761 (AIDAinnova) and from the Austrian Research Promotion Agency (FFG), project no 54547948.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Contents

1 Introduction **8**

2 The Belle II Experiment and VTX Upgrade **10**
2.1 The SuperKEKB Accelerator . . . 10
2.2 The Belle II Experiment . . . 12
2.2.1 Vertex detector . . . 13
2.2.2 Central drift chamber . . . 14
2.2.3 Cherenkov detectors . . . 14
2.2.4 Electromagnetic Calorimeter . . . 15
2.2.5 K_L - Muon detector . . . 16
2.2.6 Trigger System . . . 16
2.3 The Belle II Upgrade Proposal . . . 18
2.4 The VTX Upgrade . . . 20
2.4.1 Mechanical Structure of VTX . . . 21
2.4.2 Physics Performance . . . 24
2.4.3 Sensor Requirements . . . 26

3 Silicon Particle Detector Technology **29**
3.1 Semiconductor Basics . . . 29
3.1.1 The pn Junction . . . 31
3.2 Interaction of Particles with Matter . . . 33
3.2.1 Heavy particles . . . 33
3.2.2 Electrons and Photons . . . 34
3.2.3 Radiation damage . . . 35
3.3 Technologies of Silicon Radiation Detectors . . . 37
3.4 The DMAPS Technology . . . 40
3.4.1 General Concepts . . . 42

4 The TJ-Monopix2 DMAPS Chip **46**
4.1 Overview and Features . . . 46
4.1.1 Frontend Amplifier . . . 47
4.1.2 HitOr . . . 49
4.1.3 Calibration and Tuning . . . 51
4.2 Test-beam Campaigns . . . 53
4.3 ToT Clipping: `VCLIP` . . . 54
4.4 Efficiency Measurements . . . 56
4.4.1 Results . . . 58
4.5 Timing Measurements . . . 61
4.5.1 The TDC module . . . 62
4.5.2 Test-beam Timing Measurements . . . 64

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

4.5.3 BCID Delay Compensation . . . 70
4.6 BCID Crosstalk . . . 71
4.7 Timewalk . . . 77
4.8 Power Consumption . . . 81

5 Design and Simulation of the OBELIX Chip 84
5.1 Features and Architecture . . . 84
5.2 Voltage Regulators . . . 86
5.3 Floorplan . . . 89
5.4 Monitoring and ADC . . . 89
5.5 BCID Crosstalk Mitigation . . . 91
5.6 Services . . . 91
5.7 Sync Clock Unit and Clock domains . . . 93
5.8 Control Unit . . . 95
5.8.1 The `Sync` command . . . 97
5.8.2 The `Clear` command . . . 97
5.8.3 The `TriggerX` commands . . . 98
5.8.4 The `Write Register` command . . . 99
5.8.5 The `Read Register` command . . . 99
5.8.6 The `Calibration` command . . . 99
5.8.7 Pixel Configuration . . . 101
5.9 Trigger Unit . . . 102
5.9.1 Column Drain Readout and EoC . . . 102
5.9.2 Data Sequence and Time Order . . . 105
5.9.3 TRU Design . . . 105
5.9.4 The Trigger Group . . . 107
5.9.5 TRU Performance . . . 111
5.9.6 Peripheral Time-to-Digital Converter . . . 112
5.9.7 Triggering Strategy . . . 116
5.9.8 Maximum Trigger Rate . . . 118
5.9.9 Veto Strategy . . . 121
5.10 Transmission Unit . . . 122
5.11 Track Trigger Transmission . . . 123
5.11.1 Time Reconstruction and Resolution . . . 126
5.11.2 Occupancy and Fake Rate . . . 128
5.12 SEU Protection Measures . . . 129
5.13 Digital Injection and Test Strategy . . . 131
5.13.1 Test Suite . . . 134
5.13.2 Applications in Chip Characterization . . . 137
5.14 Debugging Ports . . . 138
5.15 Power Budget . . . 140
5.15.1 Digital Periphery: Power Simulation Scaling . . . 141

6 Conclusion 143

A OBELIX Cheatsheet 151

B Auxiliary Plots 152

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

C Additional OBELIX Material **153**
C.0.1 Timewalk Measurement with PTD 154
C.0.2 Register Table 154

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

Chapter 1

Introduction

The Standard Model of particle physics (SM) has been exceedingly successful in the recent decades, explaining 3 of the 4 fundamental interactions at small scales to unprecedented accuracy with quantum descriptions.

On the other hand, the general theory of relativity (GR) describes gravity, the missing piece in the SM. GR is a classical theory by nature, but has proven to describe many phenomena in the macroscopic world, from planetary mechanics to black holes. At galaxy scales and above, however, the GR predictions do not quite fit the observations, leading to the theory of dark matter.

Despite the efforts in the recent decades, SM and GR remain incompatible and dark matter has never been observed apart from astronomical observations.

As successful the SM is in describing the quantum world, pushing forward the experimental tests is a way to find potential anchor points for new theories. It is therefore crucial to push the limits in high energy physics. There are two frontiers to extend the parameter space of the tested SM: the energy frontier and the precision frontier, supplementing each other. Increasing the collision energy enables the creation and therefore discovery of potential heavier particles in a direct way. To facilitate this, protons, being stable and heavy charged particles, are well suited for acceleration to high energies. This is currently pursued at the Large Hadron Collider (LHC) [1] at CERN, and one of the main goals of the general purpose experiments ATLAS [2] and CMS [3] at the LHC. However, the inner structure of the proton poses limits to the achievable accuracy.

The other approach, pursued by the Belle II experiment uses electron-positron collisions at lower energies. As fundamental particles, the collisions are much 'cleaner', i.e. the typical final states consist only of a hand full of particles. This enables other types of analyses, e.g. determination of missing momentum, carried away by neutrinos or potentially undetected dark matter particles. At the same time, parameters of the SM can be measured with more precision. In particular, B mesons can be studied precisely at Belle II to map out the CP violation in the quark sector. Discrepancies with the theory might hint to unaccounted higher order virtual processes, even if potential new particles are not directly created or observed. A large number of collisions and therefore luminosity is beneficial here, as it allows the observation of rare decay processes which have not been studied with this precision so far. Therefore, the Belle II experiment relies on the nano-beam scheme with low emittance beam optics and large beam current to maximize luminosity [4].

The next flagship collider under discussion at CERN is the Future Circular Collider (FCC) [5] with about 91 km circumference. In its first stage FCC-ee, planned to start operation in the mid 2040ies, the FCC will operate as lepton collider with

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

electrons and positrons, i.e. as a precision machine but capable of higher energy than available at Belle II. It will generate vast amounts of data not only at the Z pole and WW threshold, but also as Higgs factory and at the $t\bar{t}$ threshold. As a project of unprecedented scale in high energy physics, technologies need to be developed as soon as possible. The Depleted Monolithic Active Pixel Sensor (DMAPS) technology developed in the recent years is an excellent candidate for the vertex detector of future FCC experiments and detector proposals heavily rely on this technology.

The Belle II experiment as currently largest lepton collider in the world is an important stepping stone for the FCC project. A detector on DMAPS basis will be installed for the upgrade of the Belle II vertex detector planned for 2032. The development of FCC sensor technology can directly use the Belle II developments as starting point, as well as gaining practical information about large-scale DMAPS deployment.

This thesis focuses on the upgrade of the Belle II vertex detector and in particular the development and testing of DMAPS. In chapter 2, the Belle II experiment will be introduced in detail and the upgrade plans are described. In the following, silicon sensor technology is explained in chapter 3 from semiconductor basics to DMAPS. Subsequently, two chapters describing the actual work of the author follow. Chapter 4 describes various measurements performed with the TJ-Monopix2 generic prototype chip in the laboratory and at test-beams. The last chapter 5 focuses on the development of the next generation sensor directly designed for the Belle II experiment, the OBELIX chip.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

Chapter 2

The Belle II Experiment and VTX Upgrade

The Belle II experiment [4] at the SuperKEKB accelerator [6] is an upgrade of the former Belle experiment [7] operating from 1999 to 2010 at the KEKB accelerator [8]. It is designed to operate at higher luminosity than its predecessor, increasing the capability of handling event rates by a factor of 40 at a background rate 10 to 20 times higher than in Belle. This leads to higher occupancy, radiation damage and fake rate in all detector subsystems and increases the demand on the trigger and data acquisition (DAQ) to maintain performance. At the same time, the reconstruction efficiency and vertex resolution below 50 µm needs to be maintained to resolve the vertices of B-meson decays. To allow the reconstruction of final states involving neutrinos, as well as dark matter searches via missing momentum, the detector needs to cover almost the full solid angle around the interaction region.

2.1 The SuperKEKB Accelerator

The SuperKEKB accelerator is the collider used by the Belle II experiment with a circumference of about 3 km and located in Tsukuba, Japan. An overview of the accelerator complex is shown in figure 2.1. SuperKEKB is an electron positron collider with asymmetric energies of 7 GeV and 4 GeV respectively, leading to a center of mass energy $\sqrt{s} = 10.58\,\text{GeV}$ with a Lorentz factor of $\beta\gamma = 0.28$ in respect to the lab frame. This Lorentz boost enables the precise measurement of CP violation with B mesons by creating them at a certain velocity. This means, the position of decay of B mesons has an offset in beam direction which can be used to measure their decay time. Relativistic time dilatation increases the distance in the order of percent, but this has no strong effect on the resolution.

The Electron and positron beams circulate in separate storage rings and are brought into collision in one point at the location of the Belle II experiment. The electrons circulate in the so called high energy ring (HER) and the positrons in the low energy ring (LER). Both rings can accommodate thousands of bunches of particles at a beam current in the order of 1 A and a revolution time of 10.0614 µs.

The electron beams are produced in a cold cathode radio-frequency gun. The cathode is irradiated with laser pulses to achieve low emittance, i.e. area of the phase space in transverse motion, required for high luminosity operation. The electron bunches are then injected into the HER via a 600 m linear accelerator. The positrons are produced via pair production by firing electrons at a 14 mm tungsten target. A

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

positron damping ring subsequently is used to cool down the positron bunches to achieve low emittance at an energy of 1.1 GeV. After final acceleration, the bunch is injected into the LER.

The storage rings operate at a radio-frequency of 508.887 MHz, leading to RF-buckets with about 1.965 ns spacing. Depending on the fill pattern, every second or third RF bucket can be filled with a bunch, leading to either 3.929 ns or 5.894 ns of bunch separation. For most detector systems in Belle II, this bunch spacing is less than the time resolution and a quasi-continuous beam can be assumed [6].

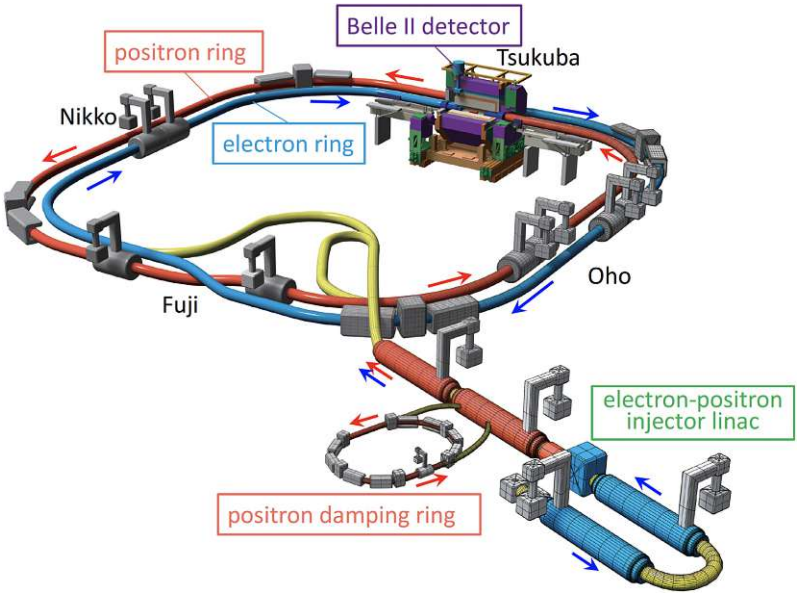


Figure 2.1: The SuperKEKB accelerator complex [6].

Luminosity

One of the main missions of SuperKEKB and the Belle II experiment is to acquire more data than previous experiments. The key parameter to quantify this is the luminosity, which can be given integrated over time or instantaneously. For a given process X with a cross section σ_X, the expected number of events N_X can be calculated by multiplication with the integrated luminosity $\mathcal{L}_{int}$:

$$N_X = \sigma_X \mathcal{L}_{int}. \tag{2.1}$$

In the same way, the instantaneous luminosity allows the calculation of the rate in which the process occurs. SuperKEKB implements the so called Nano-Beam scheme to achieve extreme luminosities. It works by squeezing the vertical beta function β_y^* at the interaction region (IR) and minimizing the longitudinal overlap of the two beams. Typically, the hourglass-effect limits the effective minimum β_y^*. When the bunch length becomes comparable to the beta function, the beam size varies along the bunch, reducing the luminosity. The effect can be mitigated by an increased crossing angle of the two beams and an extremely low horizontal emittance for both beams. The instantaneous luminosity can be expressed from parameters of the collision via:

$$L = \frac{\gamma_\pm}{2er_e}\left(\frac{I_\pm \xi_{y\pm}}{\beta^*_{y\pm}}\right)\left(\frac{R_L}{R_{\xi_y}}\right). \tag{2.2}$$

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Here, the γ, e and r_e are Lorentz factor, elementary charge and classical electron radius respectively. The suffix $\pm$ specifies either the positron $(+)$ or the electron $(-)$ beam parameters. R_L and R_{ξ_y} are reduction factors for the luminosity and vertical beam parameter arising from crossing angle and hourglass effect and are typically close to unity. This leaves three main parameters that determine the luminosity: the total beam current I, the vertical beam parameter ξ_y and the vertical beta function at the IR β_y^* [6].

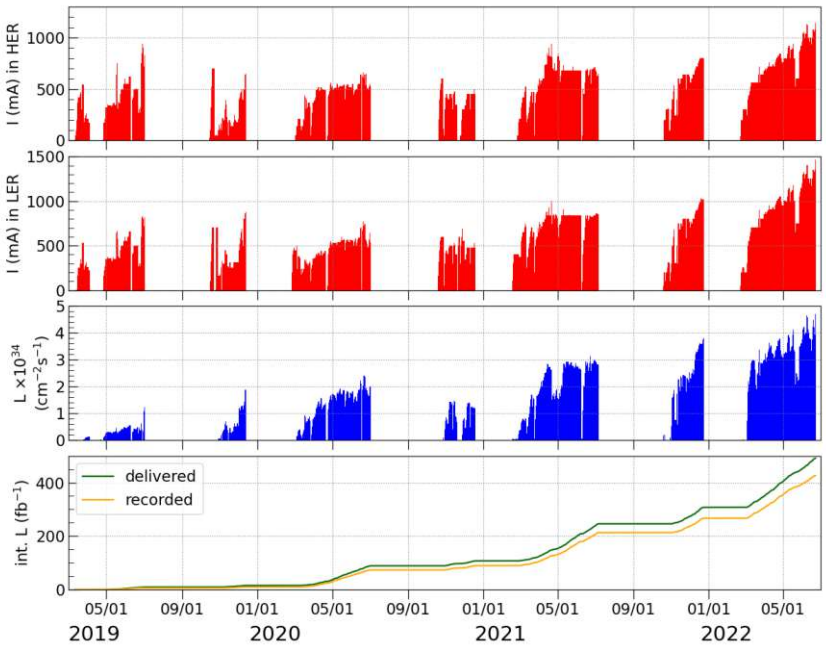


Figure 2.2: History of luminosity and beam currents before LS2 [9].

Figure 2.2 shows the beam currents, as well as instantaneous and integrated luminosity recorded by the Belle II experiment before the first long shutdown (LS1). While the SuperKEKB accelerator reached the world record in peak luminosity in 2022 with $4.65 \times 10^{34}\,\mathrm{cm^{-2}s^{-1}}$ with a $\beta_y^* = 1.0\,\mathrm{mm}$ in a physics run, it is planned to increase the luminosity even further. In the current state, SuperKEKB is expected to reach an instantaneous luminosity of $2 \times 10^{35}\,\mathrm{cm^{-2}s^{-1}}$. The final target of $6 \times 10^{35}\,\mathrm{cm^{-2}s^{-1}}$ appears not to be achievable with the current hardware and upgrades of the accelerator might be required [4, 9].

2.2 The Belle II Experiment

The Belle II detector is divided in cylindrical subsystems placed in shells around the IR, as typical for general purpose experiments in high energy physics. A general overview is shown in figure 2.3. Since the beams feature asymmetric energies, the whole detector is elongated in the beam direction of the electron beam (forward direction) to reflect the resulting Lorentz boost of the collisions. A magnetic field of 1.5 T parallel to the beam axis is generated by a superconducting magnet to allow momentum reconstruction of charged particles. The detector subsystems are presented in order from the IR outwards. The IR is covered by the beam pipe to maintain the vacuum of the accelerator. It is made from Beryllium and cooled with paraffin oil. Additionally, it is plated with a 10 µm gold layer to protect the vertex detector from synchrotron radiation [4].

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

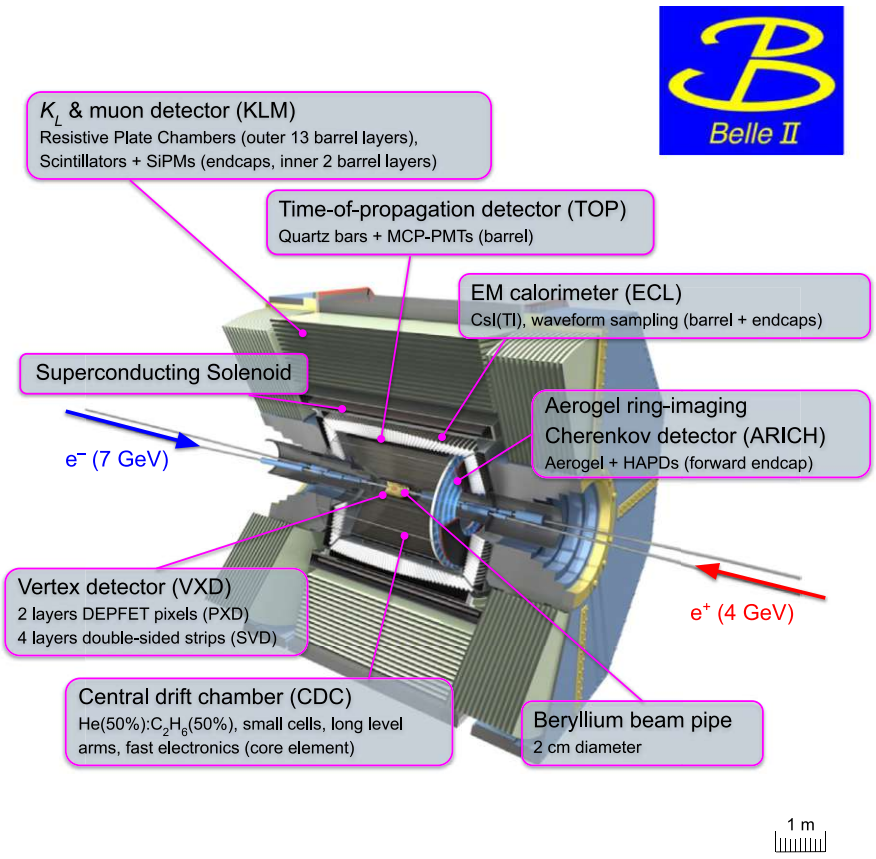


Figure 2.3: The Belle II experiment and the sub-detectors. [10]

2.2.1 Vertex detector

The vertex detector (VXD) consists of the Pixel Detector (PXD) and the Silicon Vertex Detector (SVD), and is shown in figures 2.4 and 2.5. Together, they allow the reconstruction of the vertex position and participate in the tracking and momentum reconstruction of charged particles generated by the collisions.

The PXD consists of two layers of depleted p-channel field effect transistors (DEPFET) with emphasis on low material budget with a sensor thickness of 50 µm of silicon. The PXD excels in spatial resolution, however, the integration time of 20 µs in rolling shutter mode leads to high occupancy and poses limits in the analysis. The PXD cannot participate in the track finding at the beginning of the track reconstruction because of the high occupancy. Only after a track has been reconstructed successfully, the hits on the PXD can be read out in regions of interest and subsequently used to increase the vertex resolution.

The SVD consists of four layers of double sided silicon strip detectors (DSSDs) read out by APV25 readout chips [11]. A spatial resolution of 7 to 12 µm (p-side) and 15 to 25 µm (n-side) and time resolution below 3 ns can be reached. Inherent to DSSDs, the SVD suffers from ambiguities at high occupancies. Due to the excellent time resolution, the fake rate is acceptable up to 6% occupancy and rapidly degrades above. Furthermore, the APV25 chip limits the trigger latency of the whole Belle II detector to 5.04 µs. For more sophisticated trigger algorithms, a longer trigger latency would be desirable.

For luminosity and beam monitoring, a total of 20 diamond sensors are placed around the beam pipe and inside the VXD volume [12, 13].

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

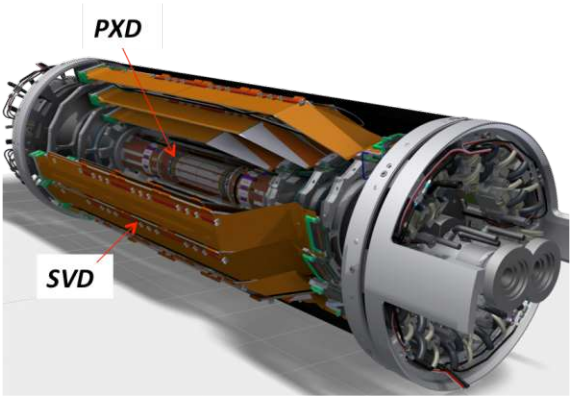


Figure 2.4: CAD rendering of VXD. [14]

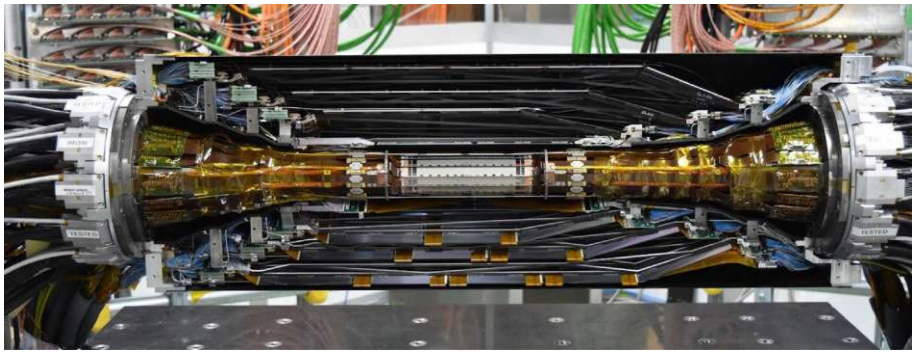


Figure 2.5: Photograph of VXD half shell after PXD completion in 2023. [13]

2.2.2 Central drift chamber

The Central Drift Chamber (CDC) is the main tracking detector of the Belle II experiment. It features 14 336 gold plated tungsten sense wires arranged in 56 layers with a diameter of 30 µm each. Additionally, 42 240 aluminum field wires are used. The CDC performs three key functions in the Belle II experiment. First, it allows the reconstruction of the particle tracks and therefore momenta. Second, it allows particle identification based on their energy loss. This is particularly important for low-momentum particles which do not reach the outer detectors. Third, it serves as the only input with tracking capability to the first level trigger of the Belle II experiment [15].

2.2.3 Cherenkov detectors

There are two detectors based on the Cherenkov effect: the Time of Propagation counter (TOP) and the Aerogel Ring imaging Cherenkov counter (ARICH). They help in particle identification, mainly for hadrons. Both detectors are able to measure the Cherenkov angle θ_c for charged particles of velocity β, exceeding the speed of light in a medium following the relation with refractive index n:

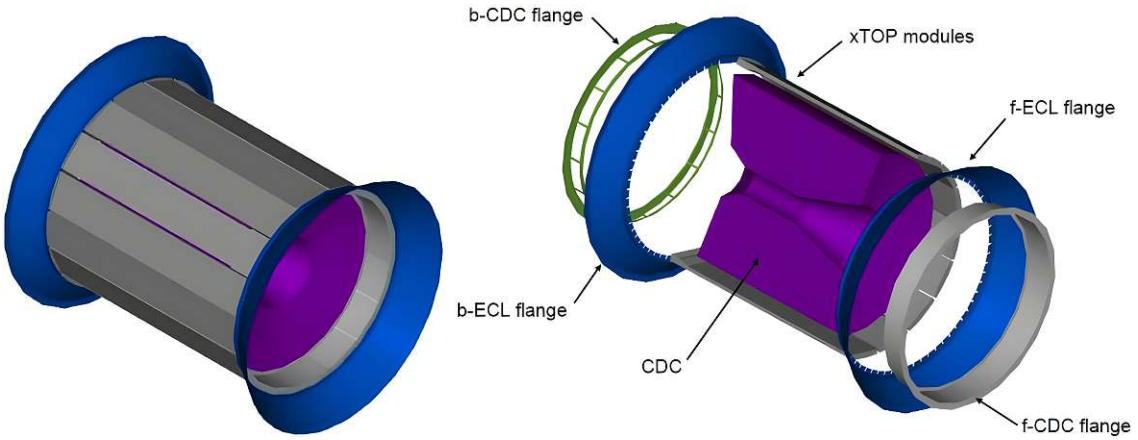
$$\cos\theta_c = \frac{1}{n\beta}. \tag{2.3}$$

The TOP detector consists of 16 quartz bars of cross section $45 \times 2\,\mathrm{cm}^2$ and a length of 2.6 m located in the barrel region just outside the CDC as visible in

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

figure 2.6. The readout is done by multi channel plate photo multiplier tubes (MCP-PMTs).

The ARICH is of disk shape located at the forward end-cap. It consists of two layers of aerogel with slightly different refractive indices and a detector plane 20 cm away. The dual layer design increases the number of photons generated while keeping the same focal plane for both aerogel layers required for spatial resolution.



(a) (b)

Figure 2.6: CDC and TOP modules together (a) and in exploded view (b) [4][1].

2.2.4 Electromagnetic Calorimeter

The Electromagnetic calorimeter (ECL) covers the barrel and end-cap regions as shown in figure 2.7. It is composed of 8736 thallium doped cesium iodide scintillator crystals. Each crystal is read out by two PIN photo-diodes. The crystals have been reused from the Belle experiment, but the photo-diodes and electronics were upgraded to cope with the increased luminosity. The amplified signal is sampled with 2 MHz and 18 bit. Due to the intrinsic time constant of the scintillator of about 1 µs, a pulse shape discrimination algorithm is required to at least partially restore performance under high luminosity operation. The ECL also provides information to the first level trigger of Belle II [4].

[1] Original image upscaled and enhanced using https://imgupscaler.com/.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

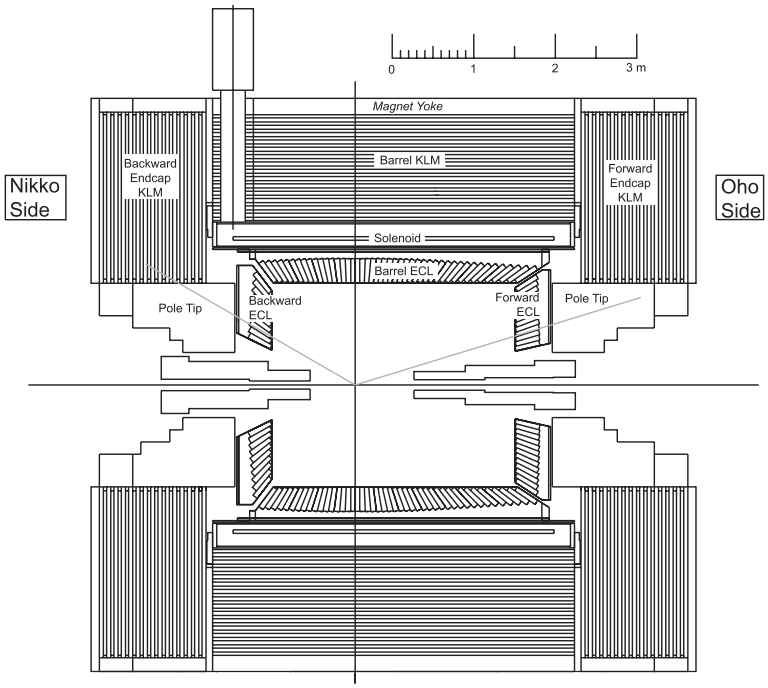


Figure 2.7: ECL and KLM systems in the Belle II experiment. [4]

2.2.5 K_L - Muon detector

The K_L and Muon detector (KLM) is located outside of the superconducting solenoid as shown in figure 2.7. It consists of 4.6 cm thick iron plates interleaved with particle detectors. The iron plates also serve as return yoke for the magnetic field. Scintillators with silicon photomultipliers and resistive plate chambers (RPCs) are used as detectors. The RPCs are a type of proportional counter like ionization chambers and are re-used from the Belle experiment. Due to their large dead times they are not suitable for high neutron fluxes expected at Belle II. Therefore, the regions of the calorimeter where high fluxes are expected were replaced with scintillators to increase the maximum particle rate [4].

2.2.6 Trigger System

The trigger system of Belle II is important to reduce the amount of data produced by the experiment. It selects ('triggers on') events that are of interest for the physics analysis while rejecting background and processes with large cross-sections such as Bhabha scattering. The decision of the trigger system needs to be sufficiently fast, since the detector systems are limited in data storage and bandwidth. The trigger system consists of two stages, an FPGA based Level 1 trigger (L1) and a software based second level, the high level trigger (HLT).

The L1 trigger has a fixed latency of 5 µs, after which a decision needs to be made with a timing precision of less than 10 ns. It is specified to a triggering rate of 30 kHz Poissonian distributed with a minimum spacing of 200 ns between two triggers. The decision of the L1 is based on inputs from the CDC and the ECL, as well as KLM. The CDC is used to reconstruct tracks, either in 2 or 3 dimensions. The 2D tracks do not extract the z information to allow faster reconstruction whereas the 3D reconstruction allows rejection of background tracks not originating from

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

the IR. The reconstructed tracks then can be matched to ECL or KLM clusters to improve event classification.

A contribution to the L1 trigger from the SVD is not possible since the readout and especially APV25 frontend chip was never designed for this purpose. However, the reconstruction and background rejection would benefit from information of the vertex detector. Therefore, the proposed upgrade for the vertex detector foresees a contribution to the L1 trigger as experimental feature. The upgrade will be discussed in detail in section 2.3 of this chapter.

The events triggered by the L1 trigger are read out from the detectors and passed to the HLT. The HLT has access to the event data of all sub-detectors and is therefore able to fully reconstruct the event. It is also responsible for identifying the regions of interest for reading out the PXD based on the reconstructed tracks. The HLT runs on 6000 CPU cores and reduces the 30 kHz from L1 to 10 kHz which is then stored and archived for analysis [16].

Top-up Injection and Trigger Injection Veto

To maintain high luminosities, new particles are injected into the main rings repeatedly during operation to compensate for particle loss due to background processes and collisions. These so called top-up injections happen every few tens of ms and inject into existing bunches. Since it is not possible to directly inject into the same volume of phase space of an existing bunch, the target bunch has largely increased phase volume at the beginning. It takes multiple milliseconds to cool down the bunch. As a result, the beam background in the detectors is largely increased when the topped-up bunch crosses the IR. These bunches are therefore referred to as noisy bunches.

Start time	Duration	Mode	Description
immediately	0.5 ms	Full veto	no data taking
0.5 ms	4 ms	Gated veto 1	3 µs veto each 10 µs
4.5 ms	10 ms	Gated veto 2	1.5 µs veto each 10 µs
14.5 ms	Until next top-up	Normal operation	no veto

Table 2.1: Trigger injection veto sequence used in previous runs, after each top-up injection from top to bottom.

A staggered vetoing scheme (the trigger injection veto) is employed to ensure physics performance, summarized in table 2.1. During the first 500 µs after injection, the full experiment is vetoed to avoid saturation effects in the detectors due to high noise. This is followed by two phases of gated vetos, where only the time of the noisy bunch crossing is vetoed. Currently, the trigger injection veto is based on time alone, but recent developments allow the use of online data from the CDC and ECL to resume data taking earlier if the bunch is sufficiently cooled in an adaptive way [17, 18].

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

2.3 The Belle II Upgrade Proposal

The SuperKEKB in the current state is expected to reach an instantaneous luminosity of $2 \times 10^{35}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$. To reach the target of $6 \times 10^{35}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$, a redesign of the interaction region including the superconducting final focus quadrupole magnet system (QCS) may be required. To facilitate this, a long shutdown (LS2) is planned around 2032 to install the necessary components to the accelerator and the experiment. To achieve the required vertex resolution, the current VXD is mechanically tightly coupled to the interaction region. Therefore, even small changes in the IR need to go hand in hand with a redesign and replacement of the VXD. This gives the opportunity to upgrade the current system to improve physics performance as well as increased safety margins against beam background. Additionally, having a vertex detector design at least partially ready provides a safety net in case the current VXD gets damaged. The LS2 also provides a window of opportunity for upgrades to other detector systems, which are mostly independent from each other and could be implemented depending on improvement of physics performance, technological readiness and funding. This thesis focuses on the upgrade of the vertex detector [9].

Beam Background

Beam induced backgrounds are the main driver for occupancy in the inner detector systems. Single-beam processes like scattering with residual gas molecules or coulomb scattering between two particles of the same bunch, known as Touschek effect [19], are possible. The scattered particles are kicked out of the stable orbit and create electromagnetic showers at collision with the beam pipe. While beam-gas background is proportional to vacuum pressure and beam current, Touschek scattering is proportional to the beam current squared, and inversely proportional to the number of bunches and beam size. Another factor is synchrotron radiation emitted at the QCS and dipole magnets close to the IR. Due to the square dependence on energy, the main contributor is the electron beam at the HER.

Another contribution considered as beam background originates from the collisions itself. Despite the fact that particle collisions are the main purpose of the SuperKEKB collider, some processes have very high cross section while being of little interest for physics measurements. Main contributors include radiative Bhabha scattering ($e^+e^- \to e^+e^-\gamma$) and the two-photon process ($e^+e^- \to e^+e^-e^+e^-$). These processes increase the occupancy and radiation dose on detectors and due to their linear dependence on luminosity, they are referred to as luminosity backgrounds. The contributions to the beam background have been studied 2021 by operating HER and LER individually and together with different beam currents to disentangle the contributions of individual processes. The results are shown in figure 2.8 [10].

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

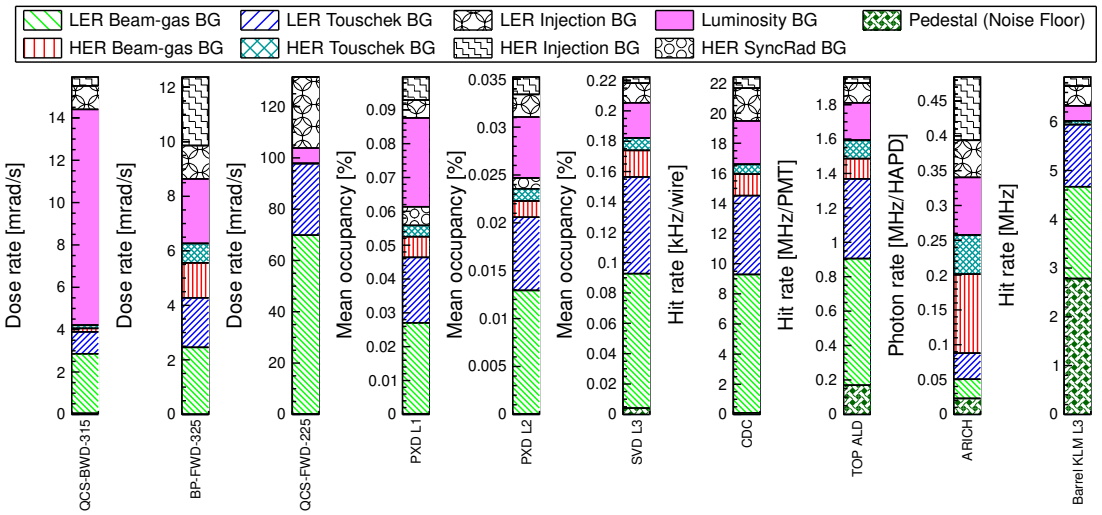


Figure 2.8: Beam background composition from a study June 2021 [10].

The extrapolation of the beam background after LS2 has high uncertainties, since the exact operating conditions and focus system is not yet known. Therefore, three scenarios have been evaluated to cover a range of possibilities:

- Scenario-1: optimistic
- Scenario-2: intermediate
- Scenario-3: conservative.

The current configuration has been measured and simulated, and the quotient of the different contributions, the so called Data/MC ratio have been determined for individual types of background. The resulting values generally stay within one order of magnitude from unity and are used to calibrate the simulations for the target luminosity of $6 \times 10^{35}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$. The scaling used is shown in table 2.2.

Background component	Average Data/MC	Scaling factor Sc.-1	Sc.-2	Sc.-3	Total scaling Sc.-1	Sc.-2	Sc.-3
Beam-gas LER	3.46	2	5	10	6.92	17.30	34.60
Beam-gas HER	0.63	2	5	10	1.26	3.15	6.30
Touschek LER	3.44	2	5	10	6.88	17.20	34.40
Touschek HER	0.18	2	5	10	0.36	0.90	1.80
Luminosity	0.81	1	1	1	0.81	0.81	0.81

Table 2.2: Background component scaling factors for the three scenarios [9].

The estimated background rates at target luminosity are shown in figure 2.9. The solid bars correspond to the default Belle II physics list whereas the crossed bars (non-stacked) include accurate neutron physics. The numbers in the boxes above the diagram correspond to the safety margin for Scenario-2. The safety margin is defined as the maximum rate over the expected background, i.e. the factor of maximally allowed increase of the background in the prediction.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

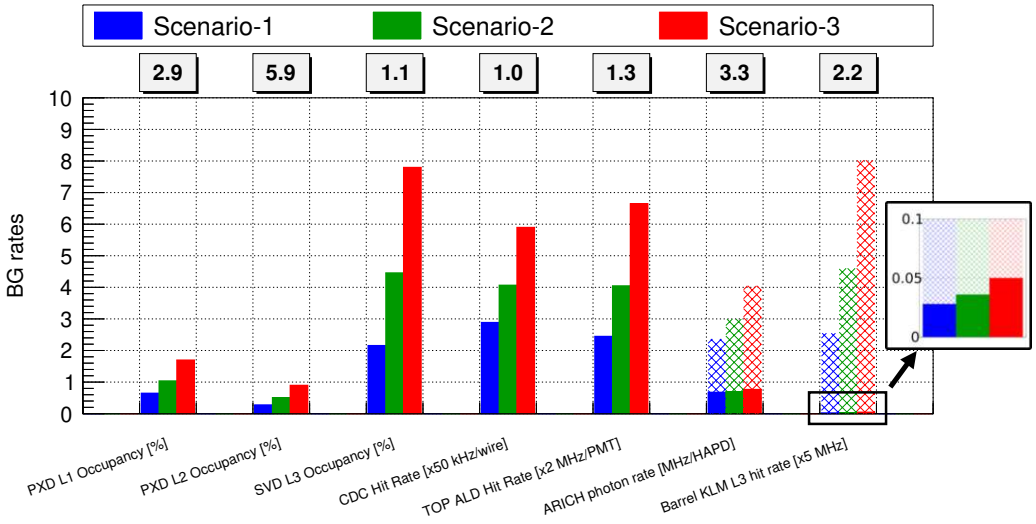


Figure 2.9: Estimated beam background at target luminosity for the current detectors of Belle II, the numbers in boxes show safety margin for scenario-2 [9].

As visible in the figure, three detector systems will operate close to the safety margin: SVD, CDC and TOP. For the TOP, operating outside the safety margin will not impact physics performance but shorten the lifetime of the MCP-PMTs, which possibly can be replaced during a regular summer shutdown. The CDC limits are based on limits defined at Belle/KEKB and might not be applicable for Belle II. This is currently under investigation, to determine the accurate background limits of the CDC. A direct upgrade to the CDC detector is considered too risky, as it is mounted tightly inside the Belle II and a crucial detector system. The upgrade is therefore limited to the CDC readout electronics, the improvements from this are still under evaluation. The upgrade of the VTX foresees that the silicon VTX takes over part of the track finding to mitigate the performance loss of the inner layers of the CDC.

For SVD, the safety factor is 1.1 in the intermediate scenario. The conservative scenario expects an occupancy of almost 8%, which would impair tracking performance. This document focuses on the upgrade of the vertex detector, which increases the safety margin to beam background, and is necessary in case of changes in the interaction region and described in the following [9].

2.4 The VTX Upgrade

Since vertex detector needs to be as close to the IR as possible to achieve best spatial resolution, it is mechanically tightly coupled with the beam pipe and QCS magnets. This means, changes to the interaction region require a redesign and replacement of the vertexing systems. However, even if no change to the IR is needed, an upgrade to the vertex detector would benefit physics performance, greatly increase safety margins and provide a backup plan in case of damage to the current VXD. Therefore, a new vertex detector, the VTX, is in the design process. The VTX is comprised of 5 or 6 straight layers with depleted monolithic active pixel sensors (DMAPS) specifically designed for this purpose. All layers use the same type of sensor, called OBELIX (Optimized BELle Ii piXel sensor) manufactured in Tower 180 nm technology [20].

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

The fully pixelated design improves tracking robustness at large beam backgrounds. Furthermore, an improvement in the tracking of low momentum particles is expected, together with a slight increase in vertex resolution while the other performance benchmarks will stay at the same level as with the current VXD.

The current baseline design foresees 5 layers and is based on the mechanical envelope of the current VXD. When the redesign of the IR is progressed to a pre-final stage, the mechanical VTX design can be adapted to suit the new constraints.

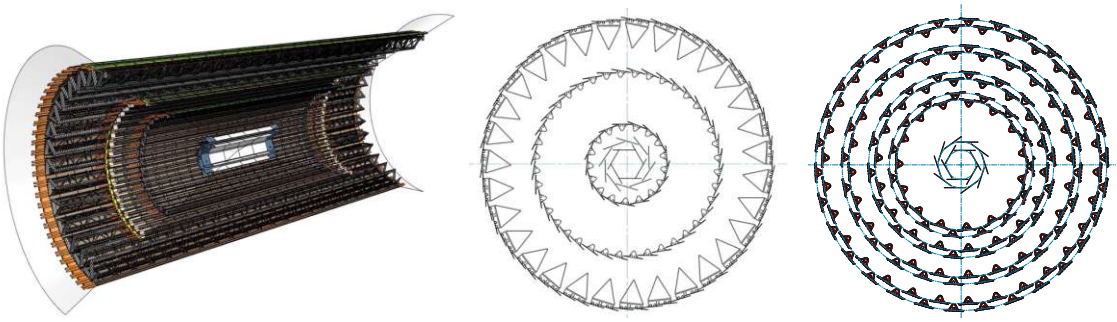
The radius of the third layer (L3) is currently under investigation. Having L3 close to the IR maximizes track impact parameter resolution and this is the current baseline design. However, moving L3 outwards closer to L4 increases the reconstruction efficiency of K_s^0 particles which decay inside the VTX volume. This alternative version of VTX is currently under investigation to evaluate the impact on other physics channels. Preliminary analysis show only a slight decrease in vertex resolution. Furthermore, a 6 layer VTX design is studied, aiming to improve the reconstruction of K_s^0 decays further by adding an additional layer.

2.4.1 Mechanical Structure of VTX

The VTX mechanical structure consists of 5 or 6 barrel shaped layers without tapered ends like SVD for mechanical simplicity. The angle of acceptance of $17° < \theta < 150°$ is maintained for all layers by elongation into the forward side. A CAD drawing of the VTX detector in is shown in figure 2.10, which corresponds to the baseline design with small-radius L3. VTX can be subdivided in an inner part (iVTX) with two layers and an outer part with 3 or 4 layers (oVTX) which have a distinct structure. Table 2.3 summarizes the properties of the individual layers for the baseline design. It has become clear in the mean time, that only air cooling as initially foreseen is not sufficient for iVTX. Therefore, a thin cooling pipe will be included, increasing the material budget to 0.3% X_0 per layer.

An alternative design using staggered layers at oVTX allows the accommodation of an additional 6th layer. Preliminary simulations show that the performance of this 6 layer staggered design is superior in all metrics except the slightly worse vertex resolution due to increased material budget. Even if this layout is new, it appears to become the favored option which mitigates the slight shortcomings of the 5 layer design. Since not all simulations are fully completed yet, the results of the old 5 layer design are shown in the next chapter. Figure 2.10c shows a cross section of the new 6 layer design. As visible, the ladders corresponding to each layer have two radii each. The properties of the new design are shown in table 2.4. The two radii quoted for each layer correspond to the odd and even ladders respectively.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.



(a) (b) (c)

Figure 2.10: 3D rendering of the baseline VTX (a) and cross section (b) [9]. Cross section of the 6 layered design (c).

	L1	L2	L3	L4	L5	Unit
Radius	14.1	22.1	39.1	89.5	140.0	mm
# Ladders	6	10	17	40	31	
# Sensors per Ladder	4	4	7	16	2 x 24	
Hit-rate (Sc-1)	14.1	5.6	1.42	0.25	0.11	MHz/cm^2
Hit-rate (Sc-2)	21.2	9.4	2.42	0.41	0.18	MHz/cm^2
Hit-rate (Sc-3)	34.0	16.2	4.05	0.68	0.28	MHz/cm^2
Material budget	0.2*	0.2*	0.3	0.5	0.8	% X_0

Table 2.3: Summary of the properties of layers for the baseline VTX. *: additional cooling pipe will increase this value to 0.3% X_0.

	L3	L4	L5	L6	Unit
Radii	62.5/69.0	82.5/89.0	108/114.5	133.5/140	mm
# Ladders	30	36	48	60	
# Sensors per Ladder	12	16	20	24	
Hit-rate (Sc-1)	0.42	0.28	0.16	0.11	MHz/cm^2
Hit-rate (Sc-2)	0.68	0.45	0.25	0.17	MHz/cm^2
Hit-rate (Sc-3)	1.12	0.76	0.40	0.26	MHz/cm^2
Material budget	0.6	0.6	0.6	0.6	% X_0

Table 2.4: Summary of the properties of layers for the new 6 Layer oVTX geometry with staggered layers and increased L3 radius.

The iVTX ladders are optimized for the lowest material budget possible, with a target of 0.2% X_0 with air cooling or 0.3% X_0 when a small coolant pipe is included. They consist of all-silicon ladders hosting four OBELIX chips each. The four chips are diced out together from the production wafer, making the ladder a monolithic, mechanically stable unit. The sensors themselves remain independent on the wafer level (no CMOS stitching or alike). Instead, a redistribution layer (RDL) is added to the ladder in a post production step connecting the sensors electrically. It consists of multiple metal layers and polymer based dielectric and is fabricated using a thin film process and photolithography. The RDL allows the placement of passive components, such as decoupling capacitors and a single connection at one end to provide all required power and data links. After RDL processing, the

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

ladder is thinned down inhomogeneously. Under the sensor, a thickness of 40 µm is targeted, while balconies of about 400 µm thickness surrounding the sensors and on the ends of the ladder are left to maintain stiffness and planarity. A sketch is shown in figure 2.11, with the active sensor area shown in blue.

A demonstrator is currently under fabrication to evaluate the production processes of the large scale RDL and inhomogeneous thinning process. It uses plain silicon wafers with heater structures to mimic the expected power consumption of the OBELIX chip. This allows thermal, electrical and mechanical evaluation after a successful production of this first prototype.

Estimates of the power consumption and fluid dynamic simulations suggest that air cooling might not be feasible. An additional small cooling pipe along one side of the ladder for liquid coolant is currently under investigation and would increase the material budget to 0.3% X_0.

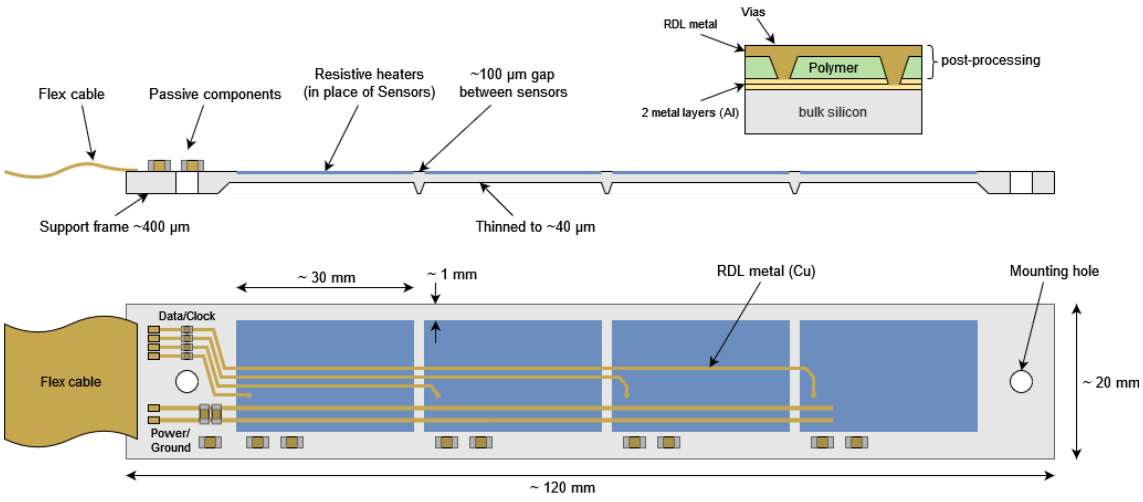


Figure 2.11: All silicon ladder concept for the iVTX [21].

The outer layers follow a more traditional approach inspired by ALICE ITS2 [22] to allow the larger span of the ladders required. The mechanical support is given by a 3D carbon fiber structure and a cold plate. The cold plate is coupled to polyimide pipes for liquid coolant circulation and the sensors are glued to the cold plate on the other side. To prevent the release of coolant inside the VTX volume in case of leaks, it is foreseen to maintain a negative pressure in the whole coolant circuit.

The electrical connections are made with flex PCBs glued on top of the sensors and connected via wire bonding. Layer 3 foresees a single flex circuit connected on the backwards side of the IR (i.e. the direction of the positron beam) to allow for more room in the forward side for machine components. The other layers foresee connection on both sides, with either two (L4) or four (L5 in baseline version) flex PCBs on each ladder.

Figure 2.12 shows the structure of the oVTX ladders. Two versions of L3 are shown, because the demands for rigidity differ depending on the length and therefore radius. The active area of the OBELIX chip is shown in yellow, whereas the periphery of the chip is shown in green. For the 5th layer, the baseline VTX 5 layer design is shown. The staggered version uses ladders of similar structure to L3 and L4 [9, 21].

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

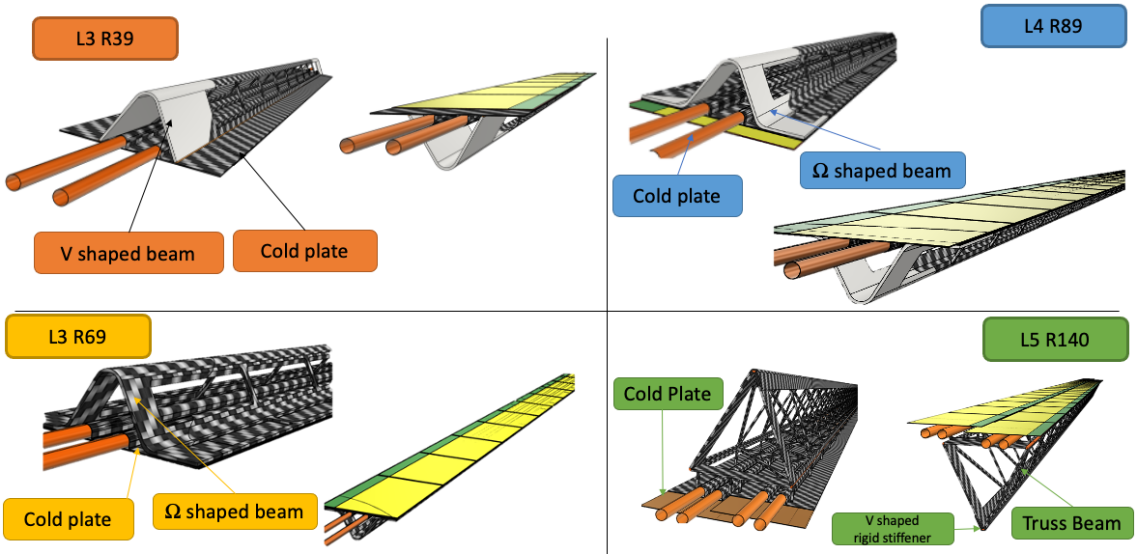


Figure 2.12: Ladder concept for the oVTX ladders, the L5 is from the baseline 5 layer design [21].

2.4.2 Physics Performance

One of the main features of the VTX upgrade is to maintain tracking performance at high backgrounds. In previous runs, the main tracking and especially track finding was done by the CDC. However, already currently, the CDC suffers from higher than expected occupancy. A replacement of the CDC is not foreseen, because of the risks attached to a modification. An upgrade to the electronics is expected to restore the efficiency, but larger improvements can be made on the algorithmic side of the analysis. The upgrade plans to use the silicon vertex detector as a main track finding detector and the use of CDC data as a second stage. Figure 2.13 shows the expected improvements when these changes are implemented with the current SVD (SVD+CDC) and the VTX detector (VTX+CDC) for the three background scenarios along with the current algorithms (CDC-only). The improved spatial and temporal resolution of the VTX would improve tracking efficiency, even in the conservative scenario only a small drop in efficiency is visible compared to the optimistic scenario.

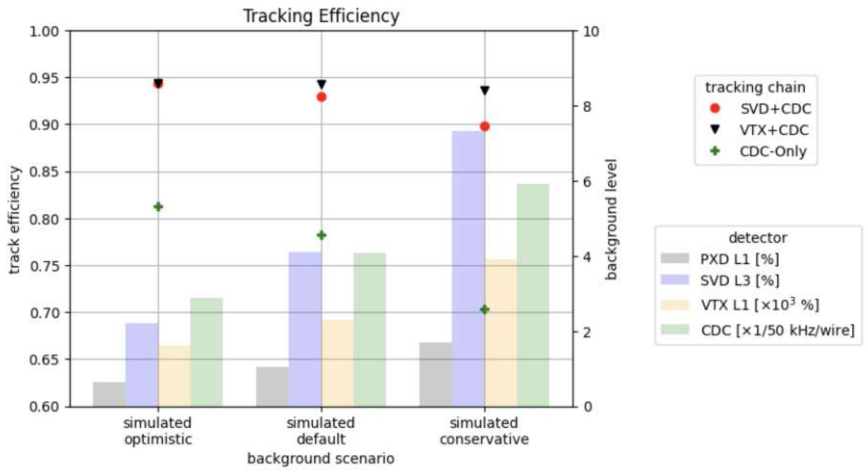


Figure 2.13: Occupancies for the three different background scenarios and corresponding general track efficiency [9].

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

	Current	VTX (5L baseline)		
	Sc-2	Sc-1	Sc-2	Sc-3
Generated events	32533	32559	32559	30255
Correctly reconstructed	10059	16913	16848	15583
Self-crossfeed	28495	51375	51826	47527
Efficiency	30.9%	51.9%	51.7%	51.5%
Purity	26.1%	24.8%	24.5%	24.7%

Table 2.5: Summary of reconstruction performance for the $B^0 \to D^{*-}\ell^+\nu$ decay chain, with $D^{*-} \to \overline{D}^0\pi^-_{\text{soft}}$ and $\overline{D}^0 \to K^+\pi^-$ for different background scenarios [9].

Table 2.5 shows the results of a simulation of the decay chain $B^0 \to D^{*-}\ell^+\nu$ with $D^{*-} \to \overline{D}^0\pi^-_{\text{soft}}$. The modes $\overline{D}^0 \to K^+\pi^-$ and $\overline{D}^0 \to K^+\pi^-\pi^+\pi^-$ have been studied, but only the first is shown in the table while the second mode has consistent results. The improvement of the results including VTX show a factor of almost 1.7 in efficiency in comparison with the current detectors, at a slightly improved purity of events. At the same time, the efficiency is robust against all three background extrapolations. The main reason for this is the enhanced efficiency for the reconstruction of π^-_{soft} candidates, especially in the low transverse momentum range.

The vertex resolution of the current detector and VTX has been studied using the $B^0 \to J/\psi K^0_s$ channel by analyzing a generated sample of 1 million events. The z-vertex resolution shows a moderate improvement from 23.0 µm with VXD to 14.4 µm with VTX baseline. The 6 layer design has a slightly worse vertex resolution of 16.1 µm in comparison with the 5 layer design due to the increased material budget of the additional layer. Those numbers are from the intermediate scenario, but show little dependence against beam background.

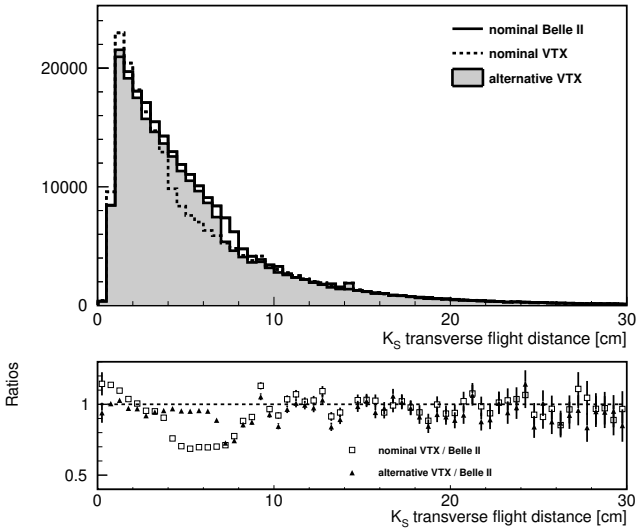


Figure 2.14: Number of reconstructed tracks (top) and VTX/current detector ratio (bottom) versus flight distance of K^0_s, background scenario 2 [9].

The same simulation sample has been used to study the reconstruction efficiency of the K^0_s decay products. Figure 2.14 shows the absolute number of events reconstructed against the transverse flight distance of the K^0_s before decay. The nominal VTX (L3 close to IR, baseline design) has a pronounced falloff at position of L3 with

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

a radius of 3.9 cm. In the alternative VTX scenario with an L3 radius of 6.9 cm, a slight drop in efficiency is visible, but less pronounced than in the initial VTX design. Beyond about 10 cm, the reconstruction is dominated by the CDC and shows little dependence on the vertex detector design.

Overall, the current vertex detector shows the best efficiency for K_s^0 reconstruction and the alternative VTX design shows a total efficiency about 5% lower. The main reason for this is the fact that the VTX only has 5 sensor layers, whereas the VXD has 6 layers. For this reason, the 6 layer VTX design is currently under investigation, which aims to improve K_s^0 reconstruction efficiency [9].

As an additional feature, the outer layers of VTX can contribute to the L1 trigger decision. Currently, only the CDC is used to build tracks at the L1 stage. At high backgrounds, this translates to a reduced trigger efficiency, which can be restored with the VTX trigger contribution. This feature of VTX is fairly new, preliminary plans together with the corresponding capabilities of the OBELIX chip will be described in section 5.11. A detailed simulation study for the VTX trigger contribution is yet to be made, but tentative simulations show acceptable results already with a VTX-only trigger. The complex interplay of a combined VTX+CDC trigger which would further improve trigger efficiency and purity is still to be investigated.

2.4.3 Sensor Requirements

The initial requirements for the sensor used for the VTX upgrade lead to the use of the TJ-Monopix2 as starting point for OBELIX development. The initial sensor specifications for the VTX upgrade are shown in table 2.6 along with the final specification of the OBELIX design. To achieve the required vertex resolution below 20 µm for VTX, which is comparable to the current vertex detector, a spatial resolution of about 12 µm is required with the proposed geometry as early simulations suggested. With the OBELIX chip of a pixel pitch of 33.04 µm, a resolution of $33.04\,\mu m/\sqrt{12} \approx 9.5\,\mu m$ is possible even without the use of charge information. To keep the material budget low, a thin sensor is required. The sensitive layer of 30 µm epitaxial silicon used for OBELIX meets this requirement. This thin active layer also enables the the possibility of thinning of the sensor down to 40 µm, given a suitable mechanical support structure, to achieve low material budget.

The conservative scenario after the Belle II upgrade estimates a hit-rate of $34\,MHz/cm^2$ for the innermost layer. An additional safety factor of 3.5 was included to set the hit-rate capacity requirement to $120\,MHz/cm^2$ for continuous operation. This allows sufficient margin, e.g. for potentially noisy pixels and injection background. While the chip should be designed to operate at an average of $120\,MHz/cm^2$, it should also withstand short pulses of increased hit-rate without saturation effects. While a prediction of the injection background is not possible at the current stage, a rate of $600\,MHz/cm^2$ for sub-microsecond bursts was defined as design target.

While the current L1 latency is 5 µs at Belle II, The trigger latency for VTX should be above 10 µs to allow for more complex trigger algorithms at the L1 stage in the future. The OBELIX chip has a configurable latency and while it is optimized for 10 µs, larger latencies are possible with minor performance degradation.

Likewise, the current trigger rate at Belle II is 30 kHz (Poissonian distributed), but an increased maximum rate would relax the requirements for the L1 trigger. Simulations show that OBELIX is capable of trigger rates up to 100 kHz with small efficiency penalties or significantly higher if the hit-rate is smaller than the design value.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

An integration time below 100 ns is required to keep the occupancy sufficiently low and perform the track finding starting from VTX. The OBELIX integration time is an integer multiple of 47.1 ns arising from the timestamp structure. Therefore, it is currently planned to integrate two successive timestamps or 94.2 ns, but this is configurable.

The VTX is designed to operate for 10 years and the inner layer are expected to receive a dose below 10 Mrad/year and a 1 MeV neutron equivalent fluence below $5 \times 10^{13}\,\mathrm{n_{eq}/cm^2/year}$. This defines the required radiation tolerance of the sensors.

Tests of TJ-Monopix2 confirmed the feasibility to meet the OBELIX specification. In particular room temperature operation was confirmed after irradiation, even if the chip was initially designed to operate at negative Celsius temperatures after irradiation.

Additionally, the timing performance of the so called HitOr lines, initially included only for testing purposes have been characterized carefully since they are used for the L1 trigger contribution and precision timing mode in OBELIX. A timing resolution below 3 ns is reachable with suitable offline corrections. This is important for the event T0 determination, currently done in SVD and possible to be kept with the VTX. However, switching on this optional feature increases power consumption and only works at lower rates, which limits its use to oVTX. A resolution of 10 ns of a VTX L1 trigger is possible when data from multiple VTX layers is combined. More detail about the measurements on TJ-Monopix2 will be shown in chapter 4. The technical details about the OBELIX design follow in chapter 5. The following chapter will focus on silicon detector technology.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

	Initial Specification	OBELIX
Spatial resolution	$< 12\,\mu m$	$< 9.5\,\mu m$
Pixel pitch	$< 40\,\mu m$	$33.04\,\mu m$
Sensitive layer thickness	$< 50\,\mu m$	$30\,\mu m$ epitaxial Si
Sensor thickness	$< 100\,\mu m$	thinning to $40\,\mu m$
Sensitive area	$\sim 30 \times 16\,mm^2$	$29.60 \times 15.33\,mm^2$
Sensor dimensions	$\sim 30 \times 19\,mm^2$	$30.17 \times 19.36\,mm^2$
Hit rate capability: burst	$> 600\,MHz/cm^2$	$> 600\,MHz/cm^2$
Hit rate capability: full chip	$> 120\,MHz/cm^2$	$120\,MHz/cm^2$, slow degradation above
Trigger latency	$> 10\,\mu s$	3 to $12\,\mu s$
Trigger rate	$> 30\,kHz$	$> 100\,kHz$
Overall integration time	$< 100\,ns$	47.1 ns, multi trigger possible
Time resolution (optional)	$< 50\,ns$	3 ns in oVTX
Total ionizing dose tolerance	100 Mrad	100 Mrad
NIEL fluence tolerance	$5 \times 10^{14}\,n_{eq}/cm^2$	$5 \times 10^{14}\,n_{eq}/cm^2$
SEU tolerance	frequently (min^{-1}) flash configuration	up to 10 Hz f. conf., resynchronization on the fly possible
Powering	through voltage regulators	on chip LDOs, $\geq 0.2\,V$ drop-out
Outputs	one at $\leq 200\,MHz$	one at 339 MHz
L1 trigger contribution	—	one extra output at 339 MHz
Max. operating temperature	room temperature	$> 40\,°C$

Table 2.6: Initial specifications for the VTX sensor and final OBELIX specification, partially from [9].

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

Chapter 3

Silicon Particle Detector Technology

In our modern world, semiconductor microchips are of uttermost importance. Advances in production processes of CMOS devices start to peek into the atomic regime with single digit nanometer scales. In high energy physics, silicon is a popular choice for radiation detection, part to the fact that commercial production is possible. As this thesis covers silicon detectors, this chapter will focus on the physical processes inside those sensors and current technologies used in particle physics.

Position sensitive semiconductor sensors are in use at high energy physics since the 1980ies. This resulted in a leap forward in measurement quality. Typical sensors provided a higher spatial resolution by two orders of magnitude in comparison with the wire chambers typically used before. This enabled the lifetime measurement of heavy fermions via secondary vertices, also exploited in the Belle II experiment [23].

3.1 Semiconductor Basics

In crystalline solids, the discrete electronic states of the individual atoms merge together and form so-called energy bands with a quasi continuum of allowed states, separated by forbidden energy regions. In respect to conductivity, two bands are of special interest: the valence and the conduction band. The valence band is the highest filled band at zero temperature while the conduction band is the lowest-energy empty band.

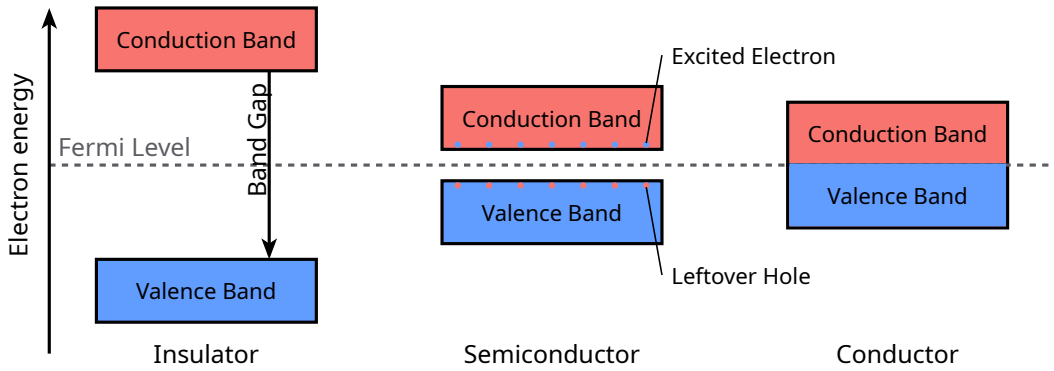


Figure 3.1: Simplified structure of electron bands close to the fermi level in solids.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

As shown in figure 3.1, the energy levels of those two bands define the electrical nature of the material and it can be categorized in three types: conductors, semiconductors and insulators. In case the bands do not overlap, the difference between the highest energy state of the valence and the lowest energy state of the conduction band is called band gap.

In conductors, the valence band overlaps with the conduction band or the valence band is only half filled. Therefore, infinitesimal energy is required to move electrons between individual states of the bands corresponding to different momenta or spatial distributions. This means that small electric fields can create a current, the key feature of a conductor.

On the other extreme, if the band gap is above a few electron volts, the thermal energies of room temperature are not sufficient to lift electrons from the valence to the conduction band. This means the valence band is fully populated and the conduction band is empty. No state transitions and therefore charge carrier movement can occur at electric fields because all states below the band gap are occupied and none are occupied above and they are separated by the band gap. Therefore there is no conduction of electricity.

Semiconductors are in the region between conductors and insulators. They have a band gap, but thermal excitation can kick some electrons to the conduction band. This results in a resistivity in the range of $10^{-3} - 10^{8}\,\Omega$cm at room temperature and the typical negative temperature coefficient of semiconductors.

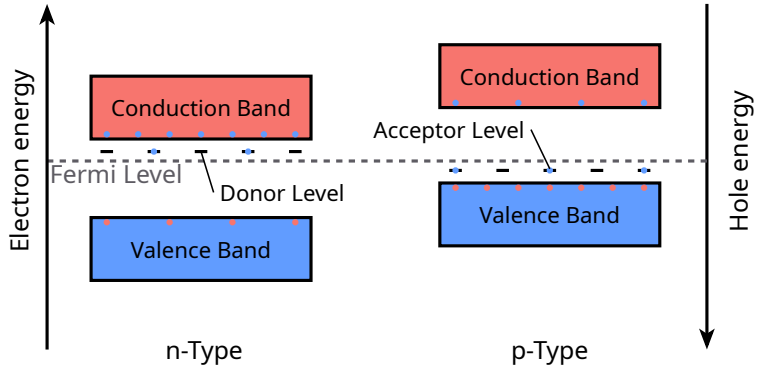


Figure 3.2: Band model with donor and acceptor states in doped semiconductors.

The full potential of semiconductors can be unlocked by artificially adding impurities to the crystal lattice in a process called doping. For silicon, each atom in the lattice shares an electron pair with four neighboring atoms. When one silicon atom gets replaced by an group V element like Phosphorus or Arsenic, an additional electron is attached to that atom. This forms a so called donor state, where the electron can be freed from the donor atom with little energy, increasing the concentration of electrons in the conduction band, as shown in figure 3.2. The leftover positive charge from the remaining dopant atoms are bound in the lattice and therefore not mobile. A decrease in the hole concentration is caused by recombination, and the material is called n-type due to the dominance of electrons as charge carriers. The recombination is similar as in chemical reactions and the mass action law:

$$n \cdot p = const., \tag{3.1}$$

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

with n and p the concentrations of electrons and holes respectively holds true for charge carrier concentrations in doped silicon. Due to the additional states close to the conduction band, the fermi level is shifted upwards in energy.

When group III elements such as Boron or Aluminum are used for doping, they can only bond with three neighbors. The resulting missing electron, the acceptor level, can be filled by an electron from another bond, effectively creating a mobile positive charge (a hole) leaving a negatively charged dopant atom in place. These at zero temperature unpopulated states shift the fermi level down in energy. In such a material, the majority charge carries are holes, electrons have a reduced concentration due to recombination. For typical dopants, almost all acceptor states are filled and donor states are excited at room temperature [24, 25].

3.1.1 The pn Junction

When a p- and n-type semiconductor are brought into contact, the corresponding majority charge carriers diffuse into the opposite side and recombine, i.e. the fermi level equalizes. This recombination just leaves the p-dopant atoms negatively charged and the n-type atoms positively charged, creating a potential barrier for the diffusion. Therefore, a region of finite thickness with no charge carriers forms, the depletion region. It has a spatial charge proportional to the doping concentration, and sometimes also called space charge region.

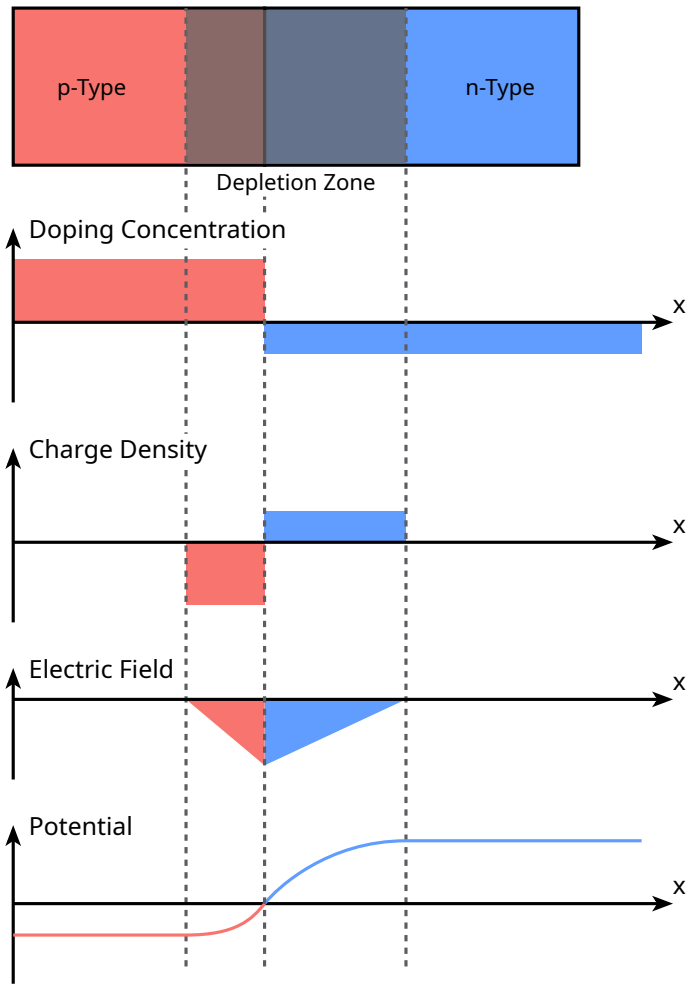


Figure 3.3: The pn junction and the depletion zone in equilibrium.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

The total charge needs to remain neutral, and therefore the width of the depletion zone parts on the n-side and p-side are inversely proportional to the respective doping concentrations. In particle detectors, the doping concentrations are typically asymmetrical, leading to a depletion region of a large volume in the low doped part and almost none in the highly doped one.

The simplest case of a planar pn junction with corresponding electrical quantities is shown in figure 3.3. The constant charge density in the depletion region creates an electric field linear in longitudinal direction and a kink in field strength as the boundary is crossed.

The potential that forms is quadratic in nature and different on the n and the p side of the junction, as it is the (negative) integral of the electric field. If an additional external potential is applied, the horizontal parts at the ends of the junction of the potential shift, and the depletion region either increases or decreases in width, depending on polarity to compensate for the potential difference. If the depletion region becomes thin enough, charge carriers can tunnel through the barrier, allowing the flow of current. This is exploited in the diode, an electric component that allows current flow in only one polarity and is based on a pn junction.

If a reverse bias voltage V is applied, the space charge density remains constant (defined by dopant concentration) but increases in width to maintain equilibrium. Assuming a highly asymmetrical doping and a bias voltage much greater than the intrinsic potential difference, the potential ϕ remains quadratic in distance x through the whole depleted region:

$$\phi(x) \propto x^2. \tag{3.2}$$

For the width of the depletion region h, this becomes:

$$V = \phi(h) \propto h^2 \Rightarrow h \propto \sqrt{V}, \tag{3.3}$$

with the depletion region proportional to the square root of bias voltage. This also remains true for the general case of arbitrary doping concentrations. Rigorous and more detailed derivations can be found in [23, 26].

The leakage current through the depleted region is mainly caused by electron hole pair creation inside. This is strongly influenced by presence of impurities creating states inside the band gap, so called generation and recombination centers. The leakage current I_L is proportional to the volume V of the depleted region via:

$$I_L = eV\frac{n_i}{\tau_g}, \tag{3.4}$$

with the charge carrier generation lifetime τ_g and the intrinsic charge carrier concentration n_i. In thermal equilibrium n_i is proportional to $T^{3/2}exp(-E_G/2kT)$ and $\tau_g \propto T^{-1/2}$ when assuming the main cause of current arises from impurity and defect levels. Typically the band gap is also re-parametrized to an effective activation energy $E_a \approx 1.21\,\text{eV}$ leading to:

$$I_L \propto T^2 exp\left(-\frac{E_a}{2kT}\right). \tag{3.5}$$

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

3.2 Interaction of Particles with Matter

3.2.1 Heavy particles

When particles traverse matter, they loose kinetic energy via interactions with the atoms in the medium. For heavy charged particles, the main contributions for this are due to ionization, excitations of atoms and bremsstrahlung. In principle, these are stochastic processes, but for macroscopic objects, the average energy loss per length can be used as approximation, usually called the linear stopping power.

The stopping power is dependent on the properties of the particle as well as on the material it is traversing. Since the density of atoms in the material is directly affecting the probability of the interactions and therefore the energy loss, the so called mass stopping power is usually given with normalized mass density to allow universal use. This also allows compound materials to be described as sum of the stopping power contributions with corresponding densities for the individual chemical elements with reasonable accuracy.

Depending on particle mass and momentum, different interaction processes dominate causing distinct regions in the stopping power. For muons in copper, the stopping power is given in figure 3.4 for 9 orders of magnitude of momentum.

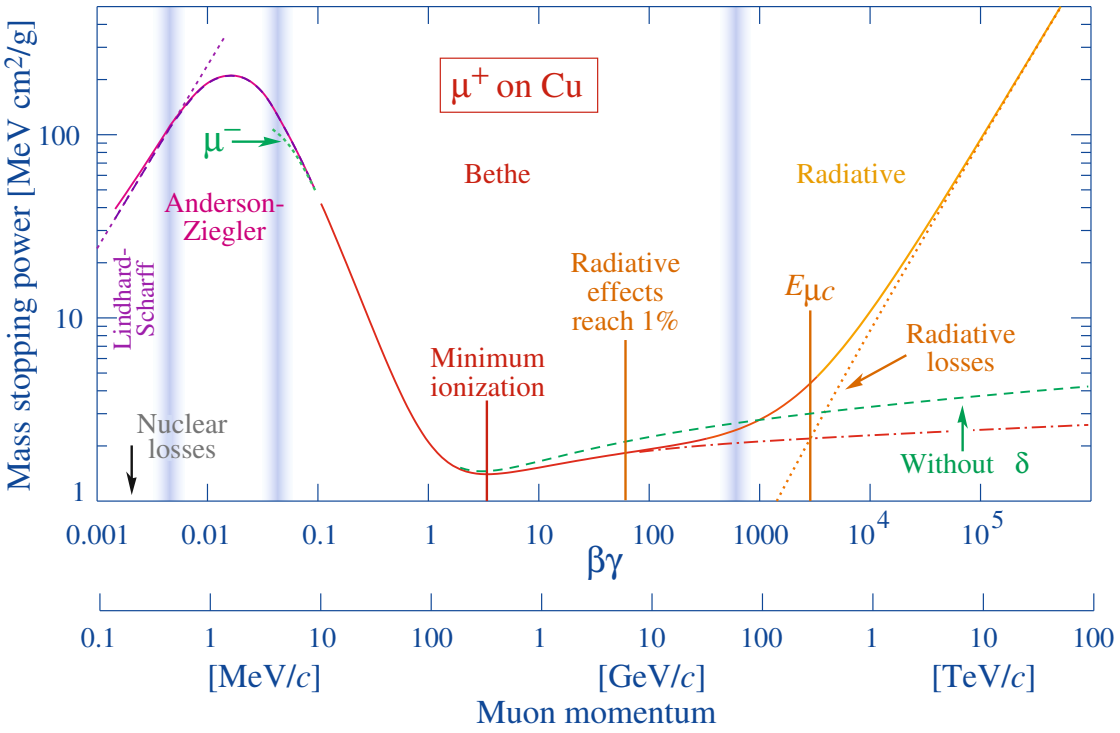


Figure 3.4: Stopping power for muons in copper [27].

In high energy physics experiments, the range between $0.1 \leq \beta\gamma \leq 1000$ is of the highest interest as this is the energy range of collision products in today's experiments. In this region, the mass stopping power $\left\langle -\frac{dE}{dx} \right\rangle$ for particles heavier than the electron can be described with the Bethe-Bloch equation:

$$\left\langle -\frac{dE}{dx} \right\rangle = Kz^2 \frac{Z}{A} \frac{1}{\beta^2} \left[\frac{1}{2} \ln \left(\frac{2m_e c^2 \beta^2 \gamma^2 W_{\max}}{I^2} \right) - \beta^2 - \frac{\delta(\beta\gamma)}{2} \right], \tag{3.6}$$

with:

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

- $K = 4\pi N_A r_e^2 m_e c^2$, where N_A is Avogadro's number, m_e is the electron mass and r_e is the classical electron radius,
- z the charge number of the incoming particle,
- Z and A the atomic number and mass of the medium,
- W_{max} the maximal energy transfer to an electron in a single collision,
- δ the density correction factor for high energies.

For intermediate atomic numbers Z, it can describe the mass stopping power up to an accuracy of a few percent. At the lower end of the Bethe-Bloch region, the $1/\beta^2$ term is dominating. At higher energies, the $ln(\gamma)$ term takes over. In the middle, a minimum is formed at around $\beta\gamma \approx 3$. Particles with this momentum are generally called minimum ionizing particle (MIPS). In particle physics experiments, MIPS pose as the worst case of signal amplitude for detectors and are therefore of special interest. The typical energies in experiments are above the MIPS, but since the $ln(\gamma)$ term increases only slowly with momentum, the energy transfer is typically close to the MIPS case for most particles in experiments.

The stochastic nature of the interaction becomes apparent at thin detectors, such as silicon sensors with thicknesses of $O(100\,\mu m)$. Since collisions can have a high energy transfer but with a low probability, the distribution contains a tail in the high transfer range as visible in figure 3.5. This energy loss can be described with the asymmetric Landau distribution. Since the mean is very sensitive to details of the high energy tail, the most probable value (MPV) i.e. the peak position is quoted typically [27, 23].

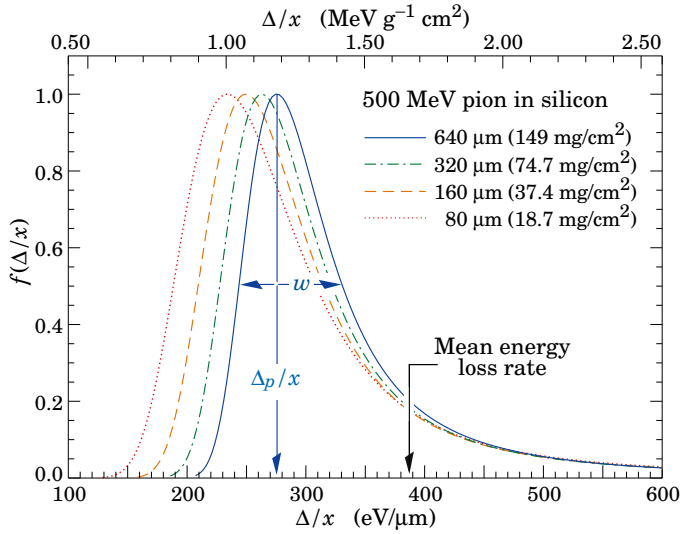


Figure 3.5: Energy transfer distribution in silicon for 500 MeV pions at different material thicknesses [27].

3.2.2 Electrons and Photons

For high energy electrons, the main effect for energy loss is bremsstrahlung in the electric field of the nuclei of the medium. For high energy photons on the other hand, e^-e^+ pair production is the main dissipation mechanism. The characteristic length of both of these processes is the radiation length X_0 in length units or

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

normalized by the density. It is the mean distance of an energy loss of e^{-1} of the electron via bremsstrahlung. At the same time, it is $\frac{7}{9}$ of the mean free path for high energy photons for pair production. This makes it the characteristic length scale for the whole electromagnetic showers in electron-calorimeters, where electrons create photons and photons create more electrons. Figure 3.6 shows different processes normalized to inverse radiation length at 1 GeV in lead.

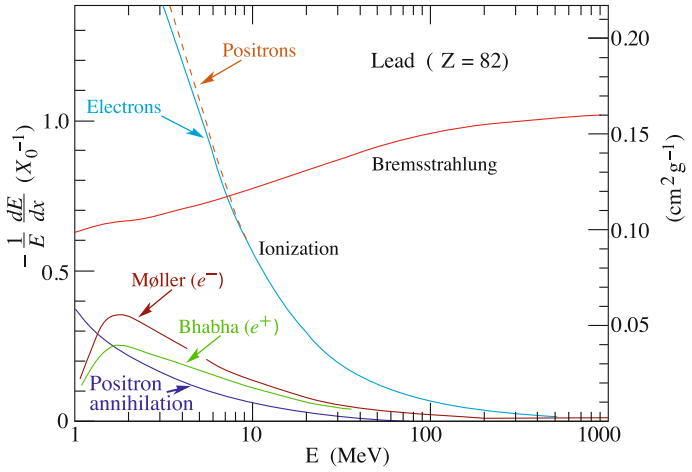


Figure 3.6: Energy loss due to different processes in units of inverse radiation length for electrons and positrons in lead [27].

3.2.3 Radiation damage

When a sensor is subjected to ionizing radiation of sufficient energy, it can cause damage to the device. For silicon devices, the type of damage can be separated into two mechanisms that can be described with one coefficient each to a good approximation: non-ionizing energy loss (NIEL) and total ionizing dose (TID). The NIEL affects the sensing element itself and degrading the performance while TID affects electronic components via changing properties of transistors.

Total Ionizing Dose

The total ionizing dose, TID for short measures the the total amount of energy deposited in the sensor. This affects the silicon oxide on top of the bulk. Electron hole pairs generated in the oxide by ionization are separated by the electric field of transistor gate bias of the electronics when it is switched on. The holes accumulate in the oxide close to the interface to the bulk, creating a net positive charge that shift the characteristics of the transistors. Furthermore, this effect is dose rate dependent due to potential recombination of electron hole pairs.

As a result, transistor characteristics slowly change when they are irradiated. The design of the electronic circuits need to take this into account, providing sufficient safety margin for correct circuit operation even for suboptimal transistor conditions. Logic gates may become slower with TID irradiation, and the behavior of transistors in the analog amplifier may change.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

Non Ionizing Energy Loss

The other mode of damage can be characterized by the so called non-ionizing energy loss, the NIEL. Massive particles interacting with the crystal lattice can create defects in the structure. The so called NIEL hypothesis states that by far the dominant process for non ionizing energy loss is the displacement of lattice atoms. Since this affects the volume of the silicon, this is referred to as bulk damage. Although different types of particles cause different types of defect distributions at different energy scales, the total effect of the damage can be quantified in a single number to a good approximation. Per convention, the scaling of the NIEL is chosen to the equivalent damage caused by a number of neutrons with 1 MeV of kinetic energy.

Some of the atomic displacements are reversible when the displaced atoms migrate back to their original position, while occasionally stable defects are created. These migrations are generally called annealing and the rate increases with temperature. While annealing initially reduces the effect of radiation damage, an optimal annealing is reached after 80 minutes and 60 °C for silicon. After this point, the so called revere annealing sets in which worsens the effects of radiation damage again. The effective doping concentration of silicon bulk material is shown in figure 3.7 as a function of annealing time to illustrate this.

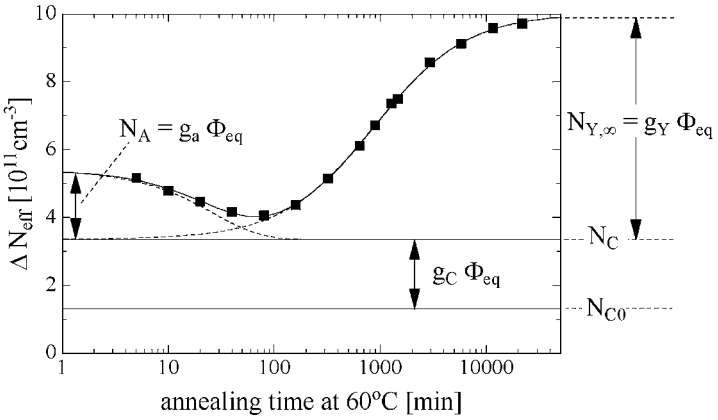


Figure 3.7: Effective doping concentration in relation to annealing time [28].

The bulk damage introduces new energy levels inside the band gap. One one hand, this increases the leakage current through the depletion zone, since the excitation of electron hole pairs does not need to overcome the full band gap at once. Instead, an electron (or hole) can jump to an intermediate energy level forming a mobile hole. At a later time, it can jump to the conduction band, completing the charge carrier pair. Due to the exponential decay of the thermal energy distribution at the high energy side, this greatly increases the chance of pair generation when states are available in the middle of the band gap.

For silicon, the NIEL hypothesis is a good first order approximation and the number of states inside the band gap and therefore increase in leakage current scales linearly with the fluence ϕ_{eq} and the depleted volume V:

$$I_L = \alpha \phi_{eq} V. \tag{3.7}$$

The conversion factor α, also called damage rate, can be determined by measurement and is generally temperature dependent and depends on annealing. Apart from these

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

two dependencies, it is a universal constant and for optimal annealing and 20 °C a value of:

$$\alpha_{80\,\mathrm{min}/60\,^{\circ}\mathrm{C}} = 4.0 \times 10^{-17}\,\mathrm{A/cm} \pm 5\%, \tag{3.8}$$

is stated in the literature [23].

The second, related effect is recombination. When charge is generated by a particle, ideally all of it is collected at the measurement electrode to achieve a large signal. However, electrons can fall into the intermediate energy levels and become unavailable for charge transport, i.e. trapped. At a later time the electron can be released again, but then it is too late to be useful for signal generation, effectively only contributing to leakage. This so called trapping reduces the charge collection efficiency and is proportional to the time the charge carriers are traveling inside the bulk with characteristic decay. For silicon and a fluence of $1 \times 10^{15}\,\mathrm{cm}^{-2} n_{eq}$, the characteristic length λ_t is in the range of 100 µm with typical lelectron drift velocities of 50 to 100 µm/ns [23, 28].

3.3 Technologies of Silicon Radiation Detectors

The first silicon sensors were developed as a replacement for the traditional wire chambers in the 1980ies, based on a depleted silicon material. When a charged particle traverses the depletion zone, it excites electron hole pairs along the trajectory. The electrons and holes are separated by the electric field and drift towards the positive and negative electrodes applying the depletion voltage. This movement temporarily creates a current flow, which can be detected. The number of charge carriers, i.e. the integral of the current can be used to quantify the amount ionization that happened at the interaction.

Strip sensors

One of the first silicon sensor technology used was the silicon strip sensor. It consists of an array of long but thin sensing elements (strips), not unlike in wire chambers, effectively giving one dimensional position information. The doped strip regions are typically implanted in a high resistivity substrate of opposite doping to allow full depletion below the strips. A metal layer on top for the readout can either be coupled directly to the doped region (DC-coupling) or via a dielectric like silicon oxide (AC-coupling) to remove the voltage offset for depletion from the readout. Each strip also needs to be connected to the bias voltage source via a resistor in the mega ohm range or silicon structures creating a high ohmic connection. This is not shown explicitly in the example illustration 3.8 for simplicity. Since typically a full depletion is desired, the backside contains a thin layer of increased doping of same polarity of the bulk. This prevents the depletion region and therefore electric field to reach the backside of the sensor, which would cause breakdowns at slight imperfections of the surface. For the same reason, the whole sensitive area is typically surrounded by guard rings, that shape the field at the edge of the sensors and define an end of depletion.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

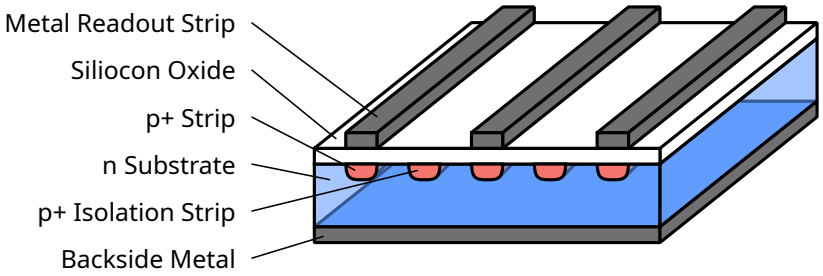


Figure 3.8: Illustration of an AC coupled silicon strip sensor.

For sensors with n-type bulk, an isolation strip as shown in figure 3.8 is optional, but can reduce cluster size. For a p-type bulk, a p-doping surrounding the n-strips is a requirement. This is because some positive charges are always trapped in the oxide due to the manufacturing process. These could form an inversion layer in the high resistivity p-substrate, effectively shorting all strips. The p doping, also referred to as p-stop prevents this by the increased doping concentration.

The readout electronics can be situated on one side of a strip sensor, allowing a flexible mechanical integration. If 2D spatial resolution is required, a second sensor can be situated on top in a rotated orientation. The angle between the two sensors can be small for maximum resolution in one axis (stereo sensors) or 90 degrees for symmetric resolution. When a particle traverses the two sensors, it activates strips on both sensors, allowing the reconstruction of the position in both axes of the sensor plane.

An improvement to this design is the double sided silicon strip detector (DSSD). Here, both sides of the same bulk material are implanted with doped strips of opposite polarity. Therefore, this integrates the two rotated strip sensors on the same wafer where positive charges are collected on one side and negative charges on the opposite side of the sensor. This type of sensor is also used in the Belle II SVD and a sketch is shown in figure 3.9.

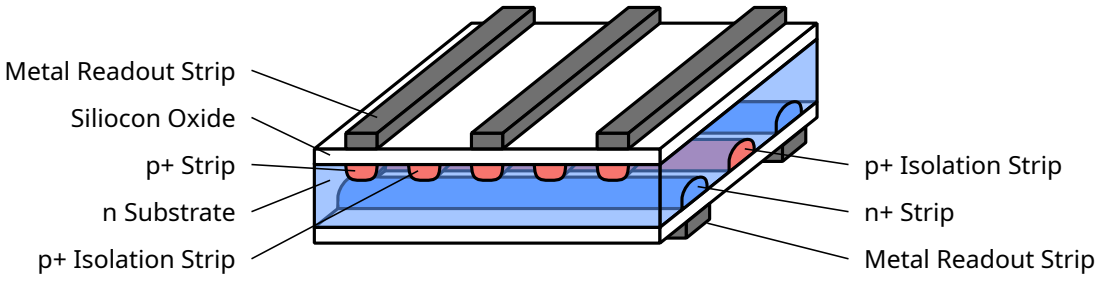


Figure 3.9: Illustration of a DSSD sensor similar to the SVD sensors.

An inherent property of strip sensors becomes apparent at high particle rates. The position is only inferred from the activated strips. If two particles cross the sensor within the time resolution, two by two strips get activated, and a reconstruction is not possible as illustrated in figure 3.10. The typical approach is to include all four hit candidates, where two are so called ghost hits, not corresponding to physical

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

particles. The ambiguities need to be resolved during tracking, which complicates the algorithms and increases computation time while potentially reducing efficiency.

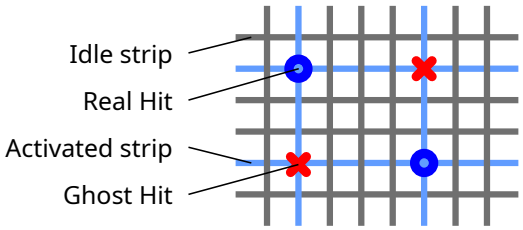


Figure 3.10: Ghost hits in DSSDs

Hybrid Pixel Sensors

Because of the ghost hit problem, pixelated sensors are favored at high rate environments. With true 2D resolution, not only the ghost hits disappear, but also the higher number of channels reduces occupancy.

The traditional approach of pixel sensors is the hybrid sensor sketched in figure 3.11. Here, the sensing element is mounted back to back with a readout chip and electrical connections are made via bump bonding, a 2D array of small metal spheres. This allows flexible design-processes, with the readout chip effectively decoupled from the sensor design. The bump bonding process, however, is difficult to control, costly, and always subject to a yield. Furthermore, the material budget is always limited, even if the chips are thinned, since mechanical structure needs to be maintained for both chips and the bump bond itself requires a metal sphere of significant size. In addition, the pixel pitch is limited by the bump bonding process. Due to their radiation hardness, hybrid sensors are widely used, for example in the inner layers of ATLAS and CMS trackers [29, 30].

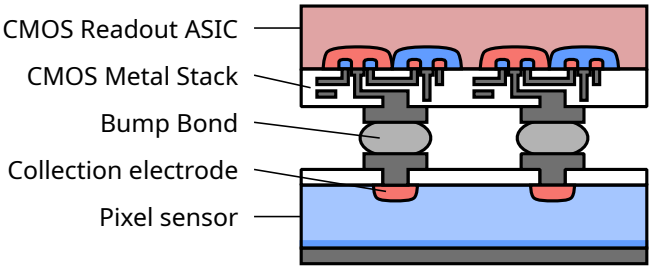


Figure 3.11: Two pixels of a hybrid pixel sensor, CMOS wells not to scale.

MAPS and DMAPS

Since both, the sensor element and the readout electronics of hybrid sensors are made from silicon, an obvious improvement would be integration into one chip. However, there are several engineering challenges associated with this modification. Only the availability of an additional extra deep p-well allows a sufficient isolation of the sensor from the readout transistors. When these became available on commercial CMOS processes, monolithic active pixel sensors (MAPS) were developed,

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

for example MIMOSA26 [31, 32]. At that time, only low resistivity substrates were available and the chips were limited to low bulk voltages. To increase depletion, a thin epitaxial layer with higher resistivity was grown on top of the bulk before fabrication.

Those chips could still not deplete the full sensing volume and partially relied on charge diffusion in the bulk as shown in figure 3.12. The fill factor, i.e. the fraction of the projection of the depleted zone over the pixel area was low. As a result, the chips were not radiation tolerant and had a time resolution in the order of microseconds, but especially in terms of spatial resolution were a significant improvement compared to hybrid sensors.

In the recent years, the availability of even higher resistivity epitaxial layers of increasing thickness and more voltage tolerant CMOS electronics, in addition with the possibility to modify some production steps allowed the development of depleted MAPS (DMAPS). Here, the full volume of the sensing area can be depleted (large fill factor), leading to a much faster charge collection via drift. Therefore, the chips are faster and more radiation tolerant. However, the radiation tolerance of hybrid sensors can still be significantly higher because the individual parts can be optimized independently and in a more flexible manner.

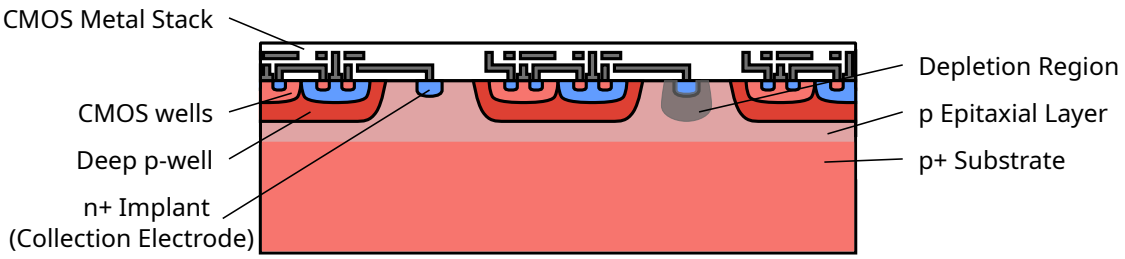


Figure 3.12: Illustration of an early MAPS sensor, depletion zone only shown in one pixel for clarity.

3.4 The DMAPS Technology

There are two approaches to achieve a high fill factor for allow for DMAPS, the small collection electrode and large collection electrode design. Both are able to do analog signal amplification and digital readout inside the pixel, with the sensing volume below the circuitry.

Large Collection Electrode

In large collection electrode designs, the full pixel electronics is embedded into a deep n-well, also forming the collection electrode. This means, the electric field is relatively linear and charge collection follows the shortest path orthogonal to the sensor plane. This poses an advantage for radiation hardness as the short path maximizes charge collection efficiency after irradiation. Additionally, a large fill factor is easy to achieve as the depletion region forms automatically under the whole pixel. A disadvantage is the high sensor capacitance. The deep n-well has a significant capacitance which increases the noise of the readout electronics. Additionally, capacitive coupling of the transistor p-wells inside the deep n-well can directly reach into the readout electrode, that is the deep n-well. Therefore, crosstalk from digital

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

circuitry inside the pixel is of special concern. A possibility is to create a deep p-well inside the deeper n-well charge collection electrode for additional shielding.

Figure 3.13 shows a schematic well layout of a large collection electrode DMAPS for two pixels including a deep p-well for shielding. For one of the pixels, the depletion region is shown. At large bias voltages, the depleted regions of the pixels can merge together, allowing a 100% fill factor. The CMOS wells are enlarged for illustration, and are typically much smaller with several 10s to 100s of transistors in each pixel. For example, this design is used by the RD50-MPW series of chips [33], aiming to push the limits of radiation tolerance of DMAPS sensors.

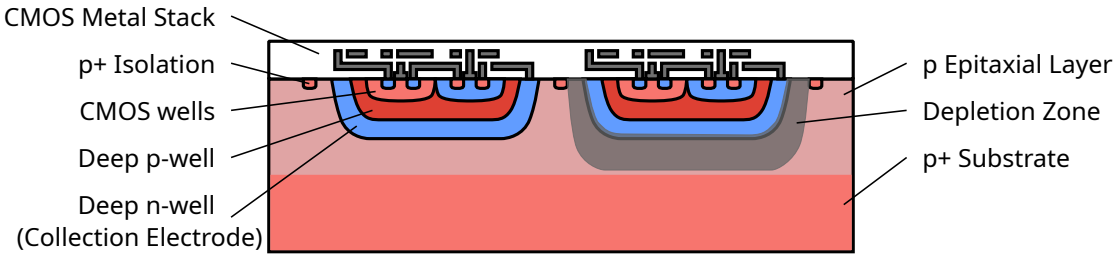


Figure 3.13: Illustration of a large collection electrode DMAPS sensor (two pixels), depletion zone only shown in one pixel for clarity.

Small Collection Electrode

The other approach uses a dedicated, small collection electrode, i.e. a small n-implant. The transistors need to be isolated inside a deep p-well for shielding. On a standard CMOS process, the depletion region would only form a blob around the collection electrode, leading to a small fill factor. However, with additional process modifications that create a low dose n-implant at the full pixel, the depletion zone can be extended to reach a high fill factor. The small collection electrode decreases the capacitance of the collection node, which is beneficial for noise as well as crosstalk suppression. The disadvantage is an elaborate, modified manufacturing process. Additionally, the radiation tolerance is generally below the large collection electrode designs, as the charge needs to follow the longer and more complicated path of the electric field up to the collection electrode.

Figure 3.14 shows a schematic layout of a small collection electrode DMAPS for two pixels including a low dose n-implant process modification. For one pixel, the depleted region is shown. The TJ-Monopix2 and therefore also OBELIX pursue the small collection electrode design. In addition to the well structure shown in figure 3.14, the TJ-Monopix2 has measures to increase isolation between pixels and therefore reduce cluster size. Two different implementations of the same principle were pursued. For the n-gap version of TJ-Monopix2, the n-implant below the pixels has a gap at the pixel edges. The second version contains an extra deep p-well at the same position, to achieve the same effect as the n-gap layout [34].

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

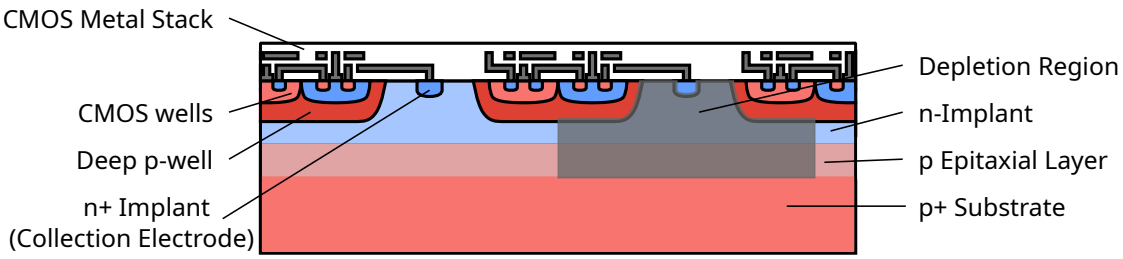


Figure 3.14: Illustration of a small collection electrode DMAPS sensor, depletion zone only shown in one pixel for clarity.

3.4.1 General Concepts

Some geneal concepts for DMAPS sensors are given here, before the specific details of TJ-Monopix2 follow in chapter 4.

Baseline, Peak Amplitude and ToT

Since modern electronics work mostly in the digital domain, the output of the analog amplifier needs to be digitized. Due to its simple implementation, a constant threshold discriminator is widely used for this purpose in DMAPS. It outputs a logic 1 if the input signal is above a certain threshold, and 0 otherwise. A sample analog waveform is shown in figure 3.15. At idle, the output voltage of the amplifier is typically at a certain nonzero value, commonly referred to as the baseline. A particle creates a pulse that is typically shaped by the amplifier as schematically drawn. The output of such a discriminator is binary, however, the pulse height can be inferred by measuring the duration of the digital pulse, the time over threshold (ToT). For a good timing resolution, a fast rise time is desired. The decay after the maximum is typically slowed down by the shaper, in order to increase the ToT and allow for a more precise measurement. A high gain of the amplifier is also beneficial, increasing the pulse height for a larger difference to the baseline.

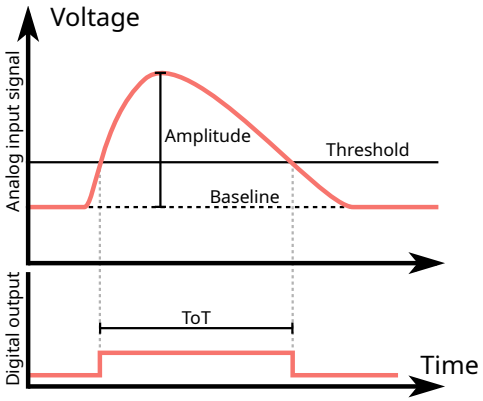


Figure 3.15: Illustration of baseline, peak amplitude and ToT.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

Since the signal generated in the depletion region is a charge Q, the voltage V_{sig} at the amplifier input depends on the total capacitance C_{tot} at the charge collection electrode via:

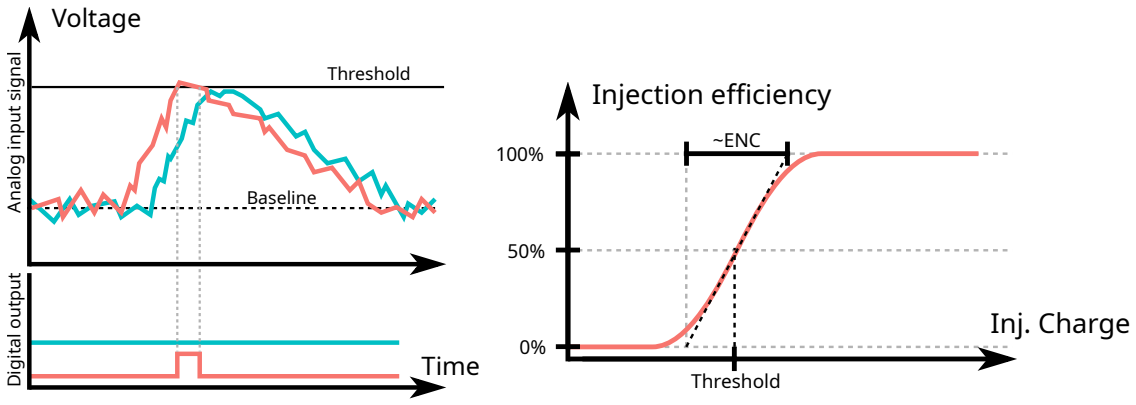
$$V_{sig} = \frac{1}{C_{tot}} Q. \tag{3.9}$$

The total capacitance is the sum of the capacitance of the collection electrode through the depleted region, input capacitance of the amplifier and potentially stray capacitance to other wires or shielding. For a large signal, a small capacitance is therefore crucial to achieve a large amplitude and good signal to noise ratio.

The S-curve

A common type of laboratory measurement is the so called s-curve. Here, increasing charges are artificially injected into the collection electrode, and the response is measured. As a result, the injection efficiency, i.e. the fraction of injections that are detected by the pixel is plotted against the charge as shown in figure 3.16b. In an ideal world, the resulting curve would immediately rise from 0% to 100% at the threshold. However, real systems contain noise. Signals close to the threshold might be registered sometimes, due to the fuzziness at peak introduced by noise. An illustration of two noisy signals of equal amplitude close to the threshold is shown in figure 3.16a, slightly shifted in time for better visualization. One of the signals gets detected, but the other not. Assuming Gaussian instantaneous noise values (white noise), the s-curve takes the shape of the Gaussian error function.

The threshold is therefore defined as the charge where a detection is occurring 50% of the time, or mean value of the noise distribution. The noise in units of charge, the equivalent noise charge (ENC) is the standard deviation of the noise distribution or $\sqrt{2\pi}$ times the inverse slope of the s-curve at the point of the threshold.



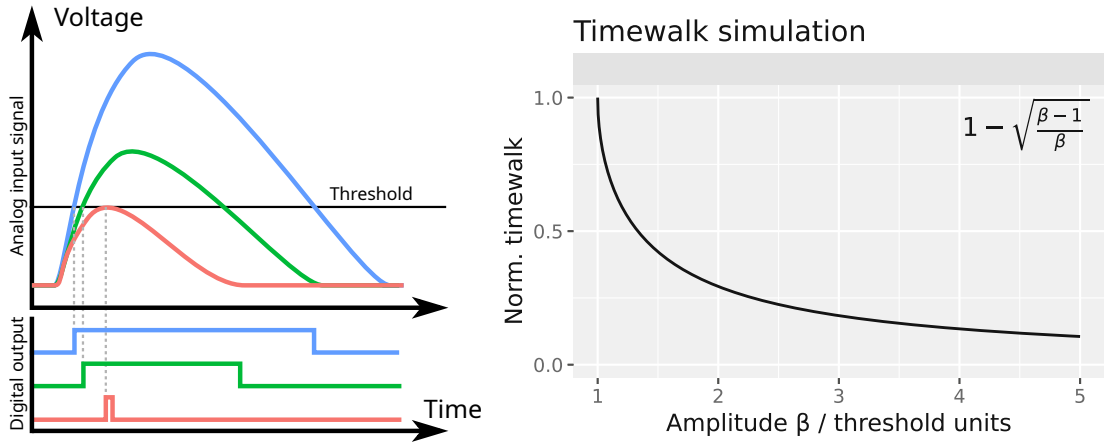
(a) (b)

Figure 3.16: Illustration of Noise (a) and typical s-curve measurement (b).

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

Timewalk

An inherent property of constant threshold discriminators is the so called timewalk. This is an effect of the signal amplitude on the time response at the digitization for pulse-like input signals. When the slope of the pulse is finite, as required for a finite bandwidth and the pulse shape is more or less independent from the amplitude, i.e. assuming a linear amplifier, timewalk occurs. A higher signal amplitude then means an earlier crossing of the threshold compared to a smaller signal. Generally this is a nonlinear relationship and particularly dominant close to the detection threshold.



(a) (b)

Figure 3.17: Illustration of timewalk for three simultaneous signals of different amplitude (a) and resulting timewalk in the ideal case (b)

Figure 3.17a illustrates the timewalk with three pulses of different amplitude but same starting time shown. Despite the equal start time, the threshold is crossed at different times, leading to delayed rising edges on the digital output of the comparator on the bottom.

To mathematically model the timewalk, only the rising part of the analog waveform is of interest. Assuming a function $\tilde{f}(t)$ with the start of the pulse at $t = 0$ and a baseline of $\tilde{f}(t = 0) = 0$, a Taylor expansion can be used up to the quadratic term, leaving a homogeneous polynomial second order:

$$\tilde{f}(t) \approx f(t) = \alpha t - \beta t^2, \tag{3.10}$$

including the negative sign of the quadratic term already. Since we assume an equal shape of the signal for all amplitudes, we can assume the peak to be on the same t position, regardless of the peak amplitude. Therefore we can set the peak position arbitrarily to $t = 1$. The derivative at the peak is zero per definition, leaving:

$$f'(t = 1) = \alpha - 2\beta = 0 \qquad \Rightarrow \qquad \alpha = 2\beta, \tag{3.11}$$

$$f(t) = \beta \left(2t - t^2\right), \tag{3.12}$$

identifying β as the peak amplitude. Let the time t^* be the crossing time of the threshold of $f(t^*) = 1$ and plugging it into the definition leads to:

$$f(t^*) = \beta \left(2t^* - t^{*2}\right) = 1. \tag{3.13}$$

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Solving for t^* and choosing the negative solution to get the first threshold crossing leads to:

$$t^* = 1 - \sqrt{1 - \frac{1}{\beta}} = 1 - \sqrt{\frac{\beta - 1}{\beta}}. \tag{3.14}$$

In this result, the signal amplitude β is in units of threshold and the timewalk is normalized to 1 at $\beta = 1$ corresponding to the maximal possible timewalk. A plot of the function is shown in figure 3.17b. Parameterizing the function to a charge q, a threshold q_{thr} and maximal timewalk t_{max} leads to:

$$\beta \to \frac{q}{q_{thr}}, \tag{3.15}$$

$$t_{timewalk} = t_{max} \left[1 - \sqrt{\frac{q - q_{thr}}{q}} \right]. \tag{3.16}$$

This means, that timewalk in general can be parametrized with two values: the threshold q_{thr} and the maximal timewalk t_{max}. The threshold depends on amplifier gain and baseline voltage, whereas t_{max} additionally depends on the rise time of the amplifier and therefore bandwidth. Therefore, the two parameters are mostly independent from each other, and can be treated separately. Since the threshold is easy to measure by other means, t_{max} is of more interest for timewalk measurements.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

Chapter 4

The TJ-Monopix2 DMAPS Chip

The TJ-Monopix2 chip is the second iteration of a DMAPS chip, initially designed as candidate for the ATLAS Inner Tracking Detector upgrade [30] in the course of the installation of the High Luminosity LHC. While the efforts for the ATLAS experiment were discontinued, the TJ-Monopix2 was chosen as base for the Belle II VTX Upgrade project. This thesis not only covers measurements of TJ-Monopix2 in this chapter, but also the development of the OBELIX chip in chapter 5 specifically for application in Belle II. Most measurements in this chapter are used to evaluate or simulate OBELIX performance.

4.1 Overview and Features

TJ-Monopix2 is manufactured in a modified Tower 180 nm imaging process [20] with small charge collection electrode design. It features a 512×512 pixel matrix with a pitch of 33.04 µm, resulting in an active area of $1.7 \times 1.7\,\mathrm{cm}^2$. Figure 4.1 shows the floorplan of the TJ-Monopix2 chip for illustration. The pixel matrix is separated in four groups of so called pixel flavors, differing in frontend amplifier design and biasing scheme. The biasing scheme differs for the coupling of the collection electrode, which can be either DC or AC depending on pixel flavor.

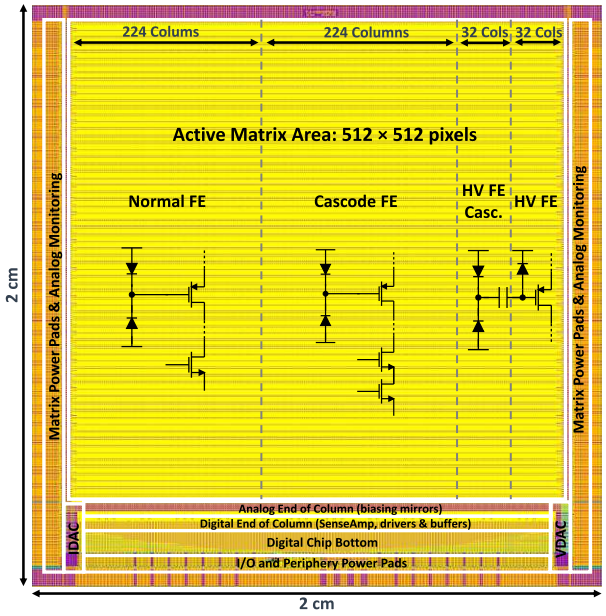


Figure 4.1: Floorplan of the TJ-Monopix2 chip [35].

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

The sensor bias can be applied via a negative potential on the substrate (PSUB) and deep p-well (PWELL) on the flavors directly connected to the frontend amplifier (i.e. DC coupled). For simplicity, in this thesis the PSUB and PWELL potential is always stated as absolute number.[1] This DC coupled type is called the Normal frontend, a slightly unimaginative name. Since the potential on PWELL directly affects the bulk potential of the transistors inside, the PWELL is limited to 6 V.

In the unirradiated state, the punch-through effect of the low dose n-implant limits the difference between PSUB and PWELL to a few hundred millivolts. With subsequent irradiation, the PSUB potential can be increased further, allowing a larger depletion zone and stronger electric field, especially in the pixel corners. At $5 \times 10^{14}\,\mathrm{cm}^{-2}$ n_{eq} VTX design fluence, PSUB can be operated at 20 V while the PWELL potential is always kept at 6 V.

A different biasing scheme is implemented in the so called HV type flavors. Here, the frontend transistor gate is AC coupled, and a positive bias is applied from the topside i.e. to the collection electrode directly. This means the full bias voltage can be applied at all times, even before irradiation. The disadvantage is an effective charge loss of approximately 40% due to the capacitive voltage divider of the coupling capacitor with the combination of gate- and stray capacitance. This design allows bias voltages of up to 40 V or more until breakdown.

The standard version of the frontend amplifier is extended by an additional cascoded transistor in the first amplification stage in the Cascode versions. The cascoded transistor allows for a higher gain, and the effect can be controlled via the `VCASC` voltage, applied to the gate of the cascode transistor.

The biasing scheme and amplifier design are independent from each other and all four possible combinations are implemented. Table 4.1 summarizes the available pixel flavors. As visible, the majority of the chip is using the Normal and Normal Cascode frontends. Only 32 columns are using the HV and HVC topside biasing scheme respectively.

Front-end	Biasing	Coupling	Amplifier	Col. range	Size
Normal (N)	PSUB/PWELL	DC	Standard	0 to 223	224×512
Normal Casc. (NC)	PSUB/PWELL	DC	Cascode	224 to 447	224×512
High Volt. Casc. (HVC)	Collection electrode	AC	Cascode	448 to 479	32×512
High Voltage (HV)	Collection electrode	AC	Standard	480 to 511	32×512

Table 4.1: Front-end flavors of TJ-Monopix2.

4.1.1 Frontend Amplifier

Each pixel contains an analog amplifier and discriminator to convert the charge signal into a digital pulse, the circuit diagram is shown in figure 4.2a. To automatically set a baseline for the pre-amplifier output, a `VCASN` generator circuit is placed in the periphery of the chip shown in figure 4.2b. It reassembles parts of the amplifier circuit and automatically compensates for PWELL bias and ground voltage variations, allowing an accurate control over the discriminator standby current [35].

The operating point of the amplifier is controlled by several voltages and currents. These are generated in the periphery by a group of 8 bit Digital-to-Analog Converters (DACs). The scaling from the raw DAC value (0 to 255) to the actual current or

[1] This is to prevent the misleading 'increased bias voltage' statements which actually mean less bias applied if the numbers are negative. A positive voltage is never applied to the PSUB or PWELL and therefore the sign does not matter. For HV, the potential is always positive.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

voltage is different for each DAC and it is most useful to treat the settings as raw 8 bit numbers in 'DAC units'. The DAC-groups are repeated 16 times along the chip periphery and supply 32 columns each. While the currents are distributed as voltages, they form half of a current mirror in the periphery and the other half in the amplifier. The **VRESET** is a potential directly applied via a diode to the amplifier input an is used to modify the input baseline.

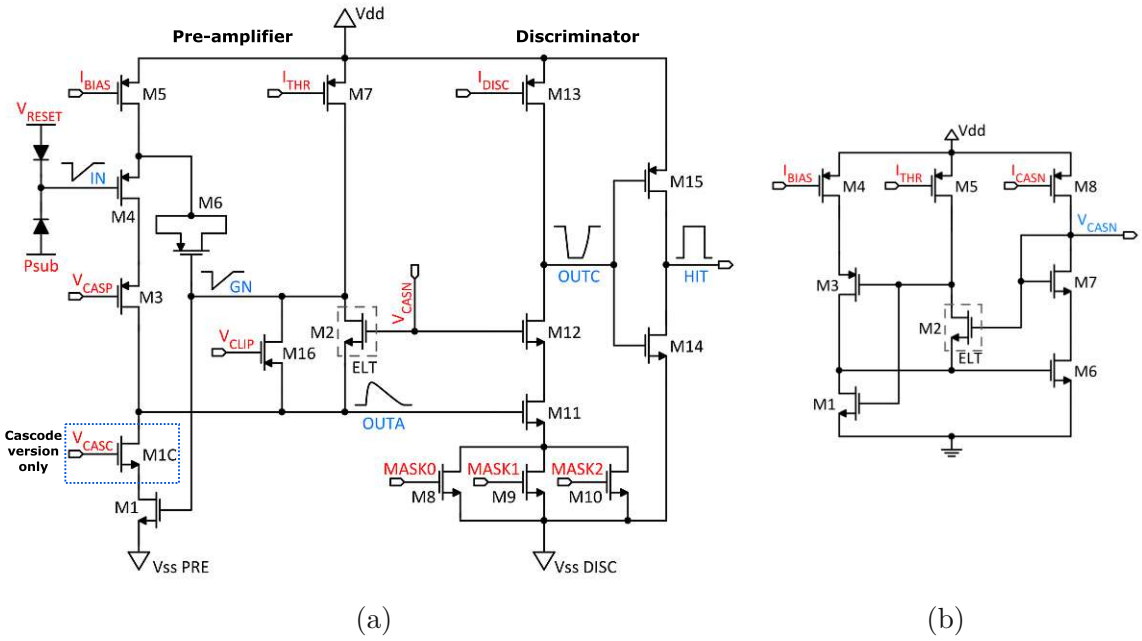


Figure 4.2: Frontend amplifier design (a) and **VCASN** generator circuit in the periphery (b) [35].

The amplifier consists of three stages with individual bias currents. The first stage is supplied with current via M5 and the **IBIAS** value. The second stage is supplied by M7 and **ITHR**. The last stage is the discriminator supplied by M13. The current of M13 varies with in-pixel tuning and is built from **IDB** and **ITUNE** as will be described in section 4.1.3. The **VCASC** (cascode frontends only) and **VCASP** change the operation point of the first stage of the amplifier.

The preamplifier output baseline is calculated from **IBIAS**, **ITHR** and a third value that can be used to modify the baseline: **ICASN**. The latter is useful to change the threshold of the frontend to a target value in a large range without affecting the rest of the circuit by only changing the baseline voltage.

The output of the analog frontend is a digital pulse on the **HIT** output that is connected to the in-pixel readout circuit. A 7 bit timestamp called BCID is distributed to each pixel. At the rising edge (leading edge, Le) and falling edge (trailing edge, Te) of the **HIT** line, the BCID is sampled and stored in latches. These so called Le-BCID and Te-BCID values are transmitted along with the pixel address by the readout over a parallel bus. Since the distribution of the BCID is subject to a delay along the columns of the chip, a compensation circuit delays also the **HIT** signal. This delay circuit is different depending on pixel location and matched to the BCID delay. However, a global tuning parameter **IDEL** allows a first-order adjustment of this delay compensation.

The difference of Le-BCID and Te-BCID allows the calculation of the Time over Threshold (ToT). The ToT is a function of the charge and can be used to reconstruct the charge deposited by an incoming particle. The **VCLIP** value can be

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

used to set a maximum ToT value in the analog domain. This so called ToT-clipping then results in roughly the same maximal ToT for all charge pulses above a certain charge. No ToT-clipping is performed at the maximal **VCLIP** of 255 and lower values subsequently reduce the maximal ToT. Further explanations and measurements of this feature will be presented in section 4.3.

All available DACs are summarized in table 4.2. The effect on threshold is denoted with + for a positive and − for a negative relation i.e. an increased value results in increase or decrease in threshold respectively. A double sign like ++ or −− signifies a large 'therapeutic index', i.e. a large range of thresholds can be selected with this value before the amplifier is driven too far outside the nominal operation point and stops working. Bold lines are deviations from the default values but have shown better performance in tests.

The values **ICOMP** and **IRAM** are also added for completeness and are used for the readout of the pixel data. **ICOMP** is the bias current for the comparator of the bit-lines of the data bus in the periphery while **IRAM** is the driving strength of the bit-lines in the pixel. Prior to reading out a pixel, the bit lines are pre charged to the **VPC** potential that is supplied externally by a voltage divider on the chipboard.

Register	Typ. value / DAC		Thr.	via		Comment
	N/NC	HV/C		Gain	Baseline	
IBIAS	100	100	−−	yes	yes	changes power dissip.
ITHR	64	64	++	yes	yes	low gain/BW if reduced
ICASN	**0 - 20**	**30 - 60**	−−		yes	used to set threshold
IDB	100	100	−−		yes	
ITUNE	**200**	**200**	(−)		yes	indirect via TDAC
VRESET	**120**	**30**	(−)		indirect	amplifier input baseline
VCASP	93	93	−	yes		
VCASC	228	228	+	yes		cascode only, min. 200
VCLIP	255	255				for ToT clipping only
ICOMP	80	80				RO comparator bias
IDEL	**50**	**60**				BCID delay comp.
IRAM	50	50				bit-line drive current

Table 4.2: Summary of the frontend parameters of TJ-Monopix2.

4.1.2 HitOr

For each pixel, the **HIT** signal is also connected to a chain of OR gates along each column. This is shown in figure 4.3, leaving out the digital control circuit for ordinary operation for a clearer picture. The chain of OR gates forms the so called **HitOr** signal, one bit per column that is subsequently reduced to one bit chip-globally via OR gates in the periphery and exposed as one bit CMOS and LVDS output on the chip. The **HitOr** signal directly connects the discriminator to the output via a combinatorial circuit only, i.e. it is asynchronous and preserves any timing information from the pixel. This allows characterization of the charge collection and amplifier with more resolution than the ordinary readout. However, due to the nature of real digital circuits, a delay is introduced by signal propagation along internal wires and through logic gates. This delay is dependent on the position of the pixel causing a different delay from the frontend to the **HitOr** output for each pixel.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

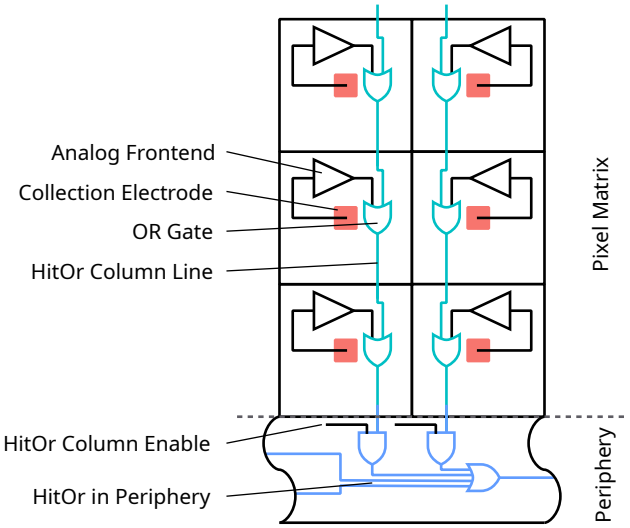


Figure 4.3: Schematic view of **HitOr** lines in two columns. HitOr Row Enable lines and regular digital readout path not shown for clarity.

As general model, the delay from the **HitOr** signal can be decomposed into a sum of two contributions with distinct properties. The first contribution is the delay arising from the signal propagation along each column inside the pixel matrix. The structure of this circuit is repetitive, i.e. each pixel contains one OR-gate of the same type and layout, adding its discriminator output to the output of the previous pixel and forwarding the output to the next pixel. Since the layout is the same, each row adds the same amount of delay to the signal, and the position dependent delay can be assumed to be a linear function of the row number to a good approximation. Additionally, since all columns (within one pixel flavor) are the same, the slope of this linear function can be assumed to be the same for all columns. This introduces some error due to local variation of gate-delay over larger distances due to Process, Voltage and Temperature (PVT) variations, but this effect is small compared to the overall delay.

The second contribution to the **HitOr** delay is the propagation from the bottom of the matrix to the **HitOr** output of the chip through the digital periphery of the chip. As the routing in this part of the chip was done in an automated way, the delays are slightly different for each column with a nontrivial pattern.

The OBELIX chip uses the column-wise **HitOr** signal for two of its special functions, which will be described in sections 5.9.6 and 5.11. Therefore, the characterization of the time-behavior of the **HitOr** circuit is of interest and has been measured in different ways presented in the following sections. Since the digital routing is different for OBELIX, the first contribution inside the pixel matrix is of particular interest here.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

4.1.3 Calibration and Tuning

This section covers the so called chip-tuning. Before any test, it is required to find settings for the various bias voltages and currents of the frontend, in-pixel threshold tuning and optionally masks for noisy pixels. All of those are saved in the same container format HDF5 (*.h5 files) for each measurement by the DAQ system. New measurements will take this tuning from the last previous measurement per default and may overwrite selected settings. There are many degrees of freedom and trade-offs between threshold, noise, amount of noisy pixels, gain and ToT-spectrum.

Injection Circuit

The injection circuit can inject a configurable charge in a selectable number of pixels in a highly flexible way. This allows calibration and performance analysis in the lab without radioactive sources. The injection works by distributing two voltages over the full matrix, controllable via two DACs, `VL` and `VH`. When an injection is performed, a voltage step from `VL` to `VH` is applied to the charge collection node in the pixel through an injection capacitor C_{inj}. At a later time, the injection signal goes back to the logical 0 state to form an injection pulse and restore the initial state to perform the next injection. It is important that this pulse length is larger than the ToT duration, in order to prevent the falling edge of the injection pulse from affecting the pixel frontend. To control which pixels are injected, column- and row-wise selection signals are distributed and each pixel performs a logical AND operation of the two. This means, each pixel where the corresponding row AND column is enabled are staged for the charge injection.

The injection circuit operates on both edges of a 160 MHz command clock, allowing a time binning of 3.125 ns. The rising edge of the injection pulse is configurable in these steps of 3.125 ns via the `PULSE_START_CONFIG[5]` setting. The range can be extended arbitrarily by adding IDLE/SYNC commands in front of the injection command. The `PULSE_STOP_CONFIG[14]` setting configures the length of the injection pulse in 25 ns units. More technical details about the command structure and injection configuration will be given in the corresponding OBELIX section 5.8 and also apply for TJ-Monopix2, but are not of importance for an understanding of the measurements of this chapter.

It is important to note that the internal buffers of `VL` and `VH` in TJ-Monopix2 are only linear up to a value of 140 DAC units. Higher values still result in higher injected charge, however, the nonlinearity prevents quantitative measurements. This will be changed in OBELIX to allow the full range of 255 DAC units of injection.

In principle, the injected charge can be calibrated in an absolute way by determining C_{inj}. This has been done with an ^{55}Fe source that has a distinctive X-ray peak at 5.90 keV. For the test-beam analysis, the conversion factors of table 4.3 have been used to allow an effective comparison, but the raw DAC value is stated for the other measurements as the conversion does not add a benefit to the understanding.

Frontends	Calibration
N/NC	9 electrons/DAC
HV/HVC	18 electrons/DAC

Table 4.3: Conversion of injection value to charge.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

Global Tuning

To allow the configuration of an arbitrary threshold, a global threshold tuning script is available. It injects a target charge into a configurable region of the matrix, and performs a bisection of the **ICASN** value to a target occupancy of 50%. This is the center region of the S-curve and therefore finds a configuration where the average threshold of the enabled pixels corresponds to the charge injected.

The value of **ICASN** is used as it provides a large dynamic range of thresholds maintaining stable operation on all the frontend flavors. Changing this value affects the baseline of the amplifier output only.

Local Tuning

Each pixel contains three bits of **TDAC[3]** configuration memory. This memory is used to individually mask pixels, as well as tuning the local threshold in a certain range. A value of 0 on the **TDAC[3]** configuration disables the pixel completely which is useful to mask individual noisy pixels. Table 4.4 summarizes the possible **TDAC[3]** values.

TDAC[3]	Threshold
0	Pixel Masked (off)
1	Lowest threshold
...	...
7	Highest threshold

Table 4.4: Local tuning for the TJ-Monopix2 chip.

The tuning range is set via the **ITUNE** DAC setting. The tuning works by selecting one of 7 possible discriminator bias currents $I_{DISC}[i]$ supplied from the analog periphery according to the **TDAC[3]** setting. These currents are created according to the formula:

$$I_{DISC}[i] = \texttt{IDB} \cdot I_{DB}^{LSB} + (i-1) \cdot \texttt{ITUNE} \cdot I_{TUNE}^{LSB} \qquad \text{with} \quad i \in \{1, 2, \dots 7\}, \quad (4.1)$$

with DAC values **IDB** for the discriminator baseline and **ITUNE** for the tuning range, and current conversion constants I_{DB}^{LSB} and I_{ITUNE}^{LSB} (currents for the least significant bits). The resulting threshold is a strictly monotonic function of this current. An example of the tuning range in terms of threshold is shown in figure 4.4 to illustrate this.

Like in the global tuning script, the local tuning algorithm uses injection at the target threshold and performs a bisection algorithm on each pixel. Since the dynamic range of in-pixel tuning is limited, an adequate threshold target needs to be set for a proper in-pixel tuning. For unirradiated samples, a range of **ITUNE** = 200 is typically required while after irradiation the maximum value of 255 is needed to achieve a uniform threshold. For the OBELIX chip, the scale of this DAC will be expanded to allow an increased tuning range.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

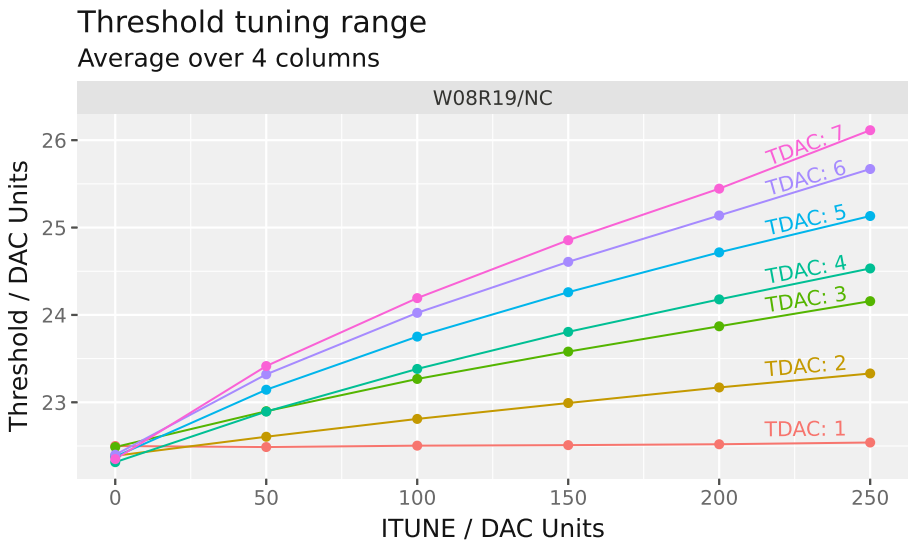


Figure 4.4: Effect of threshold dependent on **TDAC** and **ITUNE** settings.

4.2 Test-beam Campaigns

Apart from tests in the laboratory using the charge injection circuit or a simple radioactive source, some measurements require a high quality, high energy particle beam from an accelerator. These so called test-beam facilities are typically booked in time-slots to perform measurements.

#	Time	Duration	Scientific focus
1	July 2022	2 Weeks	Test of DAQ system and initial chip tests, threshold untuned
2	July 2023	1 Week + 1 Week parasitic	Tuned and low threshold operation, 1 irradiated sample at RT, **timing tests**
3	July 2024	1 Week[2]	2 irradiated samples, elevated temperature operation
4	April 2025	2 Weeks	**4 irradiated samples, elevated temperature operation w/ more precise temperature control**

Table 4.5: Overview of DESY II test-beams of the Belle II VTX collaboration.

In total, four test-beams of either one or two weeks have been performed by the Belle II VTX collaboration at the DESY II test-beam facility [36]. The DAQ system was improved and tested at the MedAustron [37, 38] cancer treatment synchrotron before each test-beam. This allowed to spot and repair issues with the DAQ system in advance, allowing fast setup times at DESY II. The four test-beams and the corresponding scientific questions are summarized in table 4.5. The bold topics are presented in detail in this thesis as they mostly superseded the previous tests.

The first test-beam was performed to test the DAQ system and general chip operation of unirradiated samples. The main focus was to verify the correct operation of TJ-Monopix2 and the suitability for the Belle II VTX upgrade. No emphasis was put on threshold and in-pixel threshold tuning and mostly default settings from

[2]Only remote participation by the author.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

the design were used, which proved to be sufficient for operation at close to 100% efficiency. At the second test-beam, the focus was on low threshold operation and in-pixel tuning of unirradiated samples and one sample irradiated to the VTX design fluence of $5 \times 10^{14}\,\mathrm{n_{eq}/cm^2}$ at room temperature. Additionally, detailed timing tests were performed and the results will be presented in detail in section 4.5.2. The timing studies were performed mostly in parasitic operation without a beam-telescope to reconstruct reference particle tracks.

The 3rd and 4th test-beams focused on operation of irradiated samples at a wider temperature range and in particular elevated temperatures. The sensitivity on temperature variations was underestimated initially, and the stability of the temperature control was not sufficient at the 3rd test-beam to provide consistent results. However, the temperature control was improved and measurements were repeated at the 4th test-beam with consistent quality. Additionally, more samples have been irradiated to different fluences in the mean time to have a better understanding of the evolution with irradiation. The results of this test-beam will be presented in section 4.4.

4.3 ToT Clipping: `VCLIP`

The pixel frontend allows the reduction of the ToT via the **`VCLIP`** setting of the analog frontend. It effectively defines a maximum ToT, and larger charges are clipped to this maximum ToT. Since this is an analog circuit, the effect is not sharp but a distribution around the maximal ToT.

For most tests of TJ-Monopix2, the ToT clipping feature is unused with **`VCLIP`** of 255 which corresponds to no effect on ToT. On TJ-Monopix2, a reduced ToT does not bring any benefit. Since the chip is intended for operation at low rates and the readout is relatively slow, the impact of the ToT on the timing is insignificant. However, for the OBELIX chip with high bandwidth readout, the ToT strongly affects the readout timing. The OBELIX performance analyses in section 5.9.5 assume different maximal ToTs, which impacts the performance of the trigger unit. This is particularly important in the first years of VTX operation, as the ToT generally decreases with irradiation and for iVTX, where the hit-rate is the highest.

Therefore, the ToT-clipping was tested and is presented this chapter. Only one unirradiated Device Under Test (DUT) was tested with different **`VCLIP`** settings. ToT histograms are shown in figure 4.5, each generated by an 4.2 GeV electron beam. Below a value of 140, the clipping shortens the ToT with lower value. For illustration, the spectra are shown for all hits, counting clusters as individual hits, and only for seed pixels, i.e. only the pixel with the largest ToT of each cluster. The peak shape and position is mainly composed out of seed-hits and therefore the same in both plots. However, the ToT values below about 20 are mostly cluster periphery, as they are almost not present in the cluster seed histogram. For the All-Hit view, it is clear that the clipping affects only large ToT values, as all histograms overlap at sufficiently low ToT.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

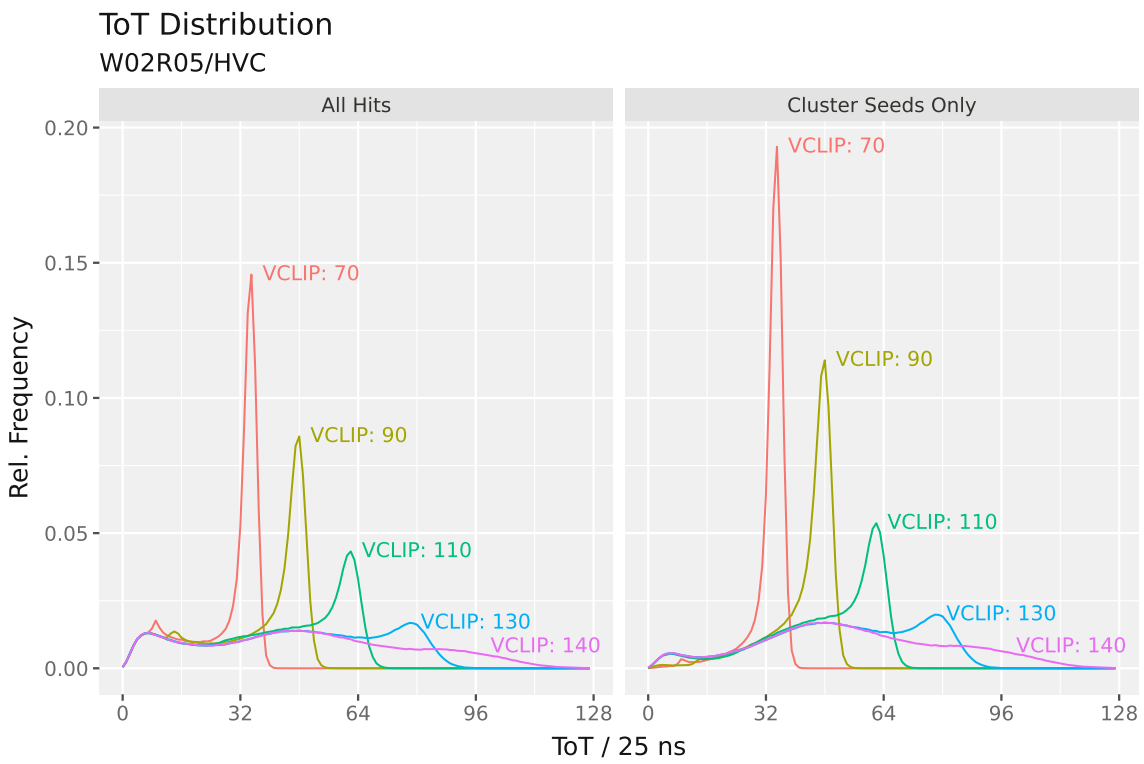


Figure 4.5: ToT spectra for different **VCLIP** settings from the July 2023 test-beam.

Figure 4.6 shows the properties of the spectra. The peak position, corresponding to the maximal ToT is shown along with the 95% percentile, i.e. 95% of all hits have ToTs below this value. For strong clipping with **VCLIP** of 30, the peak ToT can be reduced down to 12 BCID codes corresponding to 300 ns. Generally, the ToT clipping was operating as expected and **VCLIP** has a linear relationship to maximal ToT for a large range of values. It can therefore be used to set an arbitrary limit on the ToT.

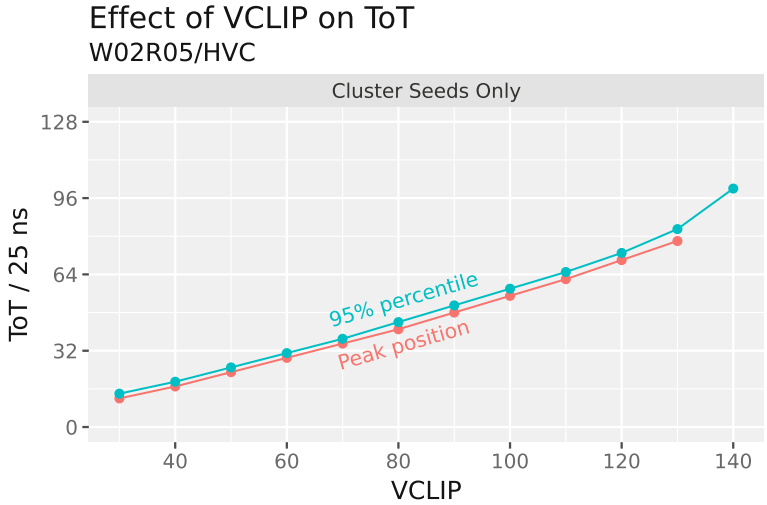


Figure 4.6: Maximal ToT for different **VCLIP** settings.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

4.4 Efficiency Measurements

Particle detection efficiency, i.e. the probability of detecting an intersecting particle, is one of the main benchmarks for a pixel sensor and nearly all particles should be detected. In collider experiments, a good tracking efficiency is needed for reconstruction of all final state particles of a collision. Since the probabilities for detecting all intersections of a track is the product of the detection probabilities in the individual planes [3], and for the full reconstruction of multiple tracks of an event again is the product for the individual tracks, nearly 100% efficiency per plane is required for good physics performance. For the Belle II VTX, a loss below 1% is targeted for the sensor.

To measure detection efficiency, the sensor (DUT) is exposed to a particle beam at a test-beam. In front and behind the DUT, reference detectors are positioned that allow the reconstruction of each individual particle trajectory intersecting the DUT, the so called beam-telescope. For each trajectory, the DUT can either detect or miss the particle, and the efficiency can be calculated. Typically, a spatial cut is applied, i.e. a detected particle in the DUT needs to be closer than a maximal distance to be associated with a track. For the analyses shown here, a spatial cut of 100 µm was applied, corresponding to about three pixels.

These test-beams are usually performed in a one or two week time-slot at the accelerator facility and are a collaborative effort due to the complexity of the measurement setup. Despite the inconsistent results, it was identified that the cascode flavors can operate at lower threshold at elevated temperature in the third test-beam in July 2024. The following efforts were therefore focused on the NC and HVC flavors.

The last test-beam in April 2025 focused on operation of several irradiated samples under elevated temperature. As the VTX detector is designed to operate under elevated temperature, it is crucial to verify that this does not impact performance.

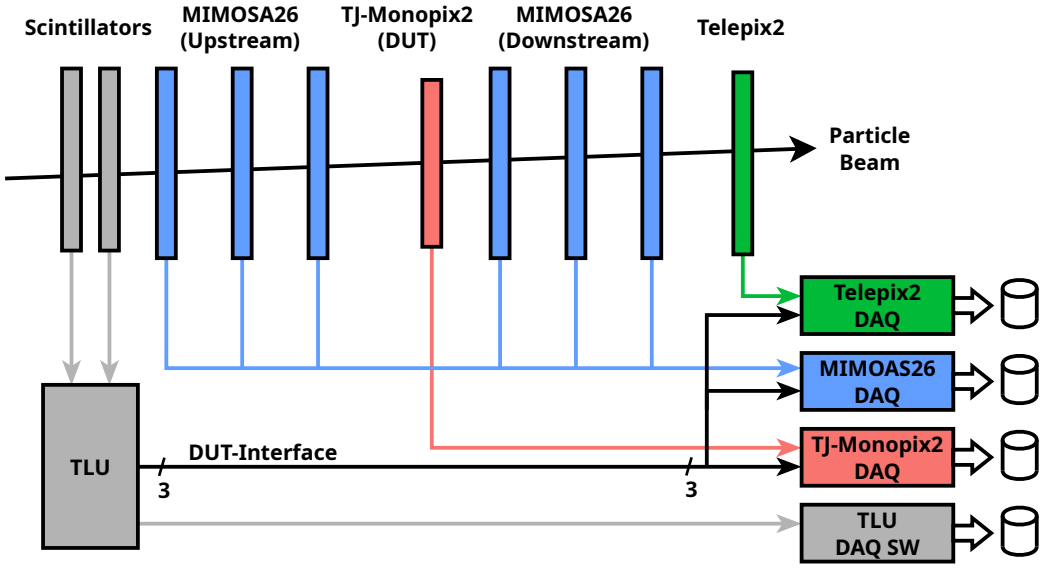


Figure 4.7: Detector setup for the April 2025 test-beam.

[3] Usually it can be tolerated if one plane misses the particle, which increases the track detection efficiency above the sole product.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Figure 4.7 shows the test-beam setup of the April 2025 test-beam. The DATURA beam-telescope with 6 planes of MIMOSA26 chips [31] was used, together with a time reference plane using a Telepix2 sensor [39]. The EUDAQ2 [40] framework was used to start and stop data taking of all the detectors, but only the Trigger Logic Unit (TLU) [41] and telescope used the EUDAQ2 data collector feature, while the Telepix2 and TJ-Monopix2 stored the data within the corresponding DAQ software. For analysis, the Corryvreckan framework [42] was used.

The first measurements used a similar DAQ system configuration as all the previous test-beams. Here, the so called EUDET mode of the TLU was used to synchronize the DUT with the other detectors. While this mode is sufficient to measure the efficiency, the time resolution is limited to the resolution of the slowest detector system. This is because a trigger number is used to match the clusters of the sensors to tracks. The trigger number is generated by the TLU and distributed to all detectors in EUDET mode. Since the MIMOSA26 chip has an integration time of about 114 µs, this was the maximum achievable.

The later runs used the so called AIDA mode of the TLU to synchronize the TJ-Monopix2 DAQ system. In this mode, the system clock is not generated locally on the FPGA board, but centrally from the TLU and distributed to all DAQ systems in AIDA mode. This synchronous operation allows the track building using timestamps. This is also the reason for the timing plane, which provides a reference time for the track which is then matched to the DUT cluster and also operates in AIDA mode. The telescope still operates in EUDET mode, but the timing reference plane is used to assign a timestamp to each track which is sufficient for this purpose.

Since the TJ-Monopix2 has a BCID width of only 7 bits, it is not possible to directly use it as timestamp. Overruns occur at the same timescale as the readout time, creating ambiguities. However, a larger timestamp was annotated on the FPGA to each hit as soon as it arrived there from the DUT.

While the transition to AIDA mode is not directly useful for TJ-Monopix2 measurements, it was important to gain experience in this mode of operation. The OBELIX chip in triggered mode requires a synchronous operation together with the TLU, and therefore the AIDA mode. It was confirmed that the system is capable to run synchronously without any detrimental effects in data taking, paving the way for measurements with OBELIX.

Serial	Irradiation particles	NIEL Fluence	Max. PSUB
W08R06	24 MeV protons	$5 \times 10^{14}\,\mathrm{n_{eq}/cm^2}$	20 V
W02R05	90 MeV electrons	$5 \times 10^{14}\,\mathrm{n_{eq}/cm^2}$	20 V
W02R17	24 MeV protons	$2.5 \times 10^{14}\,\mathrm{n_{eq}/cm^2}$	15 V
W08R04	24 MeV protons	$1 \times 10^{14}\,\mathrm{n_{eq}/cm^2}$	10 V

Table 4.6: Irradiated DUTs for the test-beam in April 2025.

Four irradiated samples were measured, as shown in table 4.6. The proton irradiation was performed at the CYRCé facility in France [43], while the electron irradiation was performed by RARiS in Japan [44]. Not only two chips were irradiated to the VTX design fluence of $5 \times 10^{14}\,\mathrm{n_{eq}/cm^2}$, but also two lower fluences were used to characterize the evolution of the irradiation. Since the isolation of the PSUB and PWELL potential increases with irradiation, the less irradiated chips only allow a lower PSUB potential before the leakage current becomes unacceptable.

For the W02R17 chip, the HV biasing was not functional, preventing operation

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

of the HVC frontend. The W02R05 chip was tested unirradiated previously. The tests after irradiation are shown in this section, however, later in this chapter more results of this chip are presented prior the irradiation campaign.

4.4.1 Results

An overview of all measurement runs with maximal bias voltage applied is shown in figure 4.8. This plot shows the achieved efficiency over the threshold that has been measured on-site for most points. At higher thresholds, the detection efficiency drops as expected and scales with irradiation. However, it is difficult to maintain a low threshold at elevated temperatures as the increased leakage current shifts the potential of the collection electrode and increases noise. The clear trend shows that the threshold is the dominant factor for efficiency. This is a valuable insight, as the threshold can be measured more easily in a lab environment. An estimation of the efficiency is achievable in the laboratory with the different frontend settings and allows the exploration of a larger space of settings without a test-beam.

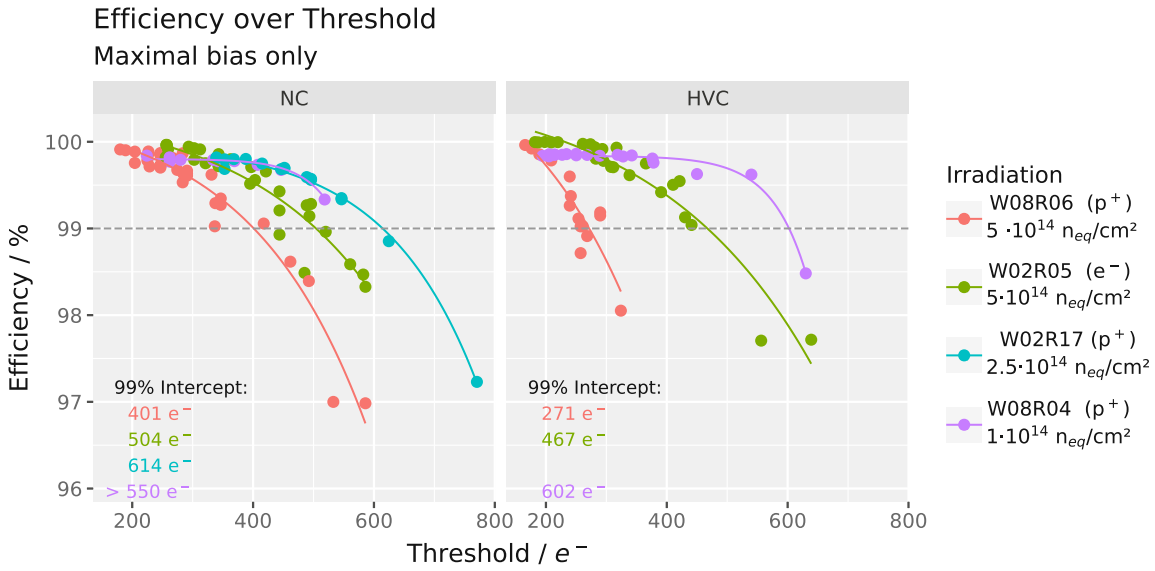


Figure 4.8: Efficiency of all measurement runs of April 2025 test-beam depending on threshold.

The data points were used to fit an exponential function and estimate the crossing of the 99% boundary. The resulting maximal thresholds for a 1% drop are shown in the plot. Generally, the HVC frontend requires a lower threshold to operate efficiently. This is explainable by the charge loss due to the capacitive voltage divider of the AC coupling capacitor with stray and amplifier capacitance. Simulations suggest a loss of about 40% of charge, which is compatible with the results. The margin until reaching an efficiency drop was smaller for the HVC frontend due to the charge loss associated with the coupling capacitor.

For W08R04/NC, it was not possible to elevate the threshold above 550 electrons without serious de-tuning of the frontend. However, since the collected data points are similar to the W02R17 chip, a value in the order of 600 electrons can be assumed. For W08R04/HVC, the two points at the high threshold end are contradicting each other slightly and have strong influence on the fit. This estimation is therefore to be taken with some caution.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

Notably, the electron irradiated chip W02R05 appears to have received less radiation damage than the proton irradiated chip W08R06, despite the same calculated NIEL fluence. This effect is not understood so far, and further investigations are required.

A subset of the runs shown in the previous figure are presented again in figure 4.9. This time, the efficiency is shown against the temperature, and only the so called 'Standard settings' i.e. generic tuning for low threshold chip operation is shown. The temperatures of the chip have been estimated from the NTC sensor on the chipboard and LV current. Reference measurements for the estimation modifying the settings to reach equal power consumption were performed in the lab with a different chip. A temperature probe directly attached to the chip while using the same setup to obtain the actual chip temperature under different conditions. Typically, the estimated chip temperature is between 6 and 9 K above the NTC temperature depending on temperature and LV current.

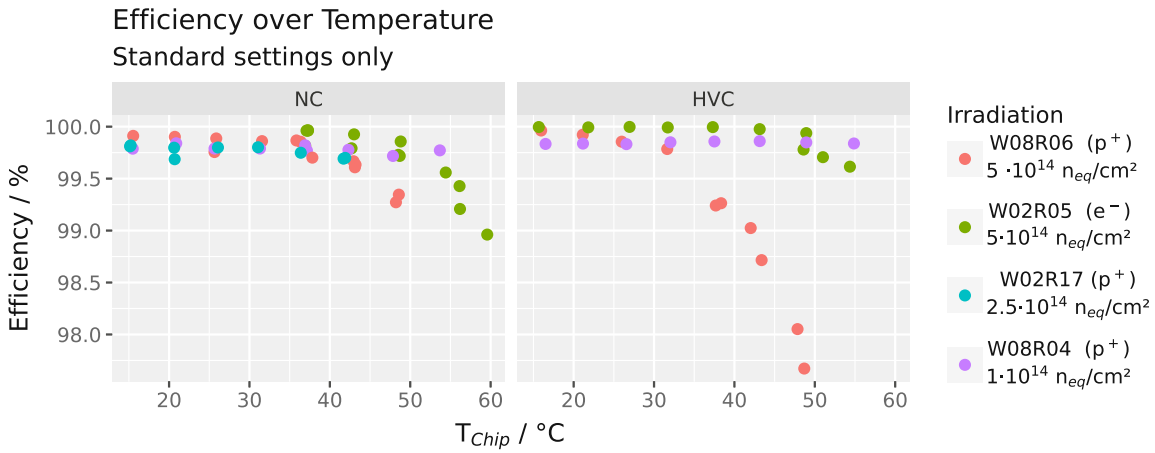


Figure 4.9: Efficiency of measurement using the typical frontend settings over the temperature.

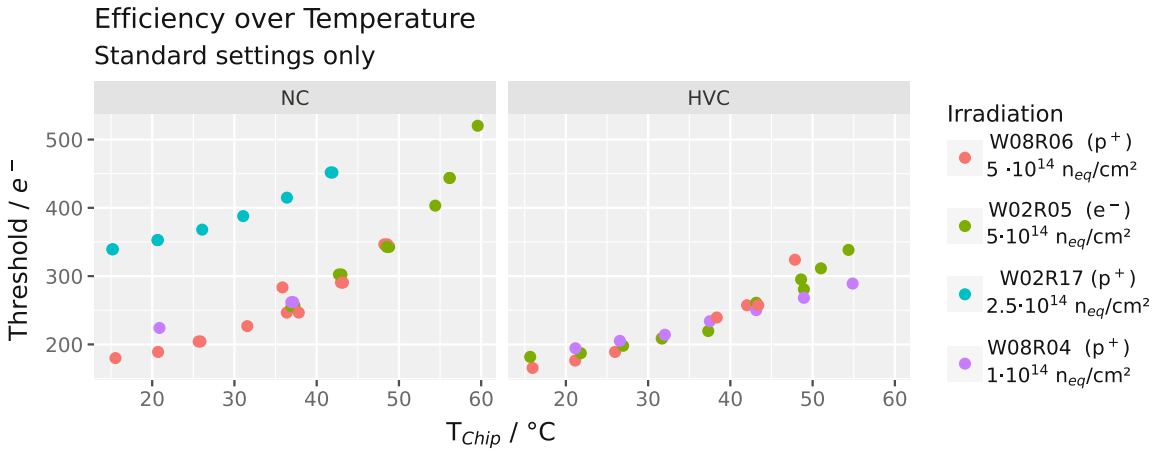


Figure 4.10: Threshold for the same data points shown in figure 4.9.

It is possible to operate the NC frontend up to a chip temperature of 50 °C without reaching the 99% boundary for all chips. At the same time, with temperatures above 35 °C, the efficiency starts to drop notably. Therefore, the recommended operation temperature should be below this point for optimal performance. For HVC

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

of chip W08R06, the onset of the drop is about 10 K lower and more pronounced due to the lower margin in threshold.

The threshold corresponding to the measurements is shown in figure 4.10. The threshold is directly affected by leakage current on the sensing electrode. As the leakage current increases, the collection electrode is discharged in a similar timescale as the amplifier can process the signal pulse. This effectively reduces the amplitude of the amplifier output and is mostly independent from AC or DC coupling. As a result, the threshold increases exponentially with the temperature.

Bias Scans and In-Pixel Efficiency

Since the telescope has a better pointing resolution than the pixel pitch of the DUT, it is possible to analyze the inner structure of the DUT pixels. The pointing resolution is the resolution of the reconstructed track at the position of the telescope and has been determined using the CERN telescope optimizer [45] to 4.2 µm. The in-pixel structure is particularly interesting at reduced bias voltages when the depletion is not complete. An example bias scan is shown in figure 4.11. At PSUB voltages below 6 V, the PWELL voltage is the same as PSUB and PWELL = 6 V otherwise. At low bias, only the center of the pixel close to the collection electrode can detect particles with high efficiency. This center region grows with bias, and the full efficiency is reached at about 9 V for W02R17/NC.

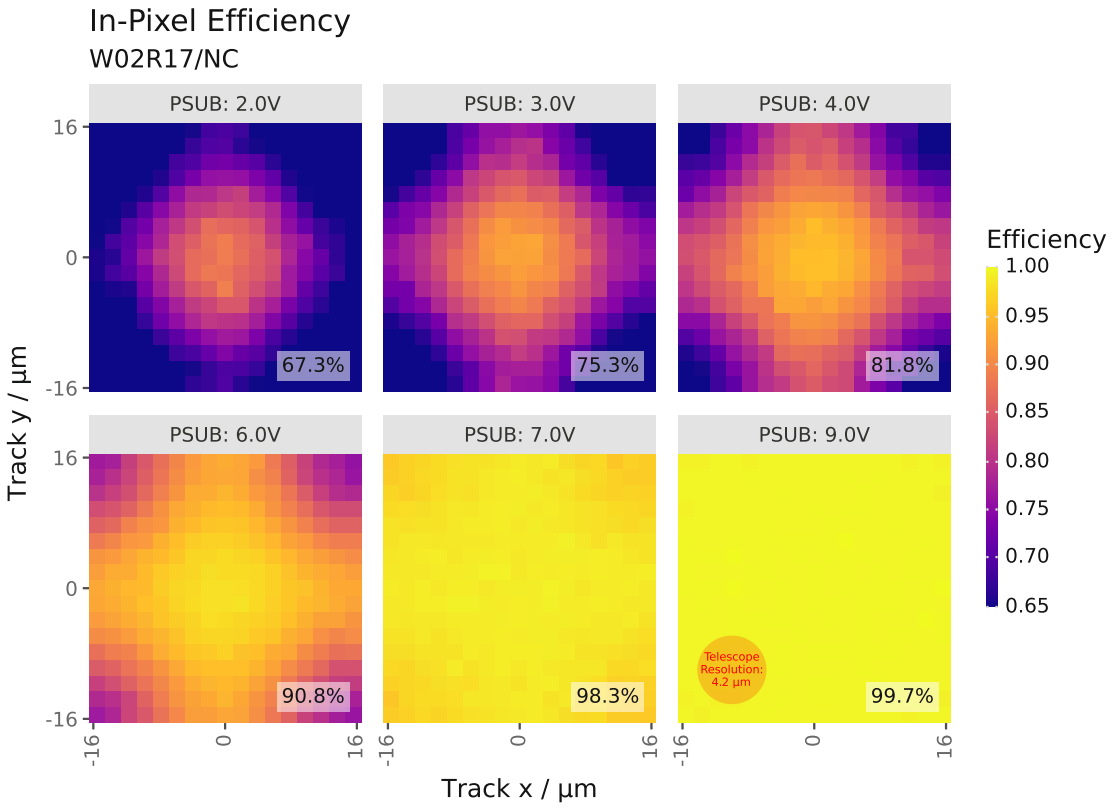


Figure 4.11: In-pixel efficiency for the W02R17 DUT, irradiated with protons to a fluence of $2.5 \times 10^{14}\,\mathrm{n_{eq}/cm^2}$.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

The bias voltage has been varied for all DUTs and the results are shown in figure 4.12. Here, bias voltage means either PSUB for the NC frontend or HV for the HVC frontend. For the NC, all DUTs reach full efficiency at least at 12 V, while for the HVC frontend a bias of 30 V is required at the highest fluence. The less irradiated chips reach full depletion and therefore full efficiency earlier.

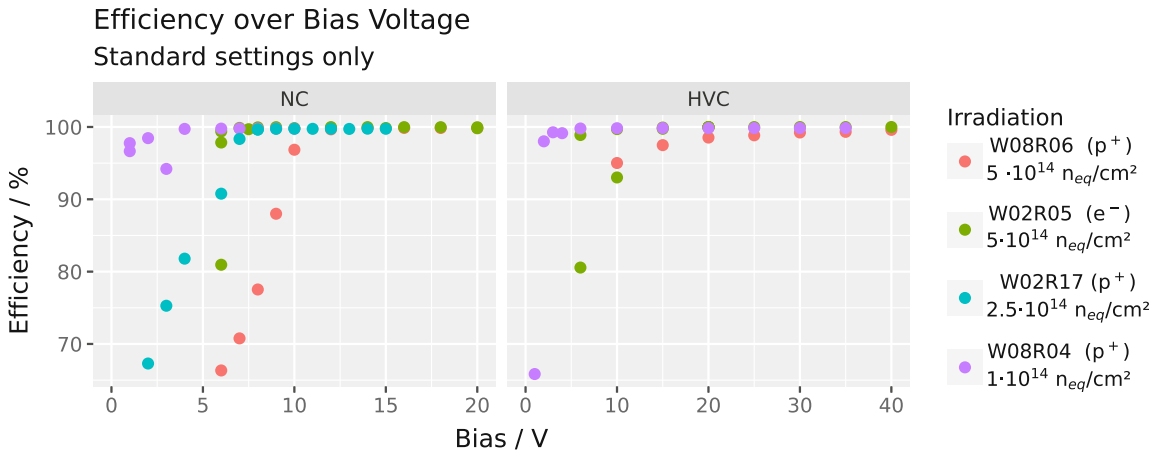


Figure 4.12: Efficiency of measurement using the typical frontend settings over the bias voltage.

4.5 Timing Measurements

Measuring the timing of individual components, such as propagation delay of specific signals is usually not possible directly in a chip with the complexity of TJ-Monopix2. Instead, for different measurements the results are sums or differences of various contributions. With multiple types of measurements, it is therefore possible to reconstruct the timing of the individual components of the chip.

To visualize these measurements, a type of time-schematic as shown in figure 4.13 is introduced. Delays are denoted as a chain of arrows, indicating the components from the start to the end point of the total delay in a similar way as voltages are denoted in electronic schematics. An arrow from start to end signifies a delay in the signal while one pointing towards the start point indicates an effectively negative delay, i.e. this delay reduces the total delay, typically by delaying the time reference.

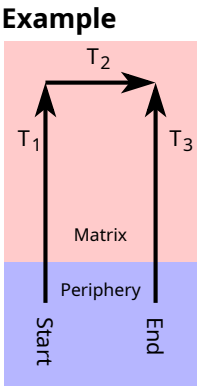


Figure 4.13: Generic example of delays with $T_{tot} = T_1 + T_2 - T_3$

Delays depending on the row address are denoted as vertical arrows, with the periphery at the bottom and the pixel matrix on the top. The length of the timing arrows has no meaning, but vertical arrows are a function of the row address in the diagrams. Occasionally it is necessary to connect two arrows at a distance, which is done with dashed lines equivalent to a 'zero delay' arrow.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

4.5.1 The TDC module

This section covers measurements regarding the timing properties of the TJ-Monopix2 chip. Most of those tests make use of the **HitOr** output, as well as a Time-to-Digital converter (TDC) module included in the FPGA firmware of DAQ system. The TDC module is a part of the Basil framework [46] and injects 32 bit words into the data-stream of the chip. These TDC words can be differentiated from other data by the 4 bit identification header (ID) of these 32 bits.

ID	TRIGGER DIST.	TIMESTAMP	TDC value
4 bit	8 bit at 640 MHz	8 bit at 40 MHz	12 bit at 640 MHz

Table 4.7: The data format of the TDC words used for the following measurements.

The TDC module exposes two inputs, a **Trigger** and a **TDC In** input. The data format of the TDC module can be configured, but only one configuration shown in table 4.7 was used for the measurements in this document. In this configuration, the TDC word is added when a rising edge of the **Trigger** input is followed by a rising edge of the **TDC In** signal.

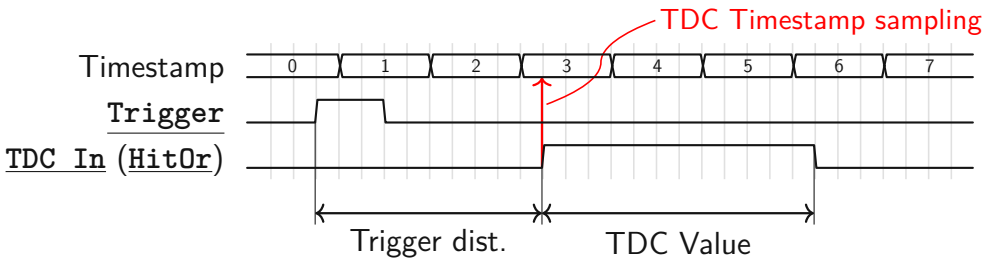
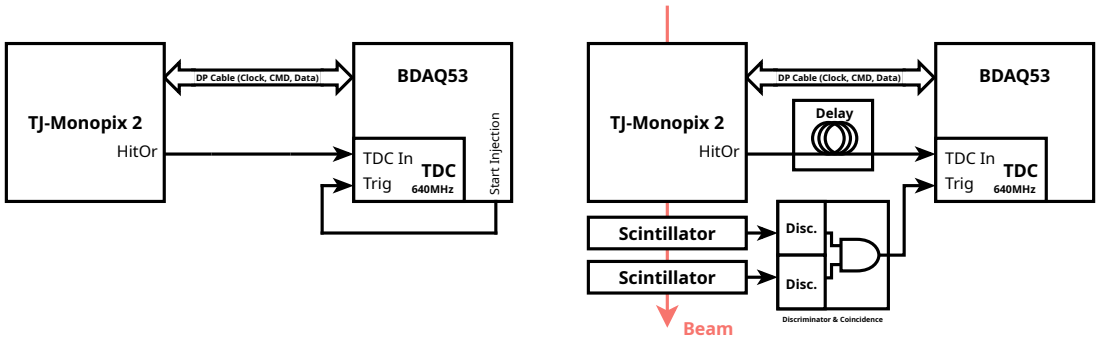


Figure 4.14: TDC Module timing chart.

The time between the two edges is measured at a clock frequency of 640 MHz with 8 bit, leading to a maximum value of 400 ns called trigger distance, with a binning of about 1.5 ns. The distance between the rising edge and the falling edge of the **TDC In** signal is measured with 12 bits at the same frequency, leading to a maximum value of 6.4 µs for this so called TDC value. A timestamp of 40 MHz is also sampled at the rising edge of the **TDC In** signal. For all measurements in this document, it is more useful to reference the timestamp to the **Trigger** input instead of the **TDC In** input. This has been done at time of analysis by up-sampling the timestamp and adjusting by the trigger distance. The timing diagram 4.14 shows the three values available of the TDC module.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.



(a) (b)

Figure 4.15: Measurement setup to measure timewalk and HitOr delay via injection (a) and at beam-test (b)

Measurements with internal time reference (Lab-tests)

For lab-tests, the charge injection circuit of the chip was used. Each pixel was injected with some charge individually, and the HitOr signal was fed into the TDC module of the DAQ-system, as shown in figure 4.15a. The TDC module was triggered by the DAQ system itself before each injection command. This allows the measurement of the time distance between charge injections in relation to pixel position and injection charge. The pixel position dependent contribution is caused by the propagation of the injection signal along the column from the periphery to the target pixel plus the propagation from the HitOr signal back to the periphery.

Since propagation delays of the injection and HitOr signals are both position dependent, only the sum of the two can be extracted together. The timewalk, however, can be fully extracted as it only depends on the charge while the other delays are constant for a given pixel.

Measurements with external time reference (Test-beam)

At a test-beam with real particles, the information available allows decomposition of the timewalk and HitOr propagation components. The measurement setup involves a trigger circuit via scintillators as shown in figure 4.15b. This allows a more direct measurement of the delay contributions but requires a more complicated measurement setup and a test-beam facility to be performed.

In principle this measurement does not require straight tracks and could be performed with a radioactive beta-source instead. This has been tested, but the measurement quality was far below the test-beam results and are therefore only the test-beam results are presented.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

4.5.2 Test-beam Timing Measurements

At a test-beam with real particles, the information available allows timing decomposition of the timewalk and HitOr propagation components as shown in figure 4.16. The measurement setup involves the TDC with a trigger circuit via scintillators as shown in figure 4.15b in the previous section.

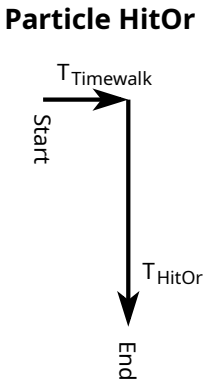


Figure 4.16: Delay diagram for test-beam HitOr measurements.

As the TDC data is injected into the TJ-Monopix2 data stream inside the FPGA, an association between hits and TDC data needs to be made offline. The association used a 10-hit search window after each TDC word. In this window, the hit-ToT was matched to the measured TDC-ToT (called TDC Value in the Basil-TDC module). Since the **HitOr** output is only one bit for the whole chip, this assures only clean data is matched, as the TDC-ToT becomes longer if multiple hits overlap and the matching fails in this case. Additionally, also the timestamp could be used for matching, but this was omitted as the resulting association was sufficiently clean.

Subsequently, three corrections were applied to correct for the HitOr delay for the column and row, as well as timewalk. Since a multi-dimensional histogram for all three properties at the same time would require a lot of memory at time of analysis, the corrections were fitted sequentially to histograms dependent on one property each. After every fit, the new set of corrections was used to recompute the histograms. In total, three such iterations were performed to obtain the optimal correction parameters. In the following, the three correction steps of column and row HitOr delay and timewalk will be explained individually.

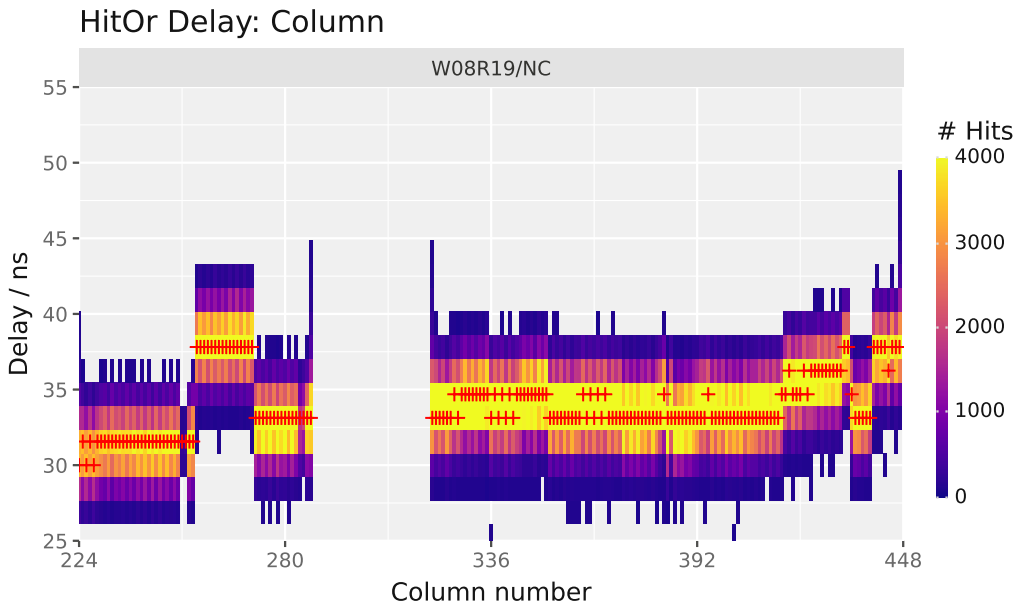


Figure 4.17: Column dependent HitOr delay, 2D histogram and used corrections (red crosses), columns 288 to 319 disabled.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

A histogram of the HitOr delay of the individual columns is shown in figure 4.17. The delay between the scintillators and the rising edge of the **HitOr** signal is shown over the column number, with the color representing the number of hits in the corresponding bin. This histogram has the required corrections for timewalk and row-delay already applied to reduce the variation on the y axis. Since the delay does not form a continuous function, the delay-bin in each column with the maximum number of hits was chosen as correction. Since the whole timing analysis was binned to the resolution of this plot (1.5625 ns), a Gaussian fit for each column-slice to increase resolution is not useful. Re-binning the data with a higher resolution could be used to take advantage of this, but was omitted for the measurements presented here. The use of the maximum bin for each column slice proved to be an extremely reliable method and the resulting maxima are shown as red crosses.

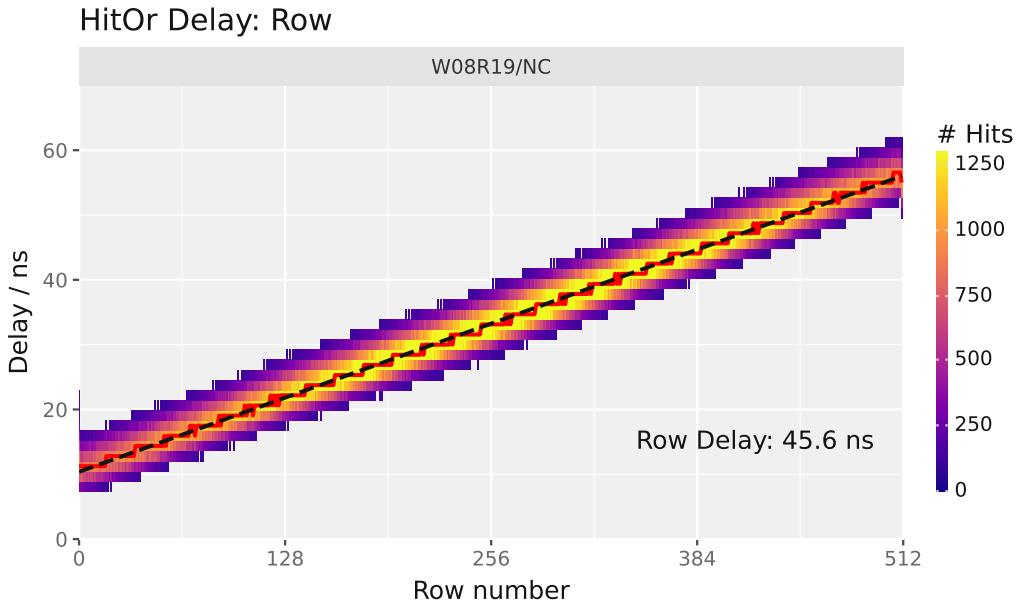


Figure 4.18: Row dependent HitOr delay, 2D histogram, maximal values (red line) and correction fit (black dashed line).

In a similar way, the row dependent HitOr delay was measured and corrected. Figure 4.18 shows a histogram with corrections for column delay and timewalk applied. Since the row delay originates from the gate delay of the OR gates in each row with the same layout in each pixel, it is a linear function of the row number. The maximum on each row slice (red line) was used to fit the linear function shown as black dashed line in the plot. The fit was used to evaluate the total row-delay, i.e. the time difference between row 0 and row 511. While again slightly more resolution could be achieved using a Gaussian fit on each slice, the maximum method proved to be more robust when analyzing large amounts of data in multiple runs.

Furthermore, a better time resolution could be achieved when each column is fitted individually and not one fit for the whole chip globally. The amount of data collected would be sufficient to prevent over-fitting also in this case, and PVT gradients across the chip could be corrected with such a method. While possible in a testing environment, in a detector with thousands of chips this is not feasible. Since the results of this chapter will be used in section 5.9.6 for the precision timing feature of the OBELIX chip, the global option was used. In VTX, the same method can be used with one number (Row delay) describing each chip. The equivalent of

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

the column delay in OBELIX will be mitigated by tighter constraints in the digital logic, and due to the smaller magnitude they can be, if needed at all, assumed the same for all chips to a good approximation.

For another feature of the OBELIX chip, namely the contribution to the Belle II L1 trigger which will be described in detail in section 5.11, it is not possible to correct for this type of delay. Therefore, the magnitude of the delay is of interest on its own too.

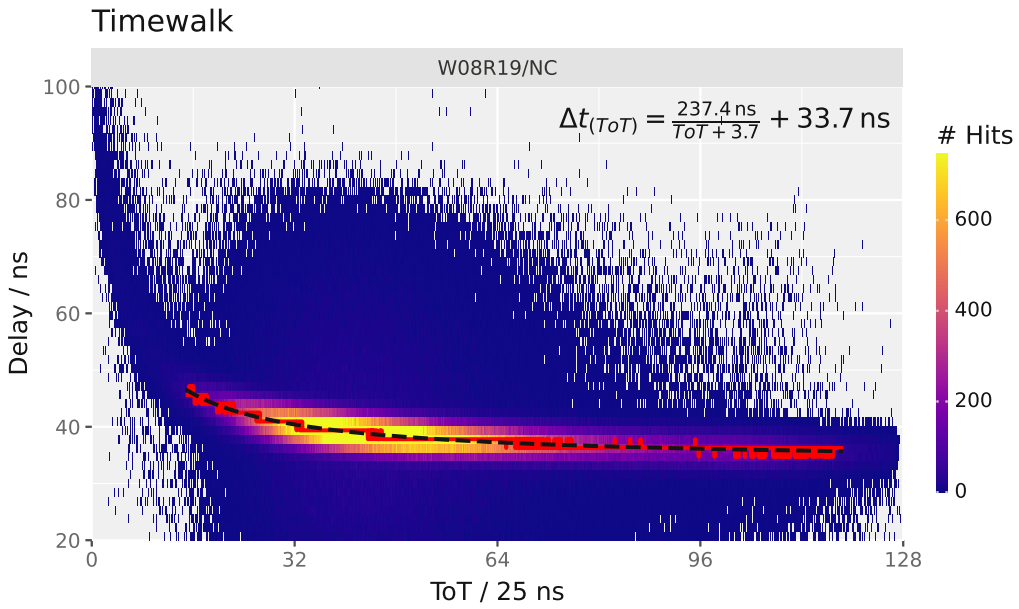


Figure 4.19: Timewalk as 2D histogram (uncleaned), corresponding correction fit.

Finally, the timewalk was measured and corrected. For the hitogram in figure 4.19, the delay of the trigger distance is plotted against the ToT with corrections for row and column delay of the **HitOr** line applied. Similar to the HitOr delay, ToT-slices with a sufficient number of hits were used to fit a generic function for the timewalk, shown in the top. The function of section 3.4.1 is not applicable, as the ToT does not directly correspond to charge, but would require calibration function which is absorbed directly into the fit shown here.

While the two previous plots were cleaned, i.e. only bins with more than 5% of the maximum hits in a bin were shown, this plot contains raw data of all hits. The halo around the mid-ToT values arises from the occasional misassociation of hits. As this is most probable at frequently occurring ToT values, a halo is formed around the real data. However, at ToT values below 20, a continuation of the timewalk curve towards zero ToT is visible. This feature arises from hits close to the edge of the sensor, when the seed pixel of the cluster would be just outside the enabled pixel region and only the periphery pixel is detected. At a close look of figure 4.17, the underestimated timewalk correction of these events is visible on the boundaries of the available data, causing the antenna-like spikes at columns 224, 287, 320 and 447. Columns 288 to 319 were disabled due to a different behavior than the rest of the matrix.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

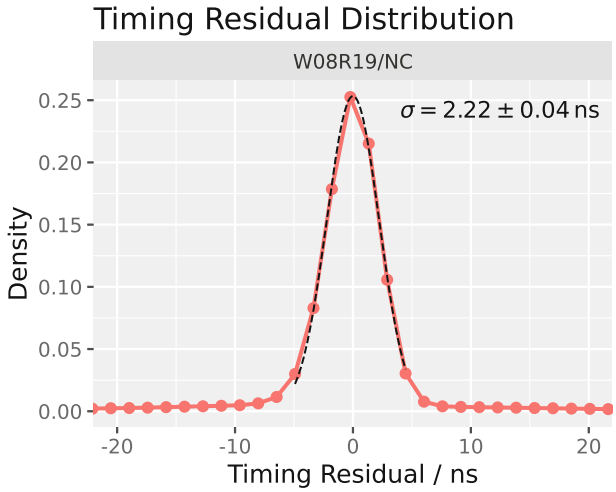


Figure 4.20: Normalized timing residual after all corrections applied.

After all three corrections applied, the timing residual is shown in figure 4.20. A Gaussian fit is also shown to estimate the timing resolution. The same types of corrections will be needed for the precision timing module of the OBELIX chip as it operates in a similar way. Therefore, this type of measurement is important to quantify the expected performance of this feature of OBELIX, which will be described in section 5.9.6.

For the measurements presented, the timing resolution of the scintillator reference is not subtracted from the estimation. Therefore, the quoted resolutions are a convolution of the timing resolution of the DUT with the time resolution of the reference. For future measurements, an estimation of the time resolution of the reference can be performed for better results, but this needs to be performed with the same setup at the test-beam itself.

Results

At the July 2023 test-beam, a total of four chips was measured as shown in table 4.8. Two chips were of the epitaxial type interesting for the VTX upgrade, while one chip was based on a Czochralski high resistivity substrate. This type is not considered by the VTX upgrade due to the cost of the high resistivity wafer material, but has been measured anyway. Additionally, one chip was irradiated with protons to an NIEL fluence of $5 \times 10^{14}\,\mathrm{n_{eq}/cm^2}$ at the CYRCé facility in France [43].

Serial	NIEL	Comment
W02R05	none	
W08R19	none	
W14R12	none	Czochralski substrate
W05R16	$5 \times 10^{14}\,\mathrm{n_{eq}/cm^2}$	proton irradiated

Table 4.8: Irradiated DUTs for the test-beam in July 2023.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

Figure 4.21 shows the standard deviations of the timing residuals of the four chips for the different frontends. The points are shifted on the x axis by random amounts to visualize the individual measurements. The results of these timing measurements have already presented in [47], but the quoted timing resolutions are off by a factor $\sqrt{2}$ due to an analysis error. This has been corrected and the residuals are correct in this analysis. The averages over the measurements are also summarized in table 4.9.

For the unirradiated samples, a timing resolution below 3 ns was measured. However, for the W08R19 chip, the HV and HVC results are significantly worse. This is caused by improper timewalk correction. For the settings used at the test, the tail of the ToT distribution reached a value above 127. This cannot be resolved with the 7 bit BCID and the number overflows. This is then detected as a low-ToT value and the timewalk correction is introducing a large error. Using the ToT-clipping feature via `VCLIP` or by a larger `ITHR` value, this could have been prevented. The measurements affected by this are excluded from the results in the summary table.

The irradiated sample W05R16 showed a slightly worse timing resolution of about 4 ns. One of the main contributors is the timewalk correction. Since after irradiation the chip operates closer to the threshold, the variation due to timewalk of typical events becomes larger. Since the correction is only done globally per chip, this increases the effect of imperfect corrections, e.g. due to variations in threshold.

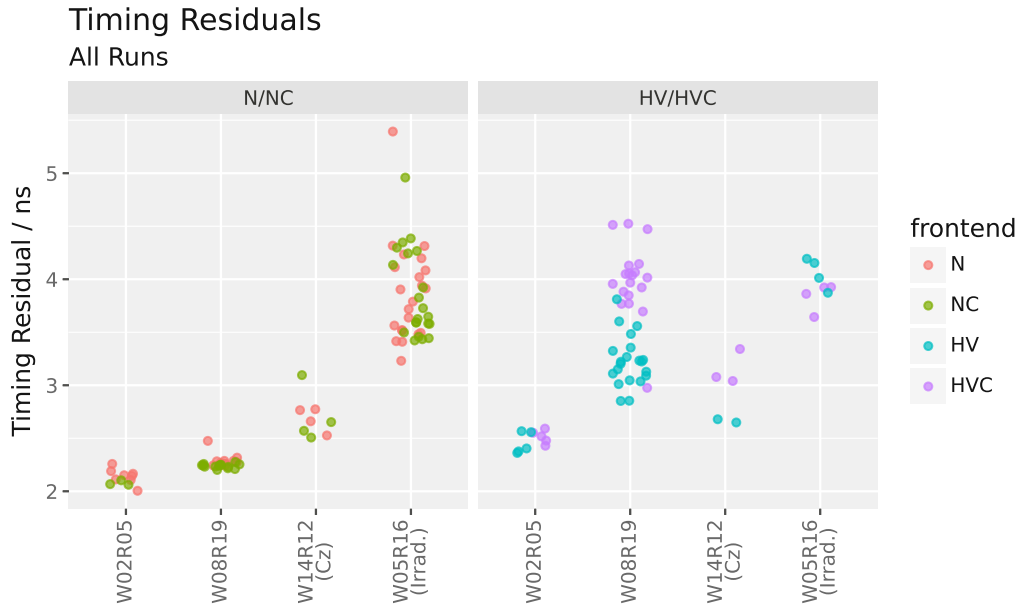


Figure 4.21: Timing residuals for different DUTs and frontends.

Serial	N/NC	HV/HVC
W02R05/W08R19	2.22 ± 0.09 ns	2.48 ± 0.09 ns
W14R12 (Cz)	2.69 ± 0.19 ns	2.96 ± 0.29 ns
W05R16 (Irrad.)	3.80 ± 0.34 ns	3.95 ± 0.17 ns

Table 4.9: Standard deviation of timing residuals for different DUTs and frontends.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

The HitOr row delay compensations used for the fits are shown in figure 4.22. While the measurements for the Czochralski chip W14R12 differ significantly, the unirradiated and irradiated epitaxial chips show similar results. For the measurements of the HV and HVC frontends, a PSUB and PWELL potential of 1 V was applied, except for some of the measurements of the W08R19 chip, which are marked with a black box. It was later found that operation at PSUB/PWELL of 0 V is preferred for the HV and HVC frontends, but this was not known at the time of measurement. Since the PWELL potential affects the transistors and therefore timing of the HitOr OR-gates, the PSUB potential changes the timing behavior. The results are summarized in table 4.10.

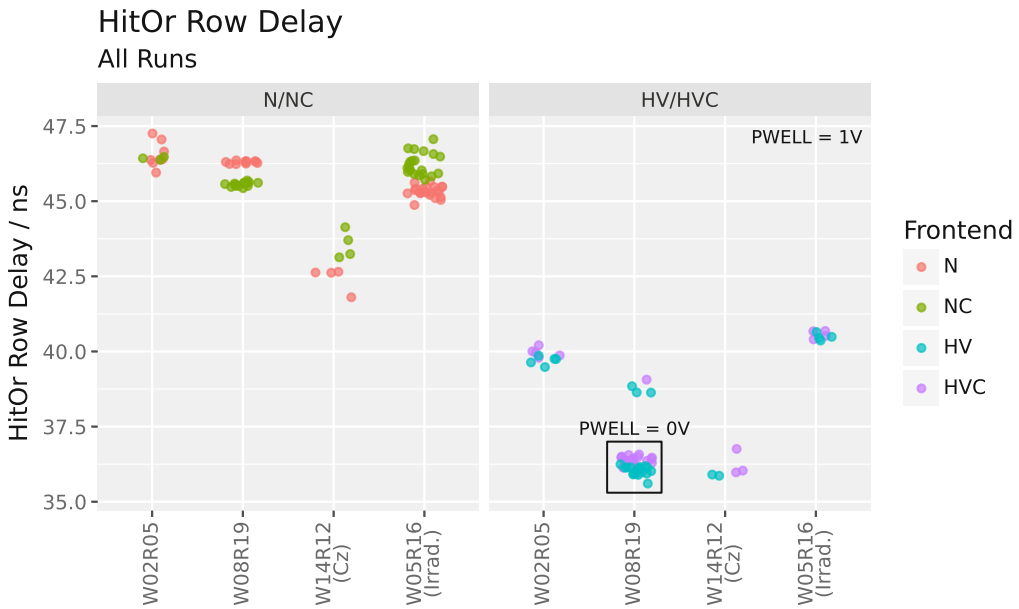


Figure 4.22: HitOr row delay for different DUTs and frontends.

Serial	N/NC	HV/HVC	
PSUB/PWELL	6 V	1 V	0 V
W02R05/W08R19	46.1 ± 0.5 ns	39.5 ± 0.5 ns	36.2 ± 0.2 ns
W14R12 (Cz)	43.0 ± 0.7 ns	36.1 ± 0.4 ns	
W05R16 (Irrad.)	45.8 ± 0.5 ns	40.5 ± 0.1 ns	

Table 4.10: Summarized HitOr row delay for different DUTs and frontends.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

4.5.3 BCID Delay Compensation

TJ-Monopix2 has an internal compensation for the row-dependent delay of the BCID distribution. Inside each pixel, a starved-current delay buffer is placed between the **HIT** output of the discriminator and the BCID sampling circuit. This means that the sampling of the BCID for Le and Te is delayed ideally to exactly counteract the delay of the BCID distribution as shown in figure 4.23. The rated delay of the circuit depends on the row number but groups of adjacent rows share the same version. In the center region where the largest change in BCID delay is happening, 16 rows share the same delay circuit whereas at the top and bottom of the chip, the BCID delay is compensated the same way for more rows.

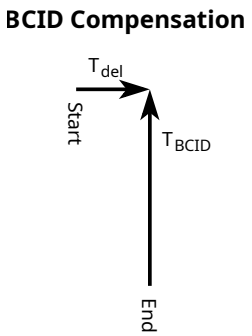


Figure 4.23: Delay diagram for BCID compensation.

A global analog signal from the **IDEL** DAC can be used to globally steer all delay circuits. A larger value of **IDEL** leads to more current in the delay circuit and reducing its delay, thus decreasing the magnitude of compensation. This allows a first order matching in case of simulation errors used to size the compensation circuits.

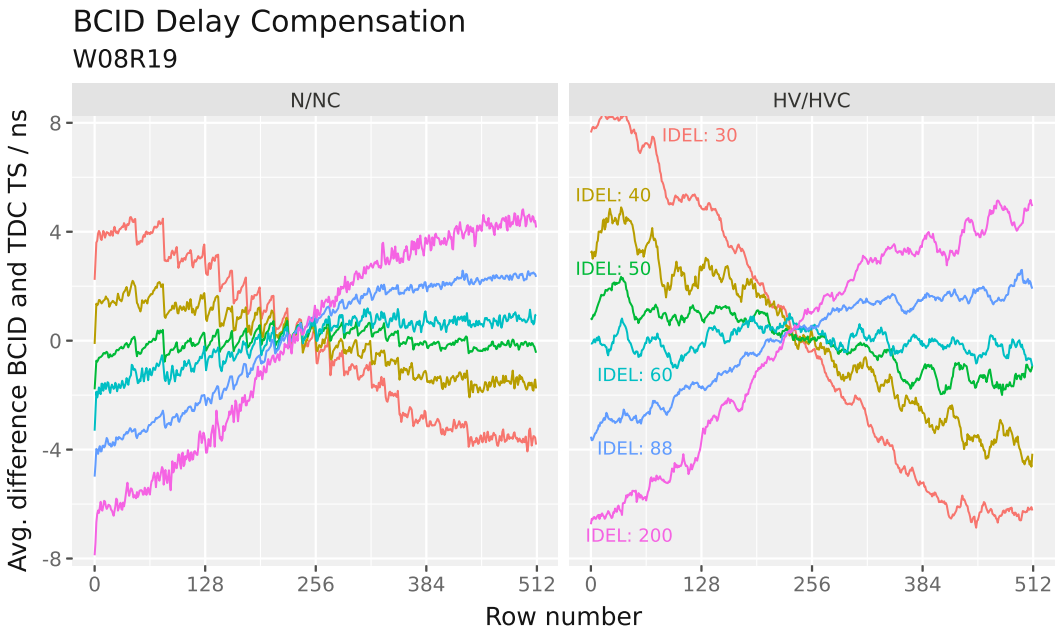


Figure 4.24: Effect of **IDEL** setting on the timing residual depending on the row number.

In a beam-test scenario, an optimal value of **IDEL** = 50 was found for N/NC flavors and 60 for the HV/HVC flavors as shown in figure 4.24. This is considerably lower than the expected **IDEL** = 88 suggesting that either the BCID delay was underestimated or the actual delay of the starved inverter was overestimated in simulations. For low values of **IDEL**, the stronger delay of the HV/HVC flavors is visible. No difference between the flavors can be seen at **IDEL** = 200, corresponding to almost no delay compensation. This suggest that the difference is due to the compensation circuit, and not caused by difference in BCID propagation time.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Most probably, the discrepancy between the two frontend types is caused by the different PWELL voltage. However, a SPICE simulation to investigate this is not possible due to the lack of transistor-models for large bulk potentials in the used technology. Furthermore, the 16 row sawtooth structure is visible for the N/NC versions in the center region of the chip. Stronger smoothing was necessary for the AC flavors because of the small number of columns, reducing the amount of data collectible in reasonable time. In particular, the amplitude of the sawtooth is decreasing with increasing setting of `IDEL` until it entirely disappears at `IDEL` = 200 with effectively no compensation.

4.6 BCID Crosstalk

It is known [48] that the analog frontend amplifier is affected by the BCID distribution. This crosstalk is particularly prominent in the 5 MHz range, as amplification of the frontend has a maximum in this frequency range. The effect is skewing the distribution of BCID codes for the Le and Te. At times where the threshold is a little bit lower, the corresponding BCID code is favored when a signal arrives close to the BCID code transition. The effect is stronger on the trailing edge, because of the slope of the analog signal shape and a the propagation is sketched in figure 4.25. The rise time is steep, but the fall time is larger, leaving a larger time window to affect BCID codes.

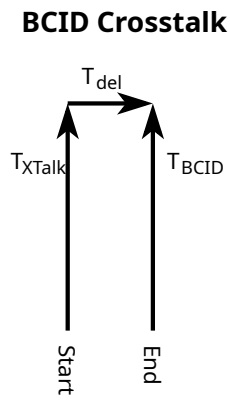
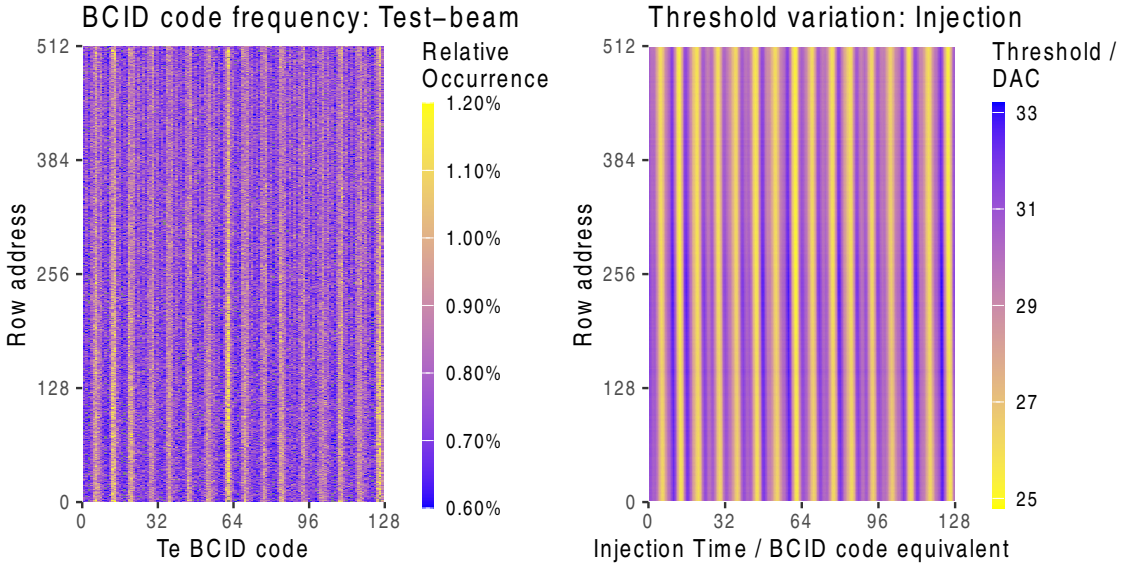


Figure 4.25: Delay diagram for test-beam BCID Crosstalk.

Figure 4.26a shows the relative number of occurrence for different BCID codes, depending on the row number. Without crosstalk, a uniform distribution is expected, however, a stripe pattern is clearly visible. The stripes show a periodicity of 8 BCID codes or 5 MHz.



(a) (b)

Figure 4.26: 2D Histogram of BCID code frequency in a test-beam depending on row number for the Te code (a) and threshold over time via charge injection (b).

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

From this graph, small differences in phase of the variation on the top and bottom rows are not perceptable. However, it is possible to fit a sine function to the individual horizontal slices. The BCID propagation and delay compensation affects this phase, as well as the propagation time of the crosstalk signal. The propagation time of the crosstalk can be directly extracted from the slope if the BCID propagation is exactly compensated which is the case at a **IDEL** = 50.

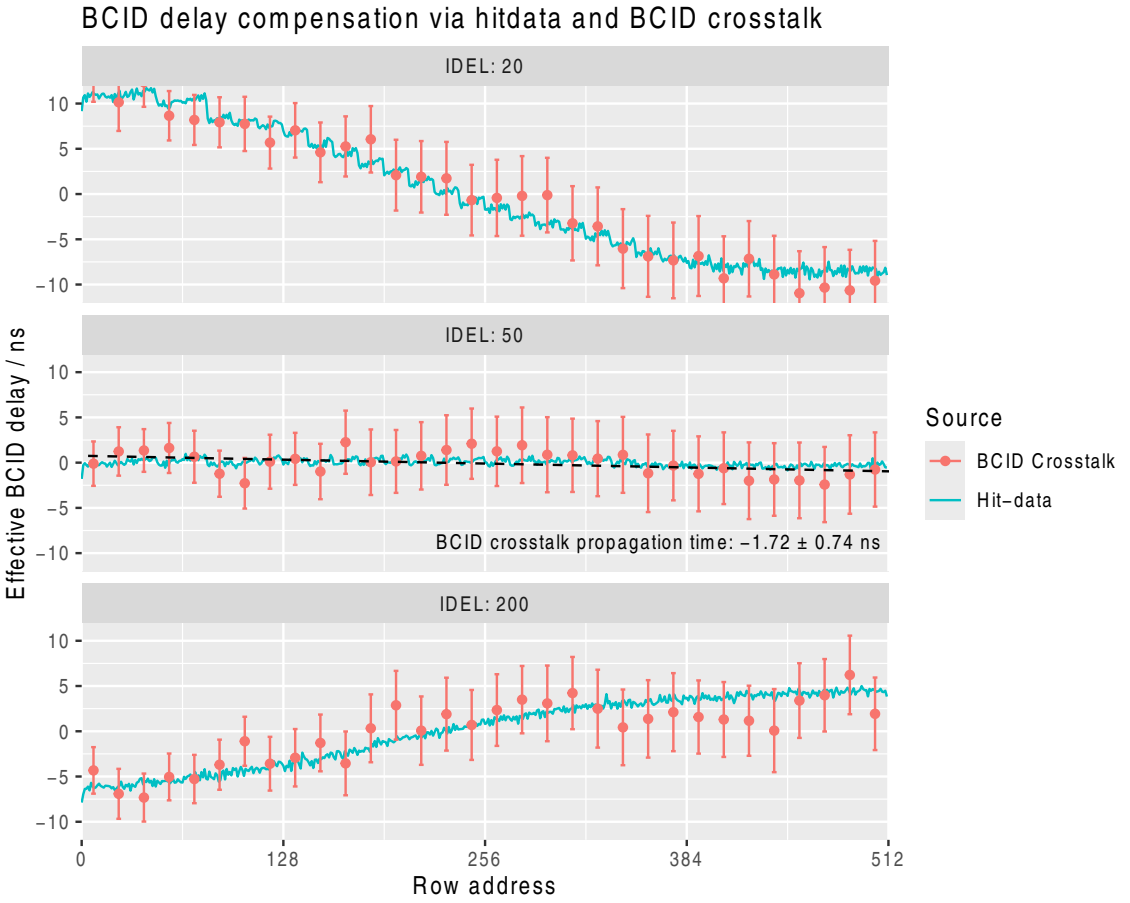


Figure 4.27: Effect of **IDEL** setting on phase of BCID crosstalk in comparison with real timing extraction using hit-timing.

In figure 4.27, the phase of the BCID code inhomogeneity calculated in groups of 16 rows is shown along the BCID-residual for three values of **IDEL**. The crosstalk tracks the BCID delay, leaving only a small propagation of the crosstalk. Despite the use of the largest data set available, the error bars are large since the effect on the BCID code frequency is small (grainy data in figure 4.26a). For the optimal compensation, the crosstalk is fitted with a linear function and shows a small and lightly negative propagation time. This excludes direct coupling of the BCID signal to the frontend and suggests coupling through the power supply lines and/or substrate, effectively simultaneous over the full matrix.

By variation of the time of injection in respect to a BCID-reset, it is possible to map the effect of this crosstalk on the threshold in a more controlled way with

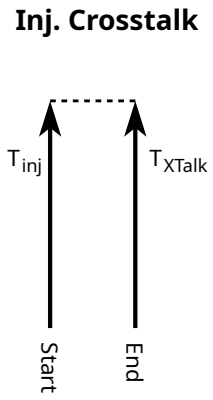
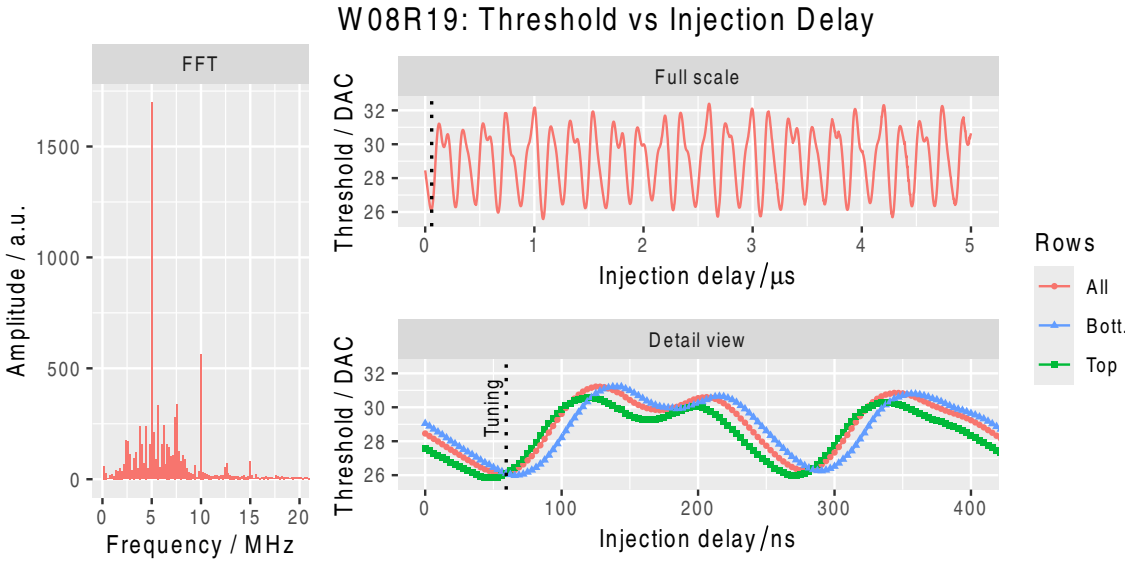


Figure 4.28: Delay diagram for BCID crosstalk via injection.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

a timing shown in figure 4.28. Figure 4.29b shows this threshold variation measured with s-curves at different injection times. As visible, the amplitude of this effect is quite severe with about 6 DAC units. An fast Fourier transform (FFT) of the measurement is shown in figure 4.29a. In TJ-Monopix2 the BCID of 40 MHz are gray encoded. The highest bit toggle frequency is a quarter of the base frequency for this encoding (10 MHz), but harmonics are also visible.

In the bottom part of figure 4.29b, data from the same measurement is shown in detail. Here, not only the total average is plotted, but also the average of the top and bottom 8 rows only. This reveals a phase difference for the top and bottom pixels.



(a) (b)

Figure 4.29: Effect of BCID Crosstalk: Threshold depending on injection time.

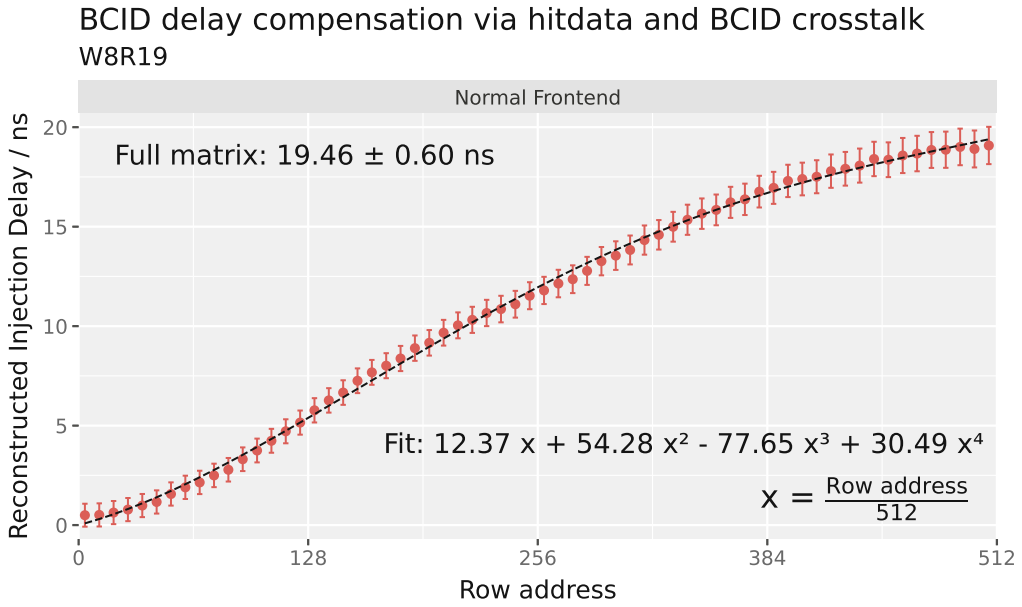


Figure 4.30: Extracted injection delay from the phase of BCID crosstalk effect on the threshold for different row positions.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Apart from the delay from the BCID crosstalk, the main source of this phase delay is the propagation of injection signal itself. This injection delay can be directly inferred by analyzing the delay for different row positions. The peak of the cross-correlation was used to fit the threshold variation curve for groups of 8 rows. All combinations of row groups were evaluated and the position dependent delay and error were calculated using the least squares adjustment method. Finally, the BCID crosstalk component was subtracted. The results are shown in figure 4.30 and a total variation of injection delay of $T_{inj} = 19.46 \pm 0.60$ ns has been determined from a polynomial fit of 4th order.

Impact of BCID crosstalk on Tuning

Since TJ-Monopix2 is capable of in-pixel tuning as described in section 4.1.3, the effect of threshold variation via the BCID crosstalk needs to be taken into account. Tuning without a fixed phase leads to non-repeatable results, because the momentary BCID phase strongly affects the threshold measurement and therefore the tuning result. Without this repeatability, a bisection algorithm as used in the tuning does not function correctly.

To restore the repeatability, a BCID reset was introduced in the measurement routine, allowing precise control over the time of injection after the BCID reset. While this immediately restores repeatability, it opens the question of the optimal time to inject. The injection delay is basically a window of about 20 ns on the threshold oscillation curve, where bottom pixels are on the lower border and top pixels on the higher border of the window.

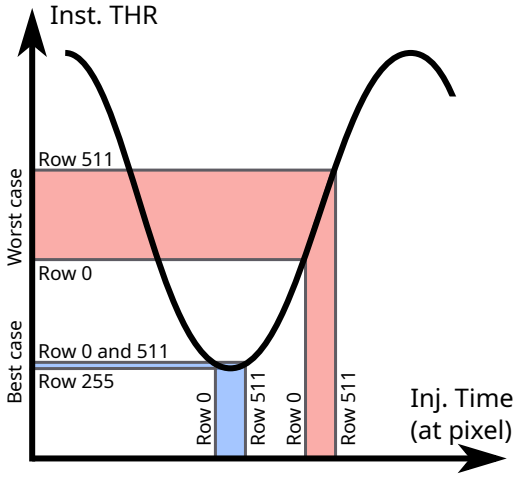


Figure 4.31: Effect of injection time on apparent threshold dispersion and therefore tuning via the instantaneous threshold, optimal case (blue) and worst case (red).

When an injection time close to the inflection point of the threshold oscillation is chosen, the variation is maximal. The measured threshold of top pixels would be significantly different from the bottom pixels, just because the time of injection at the pixel location was different due to the injection delay.

On the other hand, if an extremum of the threshold oscillation is chosen, the variation is zero to the first order. This is because the gradient at the extremum is zero by definition. Expanding the sine to a Taylor series then leaves the quadratic term to be the first nonzero, minimizing the total variation. This is schematically drawn in figure 4.31.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

Furthermore, for low threshold operation, the minimal threshold is the most important point. The minimal threshold is limited by the point where noise is mistakenly detected as particle. Therefore, the minimum was chosen for tuning over the maximum. All tunings and threshold scans were performed at the first minimum of the threshold oscillation curve in this document.

This equalized tuning is of particular interest in the timing measurements of this chapter, as it would introduce a systematic error in the row dependent timing measurements via timewalk. If the threshold at the top and bottom of the matrix is different, the timewalk would add up on one side, skewing the results. Special care was therefore taken at the time of tuning and where possible, a large charge was injected in comparison with the threshold to minimize timewalk.

Due to the leading quadratic term in the threshold systematic after tuning, the exacted polynomial shown in figure 4.30 might not represent the exact injection delay in the center of the matrix. The total variation on top vs bottom would remain unaffected, but the middle rows would have a higher threshold leading to an increased timewalk delay. This results in a slight over-estimation of the injection delay in the center region, however, this effect appears to be sufficiently small, and did not affect the measurements in the following section that build on the injection delay measurement.

HitOr Delay (via injection)

It is possible to inject charge and measure the resulting waveform on the **HitOr** output. The delay between the start of injection and the rising edge on the **HitOr** is the roundtrip time of the signal, i.e. the sum of the injection delay and the HitOr delay for this particular pixel, as depicted in figure 4.32.

This was performed for all rows and several columns individually, and the resulting waveform is shown in figure 4.33. To allow direct comparison, the timing-offset for each column was subtracted individually to equalize the zero-delay point at the end of the column. This offset contains the variation inside the digital periphery of the injection delay, as well as the column dependent HitOr delay. Since both of those are product of the synthesis tool, they are not of interest for the development of OBELIX. However, the row dependent HitOr delay is, as it is a feature of the pixel matrix. The injection delay from the BCID crosstalk measurement is subsequently subtracted from the total measured delay shown as gray area in the plot. This only leaves the row dependent HitOr delay which is evaluated using a linear fit.

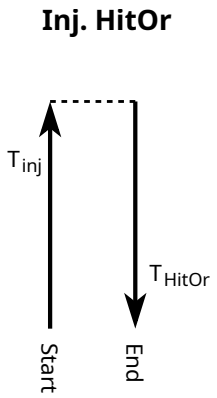


Figure 4.32: Delay diagram for HitOr delay extraction via charge injection.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

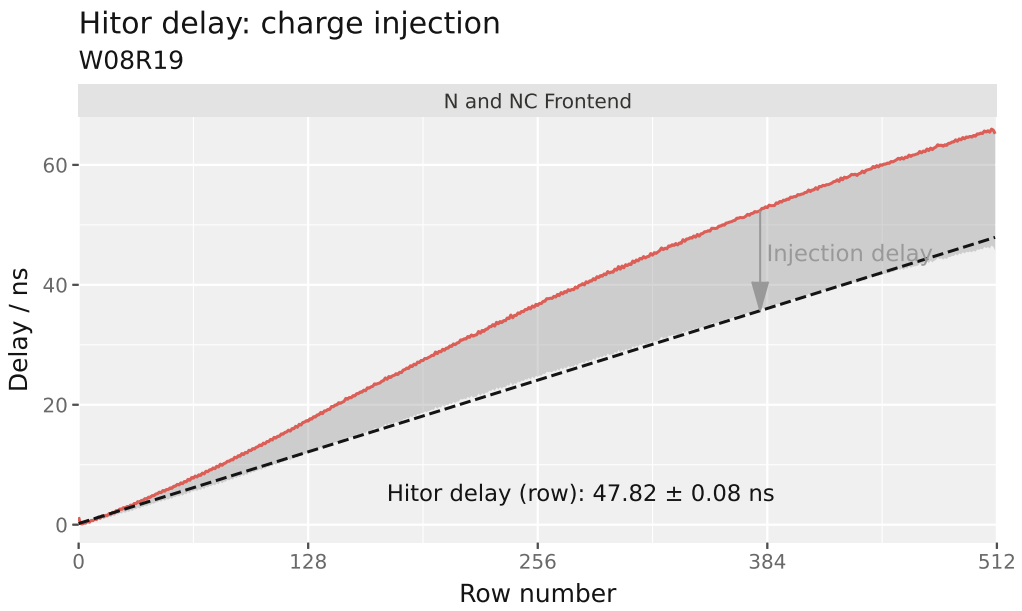


Figure 4.33: HitOr delay, measured via charge injection

Temperature and Bias dependence

The measurements of injection delay via the BCID crosstalk and the HitOr delay via the TDC module have been performed at different temperatures and PSUB/PWELL potentials. The assumption here is a constant crosstalk delay. While this is not known, the total delay of the crosstalk is less than 2 ns (see figure 4.27), so a variation of the exact value affects this measurement only slightly.

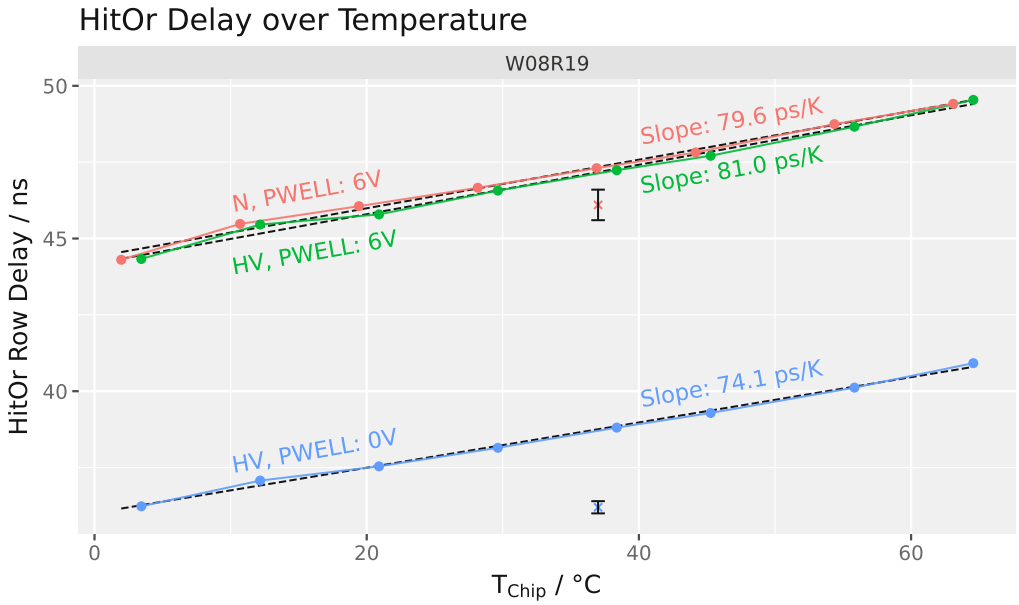


Figure 4.34: Extracted HitOr delay for different temperatures.

Figure 4.34 shows the temperature dependence of the HitOr row delay for the Normal and HV frontends, as well as for the HV frontend at a PSUB/PWELL potential of 6 V. The results from the test-beam measurement in section 4.5.2 are added as crosses with black error-bars. The indirect method with charge injection

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

clearly shows a systematic over-estimation in comparison with the direct measurement at the test-beam. However, a temperature coefficient of about 80 ps/K can be extracted. This is just an estimate and for OBELIX, this temperature coefficient should be verified at a test-beam to achieve more accuracy.

Figure 4.35 shows a similar measurement but with variation of the PSUB/P-WELL potential. Along the measurement for the normal and HV frontends, a mixed version is shown in addition. Here, the HitOr delay originates from the measurement of the HV frontend while the injection delay was taken from the normal frontend. A striking feature is the similarity of this mixed version to the results of the normal frontend. This suggests that the injection delay method is not directly applicable to the HV frontend, e.g. because the crosstalk propagation behaves differently for HV and HVC frontends. This is possible as the amplifier input has a different layout to accommodate the AC coupling capacitor. In any case, the HitOr delay is influenced by the PWELL potential and this can be further studied in test-beams with more confidence in the results.

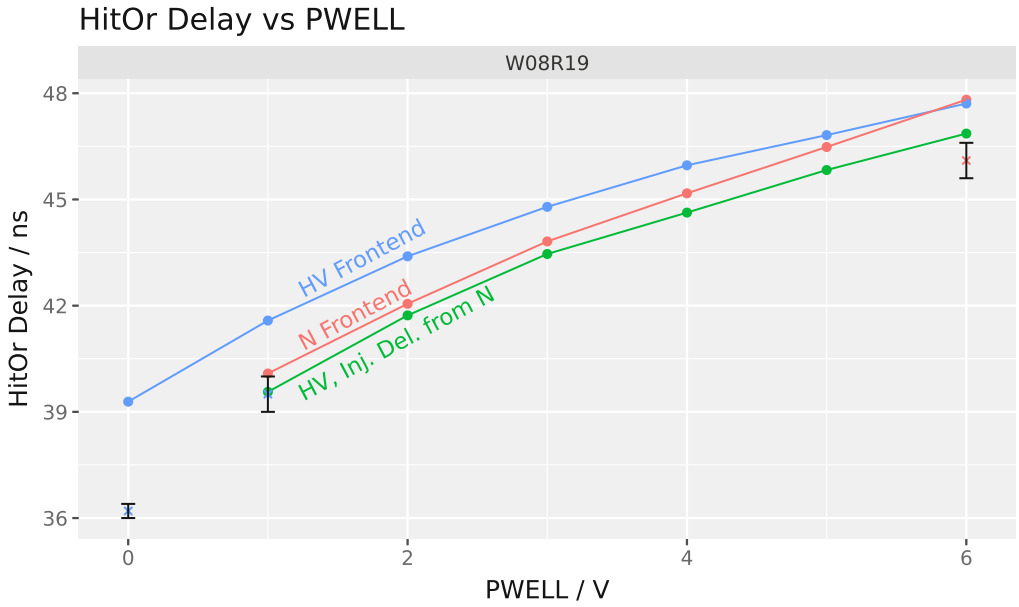


Figure 4.35: Extracted HitOr delay for different PSUB/PWELL potentials.

4.7 Timewalk

Timewalk is the unwanted property of constant threshold discriminators that the time response depends on signal amplitude. Especially small signals close to the threshold have significant delay compared with large signals. A theoretical treatment of timewalk is already given in section 3.4.1. In summary, the timewalk can be characterized with the function:

$$t_{timewalk} = t_{max}\left[1 - \sqrt{\frac{q - q_{thr}}{q}}\right], \tag{4.2}$$

where q is the signal charge, q_{thr} is the threshold and t_{max} defines the maximal timewalk. The threshold q_{thr} only depends on amplifier gain and baseline whereas t_{max} additionally depends on rise time and therefore amplifier bandwidth. Since both affect the resulting timewalk curves, the measurements have been tuned to equal

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

threshold, to allow direct comparison of t_{max}. The threshold q_{thr} can be measured more easily via an s-curve and only changes the scaling of the charge-axis.

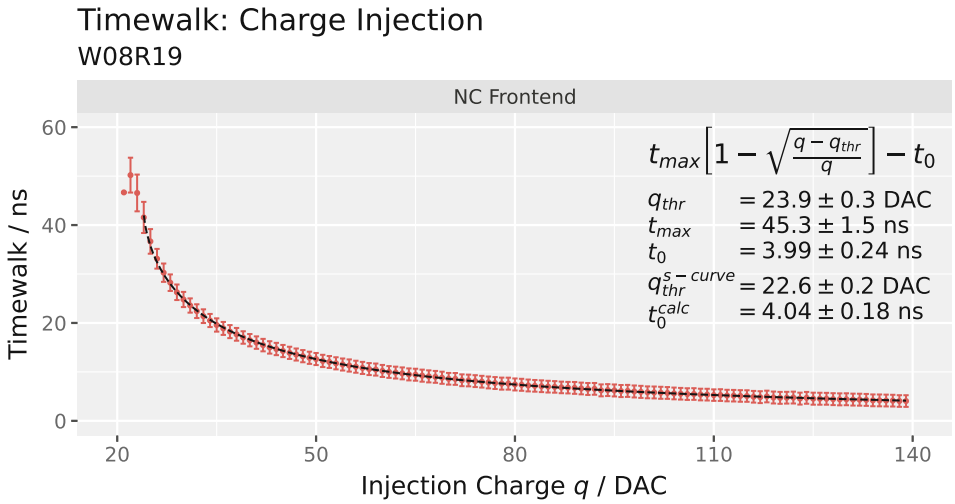


Figure 4.36: Example measurement of the timewalk with fit parameters.

To measure the timewalk in the laboratory, the typical setup with the TDC module and internal trigger is used, as shown already in figure 4.15a. Subsequently, one pixel is enabled and increasing charges are injected. For all timewalk measurements in this chapter, 16 pixels in an arrangement of 4 by 4 are averaged. For each measurement, the threshold is globally tuned via **ICASN** and locally tuned to obtain a precise and equalized threshold, except otherwise noted. An example measurement is shown in figure 4.36. The fit only includes points where the occupancy is higher than 50%, corresponding to the threshold definition of the s-curve. However, an additional effect skews the obtained threshold systematically higher: Very short ToT pulses may be registered by the pixel circuit, but are lost on the way to the HitOr output[4]. Since 50% occupancy in this measurement is now derived from registered TDC measurements, the threshold q_{thr} is slightly higher than $q_{thr}^{s-curve}$ derived from an s-curve fit. Therefore, the plots show both values for reference.

Since the absolute zero point of the delay is not known due to (constant) delays in the signal lines of injection and HitOr, an arbitrary zero was chosen at the maximally injected charge for the fit. Since the timewalk model has the zero point at infinite charge, this practice requires an additional offset parameter t_0 which is also fitted. At the same time, this offset can also be calculated from the other parameters $t_0^{calc} = t_{timewalk}(q = 140)$, which is shown in the plot. The two parameters agree within the errors for all measurements. To correctly visualize the reconstructed zero point, t_0^{calc} was used to shift the y axis in the visualization of the data and fit. This means, the y axis can be directly interpreted as absolute delay using this reconstructed zero point.

[4]This is either caused by inertial delay in the HitOr lines and/or the TDC module does not register too short pulses, but the effect was not further studied.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

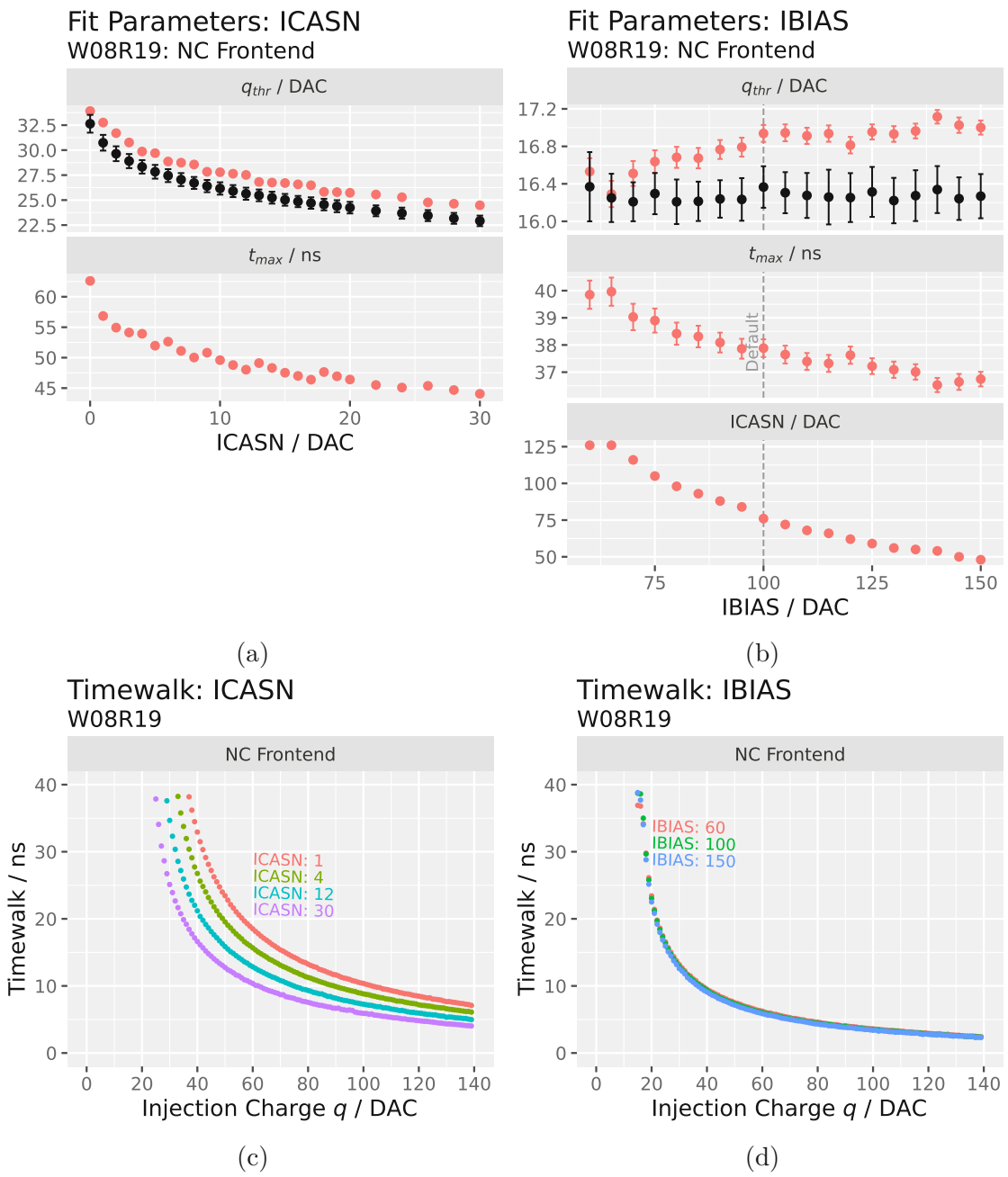


Figure 4.37: Extracted timewalk parameters for different **ICASN** (a) and **IBIAS** (b) values, corresponding timewalk curves (c and d).

Figure 4.37a shows the timewalk fit parameters for different **ICASN** values for the Normal Cascode frontend. Since **ICASN** is normally used to set the threshold, the resulting thresholds are different. This is also apparent in figure 4.37c where the timewalk curves of selected values are plotted. For lower thresholds, t_{max} is lower, but since both are affected by the baseline, no conclusion about the bandwidth can be made in this case. The red data points correspond to the extracted parameters of the timewalk fit while the black points originate from the s-curve fit.

On the other hand, figure 4.37b shows variation of **IBIAS** with an equalized target threshold of 16 DAC units[5]. As visible, the t_{max} increases with decreasing

[5]This threshold is too low to allow operation of the full matrix. However, the following scans require a low target threshold to allow for the required tuning range, and for comparability the same target was chosen for all scans. An **IBIAS** scan at higher threshold showed equivalent results.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

IBIAS. However, the magnitude of the decrease is less than 20% from the default value of **IBIAS** = 100. In the time domain shown in figure 4.37d, the difference is barely perceivable.

The value of **IBIAS** is of particular interest for the power consumption, as it is the main contributor to analog power dissipation. A significant reduction would be a suitable power reduction measure. However, this increases the threshold and reduces the bandwidth of the amplifier. The shown results suggest a small effect on amplifier bandwidth only. Therefore, a reduced **IBIAS** is a promising option for OBELIX operation, but is limited in practice due to the threshold increase.

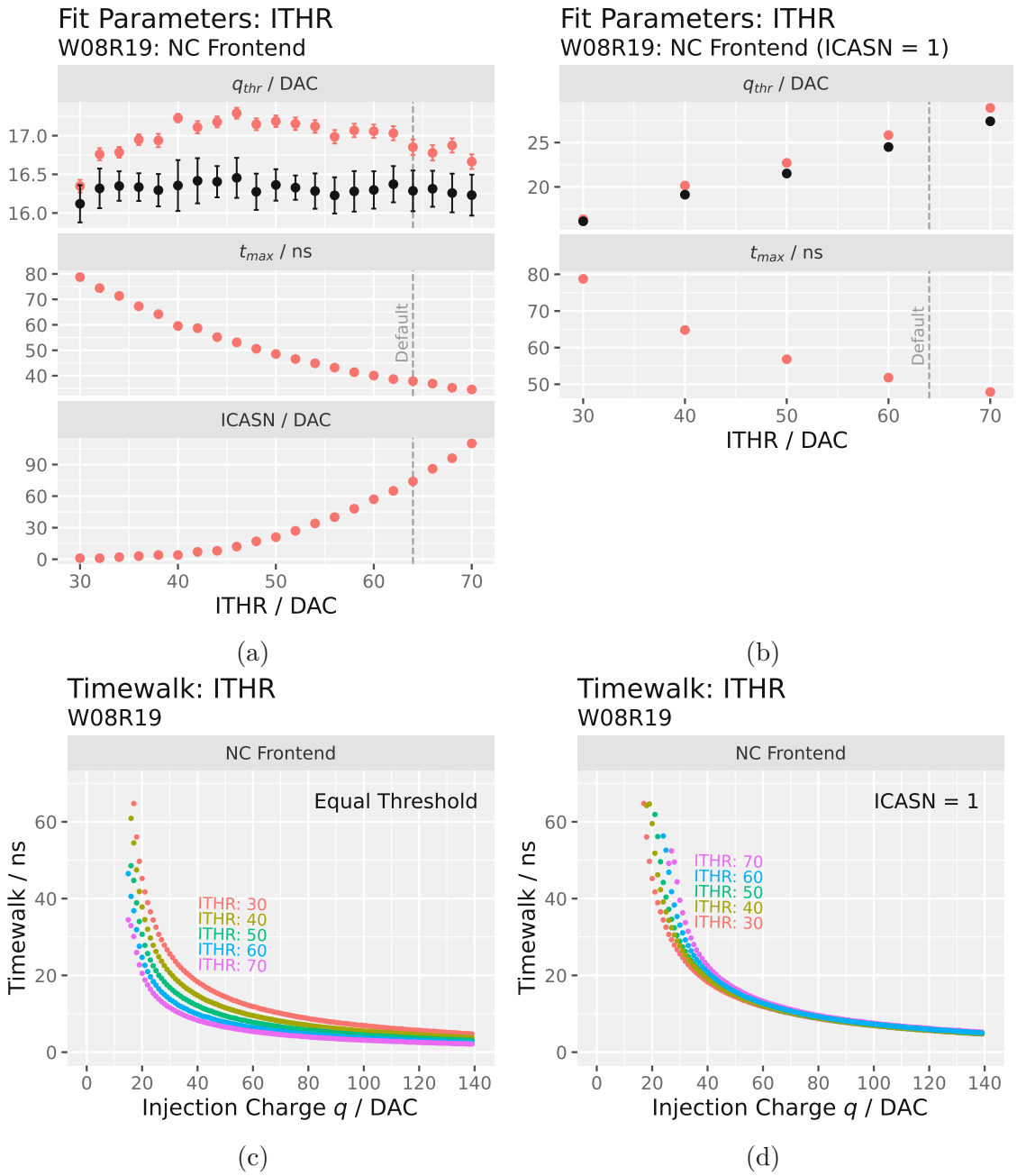


Figure 4.38: Extracted timewalk parameters for different **ITHR** values (a) and corresponding timewalk curve (c) for equal threshold and equal **ICASN** = 1 (b and d).

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

Figure 4.38a shows a similar variation, but with **ITHR**. A decreasing **ITHR** reduces the threshold which has been countered via **ICASN**. However, this results in a greatly increased t_{max}. A factor of two is visible in the shown range, also clearly visible in the time domain in figure 4.38c. Since this register is capable to reduce the threshold significantly and the threshold also affects the effective timewalk for a given charge, an additional study was performed. Here, all other registers including **ICASN** were fixed. The results in the time domain are shown in figure 4.38d. In this case, the effect of the reduced threshold overpowers the effect of the reduced bandwidth. This means, for a given charge, settings with lower **ITHR** have less timewalk, even if the magnitude is considerably larger at the corresponding thresholds.

As a result, it can still be beneficial to reduce the threshold via **ITHR**, but only if all other means are already exhausted. A reduction via **ICASN** and possibly **IDB** should have priority, but **ITHR** is a last option.

Timewalk studies on **IDB**, **VCASP** and **VCASC** have also been performed. The results show either small (**IDB**) or insignificant (**VCASP** and **VCASC**) effects on timewalk and are therefore only available in appendix B in figure B.1.

4.8 Power Consumption

For integration in the detector system, power consumption is one of the critical factors. The electrical power consumed by the chip needs to be cooled away, but increased cooling capacity inevitably increases the material budget. Due to the demands of low material budget, in particular for iVTX, the cooling needs to be precisely tuned.

To allow extrapolation and cross checks of simulations for OBELIX, the power consumption of TJ-Monopix2 was measured in various configurations. Since the bias voltage is only a few tens of volts at most and currents well below 1 mA even after irradiation, the power dissipation is dominated by the LV-supply used to operate the analog and digital circuitry of the chip.

Name	Type	Location	Typ. current
VDDA	Analog	Pixel matrix	170 mA
VDDD	Digital	Pixel matrix	7 mA
VDDA_DAC	Analog	Periphery	15 mA
VDDP	Digital	Periphery	220 mA
Total			408 mA

Table 4.11: Power domains of TJ-Monopix2 and typical currents.

TJ-Monopix2 has four power domains for the LV-supply, each with a nominal voltage of 1.8 V. They are separated for analog and digital circuitry as well as for the periphery and the pixel matrix, as shown in table 4.11 along with the supply current in typical operation. The two main contributors to the power consumption are VDDA and VDDP.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

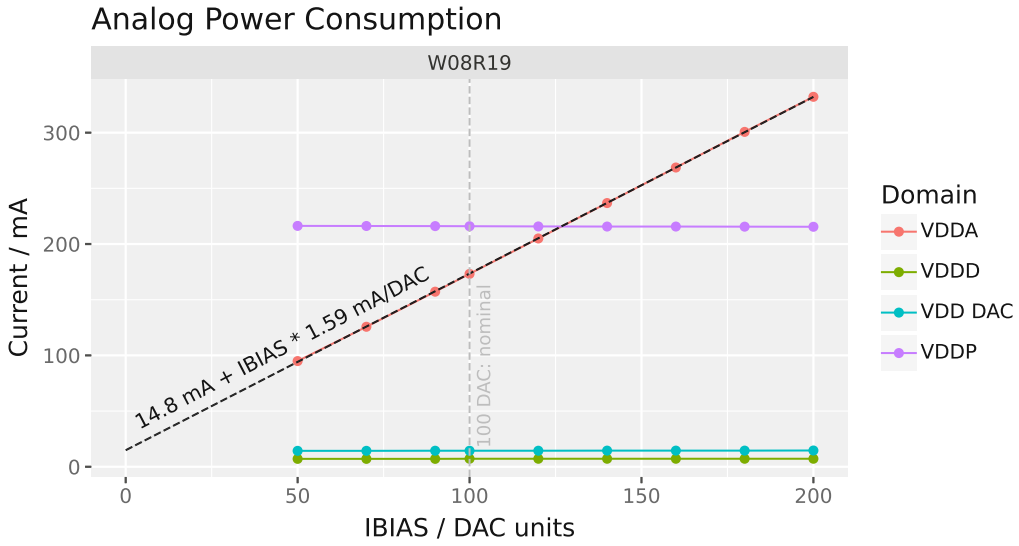


Figure 4.39: Extraction of the analog power consumption

The VDDA domain supplies all the biasing currents for the analog amplifier, however the largest one in magnitude is **IBIAS**. The current draw of VDDA was measured for different values of **IBIAS** to disentangle the contributions of **IBIAS** from the other analog bias currents as shown in figure 4.39. With the slope of the linear fit, the current contribution of all **IBIAS** transistors at nominal value of 100 DAC units can be calculated as:

$$I^{tot}_{IBIAS} = 1.59 \pm 0.01\,\text{mA/DAC} \cdot 100\,\text{DAC} = 159 \pm 1\,\text{mA}. \tag{4.3}$$

The bias current in each pixel can be calculated by dividing by the total number of pixels to:

$$I^{pixel}_{IBIAS} = \frac{I^{tot}_{IBIAS}}{512 \cdot 512} = 605 \pm 26\,\text{nA}. \tag{4.4}$$

The expected value for this current is 550 nA, which is 10% lower than what has been measured, corresponding to two times the standard error extracted from the fit. As the power consumption of the distribution network in the periphery was not included, a slightly higher than nominal value is expected. This effect increases the measured power by only about 0.4% and has not been taken into account.

The intercept of the linear fit corresponds to the power consumption of all bias currents except **IBIAS**:

$$I_{other} = 14.8 \pm 0.3\,\text{mA}. \tag{4.5}$$

The VDDP domain supplies the digital part of the periphery. The main contribution for the power consumption here is the distribution of the BCID signal across the matrix. The BCID distribution can be disabled, and this allows the separation of the BCID power from the power consumption of the rest of the digital periphery. The clock frequency of the TJ-Monopix2 chip can also be varied. This is shown in figure 4.40.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

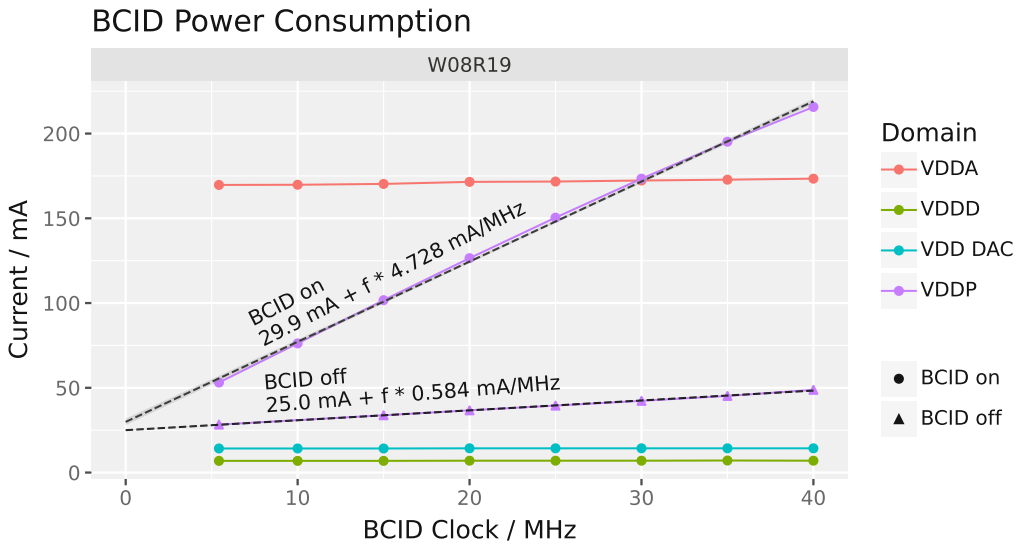


Figure 4.40: Extraction of the BCID power

Using the difference of the slopes of VDDP current vs frequency with and without BCID distribution, it is possible to obtain a value of:

$$Q_{BCID} = 4.14 \pm 0.07\,\mathrm{mA/MHz} \tag{4.6}$$

total charge stored in the BCID distribution network per BCID bit. Due to the gray encoding, only one bit changes once per clock cycle. Therefore, the calculation is independent of the number of bits in the BCID timestamp. This frequency dependent current is effectively a unit of charge in the BCID lines. Since the BCID is distributed in gray code with exactly one transition per clock cycle, for 256 DCs, the capacitance of each BCID line can directly calculated via:

$$C_{line} = \frac{C_{BCID}}{256} = \frac{Q_{BCID}}{U_{LV} \cdot 256} = 8.99 \pm 0.15\,\mathrm{pF}. \tag{4.7}$$

These measurements allow the decomposition of power consumption of the individual components of the TJ-Monopix2 chip. The results are summarized in table 4.12 and will be used in section 5.15 to estimate the power consumption of the OBELIX chip.

Name	Current	Rel.
VDDA	173 mA	41%
IBIAS	159 mA	38%
other	14 mA	3%
VDDD	7 mA	2%
VDDA_DAC	15 mA	3%
VDDP	219 mA	53%
BCID	171 mA	41%
other	48 mA	12%
Total	414 mA	100%

Table 4.12: Power consumption decomposition, measured at W08R19.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

Chapter 5

Design and Simulation of the OBELIX Chip

The OBELIX-1 chip is is the first version of a DMAPS chip specifically developed for the Belle II VTX Upgrade. A second submission, OBELIX-2 with minor feature updates and for bug-fixes is already planned, but requires comprehensive tests of the OBELIX-1 chip first. Strong emphasis is taken already for OBELIX-1 in the general layout, e.g. pad placement for electrical connections such that it can stay the same for OBELIX-2. This allows the construction of prototype modules and ladders already with OBELIX-1. In the following, the OBELIX-1 chip will be referred to as OBELIX for short.

5.1 Features and Architecture

The pixel matrix of OBELIX is inherited from the TJ-Monopix2 chip described in chapter 4, but large parts of the periphery as well as the powering scheme have been reworked. This means, the basic features of the pixel matrix are conserved, such as amplifier, readout, **HitOr**, and in-pixel threshold tuning. Also the same command interface is used to configure the chip with small extensions to adapt for the new chip, and the output data format is similar to but slightly different from TJ-Monopix2.

	TJ-Monopix2	OBELIX
Pixels	512×512	896×464
Pixel pitch	$33.04 \times 33.04\,\mu m^2$	$33.04 \times 33.04\,\mu m^2$
Active area	$16.92 \times 16.92\,mm^2$	$29.60 \times 15.33\,mm^2$
Die area	$20.20 \times 20.60\,mm^2$	$30.17 \times 19.36\,mm^2$

Table 5.1: Physical dimensions of the TJ-Monopix2 and the OBELIX chip

The size of the pixel matrix has been adjusted to use almost the maximum available reticle space and fit to the VTX geometry while optimizing the utilization of the wafer space. An overview of the physical dimensions is shown in table 5.1. The OBELIX-1 chip will include two pixel flavors of TJ-Monopix2. An equal number of columns of the the NC and HVC flavors is used and will be referred to as DC and AC in the following. At the time of this decision, it was not clear which of those two frontends would perform better. In the mean time, it has become clear that the DC frontend has superior properties especially at high temperature operation and

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

after irradiation. Since the main difference in configuration of the two frontends is `VRESET[8]`, a separate DAC is implemented for the frontend types of the chip.

In order to reduce the power consumption and allow alignment with the Belle II RF clock, the BCID frequency is reduced from 40 MHz for TJ-Monopix2 to 21.2 MHz for OBELIX. Additionally, a low-frequency mode allows the reduction by another factor of two of the BCID frequency to 10.6 MHz while keeping all other chip frequencies the same to maintain data processing throughput. This allows to save power at the expense of reduced time resolution. The name BCID (Bunch Crossing ID) was kept for the timestamp as it appears multiple times in the design, even if in Belle II it does not correspond to bunch crossings.

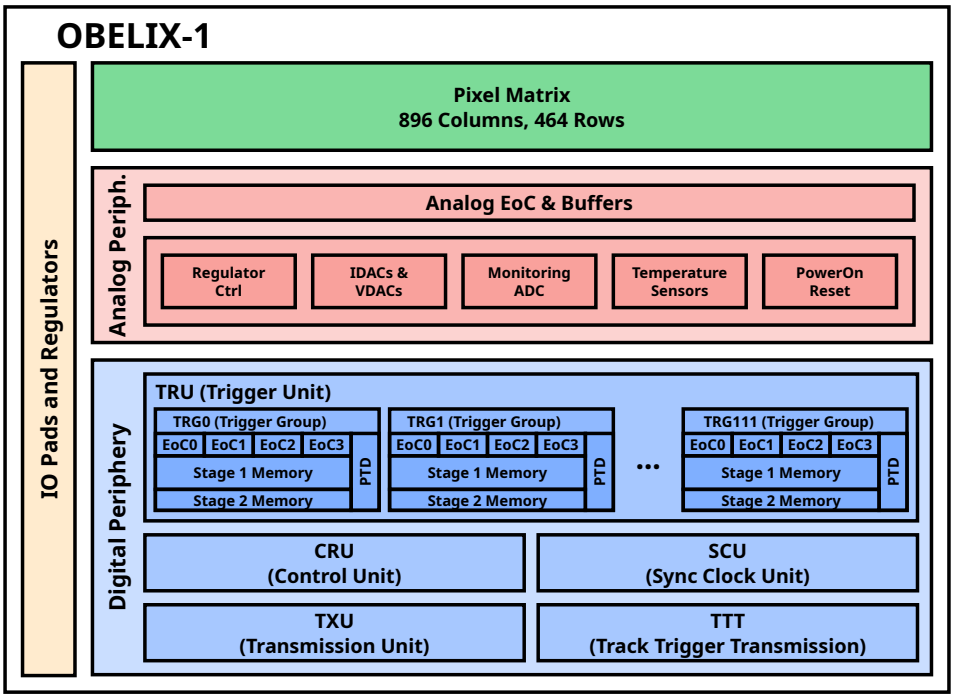


Figure 5.1: General building blocks of the OBELIX chip.

A high-level block diagram of OBELIX is shown in figure 5.1. The chip contains on chip voltage regulators to generate the supply for the analog and digital circuitry with low tolerance while allowing for an increased input voltage range.

In the analog periphery, main new features include a monitoring system with 32 temperature sensors scattered across the chip and 16 taps for supply voltage measurement. In OBELIX-2, an Analog-to-Digital Converter (ADC) will be included to directly measure these sensors along with the bias voltages and currents. While the multiplexers are already included, the ADC is not implemented and only space on the floorplan is reserved. Instead, the output of the multiplexers is available on a pad for external measurement. A power-on reset circuit is also included to automatically reset the digital parts of the chip on startup.

The digital periphery has been reworked to fulfill the VTX specifications described in section 2.4.3. This is one of the main topics of this document, as it was developed in the course of this thesis. While the Control Unit (CRU) and Sync Clock Unit (SCU) are adapted from TJ-Monopix2, the pixel readout chain from End of Column Controller (EoC) to the Transmission Unit (TXU) has been newly implemented. Notably, the Trigger Unit (TRU) was conceptualized, designed and implemented in the course of this thesis to allow for a triggered operation. The

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Track Trigger Transmission (TTT) module allows fast transmission of low granularity hit information for a VTX trigger independent from the regular readout chain. Also this module was conceptualized and implemented from scratch by the author. All individual modules are described in detail in the following sections.

As in TJ-Monopix2, the OBELIX chip configuration is stored in a set of global registers of different bit-lengths and with specific purposes. In this document the register names are stated in a highlighted font followed by the bit length in square brackets, e.g. `EXAMPLE_REGISTER[8]` for an 8 bit register. Registers can be read and written via commands, the details will be described in section 5.8. Digital signals are also highlighted with a special font like `SingleBitSignal` and if they are multi-bit also are suffixed with the bit-width in square brackets like `MultiBitSignal[5]`.

An unfortunate naming collision in digital electronics is the word register. In technical jargon, this can either mean a configuration memory like previously described or a D Flip-flop (D-FF). A D-FF is also a type of memory device and together with combinatorial logic the main building block of modern digital circuitry. D-FFs generally have a `D` input and a `Clock` input. When the clock input transits from logic 0 to 1 (rising edge), the value of the `D` input is captured and stored until the next rising edge.[1] The stored value is available as output line. Configuration registers are implemented as D-FF, but D-FFs appear also in all other blocks of the digital logic. The conceptual difference is important to understand several parts of the OBELIX design.

5.2 Voltage Regulators

An important feature of the OBELIX chip are the on chip linear voltage regulators (low dropout regulators, LDOs) with a specified input voltage $V_{in} = 2$ to 3 V. The typical circuits and logic gates of the Tower 180 nm process are specified for a supply voltage of 1.8 V$\pm$10%, the so called LV supply. In particular, the analog circuitry is influenced by smaller variations already, e.g. via the voltage DACs that are directly referenced to the supply voltage. However, in a constrained environment like a vertex detector, the cross-sections of supply cables cannot be arbitrarily increased. This means the resistance cannot be neglected and voltage drop along the cables becomes significant. While in a simple case this could be compensated, when multiple chips share a common supply rail, this is not possible any longer. The LDOs of OBELIX increase the range of acceptable input voltage variation and therefore drop to 1 V.

A simplified schematic of the OBELIX regulators is shown in figure 5.2. The higher input supply voltage passes through a pass transistor (the pass transistor M1 before distribution on the supply rails inside the chip (V_{out}). The gate of the pass transistor is controlled by an error amplifier U1 which compares the voltage on the supply rails against a reference voltage originating from a band-gap reference. This forms a control loop which creates a stable supply voltage under varying load and environmental conditions (e.g. temperature).

[1] This is the most common type of D-FF. Other common implementations may use the falling edge of the clock or can output the inverted value alone or in addition to the non-inverting output.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

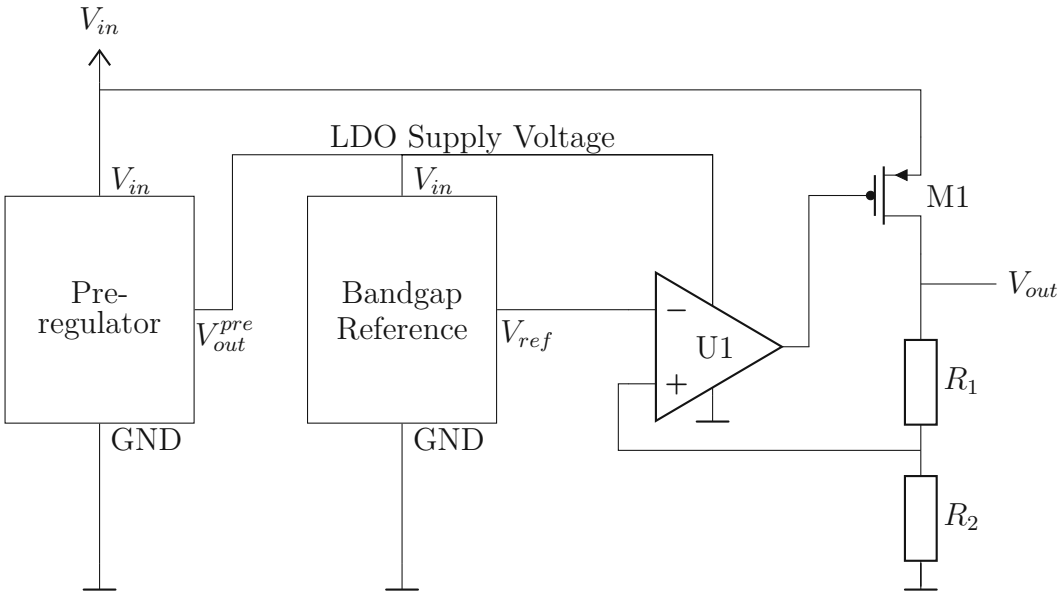


Figure 5.2: Simplified schematic for one LDO including pre-regulator.

Even if linear voltage regulators predate transistors [49], they can be fully implemented on silicon without requiring coils and have excellent noise characteristics, at the cost of low efficiency. The OBELIX LDOs require small external capacitors to assure stable operation of the control loop in all conditions. The efficiency is inherent to this type of regulator, as the pass transistor just dissipates the excess voltage at full supply current. The dissipated power in the pass transistor P_{pass} is therefore equal to the excess voltage $V_{drop} = V_{in} - V_{out}$ times the supply current I_{load}:

$$P_{pass} = (V_{in} - V_{out})\, I_{load} = (V_{in} - V_{out})\, \frac{P_{load}}{V_{out}} = \frac{V_{drop}}{V_{out}} P_{load}. \tag{5.1}$$

The second source of lost power is the power consumption of the regulator itself, i.e. of the error amplifier and voltage reference generator. This so called quiescent current is mostly independent from load conditions and can be modeled with a constant power P_q. The total efficiency η can therefore be modeled as:

$$\eta = \frac{P_{load}}{P_{tot}} = \frac{P_{load}}{P_{load} + P_{pass} + P_q} = \frac{P_{load}}{(1 + V_{drop}/V_{out})\, P_{load} + P_q}. \tag{5.2}$$

Since the quiescent power P_q is generally small compared to typical load currents, this can be simplified to:

$$\eta \approx \frac{P_{load}}{(1 + V_{drop}/V_{out})\, P_{load}} = \frac{V_{out}}{V_{drop} + V_{out}}, \tag{5.3}$$

which only depends on the output voltage and the drop voltage. The efficiency of the OBELIX LDOs is plotted in figure 5.3 against the specified input voltage range. Additionally, the factor of the load power to obtain the total power is shown. This factor is important to calculate the total power dissipation of the OBELIX chip, since the simulation tools always assume external 1.8 V supply. These numbers then need to be multiplied by this factor to obtain the real power dissipation.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

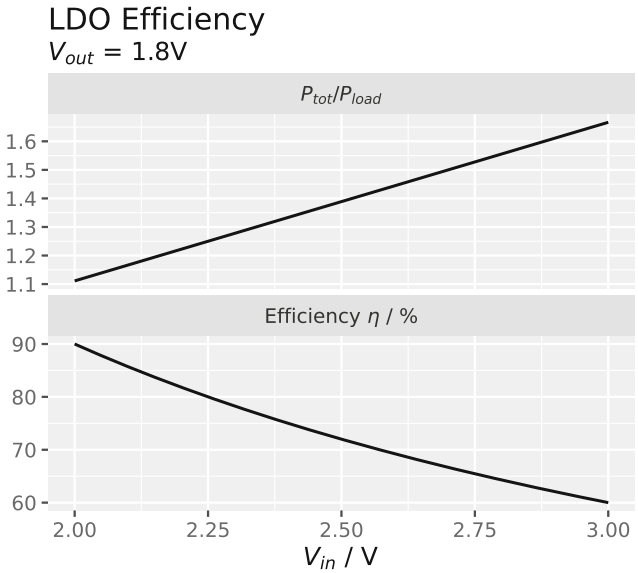


Figure 5.3: Efficiency and factor for load power for the OBELIX LDOs.

The OBELIX chip contains a small pre-regulator that creates the supply voltage for the four main LDOs. The main LDOs are supplying the analog part of the chip from the left and right side respectively (VDDAL and VDDAR) and the digital circuits from the bottom (VDDD). An additional LDO creates the pre-charge voltage for the bit-lines of the matrix readout (VPC). While the first three LDOs have a nominal output voltage of 1.8 V, the VPC output is 1.6 V nominal.

It is possible to digitally tune the output voltage of each LDO from -20% to +10%, the registers for this tuning are shown in table 5.2. The tuning range is linearly spaced with 8 bit width. The nominal 0% tuning is therefore at $255 \cdot 2/3 = 170$ while -20% and +10% correspond to values 0 and 255 respectively. While the whole circuit of the LDOs is implemented in silicon, external bypass capacitors are required in close proximity to the chip to maintain a stable control loop. The capacitors have to be placed on the RDL for iVTX and on the flex PCB for oVTX.

Register	Default	Description
`LDO_VDDAL[8]`	170	Voltage tuning for left analog LDO
`LDO_VDDAR[8]`	170	Voltage tuning for right analog LDO
`LDO_VDDD[8]`	170	Voltage tuning for digital LDO
`LDO_VPC[8]`	170	Voltage tuning for bitline pre-charge voltage

Table 5.2: Registers for tuning LDO output voltages.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

5.3 Floorplan

The floorplan of the OBELIX chip is shown in figure 5.4 in nominal orientation and is similar to the TJ-Monopix2. The power distribution of the pixel matrix is horizontal (i.e. along each row for all columns), while the data and analog bias lines are vertically routed (i.e. along each column). At the bottom, the chip periphery is located. It contains analog parts like bias voltage and current generation and distribution circuits as well as LDO control circuitry. The digital periphery contains all circuits to process the data from the pixels and control functions of the chip. The pass transistors for the analog supply of the matrix are located in two stripes on both sides of the pixel matrix, and designed as narrow as possible to reduce inactive area between neighboring chips.

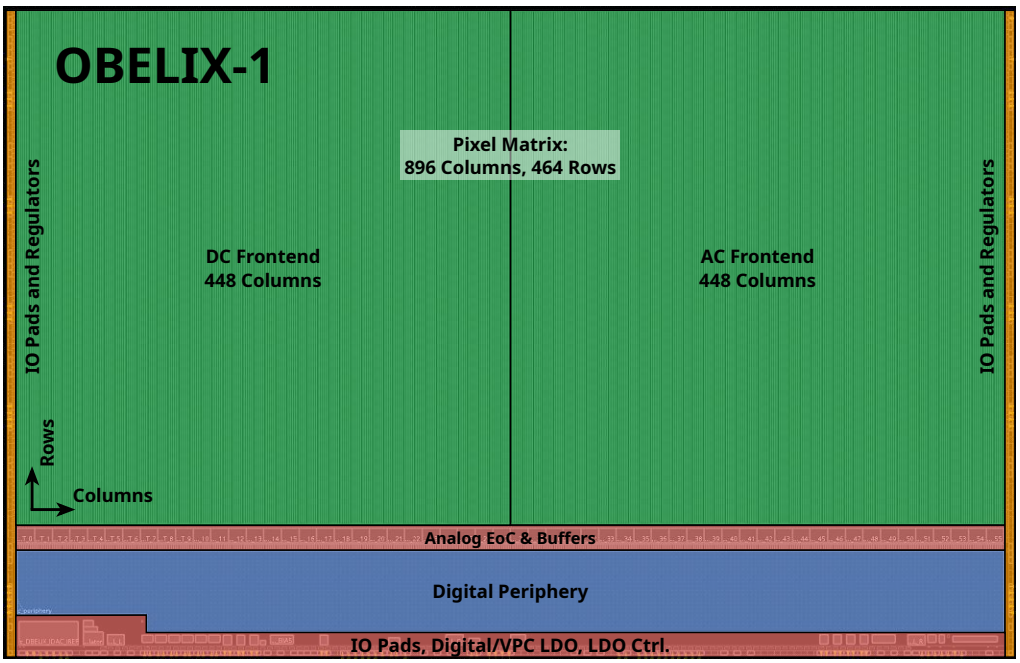


Figure 5.4: Floorplan of the OBELIX chip, top view in nominal orientation.

5.4 Monitoring and ADC

The OBELIX chip introduces a monitoring system including an ADC to measure temperature, supply voltages and currents, as well as DAC outputs for frontend bias voltages and currents. The selection of the measured signal is done with two stages of multiplexers controlled via registers shown in table 5.3. While the OBELIX-1 chip will not include the ADC, the multiplexer outputs are available for external measurement and OBELIX-2 will implement the ADC.

The `ADC_MUX[7]` register controls which type of signal is measured by the ADC. A summary of the possible values is shown in table 5.4. The selection of this register defines, which of the secondary multiplexers is active, e.g. when the current monitoring is activated `MON_I_MUX[4]`. This register then defines which of the possible currents is actually measured. For the DACs, the registers `OVR_IDAC_MUX[3]` and `OVR_VDAC_MUX[3]` allow the external overdrive of one DAC via one pin on the chip respectively. The corresponding tables for the current and voltage multiplexer and overdrive registers can be found in appendix C. The dynamic injection signal can be monitored and overdriven externally via `MON_VPULSE[1]` and `OVR_VPULSE[1]`.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Register	Default	Description
MON_I_MUX[4]	0	Monitoring selector for IDACs and LDOs
OVR_IDAC_MUX[4]	0	Channel for external overdrive one IDAC
MON_VDAC_MUX[3]	0	Monitoring selector for VDACs
OVR_VDAC_MUX[3]	0	Channel for external overdrive one VDAC
MON_VPULSE[1]	0	Monitoring of dynamic pulse voltage
OVR_VPULSE[1]	0	Overdrive of dynamic pulse voltage
MON_TEMP_MUX[6]	0	Select temperature sensor channel
MON_VDD_MUX[7]	0	Select voltage monitoring channel
ADC_MUX[7]	0	Main ADC multiplexer

Table 5.3: Registers for monitoring multiplexer configuration.

ADC_MUX[7]		
Value	Mode	Secondary MUX Register
0	Disabled	
1	Voltage measurement (DAC outputs)	MON_VDAC_MUX[3]
2	Current measurement (DAC outputs)	MON_I_MUX[4]
3	Temperature measurement non-inverting	MON_TEMP_MUX[6]
4	Temperature measurement inverting	MON_TEMP_MUX[6]
5	Supply Voltage measurement	MON_VDD_MUX[7]
≥ 6	Reserved	

Table 5.4: Possible selections of the ADC_MUX[7] register.

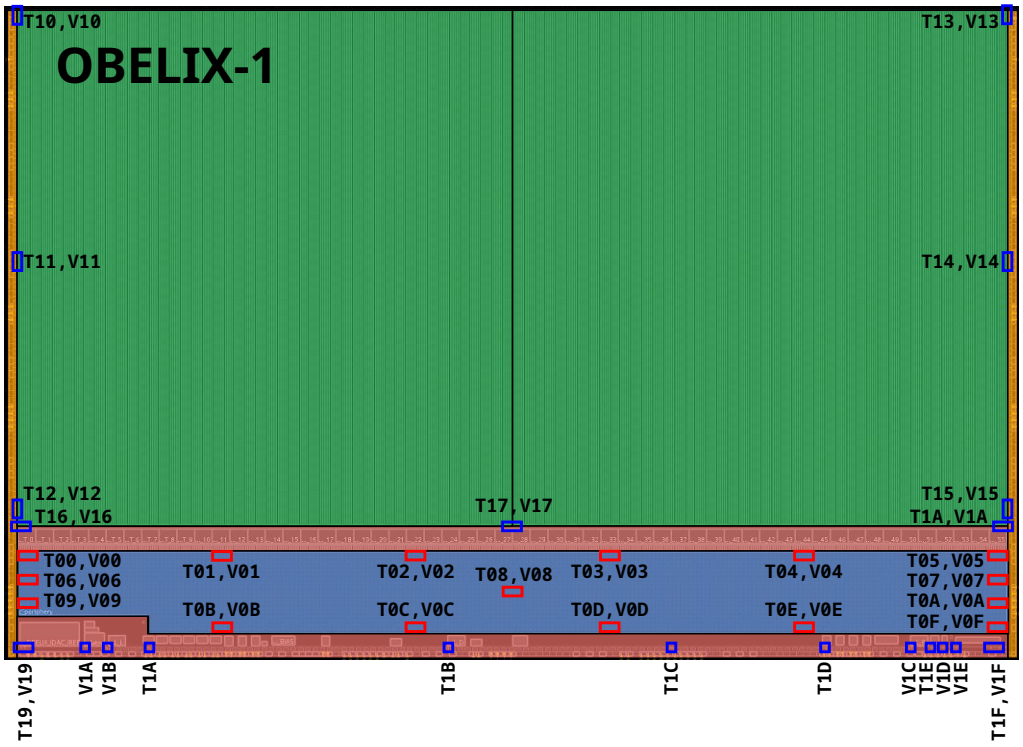


Figure 5.5: Monitoring cell locations and indices (hexadecimal) for temperature (Txx) and supply voltage (Vxx) of the OBELIX chip.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

The OBELIX chip contains 32 temperature sensors and 32 taps to measure the supply voltage. Figure 5.5 shows the locations of the sensors on the floorplan along with the number in hexadecimal notation. The numbers shown for the temperature monitoring directly correspond to the required setting of the `MON_TEMP_MUX[6]` register. Each temperature sensor has an inverting and a non-inverting output. Those need to be measured both and the temperature can be calculated from the difference of the two.

The lower 6 bits of `MON_VDD_MUX[7]` also correspond to the sensor number and the Most significant bit (MSB) of the register selects measurement of either the ground (0) or VDD (1) potential. The location of the voltage sensing also defines the voltage domain, i.e. sensors 0x00 to 0x0F measure VDDD whereas 0x10 to 0x1F measure VDDA.

5.5 BCID Crosstalk Mitigation

The BCID crosstalk in TJ-Monopix2 presented in section 4.6 is not preventing low threshold operation, but is on the borderline of causing issues. Since it is most probably caused by current spikes in the drivers, the amplitude can be expected to scale with the number of columns and therefore to be approximately twice in OBELIX without any mitigation measures. This would certainly be unacceptable and a mitigation has been developed.

On TJ-Monopix2, CMOS drivers with full voltage swing (1.8 V) are used to drive the BCID lines. Since the BCID distribution is differential, two drivers per bit drive the inverting and non-inverting line respectively.

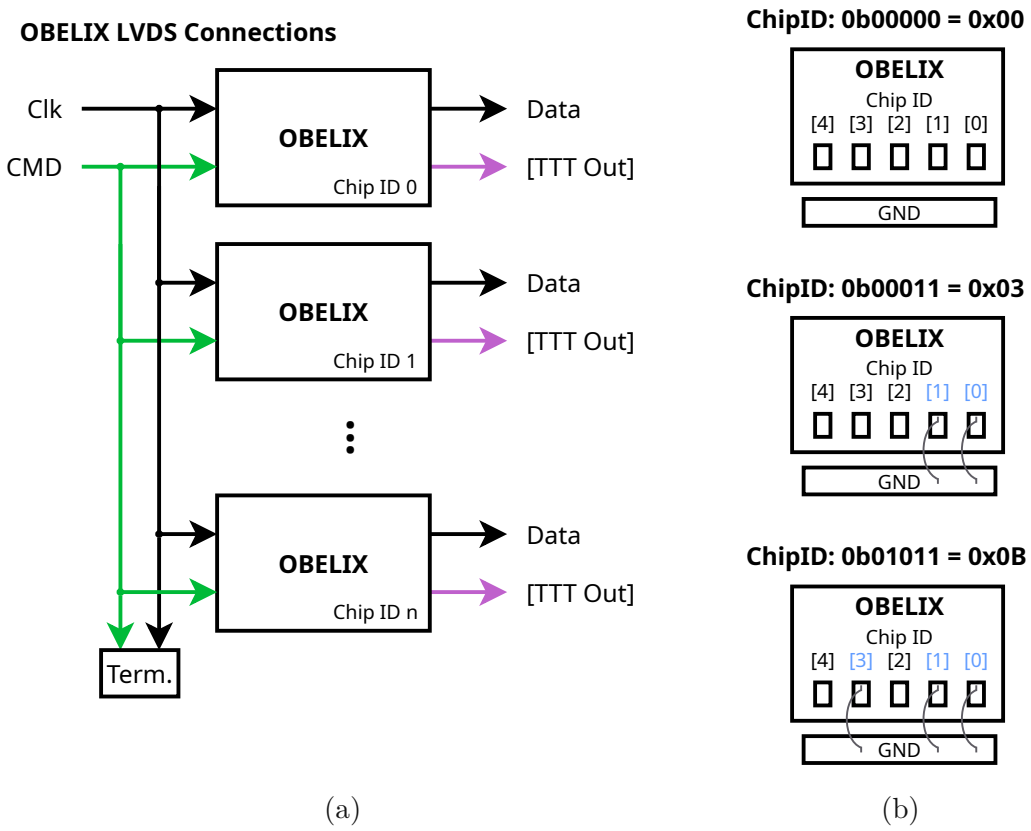
The BCID drivers on OBELIX use a complementary source follower to reduce the voltage swing to 740 mV. One one hand, this greatly reduces the current spike at transitions which is the main purpose. Additionally, the power consumption of the BCID driver circuit is reduced by about 60% in comparison to the original driver. Since the pixel circuit is comparing the inverting and non-inverting lines against each other, no modifications are needed to the latching circuit inside each pixel. However, the propagation delay is slightly higher and has a different evolution along the column, requiring an adjustment of the BCID delay compensation circuit to fit to the new type of driver.

5.6 Services

An important aspect of pixel sensors for integration into a whole detector are the connections to the outside world and their requirements for sensor operation, commonly called the services.

One of the reasons for the LDOs is that power rails of the chip can be shared across multiple chips. This requires an increased input voltage range since voltage drop will occur along the rails, but simplifies the layout of the flex PCB and cabling. In addition, the bias voltage of either PSUB and PWELL or HV needs to be supplied, but also can be shared across multiple chips. Voltage drop on the bias voltage lines is of less concern since the leakage current is small.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN



(a) (b)

Figure 5.6: LVDS services for the OBELIX chip (a) and example ChipID configurations with wire bonds (b).

The digital IO lines of the OBELIX chip are reduced to four LVDS links as shown in figure 5.6a, two inputs and two outputs. The two inputs of the chip can be shared across multiple chips, with 100 Ω termination at the end of each line. These are the clock input and the command line `CMD` that provide fast and slow commands as well as the trigger signals. Individual chips can be addressed with a 5 bit `Chip ID[5]` that can be set with wire-bonds to ground. A placed wire bond at the corresponding bit sets a logic 1 and a connection left open to 0 as shown in figure 5.6b. The address 0x1F is reserved as broadcast address, where all chips will act regardless of the configured `Chip ID[5]`. This means, up to 31 chips can share one command bus, given that the signal integrity can be maintained. In VTX, the current plan foresees to maximally connect 12 chips in such an arrangement to constitute a module.

The first output is the main `Data` output and is mandatory for each chip. It is used for the hit-data, as well as the responses to register reads and works in a packaged 8b/10b encoded format. The bit-rate is double the frequency of the input clock by utilization of both clock edges (Double Data Rate, DDR) and nominally 339.4 Mbit/s. The protocol and data format will be described in detail in section 5.10 while the available chip commands will be presented in section 5.8.

The second output of OBELIX is the `TTT Out` line, reserved for the L1 trigger contribution. It provides low-latency, low-granularity hit information directly from the matrix to the backend of the VTX. This feature is independent from normal operation, and might not be enabled in iVTX due to power constraints and high occupancy. In this case, this line does not need to be connected. Furthermore, up to four chips can share one `TTT Out` line in a star configuration at further reduced

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

spatial granularity to reduce the number of transmission lines required and allow for simpler integration. The TTT Out uses the same 8b/10b encoded DDR transmitted words as the main data output, but in a frameless, non-sparsified protocol. The details of the whole track trigger feature including the exact data format will be described in section 5.11.

5.7 Sync Clock Unit and Clock domains

The Sync Clock Unit (SCU) provides all clock signals required for the chip operation. All clocks of the OBELIX chip originate from a single clock input of nominal frequency of 169.7 MHz, which can be derived synchronously from the SuperKEKB RF clock by division of 3 in the backend electronics. Different parts of the chip require different frequencies, which are all generated in the SCU with a clock divider. Table 5.5 summarizes the clock domains of OBELIX with the rounded and exact values. The clock domains are typically referred to with frequency in the rounded form, even if the real operation frequency will be about 6% higher. These rounded numbers can therefore be seen as just labels corresponding approximately to the real frequency. The chip is designed to operate at least up to a fast clock frequency of 174 MHz i.e. a 2.5% safety margin from the frequency used in VTX is included in the design constraints.

Clock signal	Rounded	Real value	Period
Input clock	160 MHz	169.7 MHz	5.89 ns
Periphery clock	20 MHz	21.2 MHz	47.1 ns
TXU clock	32 MHz	33.9 MHz	29.5 ns
Asynchronous	—	—	—

Table 5.5: Overview of the clock domains in the OBELIX chip.

Apart from clock generation, the serial commands from the CMD line are deserialized to 16 bit words in the SCU. The command data stream contains a synchronization pattern at idle, which lets the SCU automatically synchronize to the correct start bit to correctly lock onto the received words.

In a multi-chip environment, a startup of all chips at the same time cannot be guaranteed. Since the 20 MHz `clock` is divided from the input clock, the 20 MHz `clocks` of the individual chips are shifted in random integer periods of the 160 MHz `clock` initially, i.e. with a random phase. However, an equal and defined phase is required for all chips, as the BCID is generated from this clock. Since the BCID defines the integration time, a random phase would shift start and end of the integration time for each chip differently by up to 47.1 ns, greatly reducing the time resolution. The command data stream provides the necessary means of synchronization to equalize the phase for the whole detector. When the deserializer locks onto the command stream, it also triggers a phase adjustment of the clock divider. In this way, the 20 MHz `clocks` is synchronous in the whole detector, even if each chip generates it locally. Since glitches may occur at the time of phase adjustment which generally can cause undefined behavior in digital circuits, the clock is temporarily switched off during the phase adjustment.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

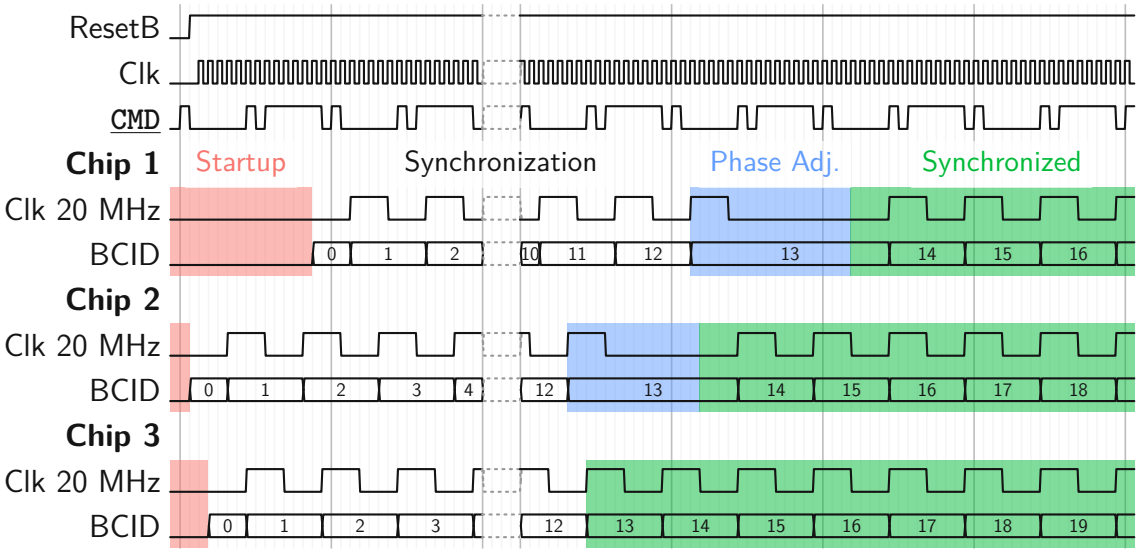


Figure 5.7: Synchronization of three OBELIX chips with different initial phases for the clock divider.

The timing diagram 5.7 shows this in detail for three chips as an example. The initial reset period shows the startup of the chips, slightly exaggerated for illustration. After the startup of the chip, the SCU will start the synchronization onto the idle pattern on the CMD line. The start of each 16 bit idle pattern is marked with the gray vertical lines. Initially, the SCU cannot detect the start of each 16 bit word but picks a random value. After 6 idle sequences are successfully received, the deserializer is adjusted to correctly align to the 16 bit words and the clock phase adjustment is initiated if necessary. For chip 3, the clock phase was correct by chance and therefore the phase adjustment was omitted. After the adjustment, the clock phase is fixed for all chips, independent of the initial starting point.

The need for this procedure becomes clear when looking at the time of BCID transitions. During the time of synchronization, each chip increments the BCID at a slightly different time, whereas in the synchronized state, the BCIDs change at the same time for all chips. Generally, the synchronization does not guarantee an equal BCID count, however, the transitions appear simultaneous for all chips after the phase adjustment. A following BCID reset can then equalize also the BCID values for all chips.

This module was adapted from TJ-Monopix2 for the lower operating frequency. Especially the clock synchronization was reworked, as the old design did not assure a glitch free phase adjustment. Digital designs are generally not capable of handling glitches in clock domains. These can create undefined states that trigger undesired behavior. Even worse, an effect called metastability [50] can be triggered - transient states of D-FF in the middle between logic 0 and 1. These metastable states are in the forbidden region of the logic levels and can cause behavior that is logically impossible in the high level description. This means occasionally, the chip could transit into an unpredictable state just after a reset.

It has been observed with TJ-Monopix2 that in about 5% of the cases, the chip stopped accepting commands until another reset is performed just after the startup. Further investigations were not made, but this is consistent with the clock glitch issue discovered in the code. For a prototype, this is not much of a problem apart from a mild annoyance to the operator, but when operating hundreds of chips together like in VTX, such issues are a major concern that could potentially prevent

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

operation as a whole. For this reason, the OBELIX chip includes an improved clock generator that prevents the creation of such glitches. [2]

For the contribution to the L1 trigger, the 32 MHz `clock` domain is also required to have the same phase for all chips in VTX. However, it is not possible to use the same technique for synchronization as in the 20 MHz `clock` domain. For the synchronization algorithm of latter, it is important that the data frames are 16 bits, providing a 10 MHz time reference. Only an integer multiple to reach 20 MHz is required. This means, individual chips can do the synchronization at different times, but the result is still an equal phase across all chips. For the 32 MHz domain, there is no integer multiple to reach the desired frequency from 10 MHz, i.e. the 32 MHz `clock` phase is not the same at the beginning of each 16 bit word which therefore cannot be used as means of synchronization. A manual synchronization via a special command before the use of the TTT module is therefore required, which will be described in section 5.8.2.

The SCU can also detect if the synchronization is lost during operation, e.g. due to bit errors in the command stream or single event upsets. The details about this will be described in section 5.8.1 after the exact structure of the command stream is introduced.

5.8 Control Unit

The Control Unit (CRU) parses the commands deserialized by the SCU and hosts the global register table. A subset of the protocol developed in the RD53 collaboration [51] is used, based on the TJ-Monopix2 control unit.

Global Register Table

The global register table is organized in physical registers on a hardware level. Each physical register is 16 bits wide and identified by a 9 bit address. The DAQ system contains a table to map the physical registers to logical addresses of arbitrary size. Each logical register is responsible for one particular function of the chip, the register names stated in previous and following sections always refer to logical registers. Typically, logical registers have fewer bits than the physical register and multiple logical registers can share one physical register. In contrast, a few logical registers for column- or row-wise control span over multiple physical registers.

While the logical registers are often desired to be manipulated as a user, only full physical addresses can be read or written as a whole. For this reason, a copy of the register table needs to be maintained inside the DAQ system, to allow specific overwriting of logical addresses while preserving the value of other logical addresses that share the same physical register. A few selected registers are shown in table 5.6 to illustrate the mapping between physical and logical registers. Each row in the table corresponds to a physical address, and the names inside the row are the locations of the individual bits of logical addresses. Reserved bits may not exist physically and therefore may always read as zero. A full table of all registers is available in appendix C.0.2.

[2] The other potential explanation for this anti-feature is an issue with the firmware of the DAQ system. It has also been observed that extreme data rates, such as when the chip is driven into the noise floor, may cause the chip to stop working. This might be either an independent problem or the root cause of the occasional problems at startup because the chip creates hit-bursts at time of configuration. In either case, the necessary counter measures are implemented in OBELIX-1.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Adr.	Register Bit Mapping															
	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
0x00	IBIAS[8]								ITHR[8]							
0x01	ICASN[8]								IDB[8]							
0x02	ITUNE[8]								ICOMP[8]							
0x2C	PIXEL_CONF_ROW_SELECT[9]										P.C.WR._SPEED[4]			reserved		
0x2D	PIXEL_CONF_COL_GROUP_SELECT[8]								reserved							
0x2E	PIXEL_CONF_PORTAL_DATA[12]													reserved		

Table 5.6: Mapping of physical to logical registers for a few example physical registers.

Some physical registers can be read only. Any write to such a register will not have an effect, but reading provides access to internal signals e.g. error counters and status flags. When registers are shown in tables inside this document, a default value of RO indicates that this is a read only register.

To prevent faults from single event upsets (SEUs), some physical registers are protected with a dual 8 to 13 bit hamming code checksum, re-using the implementation of the MIMOSIS-2 chip [52, 53]. This means, a single bit-flip can be tolerated per 8 bits of register. The register output is automatically corrected, and a write to an arbitrary register will refresh the contents of all hamming code protected registers, correcting potential bit-flips in the memory cells itself. All physical registers corresponding to logical registers with 16 bits or less are protected. There are also registers to control various features on a per-column or per-row basis. These registers are not protected, as an SEU would only affect the particular row or column. However, these registers account for 87% of the total space in the global register table and not protecting them saved significant amount of resources. Furthermore, it is possible to overwrite these registers anytime without adverse effects in a continuous reconfiguration procedure. This will be further explained in section 5.12.

CMD interface

The protocol for the command interface uses a continuous stream of 16 bit words, each subdivided into pairs of 8 bit half-words. All commands are aligned to these 16 bit words, which enable the synchronization of individual chips via this alignment. There are fast- and slow-commands. Fast-commands are only one word long and slow-commands span over multiple words. Any slow-command can be interrupted by any number of fast-commands without breaking the slow command interpretation. Therefore, fast-commands are used for timing critical, high priority purposes whereas slow-commands may shift in time if a fast command interrupts its transmission.

Multiple chips can share one `CMD` line. To identify individual chips, a 5 bit `ChipID[5]` is configurable via input bonding pads as described in section 5.6. Only 31 chips are allowed per line because `ChipID[5]` 0x1F corresponds to a broadcast address, which lets all chips respond, regardless of the configured `ChipID[5]`. This allows all chips to be configured at once if the payload is the same for all chips.

In the following, all relevant commands are discussed in detail. When 8 bits are stated, it corresponds to the actual bytes transmitted (little endian). For including binary data, a type of 5b/8b encoding is used and table C.1 in appendix C lists all symbols. This assures that the `CMD` line is DC balanced and each word can be correctly parsed. Data encoded this way have half-word descriptions in pointed brackets like <Data>. The commands have also a special font associated, like `Example Command`.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

5.8.1 The Sync command

The Sync fast-command shown in table 5.7 is used as idle pattern for the CMD stream, i.e. most ot the time in operation, the majority of commands are sync patterns with only occasional trigger commands. For a correct chip startup, more than 6 repetitions of Sync need to be received correctly.

Sync: 0b10000001	$\overline{\text{Sync}}$: 0b01111110

Table 5.7: The Sync fast-command.

During chip operation, it is also mandatory to transmit a Sync command at least every few words to ensure that all chips stay synchronized. The maximum number of non-sync commands in a row is two plus CMD_CHSYNC_THR_LOW_CONF[5] (2+31 = 33 per default). If this is exceeded, the chip transits into the desynchronized state, where commands are not executed any longer. This state is kept until two plus CMD_CHSYNC_THR_HIGH_CONF[5] Sync commands (2+4 = 6 per default) are received to resynchronize the chip. Since bit-flips may occur occasionally, it is recommended to send a Sync command at least every 16 words, which allows for one missed Sync command in the worst case before desynchronization occurs.

If desynchronization is only caused by bit-flips on the CMD line or missed Sync commands, the chip will automatically resynchronize as soon as a sufficient number of sync commands are received. While the chip may loose triggers or other commands during the time in the desynchronized state, since the 20 MHz clock keeps the correct phase the BCID code remains correct. However, if a resynchronization is required because of an error in the clock signal, the resynchronization comes along with a 20 MHz phase adjustment which will alter the BCID code in relation to the global standard. The backend system can detect and discriminate the two types by comparing BCID and trigger number which are both transmitted for each hit. In case of a wrong BCID, a full resynchronization procedure needs to be executed, which will be described along with details on the possible detection algorithms in section 5.12.

If a resynchronization happens, the CMD_SYNCLOSS_CNT[16] counter is incremented. The flag CMD_SYNCLOSS_FLAG[1] indicates a nonzero counter and a reset of the counter can be performed by writing a logic 1 to CMD_SYNCLOSS_CNT_RESET[1].

5.8.2 The Clear command

The Clear fast-command shown in table 5.8 provides a means of signaling parts of the chip in a synchronous and timed manner. The particular function of this command needs to be configured prior transmission via the CLEAR_EN_*[1] registers shown in table 5.9. All those registers are one bit, setting them to logic 1 switches on this feature. It is also possible to set multiple bits to perform multiple actions at once. The capabilities of the Clear command include the reset of the BCID to an arbitrary initial value, synchronize the TTT module across chips and reset the pixel matrix. The clear command can be targeted to a specific ChipID, but most commonly it will make use of the broadcast address 0x1F. Even more, at startup of VTX, probably a Clear command is sent to every chip in the detector simultaneously to synchronize the TTT, set the BCID and reset the matrix. If this creates an unacceptable power surge, it can also be performed in a staggered way. For the TTT sync, the clear commands need to be aligned to a period of 5 words to

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

ensure synchronicity with the odd division factor of 5 for the 32 MHz `clock` clock divider. This corresponds to 40 clock cycles in the 160 MHz domain which is the least common divider for 5 (32 MHz `clock` domain) and 8 (20 MHz `clock` domain and basis for the `Clear` command).

`Clear`: 0b01011010	<`ChipID[4:0]`>

Table 5.8: The `Clear` fast-command.

Register	Action
`CLEAR_EN_SET_BCID[1]`	Set the BCID counter to `BCID_INITIAL_VALUE[9]`
`CLEAR_EN_SYNC_TTT[1]`	Reset the clock divider for the 32 MHz `clock`
`CLEAR_EN_RESET_MATRIX[1]`	Clear data in the pixel-matrix
`CLEAR_EN_RESET_TRIG_ID[1]`	Reset the `Trigger ID`

Table 5.9: Different registers to control actions of the `Clear` command

5.8.3 The `TriggerX` commands

The chip is notified about triggers via fast-commands, each BCID code can be individually triggered exactly after the configured latency. Since a fast-command takes 16 cycles of the 160 MHz `clock` to transmit ($\sim$ 100 ns), two BCID codes ($\sim$ 50 ns each) correspond to one fast-command. For this reason, there exist three distinct trigger commands, as shown in table 5.10: The first `Trigger1` command corresponds to the first BCID, the second `Trigger2` to the second BCID code and the third `Trigger3` to both BCIDs. To illustrate this, a sketch is shown in figure 5.8. In principle, each `TriggerX` command could be targeted to a specific chip via the `ChipID[5]`. In practice, the `TriggerX` commands will always use the broadcast address 0x1F. Each triggered BCID code increments the internal `Trigger ID[5]` counter. This value is also transmitted by the TXU and can be used to check that all previous `TriggerX` commands have been successfully received and processed.

`Trigger1`: 0b00101011	<`ChipID[4:0]`>
`Trigger2`: 0b00101101	<`ChipID[4:0]`>
`Trigger3` 0b00101110	<`ChipID[4:0]`>

Table 5.10: The `TriggerX` fast-command family.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

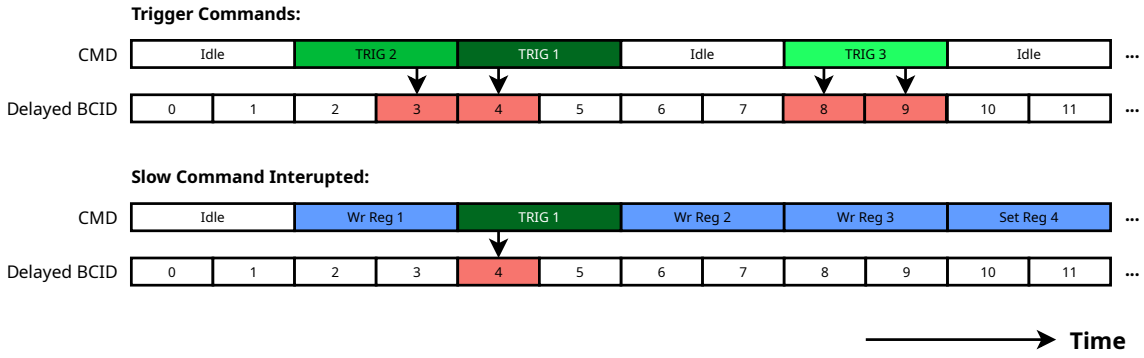


Figure 5.8: Illustration of command stream with focus on trigger commands and corresponding delayed BCID code, i.e. the BCID where a trigger decision is due.

5.8.4 The `Write Register` command

The `Write Register` slow-command shown in table 5.11 writes to the global register table. An address of 9 bits and 16 bits of data is required to write a physical register. It is not possible to write a particular logical register. Write commands to read-only registers will be ignored.

`Write Register`: 0b01100110	<**`ChipID[4:0]`**>
<0b0, Address[8:5]>	<Address[4:0]>
<Data[15:11]>	<Data[10:6]>
<Data[5:1]>	<Data[0], 0b0000>

Table 5.11: The `Write Register` slow-command, each row corresponds to 2 times 8 bits.

5.8.5 The `Read Register` command

The `Read Register` slow-command shown in table 5.12 reads from the global register table. An address of 9 bits is required to read a physical register. It is not possible to read a particular logical register. The response is a special frame on the **`Data`** output containing the address and value of the register requested to read. Details on the response format will be described in section 5.10.

`Read Register`: 0b01100101	<**`ChipID[4:0]`**>
<0b0, Address[8:5]>	<Address[4:0]>

Table 5.12: The `Read Register` slow-command.

5.8.6 The `Calibration` command

The `Calibration` slow-command shown in table 5.13 allows charge injection into the pixels. Prior to the injection, the targeted pixels need to be enabled for injection via the **`EN_INJ_COL[896]`** and **`EN_INJ_ROW[464]`** registers. To enable injection in a pixel, both, the row and column of the pixel need to be enabled for injection. The injection command has two parameters, **`PulseStartConfig[6]`** and **`PulseStopConfig[14]`**, these define the injection pulse sent to the pixel. The **`PulseStartConfig[6]`** selects

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

the timing of the rising edge of the **Pulse** signal, as shown in timing diagram 5.9. The **Pulse** signal is sent to each pixel, performing the injection via a voltage step through an injection capacitor. The unit for **PulseStartConfig[6]** is half clock cycles in the 160 MHz **clock**, i.e. the injection pulse is DDR. This doubles the time resolution of the injection pulse onset.

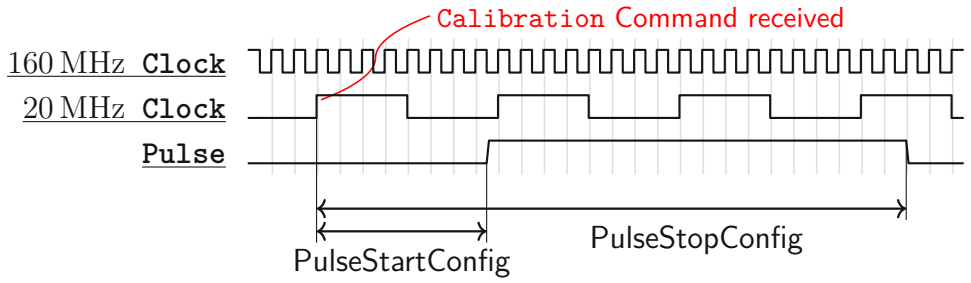


Figure 5.9: Injection pulse from `Calibration` command.

The **PulseStopConfig[14]** value defines the location of the falling edge in relation to the reception of the command. This is also in units of half cycles of the 160 MHz **clock**, i.e. allowing for a maximum pulse length of about 50 µs. Theoretically the falling edge of the injection pulse does not matter for the injected charge. However, it may disturb the amplifier if the injection pulse is shorter than the ToT, as the falling edge changes the potential on the input of the amplifier. On the other hand, setting the pulse too long slows down the injection as the DAQ system needs to wait for the command to complete. A value of about 3.2 µs has been used for TJ-Monopix2, providing a sufficient safety margin for the ToT while having low impact on injection speed.

`Calibration`: 0b01100011	<**ChipID[4:0]**>
<**PulseStartConfig[5:1]**>	<**PulseStartConfig[0]**, **PulseStopConfig[13:10]**>
<**PulseStopConfig[9:5]**>	<**PulseStopConfig[4:0]**>

Table 5.13: The `Calibration` slow-command

Configuration of the CRU

Several error counters exist in the control unit. Each counter comes with a flag indicating a nonzero counter and a reset bit to clear the counter. The reset register does not change itself back automatically, i.e. it needs to be set to 1 followed by a set back to 0. The **CMD_BITFLIP_WARNING_CNT[16]** indicates that a bit-flip was detected, but the command was still recovered and executed. For example, a `Sync` may be still understood in case of a bitflip under certain circumstances. The **CMD_BITFLIP_ERROR_CNT[16]** register indicates a bitflip where ambiguities prevented a recovery of the command and therefore the command was dropped. In case the command structure as a whole is violated, e.g. the parameters do not fit to the command, the **CMD_ERROR_CNT[16]** is incremented. It is also possible to read the last values of **PulseStartConfig[5]** and **PulseStopConfig[14]** for debugging purposes. All registers related to the CRU are shown in table 5.14.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Register	Default	Description
CMD_SYNCLOSS_CNT[16]	RO	Counter for resynchronization events
CMD_SYNCLOSS_FLAG[1]	RO	Nonzero CMD_SYNCLOSS_CNT[16]
CMD_SYNCLOSS_CNT_RESET[1]	0	Reset CMD_SYNCLOSS_CNT[16]
CMD_CHSYNC_THR_HIGH_CONF[4:0]	4	Sync threshold high
CMD_CHSYNC_THR_LOW_CONF[4:0]	31	Sync threshold low
CMD_BITFLIP_WARNING_CNT[16]	RO	Count recoverable CMD bitflips
CMD_BITFLIP_WARNING_FLAG[1]	RO	Nonzero counter flag
CMD_BITFLIP_WARNING_CNT_RESET[1]	0	Reset counter
CMD_BITFLIP_ERROR_CNT[16]	RO	Cnt. unrecoverable CMD bitflips
CMD_BITFLIP_ERROR_FLAG[1]	RO	Nonzero counter flag
CMD_BITFLIP_ERROR_CNT_RESET[1]	0	Reset counter
CMD_ERROR_CNT[16]	RO	Command framing error
CMD_ERROR_FLAG[1]	RO	Nonzero counter flag
CMD_ERROR_CNT_RESET[1]	0	Reset counter
CMD_LOCKLOSS_CNT[16]	RO	Legacy counter for desync events
CMD_LOCKLOSS_FLAG[1]	RO	Nonzero counter flag
CMD_LOCKLOSS_CNT_RESET[1]	0	Reset counter
CHIP_ID_READ[5]	RO	Read back the ChipID
PULSE_START_CONFIG_READ[6]	RO	Read back charge injection conf.
PULSE_STOP_CONFIG_READ[14]	RO	Read back charge injection conf.

Table 5.14: Registers related to the CRU.

5.8.7 Pixel Configuration

The registers related to the configuration of the pixel matrix are shown in table 5.15. To configure the three bits of in-pixel memory, the registers PIXEL_CONF_* are used. Four pixels in one row are configured at once, leaving 12 bits of payload data. At first, the row and column group need to be selected via the PIXEL_CONF_ROW_SELECT[9] and PIXEL_CONF_COL_GROUP_SELECT[8] registers. The column group is the column divided by four, as the four pixels of the group are configured together. Subsequently, a write command to PIXEL_CONF_PORTAL_DATA[12] triggers the configuration of the selected pixels.

Register	Default	Description
PIXEL_CONF_ROW_SELECT[9]	0	Pixel portal row selector
PIXEL_CONF_WRITE_SPEED[4]	3	Pixel conf. write speed
PIXEL_CONF_COL_GROUP_SELECT[8]	0	Pixel portal column group select
PIXEL_CONF_PORTAL_DATA[12]	0	Pixel portal data fro 4 pixels
EN_HITOR_COL[896]	'1	HitOr column enable
EN_HITOR_ROW[464]	'1	HitOr row enable
EN_INJ_COL[896]	0	Injection column enable
EN_INJ_ROW[464]	0	Injection row enable
EN_RO_CNF[448]	'1	Enable readout per DC
EN_BCID_CONF[448]	'1	Enable BCID driver per DC
EN_RO_RST_CONF[448]	'1	Enable of DC readout reset

Table 5.15: Registers related to the pixel configuration.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

The HitOr and charge injection can be enabled with corresponding registers per row and column. For a given pixel, the column and row signals need to be enabled to activate the HitOr or injection. The EN_RO_CNF[448] register allows to selective disable the readout on particular Double Column (DC)s, EN_BCID_CONF[448] the BCID distribution and EN_RO_RST_CONF[448] the distribution of the pixel reset signal.

5.9 Trigger Unit

The Trigger Unit (TRU) is one of the main additions and changes in comparison to TJ-Monopix2. It replaces the readout and data handling to allow for higher rates and triggered operation. Since TJ-Monopix2 was inherently untriggered, i.e. all detected hits were sent out immediately, the pixel readout was controlled by a global readout controller reading one pixel after the other for the whole chip and preparing them for transmission. In OBELIX, each DC is read out independently by an End of Column controller (EoC), greatly increasing the readout bandwidth for the pixel matrix. The data of four EoCs is combined and passes two stages of trigger memory until it may be prepared for transmission in the transmission unit. The OBELIX chip supports triggered and untriggered modes of operation which can be configured via global registers. In the following chapters, the readout scheme of the matrix is explained, together with the design and features of the TRU and the relevant performance simulations.

5.9.1 Column Drain Readout and EoC

The analog amplifier and discriminator contained in each pixel has already been described in section 4.1.1. In addition to the analog circuitry, each pixel contains a digital control circuit derived from the FE-I3 chip [54] shown in figure 5.10. The discriminator output HIT is firstly passed through the hit delay compensation in the left bottom corner of the schematic. This circuit compensates for the delay of the BCID line and has been characterized in TJ-Monopix2 as described in section 4.5.3. Subsequently the resulting DEL_HIT signal is passed through an edge detector circuit. This circuit creates a short pulse at the rising and falling edge of DEL_HIT, also known as the leading edge (LE line) and trailing edge (TE line) of the pulse.

The current BCID code is latched into corresponding memories when the pulses on LE and TE appear respectively. Additionally, the TE pulse sets the TE latch, which prevents successive hits to be processed by disabling the edge detector and marks the pixel as hit.

The EoC controls the readout of hit pixels with two control lines, Freeze and Read that are distributed to each pixel in the DC, and one Token line propagating through each pixel (Token In to Token Out) arriving at the periphery. The actual pixel data is transmitted on a parallel bus arbitrated by the EoC.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

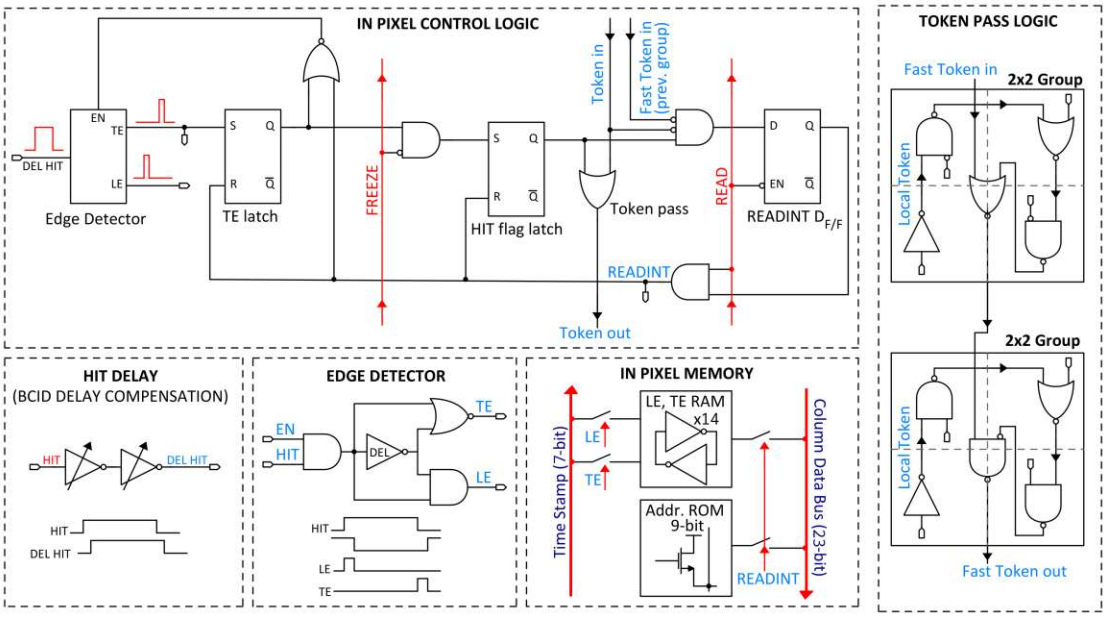


Figure 5.10: In pixel digital circuit of TJ-Monopix2 and OBELIX, adjusted for OBELIX from [35].

A readout cycle is initiated when a token is detected at the EoC, i.e. a logic 1 at the **Token Out** line of the last pixel. This happens when at least one **TE latch** is set and the **Freeze** line is 0, which is the idle state. This condition sets the **HIT flag latch** and asserts the token on the **Token Out** line. Since **Token Out** is the **Token In** line of the next pixel and they are connected via an OR gate, the logic 1 propagates through each following pixel to the EoC, notifying it about a hit. Two readout cycles are shown in timing diagram 5.11 to illustrate this.

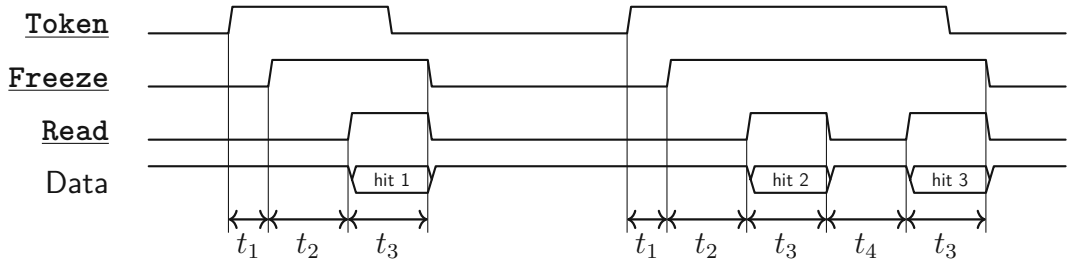


Figure 5.11: Example readout of one followed by two activated pixels. Gray areas symbolize the propagation of signals from source to sink.

The EoC subsequently asserts the **Freeze** signal and precharges the data bus to the VPC potential. While the **Freeze** is asserted, subsequent hits are blocked from accessing **HIT flag latch** and therefore **Token** line which prevents data collisions. After delay t_2 to allow the token to settle, the **Read** signal is asserted by the EoC. This notifies the activated pixel with the highest priority to transmit the hit-data in parallel by discharging corresponding bits on the data bus. It also clears the token from this pixel by resetting the **HIT flag** and **TE latches**. The pixel with the highest priority is the first pixel which has the **HIT flag latch** set and a logic zero on the **Token In** line. All other pixels with the **HIT flag latch** set can detect the token from this pixel on their corresponding **Token In** line and therefore have to wait with transmission.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

After a delay t_3 to allow the discharge of the data lines and signal propagation, the data is latched by the EoC. If the token is zero, i.e. no more pixels are staged for readout the **Freeze** signal is de-asserted. In case the EoC still detects a token, the **Read** signal is pulsed again. This repeats until no token is detected and the EoC finishes the cycle by de-asserting the **Freeze** signal.

The **Freeze** signal only prevents access to the token, so any hits during a readout cycle can still be captured by pixels containing no data. As soon as the **Freeze** signal is de-asserted, such pixels gain access to the token and another readout cycle can start immediately.

The readout can be slowed down via registers shown in table 5.16. All `EOC_T_*[]` registers are in units of cycles of the 20 MHz **clock**. Those registers allow to slow down each part of the readout sequence separately. For `EOC_T_FREEZE[4]`, `EOC_T_PRECHARGE[4]` and `EOC_T_READ[4]`, one clock cycle is the minimum waiting period for t_2, t_3 and t_4 respectively and a nonzero setting only adds additional waiting time. Simulations show that the default values provide sufficient waiting periods to allow for signal propagation. However, since the readout of TJ-Monopix2 was considerably slower because of longer propagation times of the token, this has not been verified experimentally so far. For this reason, it is possible to account for potential errors in the analog simulation of signal propagation by slowing down the readout of the EoC.

Register	Default	Description
`EOC_T_START[4]`	0	Time from token to **Freeze** assertion (t_1)
`EOC_T_FREEZE[4]`	0	Extra delay from **Freeze** to **Read** assertion (t_2)
`EOC_T_READ[4]`	0	Extra delay of **Read** high period (t_3)
`EOC_T_PRECHEGE[4]`	0	Extra delay **Read** low period (t_4)
`TRU_STOP_EOC`	0	Stop readout globally after current cycle
`EN_RO_CNF[448]`	'1	Enable readout per DC
`EN_BCID_CONF[448]`	'1	Enable BCID distribution per DC

Table 5.16: Registers related to the EoC and readout, '1 is Verilog notation for all bits logic 1.

As data from four EoCs is combined without a buffer memory, the configured waiting period of `EOC_T_READ[4]` is the minimum and can be longer. When an EoC is not allowed to immediately pass the data down to the trigger memory because another EoC is sending at the moment, it may wait with the **Read** signal asserted until the latched data can be forwarded to the memory.

The register `EN_RO_CNF[448]` can be used to individually disable EoC circuits, e.g. in case of manufacturing defects in the corresponding DC. Additionally, the `TRU_STOP_EOC[1]` register halts the readout of all EoCs while still allowing already started readout cycles to complete. This can be used as veto mechanism to save power and will be described in detail in section 5.9.9.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

5.9.2 Data Sequence and Time Order

While the readout sequence of the OBELIX EoC is similar to the TJ-Monopix2 global readout controller, the reduced readout time reveals a complication in data processing. The edge detector in each pixel creates pulses to sample the current BCID at the leading and trailing edge of the discriminator. Only after the trailing edge, i.e. after the ToT has passed, the **`HIT flag latch`** is activated and the pixel proclaims that it contains data to be read out.

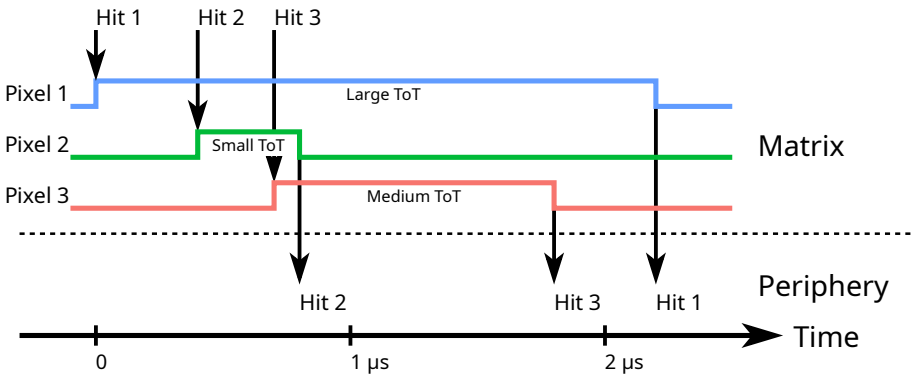


Figure 5.12: Readout of three hits with different ToT, colors correspond to different pixels.

Figure 5.12 shows an example sketch for three hits in three pixels. The time to actually read out a hit is small (typ. 141 ns for a single hit) compared to typical ToTs (1 µs or more) and not explicitly shown in the sketch. Three pixels are respectively hit by one particle each in order 1, 2 and 3 in this example. Since hit 1 has a large ToT due to large charge deposition, it is read out last, and hits 2 and 3 are read out first. This behavior is independent from the locations of the three hits on the pixel matrix and creates a local time-disorder in the data stream.

At higher hit-rates, the priority logic in the DC may also cause additional disorder of the hits. If two or more pixels are ready to be read out, the priority chain forces readout from top (high row number) to the bottom of the sensor. However, the majority of the time disorder arises from the first effect. Only at extreme hit-rates close to the specification limit, the priority logic contributes to the disorder but is still the smaller effect.

5.9.3 TRU Design

The TRU is organized in 112 Trigger Groups (TRGs), that process the data of 4 DCs each. Two group-level priority chains for the left and right half of the chip are implemented to drain the data from the TRGs and forward it to the TXU.

The TRU can be configured in three operation modes called Triggered, Triggerless and Triggerless sorted via the **`TRU_TRIGGER_MODE[2]`** register as shown in table 5.17. After changing this register to a mode requiring a configured trigger latency, a BCID reset needs to be performed.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

TRU_TRIGGER_MODE[2]		
Value	Mode	Requires latency
0	Triggered mode	Yes
1	Triggerless sorted mode	Yes
2	Triggerless mode	No
3	Reserved	—

Table 5.17: Modes of TRU configuration depending on **TRU_TRIGGER_MODE[2]** register

In Triggered mode, data is automatically discarded after the configured trigger latency has passed. When a `TriggerX` command to the corresponding Le-BCID is received, the data is marked as triggered and forwarded to the TXU.

In Triggerless mode, no trigger is required and data is immediately sent to the TXU. In this case, the TRU is configured as FIFO memory and the trigger latency has no effect. The hits may arrive disordered chronologically due to different ToT values. This is similar to TJ-Monopix2, but with additional buffer memory which increases the acceptable data rate. Despite this, the total rate is still limited to a few MHz, since the TXU is limited in bandwidth to about 4.8 MHits/s.

In Triggerless sorted mode, no trigger is required but the TRU is used to chronologically sort the hits. This requires a trigger latency larger than the maximal expected ToT and also delays transmission by this latency. Due to internal limitations, this mode does not honor the two least significant bits of the Le-BCID code. Therefore, the 2 least significant bits are not included in the sorting. This mode is a hybrid of the other two modes and mainly intended for debugging potential issues with the Triggered mode without requiring trigger commands.

The trigger latency of the TRU module can be configured via the **TRU_LATENCY[9]** register to an arbitrary integer number of cycles of the 20 MHz **clock** domain. After changing the latency, the BCID needs to be reset to activate the change, either via the `Clear` command (preferred) or the **TRU_RESET_BCID[1]** register. The **TRU_RESET_BCID[1]** register keeps the value, i.e. it needs to be set and then reset via two separate `Write Register` commands. When the BCID is reset, the initial value of the BCID counter is configurable via the **BCID_INITIAL_VALUE[9]** register.

As an option, the **TRU_SLOW_BCID[1]** register allows the reduction of the BCID frequency by a factor of two. This saves power, but decreases the time resolution accordingly. The TRU itself keeps operating at the same frequency, so no adverse effects on processing speed are expected in this mode. Although the chip is not designed for this mode, it is intended as a fallback option if the power consumption exceeds the limit of the cooling capacity. If desired, this bit should be set before a BCID reset.

At nominal configuration, no data frame is transmitted when a trigger command is received but no corresponding hit data is available. By setting the **TRU_ENABLE_EMPTY_TRIGGER[1]** register, an empty dummy frame is generated in such a case instead. This can be useful in debugging but is not intended for VTX operation since it causes unnecessary load on the TXU. It can only be used in Triggered mode and the dummy frames contain only logic 1s in the Le and Te-BCID, as well as row and column address. In the case that it is not possible to receive hits in the Triggered mode at the testing stage for an unknown cause, this feature can be enabled. This allows to see the correct interpretation of the `Trigger` commands at roughly the right time. Another use case is to count the number of

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

triggers received. This is also available as the `Trigger ID[5]` information, but if discrepancies are seen, this mode can help to diagnose the problem.

The `PTD_ENABLE[1]` register switches on the precision timing module which will be described in detail in section 5.9.6.

Register	Default	Description
`TRU_LATENCY[9]`	0	Trigger latency for the TRU
`TRU_RESET_BCID[1]`	0	Reset BCID without `Clear` command
`TRU_TRIGGER_MODE[2]`	0	Operation mode for the TRU
`TRU_ENABLE_EMPTY_TRIGGER[1]`	0	Transmit a dummy frame if no hits are triggered
`TRU_SLOW_BCID[1]`	0	reduce BCID clock by a factor of 2
`BCID_INITIAL_VALUE[9]`	0	Initial BCID code after a reset
`PTD_ENABLE[1]`	0	Enable precision timing module

Table 5.18: Registers related to the TRU configuration.

5.9.4 The Trigger Group

The Trigger Group (TRG) performs the actual data buffering and association with the trigger. An overview of the internal structure of the TRG is shown in figure 5.13. The hit-data of four DCs is read out by the EoCs. It is subsequently combined into a common data stream by a round robin arbiter and stored in the two stages of the trigger memory. The arbiter adds two bits to the pixel address to identify the source DC inside such a 4-DC group. Subsequently, hits to be transmitted are sent to the TXU via the group-level priority chains. The whole circuit operates at the 20 MHz `clock`. The choice of 4 DCs for each TRG arises from the bandwidth requirements as shown in table 5.19. It takes at least two clock cycles to read data from a pixel by the EoC, so the maximal bandwidth is half of the clock frequency whereas the round robin arbiter can process one hit per clock cycle. The expected data rate at 120 MHz/cm^2 leads to an average utilization of about 25%, giving sufficient margin to prevent congestion which would increase the time-disorder.

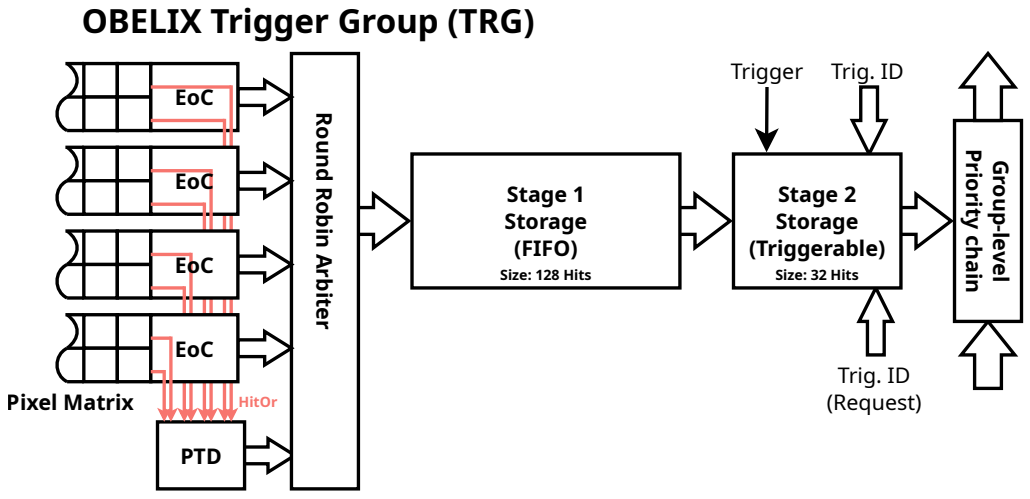


Figure 5.13: Overview of one Trigger Group (TRG).

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

Hit-rate	DCs	Bandwidth	Data-rate
120 MHz/cm^2	1	10.6 MHits/s	1.22 MHits/s
	4	21.2 MHits/s	4.86 MHits/s
60 MHz/cm^2	1	10.6 MHits/s	0.61 MHits/s
	4	21.2 MHits/s	2.43 MHits/s

Table 5.19: Calculation of hit-rates and bandwidth, $A_{DC} = 10.13 \times 10^{-3}\,\mathrm{cm}^2$

BCID Extension

In the pixel matrix, the BCID for Le and Te are 7 bit wide. This means that with a clock frequency of 21.2 MHz, an overrun occurs roughly every 6 µs and a new matrix epoch begins with repeating timestamps.

After the pixel-data is read from the pixel matrix, the EoC decodes the 7 bit Le-BCID from Gray code to binary, and extends it by 2 bits to increase the distance between epochs in the periphery to 24 µs. The extension compares the hit Le-BCID to the corresponding bits in the current BCID. The result decides if the extension bits are taken from the current or the previous matrix epoch. This ensures the correct extension even at the epoch-transitions, but reduces the maximum acceptable ToT period by the maximum pixel readout delay. A visualization of this moving window extension is shown in figure 5.14. The horizontal axis shows the time in BCID-units as hits arrive at the BCID extension circuit in the EoC. The vertical axis shows the Le-BCID code of the hits. The red arrow shows the delay from the ToT and the readout that causes the hits to leave the diagonal color-boundaries and the shift to the right.

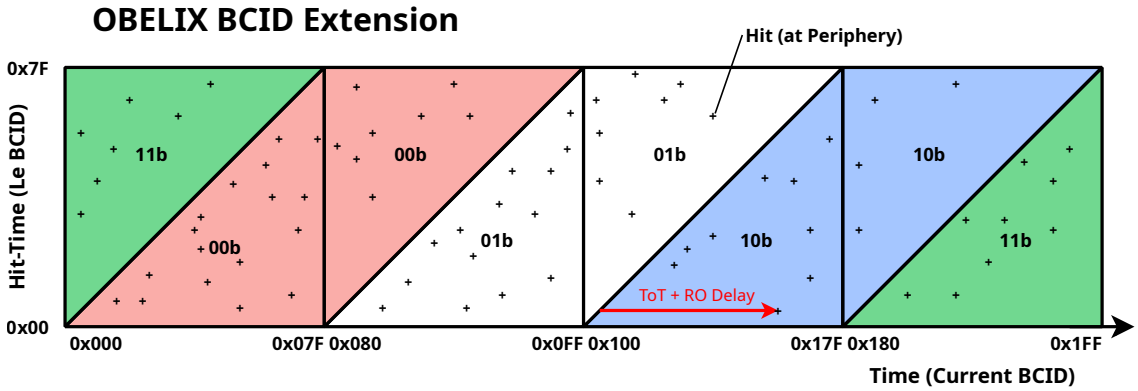


Figure 5.14: BCID extension time diagram for the duration of one periphery epoch. Colored areas correspond to different extension-values.

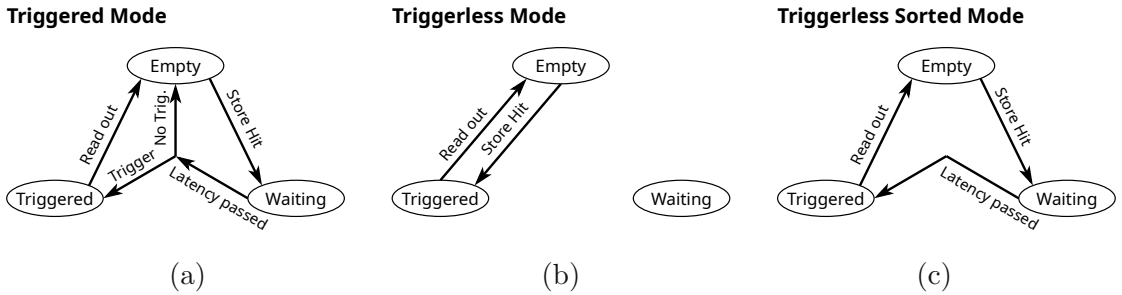
Stage 1 and Stage 2 Memory

The data subsequently transferred to the trigger memory, consisting of two stages. Stage 1 (S1) is capable of storing 128 hits and uses an SRAM block available in the technology configured as a FIFO. Stage 2 (S2) consists of 32 cells for one hit each and as soon as a cell is empty in S2, a hit is transferred there from S1. The S2 performs the actual association with the trigger, but S1 is more efficient in terms of power and area. When a trigger arrives, only data already residing in S2 can be associated. Therefore, the maximum time-disorder defines how early hits need to arrive in S2. Since the time disorder is in the range of the maximum ToT (Landau

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

tail up to 3 µs, depending on the settings) but the target latency is 10 µs, the data can initially stay in S1 for more efficient resource utilization.

Each memory cell in S2 contains a small finite state machine (FSM). While the three states Empty, Waiting and Triggered are always the same, the transitions depend on the `TRU_TRIGGER_MODE[2]` setting. Figure 5.15 shows the three possible configurations in detail.



(a) (b) (c)

Figure 5.15: Different configuration of the FSM inside each cell of S2.

In Triggered mode, S2 performs the actual trigger matching. To do this, a delayed BCID is generated by a second BCID counter. It provides the current BCID minus the configured trigger latency to all cells of S2. At each clock cycle, every cell with hit data (i.e. in the Waiting state) compares its Le-BCID with the delayed timestamp. When the two values are equal, this means that this hit has exactly passed the trigger latency and a decision for the future of the hit needs to be made.

At the time of match, the cell looks up the value of the global `Trigger` line. This line is set to logic 1 for one clock cycle if a `Trigger` command has been received and 0 otherwise. In the absence of trigger, the data in the cell is discarded and the state transits to the Empty state. From there, it can receive another hit from S1 to continue in the Waiting state.

On the other hand, if `Trigger` is asserted, i.e. a `Trigger` command has been received, the Waiting state transits into the Triggered state. Additionally, the globally distributed `Trigger ID write[5]` value is stored along with the hit for identification. This ID is generated by the trigger ID controller of the TRU and incremented after each triggered BCID code, as shown in the overview diagram 5.16.

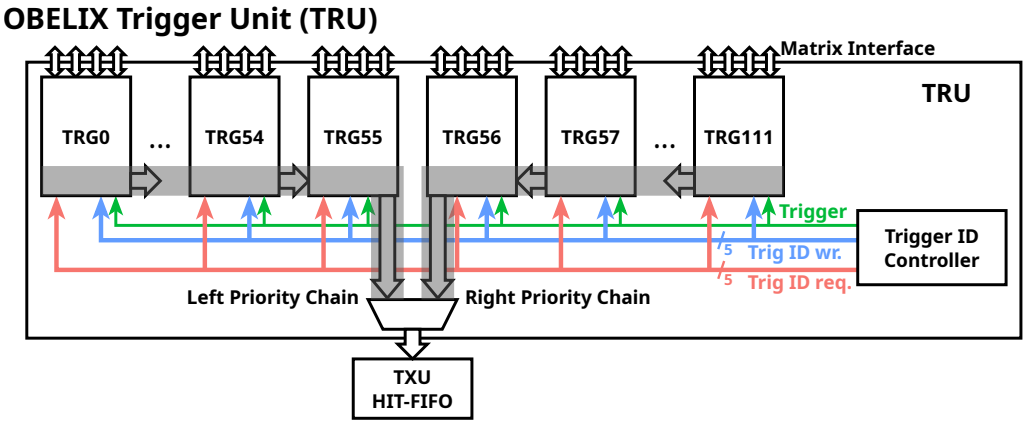


Figure 5.16: Overview of the Trigger Unit (TRU).

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

The trigger ID controller has also a second trigger ID counter, the **Trigger ID request[5]**. After a Trigger command, the **Trigger ID request[5]** selects the corresponding S2 cells in the Triggered state with a matching Trigger ID. These cells are sequentially read out and the data is sent to the TXU for transmission. This frees up the cells which subsequently transit into the Empty state. Switching to the next **Trigger ID request[5]** value takes 3 clock cycles if no data is available to allow propagation of signals through the individual priority chains inside the S2 in each TRU and the group level priority chains. Reading a hit takes between 1 and 4 clock cycles, depending on distribution across the matrix. The individual TRGs each stage one hit for readout in the group level priority chain. In the high occupancy case, the left and right priority chain can interchangeably send data and each TRG can send one hit every fourth clock cycle. This means, every second clock cycle one half of the chip can be read and if two TRGs contain data, they can again interchangeably send data each second clock cycle, so each of the every fourth clock cycle. At hit-rates close to the specified maximum, the transmission bandwidth of the TXU also limits the speed, slowing down the reading from the TRGs.

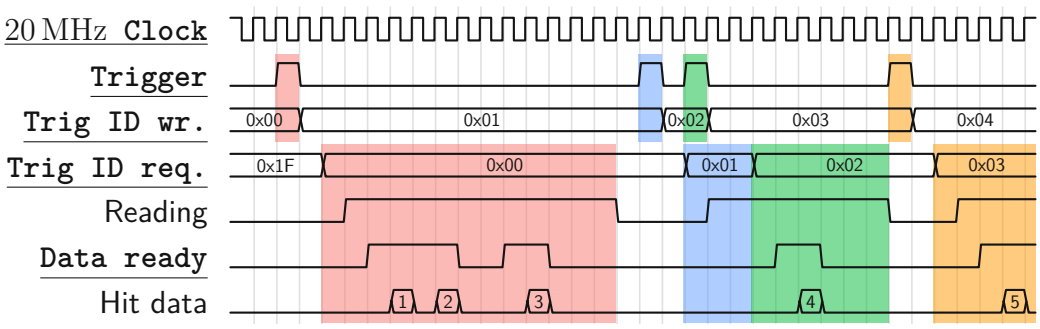


Figure 5.17: Management of the trigger IDs.

This trigger ID logic assures a chronological transmission of the hits, even when multiple triggers are received in quick succession and the reading of a large number of cells in Triggered state takes time. The timing diagram 5.17 shows this behavior in detail. After the first trigger with ID 0x00, three hits are read out (hits 1 to 3) and no additional trigger is received. The next two triggers with ID 0x01 and 0x02 arrive in quick succession, and even though trigger 0x01 has no hits, it shifts the reading of the 0x02 trigger with one hit (hit 4). It should be noted that the current plan foresees to trigger two subsequent BCID codes for each Belle II trigger. This is treated as two independent triggers by the TRU, but single BCIDs are triggered in the timing diagram for clarity.

Since the trigger ID lines are 5 bits, a maximum of 32 different triggered BCID codes can reside in S2, i.e. the **Trigger ID req[5]** needs to be able to catch up with **Trigger ID write[5]**. If this limit is exceeded, e.g. because of too high trigger frequency or too many hits in each trigger, the trigger ID controller will start to ignore triggers. When the number is reduced again because the transmission of one event finished, the chip resumes to process Trigger commands. Since the trigger ID is also transmitted, this case can be immediately detected in the backend system with at the next successful trigger.[3] In principle this is no reason for concern if the Le-BCID code is still correct. However, the backend system might still reset the

[3]We assume that the next trigger actually contains at least one hit, otherwise nothing is transmitted. Since this overwind only happens at high rates, this is a reasonable assumption.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

trigger ID for the particular chip when the next trigger ID of the other chips is zero so that the trigger IDs of all chips are harmonized again. Generally, this will not happen at the specified hit- and trigger-rates, but needs to be considered if the chip is operated above those limits.

In the Triggerless mode, the hits are directly stored in the triggered state and both trigger ID values are constant zero. This means, the hits are immediately staged for transmission, but S1 and S2 both serve as a buffer memory.

The Triggerless Sorted mode corresponds to a constant triggering of all BCID codes. Since the chip cannot process this sufficiently fast, the lower two bits of the BCID code are masked at the comparison in this case. Therefore, the hits are only time ordered in the higher 7 bits, and the `Trigger ID write[5]` is incremented every fourth clock cycle. This limits the maximum hit-rate below the hundred kHz range, but is sufficient for a test-beam scenario. While the use case for this mode is limited to debugging the TRU, it costs almost no chip space. All the functionality for this mode needed is already available from the other two modes and just a few signals are re-routed when this mode is enabled.

5.9.5 TRU Performance

To verify function and performance of the TRU, extensive simulations of the Triggered mode have been performed. Since the time order of the hits is of particular interest, a ToT spectrum was simulated, using landau distributed charge values and decomposing them into clusters with ToT conversion. This creates a spectrum that are similar to ones from TJ-Monopix2. The simulations have been performed for three ToT-clipping scenarios, without clipping, moderate clipping to 1.2 µs and strong clipping to 0.3 µs.

The behavior of the matrix is modeled in a high level functional model, and Poissonian distributed random hits are injected to the model of the matrix. The chip receives random triggers, also Poissonian distributed. The hits corresponding to the triggers are subsequently compared against the transmission from the design under test and the fraction of received hits is compared to the number of triggered hits injected, i.e. the efficiency is calculated.

The results of this simulation are shown in figure 5.18 for the three ToT-clipping scenarios. Three different trigger latencies were evaluated: 5, 10 and 12.5 µs. In principle there are two different types of overload to the TRU. On one hand, the S2 can be too small to cover the full variation of time-disorder. In this case, the efficiency slowly decreases with rate, as only the extremely late hits are lost. For example, this is the case for the 5 µs latency curve at hit-rates above the specification of 120 MHz/cm^2 without clipping. Since stronger clipping reduces the ToT diversity and therefore time disorder, the efficiency can be recovered with stronger clipping in this case.

The other type of overflow is an insufficient total memory size. This happens for the 12.5 µs latency case with a steep falloff between 100 and 120 MHz/cm^2 mostly independent from clipping.

For the design case of 10 µs latency, both effects play a role. Without any clipping, the tail of the landau distribution creates a penalty in performance for high rates. However, even moderate clipping to 1.2 µs recovers the efficiency to basically zero loss at design hit-rate of 120 MHz/cm^2. For all simulations, the data loss due to buffer overflows is well below 1% for the design hit-rate and below, even for the 12.5 µs latency case.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

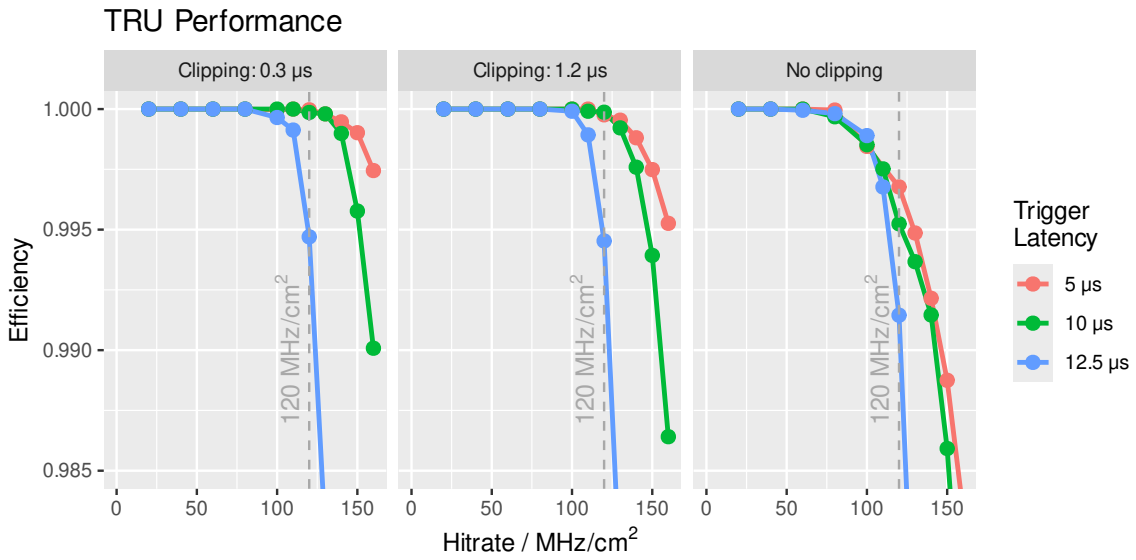


Figure 5.18: Results for the TRU performance simulation for different ToT-clipping and latency settings.

Since the triggered hits reside in S2 longer than untriggered hits because they are waiting for readout, they reduce the effective size of S2. To limit the computation time of the simulations, only 16 DCs were simulated. At the same time, the trigger rate was increased to 1 MHz to increase the amount of data acquired during the same interval of simulation time. Despite the increased trigger rate, the readout of data from S2 to the transmission unit still plays a minor role in the case of 16 DCs. Therefore, the effect of triggered hits accumulating in S2 is generally underestimated and the results still represent the low-trigger-rate limit of operation. Simulations in section 5.9.8 will explore the required corrections for different trigger rates of the full chip with 448 DCs, including the bandwidth limitations of transmission. This approximately results in the same performance as a certain increase in hit-rate, and these correction factors will be calculated in section 5.9.8.

5.9.6 Peripheral Time-to-Digital Converter

The Peripheral Time to Digital Converter (PTD) unit is part of each TRG in the OBELIX-1 chip. It provides increased timing resolution to mitigate and surpass the effects of the reduced BCID timestamping frequency compared to TJ-Monopix2. As this requires a substantial amount of power, it can be disabled for the two innermost layers of the VTX (iVTX) which rely on cooling with the least material budget possible. In section 4.5.2 it has been shown with TJ-Monopix2 that the matrix design is capable of less that 3 ns timing resolution with adequate corrections applied (4 ns after irradiation to the VTX design fluence). The PTD module implements everything needed to measure at such a timing resolution on-chip, at least for sufficiently low hit-rates and high power budget.

The PTD module utilizes the per-column `HitOr[896]` signals, sampled at both clock edges of the 160 MHz `clock`. Groups of 8 columns, i.e. one TRG share a one PTD block, a simplified functional diagram is shown in figure 5.19. Some complexity is hidden in the DDR-Edge-Detectors which effectively determine the Least significant bit (LSB) of the counter result allowing a resolution of two times the clock frequency.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

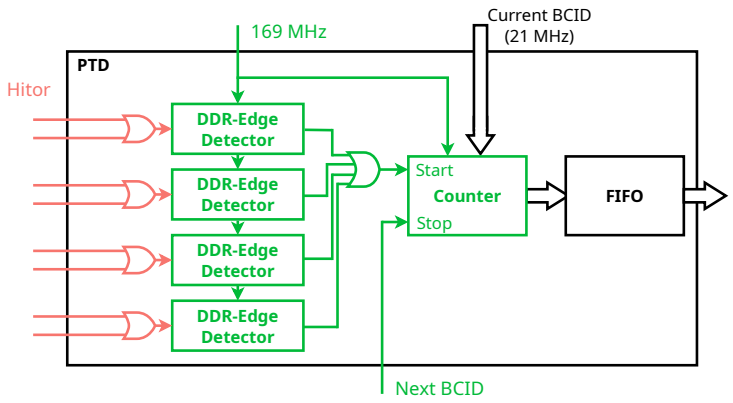


Figure 5.19: Internal structure of the PTD module in each TRG.

The PTD data, i.e. the output of this block is independent from the actual hit data received from the matrix. The association of PTD- and hit-data is not done on-chip. Instead, they are sent out individually and have to be correlated offline. Ideally, this is included in the clusterizer algorithm which then also annotates each cluster with the precision timing information, if available.

The data from the PTD block is injected to the ordinary hit data-stream at the arbiter stage as so called PTD-hits. Such PTD-hits use the same data-format as regular hits, with a row address of 511. There are no pixels with this row address, so PTD-hits can be uniquely identified at the data analysis. The rest of the TRU treats PTD-hits and ordinary hits in the same way, reducing complexity of the following modules. The usage of the individual fields of ordinary and PTD-hits is shown in table 5.20. The PTD counter data is 4 bits which are stored zero-padded at the Te-field.

The PTD module itself has no spatial resolution. Since each TRG contains one PTD module, the column address is available up to TRG level resolution, but the PTD module cannot provide data about the individual columns inside each TRG. Therefore, the 3 lowest bits of the column address are always zero for PTD-hits.

Parameter	Normal hit function	PTD-hit function
Le-BCID	Leading edge	BCID code of **HitOr** leading edge
Te-BCID	Trailing edge	PTD count (4 LSBs only, zero padded)
Row address	Pixel row	Always 511
Column address	Pixel column	Address of the PTD block, 3 LSBs zero

Table 5.20: Functions of normal hit and PTD data

The operation of the PTD module can be enabled by setting the **PTD_ENABLE[1]** register, which is the only configuration for this module. When disabled, all clocks of the PTD components are disabled so there is no dynamic power consumption. The PTD module is disabled by default after a Reset or startup of the chip.

The PTD module itself consists of two parts. The first part is a DDR-edge detector (ED), which consists of two ordinary edge detectors operating on rising and falling edge respectively. These synchronize the **HitOr** signal to the 160 MHz **clock** and clip the signal from the ToT period to one clock cycle. This design allows a time binning of the double clock frequency (2.95 ns).

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

The synchronized signal then starts the PTD counter running with 160 MHz and also samples the timestamp of the BCID counter. On the next rising edge of the 20 MHz `clock`, the PTD counter is stopped and the PTD-hit is transferred into a FIFO for injection into S1 via the round robin arbiter. As a PTD-hit waiting for the round robin arbiter to allow transmission would prevent further operation and cause increased dead time, the FIFO of size 4 immediately accepts the data and subsequently handles the arbitration while freeing the PTD module.

One PTD counter can be connected to multiple DDR-EDs, which causes a two-staged dead time for the PTD circuit. In the first dead-time step, none of the 8 columns in one TRG can process the PTD information of another hit as the counter is counting. This lasts between one and two clock cycles of the 20 MHz `clock`, a first while the counter is counting (part of a full clock cycle depending on arrival time) and the second full cycle when the result is transmitted to the FIFO and the counter is cleared.

After that, a second dead time affects only the columns connected to the activated ED, and lasts for the ToT of the initial hit. This is caused by the logic level of the `HitOr` which stays high for the entire ToT period, so any other hit cannot be seen by the same ED. Additionally, if another hit is received during the ToT period, it prolongs the dead time further, as the `HitOr` just stays at the logic 1 level. These stages are illustrated in figure 5.20.

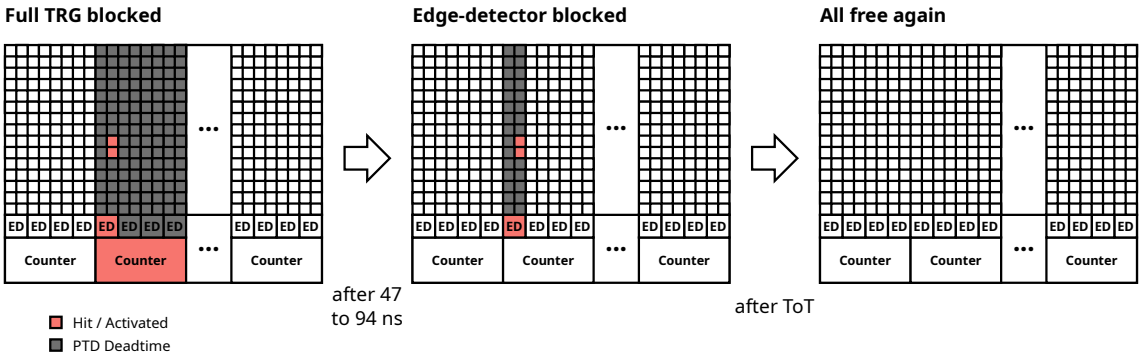


Figure 5.20: Illustration of the deadtime stages of the PTD module in 2 Cols/ED implementation.

Furthermore, an association of the PTD data is not possible if more than one particle in the same TRG have the same BCID codes. In this case, there would be two ordinary hits and one PTD-hit, but no information to which hit the PTD info belongs. For hits belonging to the same particle (i.e. inside the same cluster), this is not an issue since timewalk assures that the PTD hit belongs to the seed pixel i.e. the hit with the longest ToT.

In the performance simulation described in the following, all PTD hits with such multi-hit ambiguities were dropped and therefore count as lost. However, in a multi-layer detector like VTX, there might be a possibility that the ambiguities are resolvable in some cases. Therefore, the presented values are a lower limit to what actually might be achievable.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

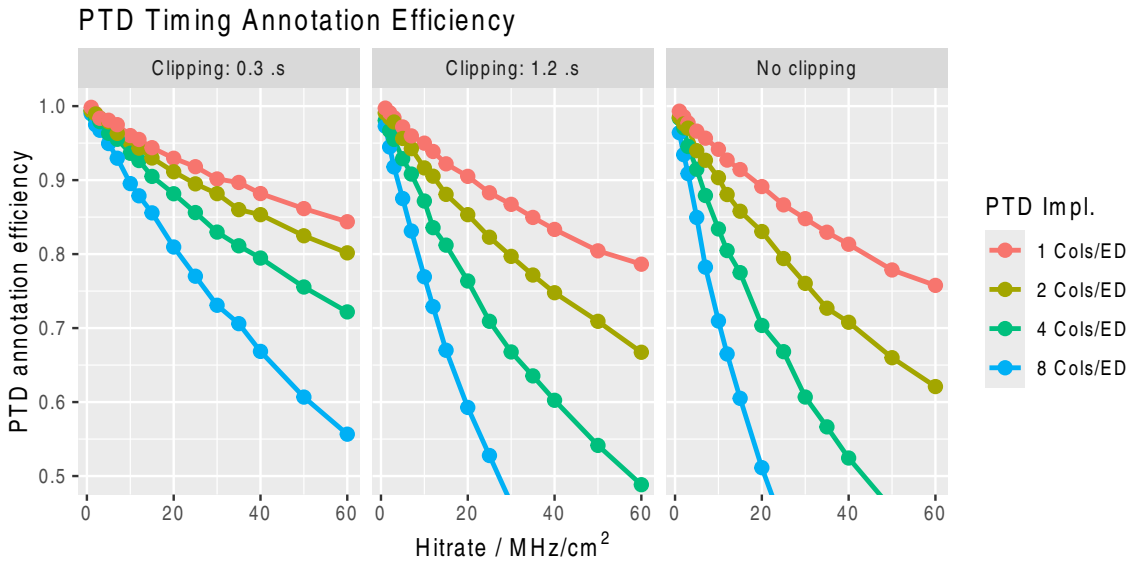


Figure 5.21: Timing annotation efficiency for different PTD implementations and clipping settings.

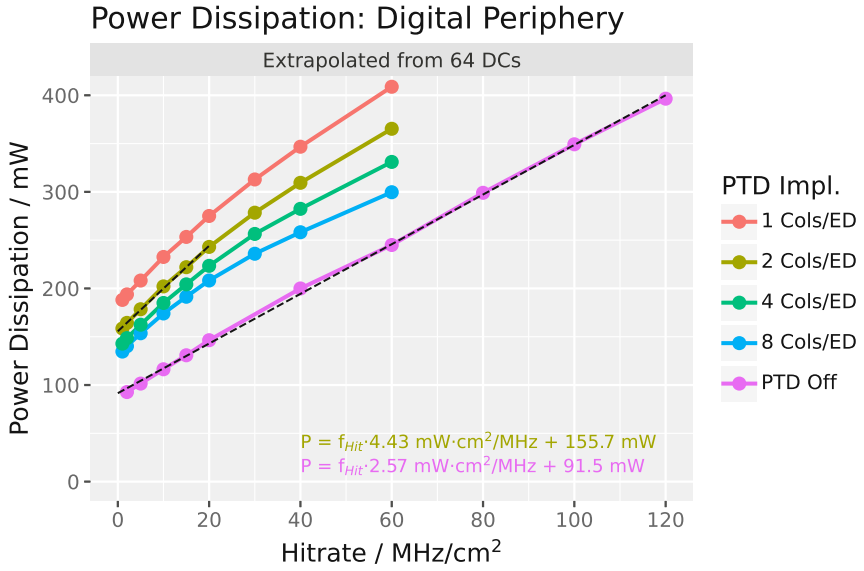


Figure 5.22: Power dissipation of the digital periphery for different PTD implementations and without the PTD module enabled for comparison, the power from all other parts of the chip is 606 mW.

Four different implementations of the PTD module were simulated with 1, 2, 4 and 8 columns connected to one ED respectively. This design choice is a trade-off between PTD annotation efficiency and power consumption. The results of performance- and power simulation at the post-routing level are shown in figures 5.21 and 5.22 respectively. The annotation efficiency is the fraction of hits that can uniquely be associated with the corresponding PTD-hit.

The first layer with PTD enabled will be L3. In the pessimistic scenario, the expected hit-rate is 1.12 MHz/cm^2. When the rate is scaled with the equivalent factor to reach 120 MHz/cm^2 in L1, this results in a rate of about 4 MHz/cm^2.

For the 8 cols/ED implementation, the dead time for all 8 columns in each TRG extends to the full ToT period. Therefore, this implementation is strongly affected

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

by ToT clipping. Without clipping, the performance drops to 50% at 20 MHz/cm^2. Nevertheless, this implementation contains the fewest high frequency D-FF and is therefore the most energy efficient. The other extreme is one individual ED for each column. This is the best performing scenario with little improvement for clipping, as the efficiency is limited by the dead time caused by the counter in this case, which is independent from the ToT. The power consumption of this version is considerably higher, more than twice of the 8 cols/ED version when the baseline from the PTD-Off simulation is subtracted.

For OBELIX-1, the implementation of 2 Cols/ED was chosen. The performance gain of going down to 1 Col/ED at 20 MHz/cm^2 is only in the single digit percent range, while the power consumption would increase significantly, even in idle.

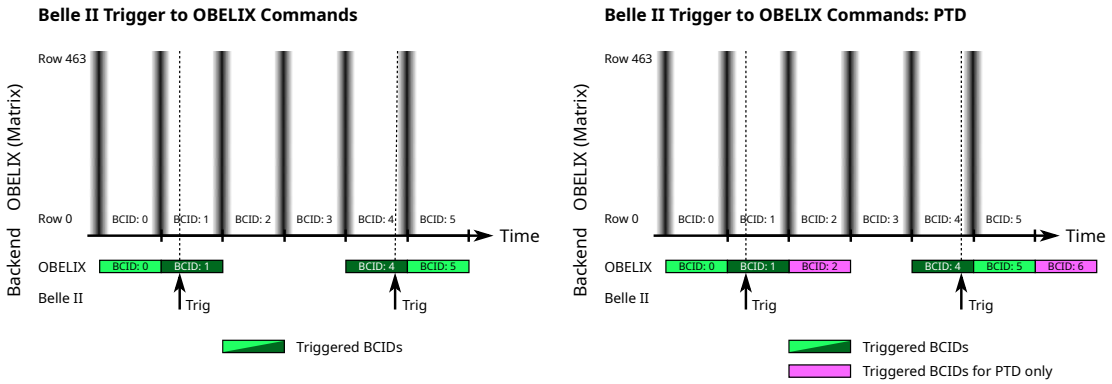
When compared with hit efficiencies, where 99% is considered 'bad' already, the PTD annotation efficiencies are very low. However, to associate precision timing to a full track, it is sufficient to associate a single hit of the track with PTD data whereas for hit efficiency, hits are required on (almost) all detector planes.

When enabling the PTD module on three oVTX layers, the probability of not being able to associate any PTD information of three layer intersections is well below 1%. While any additional timing information is not strictly required, it helps to improve the track timing resolution but also simplifies the reconstruction of the track itself.

With the PTD module, it is also possible to measure timewalk directly for a large number of pixels without additional external components. Some technical details on this are available in appendix C.0.1.

5.9.7 Triggering Strategy

The VTX backend system needs to convert the Belle II L1 trigger signals to adequate trigger commands for OBELIX. The L1 trigger has less than 10 ns time resolution (sigma) and is based on a frequency of 127.2 MHz or RF/4, corresponding to a binning of 7.86 ns [16]. The BCID of OBELIX has a frequency of RF/24, so that 6 Belle II trigger bins correspond to one BCID code.



(a) (b)

Figure 5.23: Visualization of trigger and hit timing uncertainty on triggered BCID for iVTX (a) and oVTX with PTD mode (b)

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

Since also the OBELIX chip has a finite time resolution, it is currently planned to trigger on two subsequent BCID codes per Belle II trigger, corresponding to a constant integration time of 94.3 ns. A schematic of this is shown in figure 5.23a. The black gradients represent the total uncertainty in time, mainly from trigger jitter and OBELIX time resolution. The DAQ system needs to select the two BCID codes to trigger on according to the trigger time that is available at higher resolution.

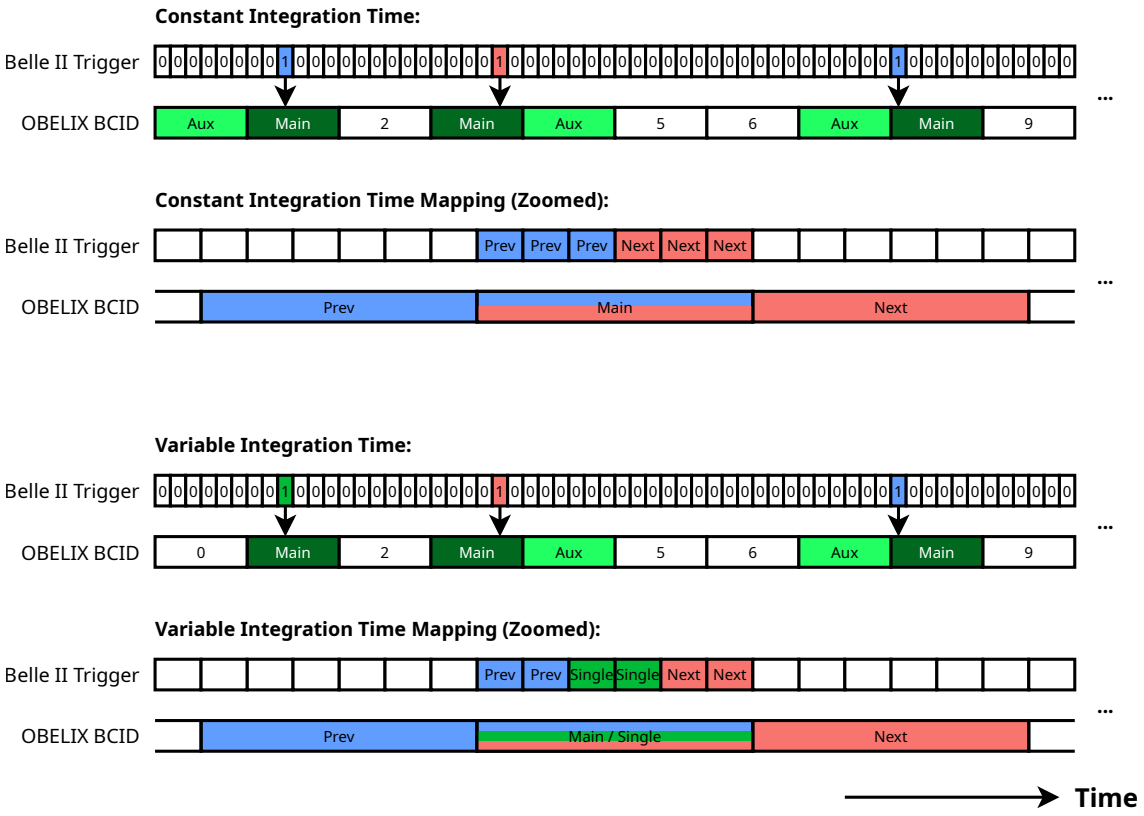


Figure 5.24: Possible trigger mappings of the DAQ system.

This behavior is visualized more precisely in figure 5.24. When a Belle II trigger arrives, it uniquely corresponds to one BCID, marked with Main/dark green in the graphic. The backend system needs then to decide if the previous or the following BCID code is triggered in addition, marked with Aux/bright green. The second block shows the general mapping of the constant integration mode for all triggers possible inside one BCID frame, the time axis is scaled for better visualization. Depending if it is on the left or right side of the BCID code, the previous (Prev) or next (Next) BCID is triggered as Aux.

An alternative option with variable integration time is shown in the lower half of the graphic. Here, The center bins would only cause one BCID to be triggered on, whereas two BCIDs are triggered closer to the transitions. In the case shown, the selection is 2-2-2, effectively reducing the trigger load of OBELIX by 1/6 or 17%. If a higher time resolution can be achieved, the window could be increased which would further decrease the trigger frequency of OBELIX. In return, the maximum trigger frequency of Belle II could be increased by the same factor.

In case the time resolution of the total system is not sufficient, it might be necessary to trigger on three BCID codes. The full implementation of this behavior would always trigger on a third BCID code either before or after the main code. A less strong option would be to trigger the previous and the next BCID code when

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

the trigger is in the center of the BCID, corresponding to a non-constant integration time.

Even though an increased number of BCID codes to be triggered causes more load on the chip, the time resolution is not lost. At time of analysis, it is always possible to put narrower cuts on the BCID codes used for tracking, e.g. if a higher time resolution on the event T0 is available. Since the occupancy scales linearly with integration time, a small integration time is desirable for tracking performance.

For layers using the PTD module, an additional BCID code needs to be triggered following the regular trigger integration time. The hit data of the additional BCID can be immediately discarded, possibly already in the VTX backend. The PTD information is stored in the same format like real hit data (PTD-hits) and treated as such by the TRU. However, the corresponding Le-BCID to these PTD-hits might be one code after the actual hit-data due to the **HitOr** delay, requiring an additional BCID code to be triggered to extract the PTD information, as illustrated in figure 5.23b.

This does not reduce the maximum trigger frequency overall, as the PTD feature is only used by oVTX, where the hit-rate is significantly lower. The main limit in trigger frequency comes from iVTX with the highest hit-rates.

This chapter so far was describing the triggering strategy in the VTX scenario. For the test-beam scenario where a prototype chip is used with a beam-telescope, a similar scheme can be used. Since the particle rates at test-beams usually are far below 1 MHz, the trigger integration window can be generously increased, especially at the beginning of the tests. Theoretically up to 32 successive BCIDs can be triggered given a sufficient veto period after each trigger to process all those triggered BCIDs. This leaves about 1.5 µs of integration time to 'strike the right spot' in terms of timing without worrying too much about things like cable lengths. Since the Le-BCID is transmitted along with the trigger ID for each BCID, tighter time cuts can be easily applied offline during the analysis.

When the system runs in a stable way and is well understood, it is possible to reduce the integration window to have a scenario more similar to VTX. It is also possible to artificially simulate VTX beam background during a test-beam using the digital injection module which will be described in section 5.13.

5.9.8 Maximum Trigger Rate

The simulations of the trigger unit hit capacity presented in section 5.9.5 used a scaled version of the chip to reduce computing power requirements. Therefore, they are effectively in the low-trigger-rate limit of operation where the time of hits residing in S2 after triggering is small.

In reality, the TXU has a finite transmission bandwidth. There exists a transmission FIFO memory, but the size of 8 words is only intended to mediate the clock domain crossing, not to actually buffer hit data. This means that triggered hits stay longer in S2 than untriggered hits, reducing the size of S2 that is usable for storing new hits. Effectively, the space in S2 is reduced by these triggered hits. The processing capability of this reduced S2 size is approximately the same as a higher hit-rate. Therefore, this simulation calculates a correction factor that needs to be applied to the x-axis of figure 5.18 in section 5.9.5, depending on the trigger rate.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

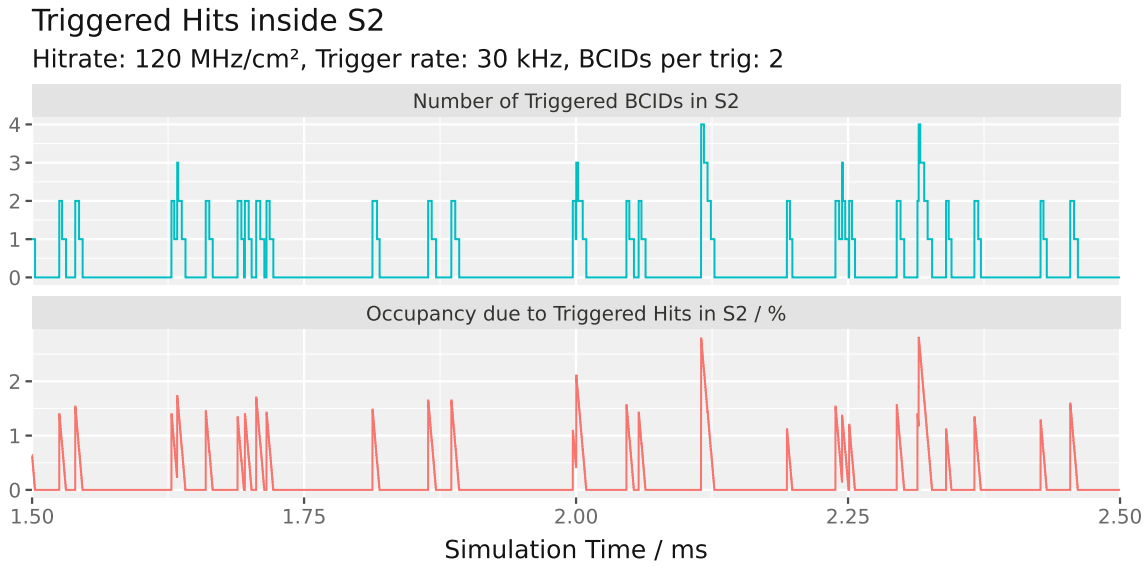


Figure 5.25: Result of the queue simulation in the time base.

A high-level simulation has been performed to model and study this behavior and calculate the correction factors. It assumes hits and triggers with Poissonian distribution and models the bandwidth of the TXU. The S2 in this model behaves as queue, with only triggered events considered. In total, 2 seconds have been simulated. A small section of the results in time base are shown in figure 5.25, which use the specification limits of OBELIX. The upper plot shows the number of triggered BCIDs/Trigger IDs that currently reside in the S2 memory. The bottom plot shows the usage of the total S2 memory to buffer hits that already have been triggered, waiting for readout and therefore are unavailable for storing new data. After a trigger, this value jumps to the number of hits for this trigger and gradually decreases as hits are transmitted. Since the simulation assumes two BCID codes per Belle II trigger, the number of triggered BCID codes in S2 increases by 2 for each trigger. Most of the time, the triggers are spaced sufficiently to fully transmit the event, however, occasionally two or more triggers appear in quick succession, increasing the occupancy further.

The results of these occupancies were subsequently used to fill the histograms in figure 5.26. The upper histogram shows again the number of triggered BCID codes residing in S2. Since the OBELIX chip can maximally process 32 triggered BCID codes simultaneously in S2, it is important to have sufficient margin, which is the case in this scenario with a maximum of 10 BCID codes during this 2 s simulation with a frequency below 10^{-6} in BCID units. The bottom plot shows the histogram of occupancy in S2 due to the triggered hits. The 99% percentile for the full histogram is marked, which gives an upper limit on how much of the memory is unavailable for data storage, i.e. 99% of the total time, the occupancy is below this value.

The mean occupancy is also marked, but the calculation excludes the zero bin, i.e. it is the mean occupancy after a trigger until the transmission line idles again. This gives a more accurate estimation of the average case, but may underestimate the effect of large occupancies occurring rarely. The occupancies can be directly interpreted to result in an equivalent performance of hit-rate increase in the low-trigger-rate limit in the simulations shown in section 5.9.5.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

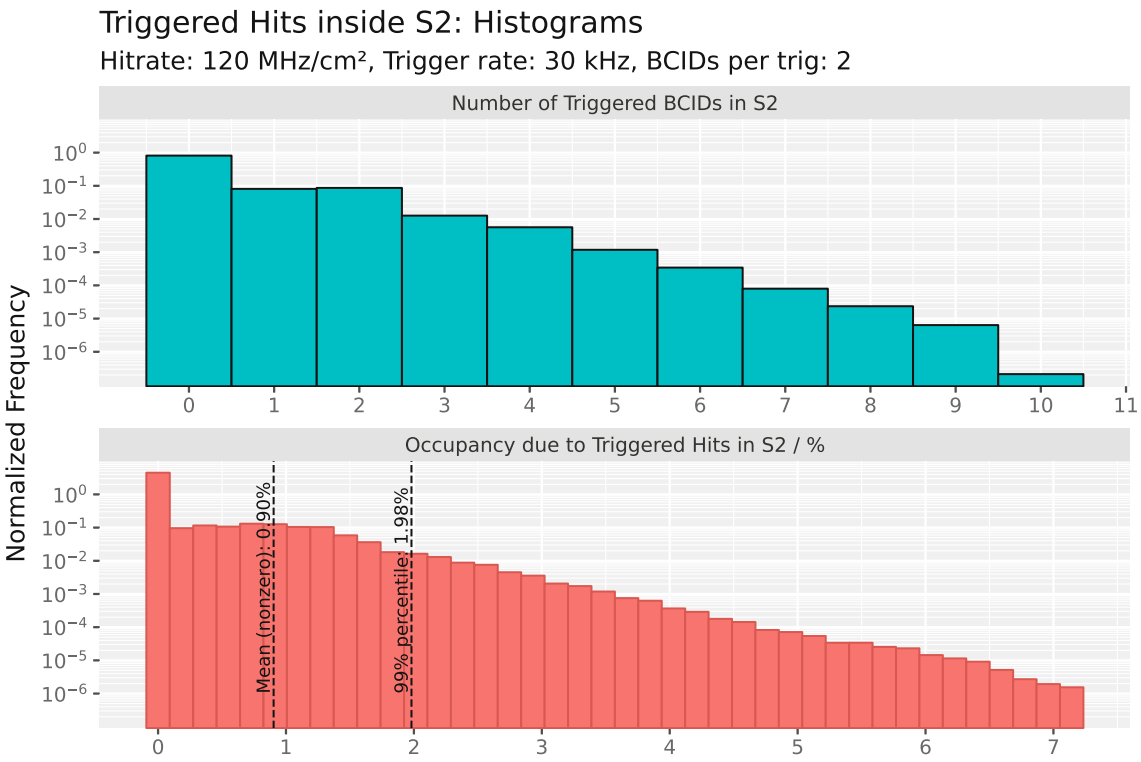


Figure 5.26: Histogram of the number of triggered BCID codes and occupancy for triggered hits in S2.

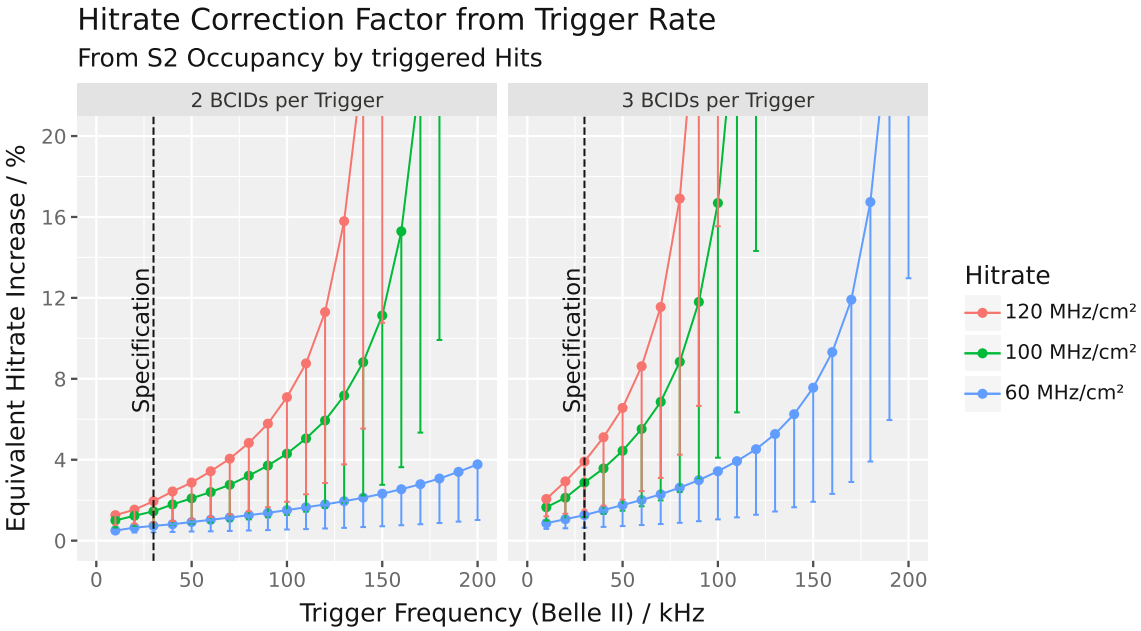


Figure 5.27: Hit-rate increase factor for equivalent TRU load for high trigger rates.

The simulation described above has been performed multiple times for different trigger frequencies and hit-rates. The results are shown in figure 5.27. The 99% percentile has been used as upper limit and main data point, with the error bars spanning down to the average case. The results diverge when the transmission bandwidth of the TXU is approached. The TRU is designed to tolerate more than 10% increased hit-rate with losses in the per-mill range. For the target 120 MHz/cm^2,

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

it is therefore possible to trigger at a frequency of 100 kHz and even more at lower hit-rates. An increased integration time, however, will reduce the allowed trigger rate linearly with the integration time.

Furthermore, above a factor of about 10%, chips will occasionally lose triggers, when the limit of 32 simultaneous trigger in S2 is reached. While the total loss due to these lost triggers is below 1% for the range of the plot, it might cause issues in the data processing. This is because the trigger IDs between chips do not match anymore. Potentially, this can be mitigated by limiting the number of triggers in a sliding time window, but detailed simulations would be needed to confirm this. Another option is to accept the discrepancy of Trigger IDs and maintain an offset for each chip in VTX. Since this is far outside the specifications of OBELIX, further analysis of possible mitigations have not been performed.

5.9.9 Veto Strategy

The Belle II experiment implements a staggered trigger veto after top-up injections as described in section 2.2.6. Directly after the injection, the whole experiment is vetoed for half a millisecond (full veto). After this, only a few microseconds assert the veto whenever the newly injected bunch is passing the detector, the so called gated veto.

The instantaneous hit-rate is less of concern for the TRU as the average hit-rate during the time of one trigger latency time period. Therefore, hit-rate spikes from injection background can be tolerated as long as the total average rate stays below 120 MHz/cm^2. This mode is foreseen for the time of the gated veto.

For the full veto, however, the hit-rate is likely too high to be processed. In principle, the TRU can recover from extreme rate events within a few tens of microseconds if no triggers are sent. Since there is no limitation of event size[4], a trigger during an extreme rate event might initiate a long frame effectively blocking the TXU. The theoretical maximum for such a transmission is still below 1 ms corresponding to all hits in all S2 memory cells to be read out. Potential analog saturation effects in the pixel matrix have not been studied so far, as without a trigger mechanism, TJ-Monopix2 is not capable to handle sufficiently high rates.

However, since the power consumption of the trigger unit increases with the hit-rate, a veto mechanism for the full veto is still desirable to prevent excess power consumption without physics benefit. For this reason, the `TRU_STOP_EOC[1]` register is available. If this it is set, the EoCs will complete any running readout cycle but subsequently stay in idle and therefore following hits will accumulate only in the matrix and not be read out. This can be used as a soft-veto mechanism to reduce power consumption of the periphery during top-up injection. The other parts of the TRU are unaffected by this and any hits in storage are still processed according to potential triggers. Before clearing `TRU_STOP_EOC[1]`, it is advised to clear any hits that accumulated in the matrix by using the `Clear` command or `RESET_PIXEL_MATRIX[1]` register as described in section 5.8.2. Because a pixel initiates the readout after the trailing edge, hits can be lost for a short time period before setting `TRU_STOP_EOC[1]`. The duration of this period is the maximum allowed ToT time. Only when this time has passed after the veto started, the `TRU_STOP_EOC[1]` should be set.

[4]This might be a useful feature for OBELIX-2: a configurable maximum after which hits are dropped quickly from a single trigger to prevent a long transmission and accumulation of data.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

5.10 Transmission Unit

The transmission unit (TXU) packs the data in frames and sends them over the serial output link to the backend system. There are two types of frames for the two types of data the chip can send. The hit frames contain hit data including Le-BCID, Te-BCID, address and auxiliary information. The `Read Register` response frames contain address and value of the requested register.

The serial output is a differential LVDS link sending two bits per input 160 MHz clock cycle ($2 \cdot 169.7$ MHz nominal) i.e. is using DDR. The transmission uses 8b/10b encoding [55]. Therefore, it takes 5 full clock cycles of the 160 MHz clock to transmit one byte of payload data, corresponding to 10 bits after encoding. If there is no data to be transmitted, control symbol K28.1 is transmitted as idle pattern. The K28.1 word allows synchronization of the deserializer in the backend system, i.e. to find the first bits for each byte transmitted.

Data Frame Structure

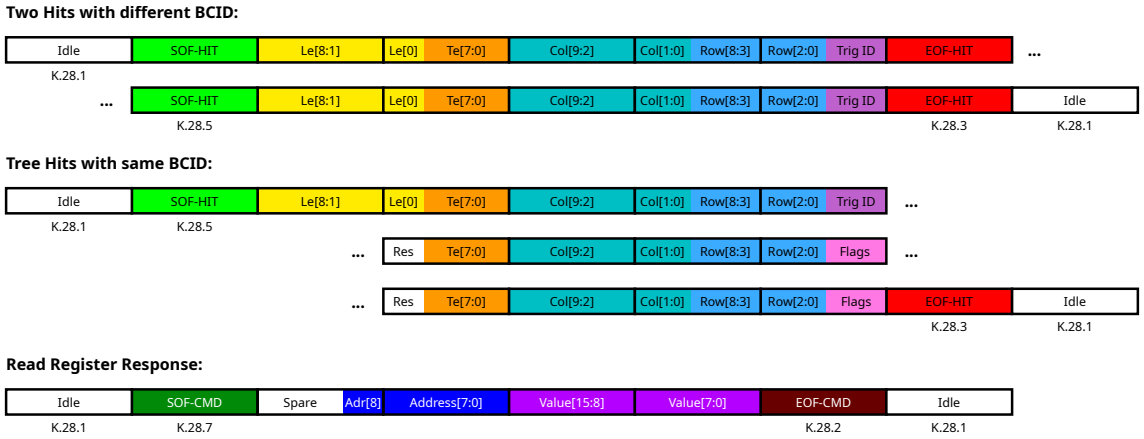


Figure 5.28: Frame structure for hits and read-register responses.

The structure of the two types of frames is shown in figure 5.28. A hit frame is initiated by transmitting control symbol K28.5 (SOF-HIT, start of frame). Then the 9 bit Le-BCID of the first hit is transmitted, followed by the 7 bit Te-BCID and 19 bit address of this hit. If there are subsequent bits to be sent, Te-BCID and address of the subsequent hits are transmitted. To finish the frame, a control symbol K28.3 is transmitted (EOF-HIT, end of frame). The chip assures that the Le-BCID code of all hits within one frame are the same, and will therefore not re-transmit this information. However, the chip may split subsequent hits of the same Le-BCID into more than one frame, if the internal fetching of data takes longer than the transmission. For hits of different Le-BCID, a new frame is always started.

The Le-BCID of 9 bits, Te-BCID of 7 bits and address of 19 bits leave 5 bit empty to fit into the 8 bit words. These spare bits contain auxiliary information and can be configured using the `TXU_AUX_SELECT[2]` register according to table 5.21. Per default, the spare bits of the first hit include the Trigger ID for the frame. Subsequent hits contain the flags of error counters from the SCU. The Trigger ID is redundant with the Le-BCID in principle. However, this allows the detection of desynchronization errors in the backend system and allows quick actions to resynchronize the chip as will be described in section 5.12. Similarly, the error flags are

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

useful for debugging as they allow a quick detection of bit errors in the command stream which might be an indication of suboptimal data link quality.

If a read register command is sent to the chip, it responds with a read-register response frame, each frame is started with K28.7 (SOF-CMD) and ended with K28.2 (EOF-CMD). Inside are two bytes containing the register address (zero padded) and two bytes of register value. This may repeat if multiple registers are requested in quick succession. This type of request is given priority, i.e. hit frames might be delayed when a register read command is sent, however, a frame will never be interrupted by another.

`TXU_AUX_SELECT[2]`	
Value	Description
0	Trigger ID
1	Error flags
2	Reserved
3	**First hit in frame: trigger ID, subsequent hits: error flags**

Table 5.21: Spare bit mode selection depending on `TXU_AUX_SELECT[2]` register.

5.11 Track Trigger Transmission

Currently only the CDC, ECL and KLM detector systems contribute to the Belle II L1 trigger system, the current vertex detectors PXD and SVD do not provide data. The VTX should bring a trigger capability with tracking included, either standalone or together with the CDC. The functionality of this Track Trigger Transmission (TTT) feature of OBELIX has been already described in publication [56], but will also be described in this chapter in greater detail. The VTX track trigger improves the robustness of the trigger system by providing a decision based on a different detector and could improve accuracy and fake-rate. In particular, as the inner layers of the CDC lose in performance at higher backgrounds, the trigger efficiency could be restored with the contribution from VTX and merging with the CDC data. However, currently tentative simulations show that even a VTX only trigger can detect high transverse momentum particles with reasonable resolution and fake rates. A more detailed simulation with more realistic scenarios, as well as a simulation combining the CDC with VTX data is still pending. While the latter brings the largest opportunity in accuracy and fake rate, it is also the most complex option, and even just a simulation will require significant effort. Furthermore, the required calculations are significant, and it is unclear if this is even feasible on FPGAs within the targeted 10 µs L1 trigger latency.

The OBELIX TTT Module

An independent data output is available on the OBELIX chip to allow a low-latency, low-granularity transmission of the hit data in parallel and independent from the regular data output. The module on the OBELIX chip providing this is the TTT module.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

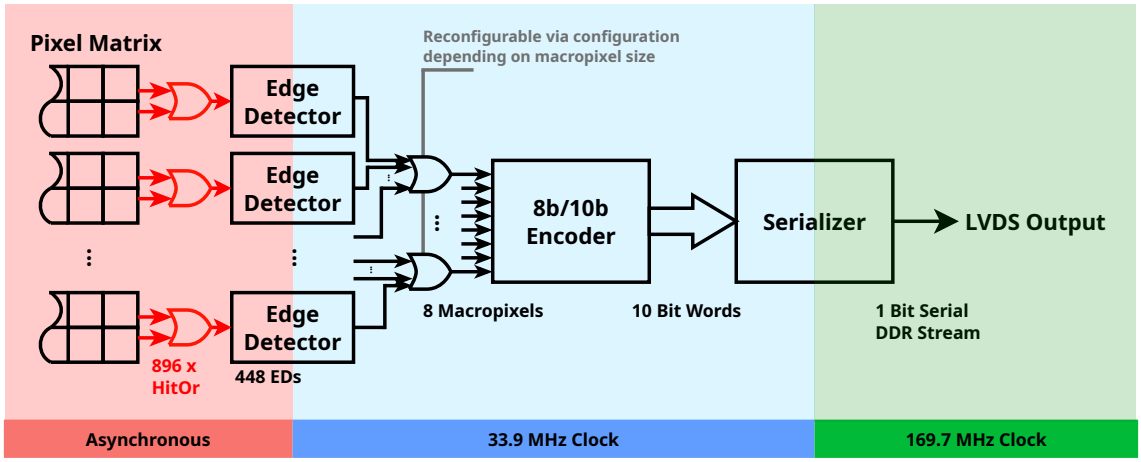


Figure 5.29: Simplified schematic of the TTT module.

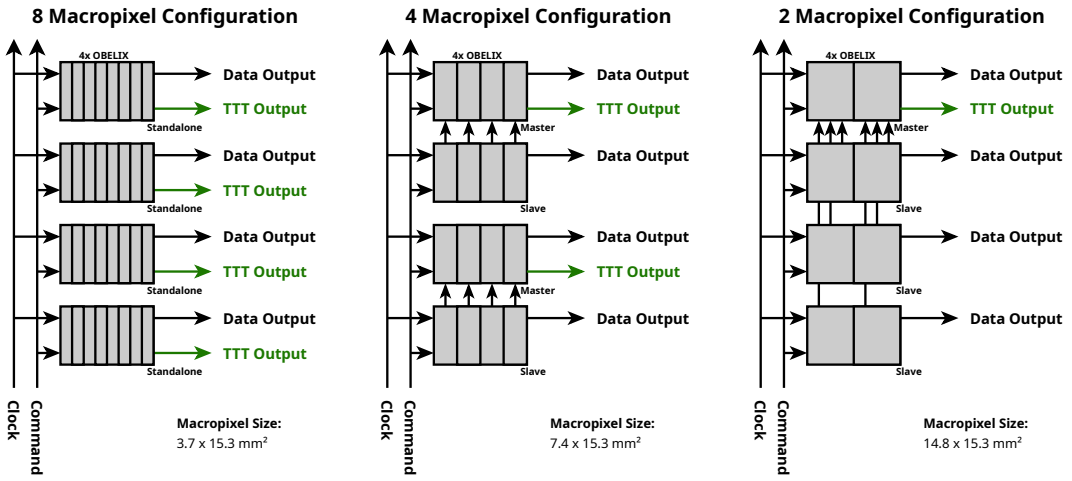
The TTT system inside the OBELIX chip is schematically shown in figure 5.29. It uses the column-wise `HitOr` signals, just like the PTD module. This allows a well-defined and small latency of three cycles of the 32 MHz clock (88.5 ns) after the edge of the `HitOr` plus about 20 ns of average `HitOr` propagation. Due to the organization of the columns, a distinction of the row position is not possible for the TTT system. Multiple columns are subsequently grouped into a so called macropixel. Only binary hit information is available for each macropixel, greatly reducing the bandwidth need for transmission. To detect the rising edges in the `HitOr` signals, an edge detector (ED) clocked from the 32 MHz clock domain is used, each shared between two columns. During the ToT of one hit, the full two columns that share an ED are blind against other incoming hits, causing a dead-time for the duration of the ToT and possibly longer if subsequent particles keep the `HitOr` on logic 1.

The outputs from the EDs are further OR-ed together in groups to form up to 8 macropixel signals. The macropixel signals are subsequently 8b/10b encoded and serialized to a DDR data stream in a similar way as the main data output. If there are no hits in any macropixel, the data is replaced with an idle pattern K28.1 to allow synchronization in the backend system.

Since the time-quantization is done on the 32 MHz clock by the EDs, this clock needs to be synchronous on all chips inside the detector, like the 20 MHz BCID clock for the normal data processing. A suitable `Clear` command is therefore required prior operation, as described in section 5.8.2.

In contrast to the main data output, the TTT data has no internal frame structure. Each 10 bit word stands for its own and can carry the information of 8 bits corresponding to 8 macropixels without zero suppression. This leads to an intrinsic time binning of 29.5 ns and the maximal number of macropixels of 8 per chip. The lack of zero suppression allows activation of an arbitrary number of macropixels within the same TTT time-bin while maintaining a constant latency. However, it limits the number of macropixels at the available bandwidth of the transmission line.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.



(a) (b) (c)

Figure 5.30: Macropixel configurations of the OBELIX TTT module.

To reduce the number of required LVDS links inside VTX, it is possible to increase the size of the macropixels further and share one TTT link across multiple chips. In the schematic 5.29, this is depicted by the gray line between the edge detectors and the 8b/10b encoder. If multiple chips share one transmission line, this is the location where signals are transmitted from chips without a connected TTT output (Slaves) to a neighboring chip that has the `TTT output` connected (Master). The links are point-to-point LVDS links directly connecting the outputs of macropixel-OR gates of the slave chips to the 8b/10b encoder of the master chip in a star topology. Since the transmission from slave to master is synchronous and D-FFs are located on either side of the links, no degradation of hit-timing arises from the transfer from one chip to the next. As the LVDS lines have a length of only a few centimeters and are relatively low speed, it is possible to increase the LVDS termination resistance while reducing the bias current via configuration for those links in order to save power.

Register	Default	Description
`TTT_EN_MASTER[1]`	0	Enable Master/Standalone mode
`TTT_EN_SLAVE[1]`	0	Enable Slave mode
`TTT_PIXELSIZE[2]`	0	Macropixel size selector
`LVDS_TTT_ENABLE[1]`	1	Enable TTT-Ouptut LVDS Driver
`LVDS_TTT_MASTER_RX_ENABLE[6]`	0	Enable Master interchip inputs individually

Table 5.22: Registers related to the TTT module.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

TTT_PIXELSIZE[2]		
Value	Description	Macropixel size
0	8 Macropixel mode (Standalone)	$3.7 \times 15.3\,\mathrm{mm}^2$
1	4 Macropixel mode (1 Master, up to 1 Slave)	$7.4 \times 15.3\,\mathrm{mm}^2$
2	2 Macropixel mode (1 Master, up to 3 Slaves)	$14.8 \times 15.3\,\mathrm{mm}^2$
3	Reserved	—

Table 5.23: Number of Macropixels depending on **TTT_PIXELSIZE[2]** register.

Figure 5.30 shows the three possible configurations of either 8, 4 or 2 macropixels per chip. In 8 macropixel mode, each chip works independently (Standalone) and requires a connected **TTT output**. For 4 and 2 macropixel mode, each master requires a connected **TTT output** while it can be omitted for 1 or 3 slave chips respectively. The configuration registers for the TTT module are summarized in table 5.22. To enable the TTT output for standalone or master operation, it is required to set **TTT_EN_MASTER[1]** and **LVDS_TTT_ENABLE[1]**, while **TTT_EN_SLAVE[1]** needs to be set to enable slave mode. The number of macropixels is defined via register **TTT_PIXELSIZE[2]** according to table 5.23, which also determines if the chip operates in standalone or master mode. The table also shows the resulting sizes of the macropixels.

Since the number of chips in one VTX ladder may not be divisible by 2 or 4, it is possible to have fewer slave chips than theoretically possible on the ends of a ladder. The corresponding inputs on the master chip can be individually disabled via the **LVDS_TTT_MASTER_RX_ENABLE[6]** register.

5.11.1 Time Reconstruction and Resolution

The intrinsic time-binning of 29.5 ns of the **TTT output** is limited by the implementation. Since the **HitOr** lines are used, the signal propagation further limits the time resolution. The three main sources of uncertainty are summarized in table 5.24. The time binning, i.e. the real particle time versus the edge of the 32 MHz clock transition, is a global property of the particle track and therefore common for all layers of VTX. It is equally distributed with a width of 29.5 ns. In contrast, the **HitOr** delay depends on the row-number of the intersection on the individual chips, which is assumed uncorrelated on each layer. The resulting equal distribution depends on the **HitOr** delay which has been measured with TJ-Monopix2. Since the OBELIX chip has fewer rows than TJ-Monopix2, a scaling factor of 464/512 was applied to the measurement for the DC frontend resulting in a width of $464/512 \cdot 46\,\mathrm{ns} \approx 42\,\mathrm{ns}$. All other contributions like time-walk, chip-to-chip variations of timing and clock jitter are collectively modeled with a Gaussian distribution of $\sigma = 5\,\mathrm{ns}$ individually for each chip.

Cause	Chip/Track	Distribution	Scaling
Time binning	Track	equal	width = 29.5 ns
HitOr delay	Chip	equal	width = 42 ns
Timewalk and jitter	Chip	Gaussian	$\sigma = 5\,\mathrm{ns}$

Table 5.24: Contributions of uncertainty for the TTT module and simulation basis.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

Since the uncertainties per chip are larger than the time-binning, a VTX track trigger algorithm needs to search tracks integrating two subsequent time-bins together in a sliding-window fashion. Nevertheless, after identifying a track with a suitable algorithm, the individual time-bins can be used to improve the time resolution of the track. Figure 5.31 illustrates this for three layers in an example event with time flow in the horizontal axis. The three bars correspond to the macropixel activation of the identified track, with the time bins as boxes and the dashed gray lines signifying the center of each time-bin. The real particle time is marked with the left black line. For each layer, arrows from the real particle time signify the local **HitOr** (green) and Timewalk/Jitter (blue) delays, and point to the macropixel time-bins. For clarification, the distribution of **HitOr** delay is centered around zero. In reality, it has a positive offset that shifts the whole reconstruction, which needs to be subtracted.

The Track trigger algorithm can only use the activation of the time-bins as basis for the time reconstruction. A suitable algorithm is the average of the bin-centers. This is depicted by the black arrows pointing from the bin-centers to the TTT-estimate line. The difference in time of the particle and TTT estimate is the timing error of the trigger (trigger jitter). While an average is the simplest case, other algorithms could exploit the fact that the **HitOr** delay is an equal distribution providing a small timing improvement. The possible combinations of activated bins are limited, so even a lookup table with an optimized estimate for each combination is possible. This remains to be simulated when a final number of contributing layers is defined.

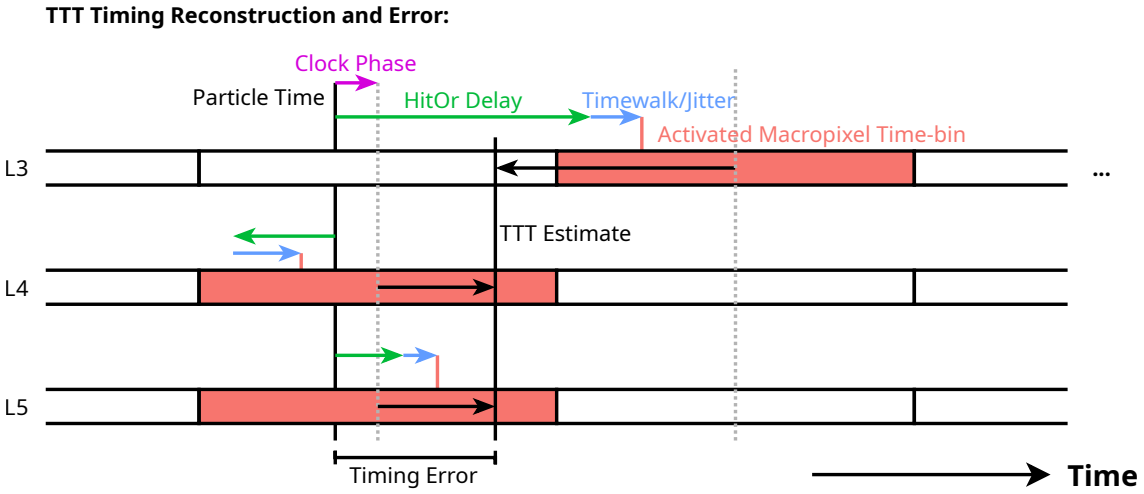


Figure 5.31: Example of time reconstruction from TTT bins.

A simulation of the averaging case was performed using random numbers according to table 5.24. Each simulated particle has a global clock phase and local **HitOr** delays and Timewalk/jitter values independent on each chip. The resulting error distributions are shown in figure 5.32 for different numbers of layers contributing to the track trigger. While the original distribution is shown in gray, the colored Gaussian fits are almost indistinguishable due to the contributions of the multitude of distribution functions. Despite the large uncertainties of the individual layers, already for the 3 layer case a time resolution slightly below 10 ns is reached. This is the target resolution for the L1 trigger with improvements for each layer added.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

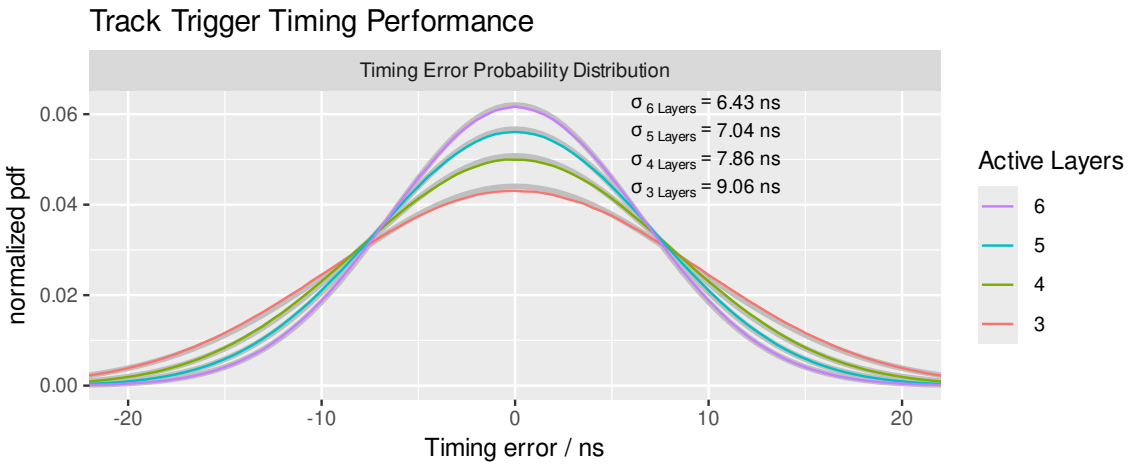


Figure 5.32: Timing resolution simulation of a VTX-only Trigger, assuming different number of layers contributing.

5.11.2 Occupancy and Fake Rate

The hit-rate estimates from section 2.4.1 were used to calculate the occupancies for macropixels for the individual VTX layers of the 6 layer VTX and shown in table 5.25. Different macropixel numbers per chip were assumed for different layers, generally increasing macropixel size with layer number, keeping the number of channels and therefore transmission lines roughly constant starting from L3. The particular choice of macropixel number shown along with the resulting number of transmission lines is an example and depending on the space available for wiring on the flex PCB and cables to the backend, different macropixel sizes could be chosen. Smaller macropixels mean higher spatial resolution and less occupancy, improving track quality while reducing the fake-rate at expense of additional transmission lines required. For example, a suitable candidate would be to use 4 macropixel mode in L5, as this layer has the highest occupancy in oVTX, which would increase the number of TTT lines by 240 lines.

Layer	Macro-pixels	TTT Lines	Macropixel Occupancy			Two-bin Occupancy		
			Sc-1	Sc-2	Sc-3	Sc-1	Sc-2	Sc-3
L1	8	24	20.9%	29.8%	43.3%	37.4%	50.7%	67.9%
L2	8	40	8.9%	14.5%	23.7%	17.0%	27.0%	41.7%
L3	8	360	0.7%	1.1%	1.9%	1.4%	2.2%	3.7%
L4	4	288	0.9%	1.5%	2.5%	1.9%	3.0%	5.0%
L5	2	240	1.1%	1.7%	2.7%	2.2%	3.3%	5.3%
L6	2	360	0.8%	1.1%	1.7%	1.5%	2.2%	3.4%

Table 5.25: Estimated occupancies for TTT macropixels on different layers for three background scenarios.

Since the probability of multiple particles hitting the same macropixel is significant at the inner layers, Poissonian statistics needs to be taken into account in the calculation. The macropixel occupancy is the expected fraction of activated macropixels per time unit. Since the trigger algorithm needs to search for tracks in two subsequent time-bins, the effective integration time is doubled at track finding

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

stage. The corresponding occupancies are shown as 'Two-bin Occupancy' in the same table.

Currently it is unclear if the iVTX layers L1 and L2 will be capable to contribute to the track trigger. However, it has to be noted that as the tracks in the small iVTX volume can be assumed straight for the relevant energies, the search cone for tracks will include only a small number macropixels, possibly just one. In this case, the fake rate can be directly reduced by the occupancy factor, e.g. a 30% occupancy leads to a reduction of fake rate by a factor of 0.3, which can be a significant contribution. But also when a few macropixels are included in the search cone, a notable reduction could be achieved, at least for L2.

To sum up, rigorous simulations of a VTX track trigger are required to estimate fake-rates and efficiency that can be achieved. Depending on the results, either a standalone VTX trigger or a combined VTX+CDC trigger (or both) can be implemented. Such a simulation can also be used to optimize the number of macropixels on each layer. In particular, the fake rate will be directly influenced by the occupancy, and reducing the macropixel size in the outer layers leads to a reduction in occupancy. The trade-off of mechanical integration versus performance is still to be made by an informed decision. The OBELIX chip is flexible with the three options available.

5.12 SEU Protection Measures

The focus of OBELIX-1 is the implementation of the required features, while Single Event Upset (SEU) protection was of lower priority. SEUs are errors in the chip operation due to ionizing particles crossing the transistors of the digital logic. They can affect any logic gate, but most prominently they can change the value of D-FFs. Potentially, this can even leave the logic in states that are unreachable during ordinary operation, making SEU hardening a difficult task.

It is foreseen to test the stability of the chip and potentially implement further protection measures in OBELIX-2. While true mitigation measures on the chip such as triplication typically come along with significant increase in power and area, they are the direct and effective approach to prevent SEUs.

Another option to mitigate the occasional SEU in the detector is a reconfiguration of the affected chip during operation without interrupting data taking. When the logic of OBELIX was developed, this was considered and therefore all necessary measures were taken to support this. Of course, this requires the possibility to recover from SEUs without a hardware reset.

While potential SEU protection is most important for the global control logic, the actual hit data will remain unprotected in any case. Protecting the data path with triplication would likely increase the area of the digital periphery and power consumption by a factor greater than 3. The effect of an SEU in the data path is a transient loss in efficiency and increase in noise rate, as the affected hit changes in properties, for example pixel address. Hit-data gets replaced quickly with new hits, flushing out potential SEUs without long term effects. The efficiency drop is therefore small and does not justify the expenses of data protection.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Continuous Reconfiguration

The global register table of the OBELIX chip is partially protected with hamming codes as described in section 5.8. This protects the value of protected registers against a single SEU per 8 bits. When any register is written, all potential errors are corrected for all protected registers. However, only the registers that have a global effect on the whole chip include the hamming code protection, which accounts for about 13% of the total register table. The other registers control individual columns, DCs or rows. This means, a potential SEU only affects this particular fraction of the pixel matrix. However, if no action is taken, the errors in these registers might accumulate during chip operation and therefore some kind of mitigation is still required.

The OBELIX chip supports `Write Register` commands at any point during operation. It is therefore foreseen to rewrite the unprotected registers in a certain interval. This continuous reconfiguration restores the original value in case on an SEU. All unprotected registers can be changed in value without special measures or adverse effects on data taking. Rewriting all 254 unprotected registers takes about 120 µs per chip or 1.44 ms for all 12 chips sharing one `CMD` line, sequentially. This reconfiguration can either happen at the top-up injections or during data taking, as `Trigger` commands can interrupt `Write Register` commands at any time without the loss of any command. The numbers are summarized in table 5.26.

Another unprotected configuration memory is the in-pixel tuning. Here, an SEU will only affect a single pixel, but it is possible that this pixel becomes noisy as the tuning is changed. In such a case, this has adverse effects on the efficiency of the whole TRG, but only because the buffers are loaded with more noise which still allows for regular operation. Also the in-pixel configuration can be written during data taking, and can therefore included in the continuous reconfiguration procedure. However, since this takes a significantly longer time of about 100 ms per chip while the impact of an SEU is smaller, the time interval between reconfigurations can be larger than for the global registers.

A possible scheme would be to interleave the pixel reconfiguration with multiple repetitions of register reconfiguration. In this way, the registers can be refreshed at a higher frequency while still refreshing the in-pixel configuration.

Operation	Calculation	Time
Single `Write Register`	$10 \cdot 47.2\,\text{ns}$	0.47 µs
Writing all unprotected registers	$254 \cdot 0.47\,\mu\text{s}$	120 µs
Writing pixel config	$208\,800 \cdot 0.47\,\mu\text{s}$	100 ms

Table 5.26: Time durations of various chip operations related to continous reconfiguration.

Resynchronization

A SEU in the command or clock path might trigger a resynchronization of the chip, as described in section 5.7. This will in general cause a mismatch of the BCID code of the chip in respect to the VTX global BCID. The backend system can immediately detect such a failure because each transmitted hit contains the Trigger ID and the corresponding Le-BCID. When a mismatch is detected, the chip needs to undergo a resynchronization procedure. This procedure can be performed while all other chips

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

are taking data and does not require a halt of operation, targeting the specific chip to be resynchronized. The procedure includes several steps:

1. Register rewrite of the protected registers:

   While this is optional, it is a safe precaution to rewrite the hamming protected registers. In particular, the fault could have been caused by a double-SEU that overloaded the hamming code protection.

2. BCID reset:

   This can be best done with a precisely timed `Clear` command to align the BCID with the rest of the detector. Either the current BCID is written with a suitable offset to `BCID_INITIAL_VALUE[9]` followed by a `Clear`, or the backend waits for a BCID overflow and then resets the BCID to zero. While the first approach is faster, the second approach is easier to implement.

3. TTT Resynchronization:

   Since the 32 MHz `clock` domain also needs to be synchronized for all layers using the TTT feature, a `Clear` command needs to be used to align this to the global TTT clock phase.

4. Reset of Trigger ID:

   Since the Trigger ID is used by the backend to check the BCID codes, a harmonized Trigger ID for all chips is beneficial. This is not strictly required, but otherwise the backend needs to keep track of the different Trigger IDs of all chips individually. The easier approach is to reset the Trigger ID with a `Clear` command when the global Trigger ID is zero.

In principle, this is the same procedure as at the startup of data taking. However, it requires more attention to achieve a correct alignment, as it needs to be inserted into the existing run. While requiring some effort for the backend system, such a on-the-fly resynchronization would make the VTX extremely resilient in operation, as nearly all possible faults can be corrected on-the-fly with such a procedure.

5.13 Digital Injection and Test Strategy

In industrial CMOS manufacturing, the automation of testing after production is crucial. Defects, such as short circuits between signals and stuck at 0 or stuck at 1 faults can happen during manufacturing and a reliable detection is required to prevent the shipping of faulty devices. The industry standard to enable such tests is the use of so called scan chains. A scan chain adds a testing mode to the design, that rewires all registers (i.e. D-FF) to one or more large shift registers. This mode, opposed to normal operation is usually called scan mode and a simplified diagram is shown in figure 5.33. The actual functionality of the design is hidden in the clouds of combinatorial circuits. In normal operation, all multiplexers are on the 0 branch utilizing the combinatorial circuits. In scan mode, a bypass connects all registers to form a large shift register allowing access to all D-FFs.

When running a test, test patterns are shifted into these shift registers with scan mode enabled. Subsequently, one clock cycle in normal operation of the device under test (DUT) is executed. This pushes the test pattern through the combinatorial

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

circuits of the DUT and the result is captured by the same or other registers. The result is then shifted out of the DUT in scan mode again, and compared against the expected result. This is repeated multiple times with typically auto-generated and optimized test patterns to obtain very high coverage, typically above 99% of all signals, with as few tests as possible. A big advantage is the short amount of time required for running such a test collection, ideal for industrial mass production, while allowing high coverage of the circuits inside the DUT [57, 58].

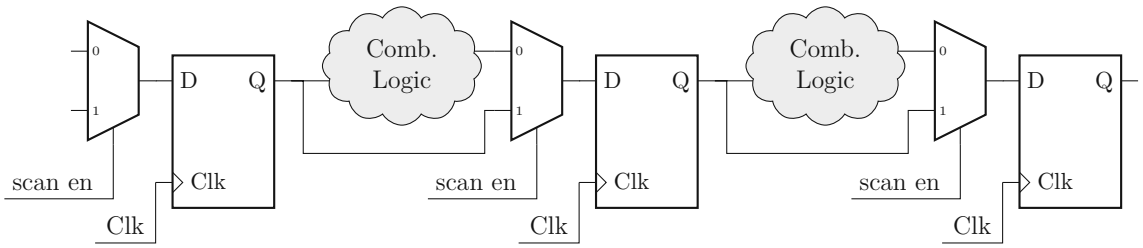


Figure 5.33: Simplified circuit with a scan chain

However, the introduction of the scan mode also comes at a cost. In particular, each register requires an additional multiplexer at its input to rewire it as shift register when scan mode is activated. The combination of DFF and input-multiplexer is typically provided in the PDK as optimized standard cell due to its common usage. For the used Tower PDK, these scan chain versions of flip flops occupy about 25% more area than the non-scannable versions. Furthermore, the introduction of the scan chain occupies a significant amount of routing resources as every register needs a connection to the global scan enable wire, as well as the connections to the previous and next register.

Therefore, this is not feasible in a highly constrained environment such as a DMAPS chip operating at the very limit of the technology. In the particular case of OBELIX, most of the area of the digital periphery is occupied by the TRU or more precise the S2 memories, which have an unusually high density of DFFs compared to typical digital circuits. This is because the primary task of S2 is to store hit information which is facilitated by DFFs and the combinatorial part used for data feed-in, trigger association and data draining is comparatively small. When a scan chain is introduced in this design, the increase in area is significant. Additionally, the power consumption is increased also in normal operation, as longer wires with increased capacitance need to be driven. For these reasons it was decided to not use the scan chain technology in OBELIX.

However, it is still required to verify the chip functionality, but those tests are allowed to take more time as in common industrial settings, since the number of chips is in the order of 1000 for VTX, which is extremely low compared to industry production. A small module is introduced in the digital circuit to aid in the testing of the chip, the digital injection module.

The Digital Injection Module

The digital injection (DI) module is a dedicated module in the digital periphery to allow and simplify certain types of tests. When enabled, it injects predictable pseudo-random hits at the level of the round robin arbiter inside each TRG. It utilizes the same port as the PTD module and the two can therefore not be used at the same time. The multiplexer for this port, as well as the switch-on of the clock

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

of the DI module, are controlled via the `DIGITAL_INJ_EN[1]` register. This register needs to be set to logic 1 before any other action on the DI module is taken. The injection sequence can be started and stopped via the `DIGITAL_INJ_RUN[1]` register. An overview of available registers is shown in table 5.27 [59].

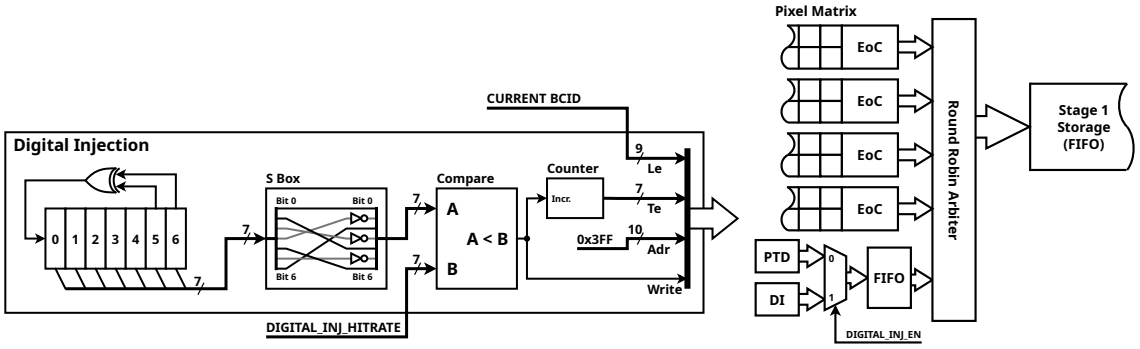


Figure 5.34: Internal structure of the DI module (a) and position in TRG (b)

Register	Default	Description
`DIGITAL_INJ_EN[1]`	0	Global enable of DI module
`DIGITAL_INJ_RUN[1]`	0	Inject hits with DI module
`DIGITAL_INJ_HITRATE[7]`	0	Hitrate selector in units of 4.1 MHz/cm^2
`DIGITAL_INJ_SEED[7]`	0	Manual seed value overwrite
`DIGITAL_INJ_SEED_SELECT[7]`	0	TRG adress, writing triggers seed write

Table 5.27: Registers related to the digital injection module.

At the core, each DI module contains a 7 bit linear feedback shift register (LFSR), operating on the 20 MHz `clock`. This is a shift register, where the shift-in value is calculated from the current state via a linear function, typically XOR gates. The particular function used in OBELIX is an XOR of bits 5 and 6. This choice forms a so called Fibonacci-LFSR [60], which means that the period after the generated numbers are repeating is of the maximum possible length, for this 7 bit implementation 127 cycles. The state of all-zeros is not possible, as the shift register would stay in this state indefinitely.

The value of the LFSR is fed through an S-Box which permutates the individual bits and inverts some of them. This obfuscates the correlation of the generated number sequence originating from the shift nature of the LFSR. After the S-Box, the result is compared with a constant threshold defined in the `DIGITAL_INJ_HITRATE[7]` register. If the S-Box output is smaller than the value set in this register, a hit is injected into the PTD FIFO with a Le-BCID of the current value. A schematic drawing of this is shown in figure 5.34.

This means, the DI hit-rate can be configured to emulate hit-rates in integer multiples of 1/127 times the clock frequency of 21.2 MHz for 4 DCs, resulting in hit-rate steps of 4.1 MHz/cm^2. The address of the hits is row 511 and column 7 plus TRG number, not corresponding to any physical pixels. A simple counter is added to allow tracking of the injected hits and the result is available at the Te-BCID field.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

After a reset, the individual DI modules are initialized with unique seed values. If desired, individual LSFRs can be initialized with custom seeds by first writing the address of the LSFR (i.e. TRG number) to `DIGITAL_INJ_SEED_SELECT[7]` and then the desired seed to `DIGITAL_INJ_SEED[7]`. This can be repeated for some or all of the DI modules and then the `DIGITAL_INJ_RUN[1]` can be set in order to start the injection process. Before any attempts to write a seed value, the `DIGITAL_INJ_EN[1]` bit needs to be set to enable clock distribution in the necessary parts of the chip. A value of `DIGITAL_INJ_SEED_SELECT[7]` = 0x7F is interpreted as broadcast address to set all seeds simultaneously to the same value. Since the lowest rate setting of the DI module corresponds to 4.1 MHz/cm^2, it is only useful to enable the DI module when the TRU is configured in Triggered mode. For the triggerless modes, even the lowest rate setting would quickly overflow all buffers as the transmission bandwidth is insufficient.

BCID	TRG0			TRG1		
	LFSR	S Box	Inject	LFSR	S Box	Inject
0x000	0x01	0x2B	No	0x02	0x3A	No
0x001	0x02	0x3A	No	0x04	0x22	No
0x002	0x04	0x22	No	0x08	0x28	No
0x003	0x08	0x28	No	0x10	0x6A	No
...		...			...	
0x011	0x14	0x62	No	0x28	0x08	Yes
0x012	0x28	0x08	Yes	0x51	0x6F	No
0x013	0x51	0x6F	No	0x23	0x1B	No
...		...			...	

Table 5.28: DI sequence for a few steps, in this case the two sequences have a phase shift of one, `DIGITAL_INJ_HITRATE[7]` = 0x08

Since all LSFRs and S-boxes are the same, all DI modules undergo the exact same sequence. However, due to the different seeds, each DI module has one unique phase from the 127 possible phases, when the default seeds are used. Table 5.28 shows the LFSR and S-Box outputs for the first two TRGs. Since they have the seeds 1 and 2, this means the phase shift of those two modules is exactly 1 clock cycle in this particular case. For the other TRGs, the phase shifts are not ordered like this.

5.13.1 Test Suite

This section contains some preliminary ideas of chip testing at a larger scale. While multiple tests are available and ordered with increasing time, not all of them might be necessary or other tests may be added. In principle, all these tests are possible with a wafer probe-card, e.g. to select wafers that go to iVTX ladder production. The probe card needs to provide the supply voltages, connections to digital IOs and bypass capacitors for LDO operation.

The BDAQ53 board is a suitable control board for these tests and has already been used for this purpose by the RD53 collaboration [61]. The numbers and pass criteria stated here are suitable starting points, but in practice will need to be tested to achieve a certain yield. If the criteria are too strict or too loose, too few or too many chips will pass the tests. It is also possible to grade chips with the metrics presented in the following to allow a finer distinction than pass/fail.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Power on

As first step of testing, power is applied to the LV domain and the bias potential (PSUB/PWELL or HV) is ramped to the target value. The current on LV and bias is monitored and compared to a suitable threshold. Additionally, the output voltage on the LDOs which is available on the external bypass capacitors can be measured to verify correctness. Depending on the ramping speed, this test can be finished in a matter of seconds.

Register read/write

After the initial power up, a simple test to verify correct SCU and CRU operation would write and read back all registers. As a first step, all available registers can be read and checked against the default values. Subsequently, most registers can be written with test patterns that are read back to check for errors. Suitable periodic patterns are 0x0000 (all zeros), 0xFFFF (all ones), 0x5555 and 0xAAAA (binary 0 and 1 alternating). This could be followed by a few iterations of pseudo-random numbers to increase the test coverage. Reading and writing one physical register is possible in less than a microsecond. With 300 physical registers and 4+5 periodic and pseudo-random patterns, this test takes only a few milliseconds. However, in practice the DAQ system and especially the software will not be able to send all commands in the maximum possible speed when a similar framework as in TJ-Monopix2 is used. Nevertheless, this test should only take a few seconds in total and an optimization is not needed.

While the majority of registers can be written without caution, a few registers need to be excluded form the write test. These include the read-only registers, but also the LDO tuning registers `LDO_V*[8]`. Furthermore, switching on the DI module accidentally might swamp the output line with DI hits and changing `CMD_CHSYNC_THR_*_CONF[5]` randomly might desynchronize the chip. [5]

Analog scan in triggerless mode

This test is intended to verify the analog circuitry of the chip. The chip frontend is configured with suitable values and all pixels are injected with a suitable charge with TRU in triggerless mode. As a result, the injected hits are received and statistics are calculated and compared to thresholds to pass the test. A suitable requirement would be to have $> 99\%$ of all pixels with exactly the same number of hits as had been injected. This criterion verifies that most pixels are functional. While allowing for the occasional cosmic ray to pass the matrix, it also detects if the chip contains a large number of noisy pixels and ignores a small number of noisy pixels. Another additional metric would be to expect $> 99\%$ of all received hits to have a ToT in a certain range to verify the BCID distribution of the chip and that pixel amplifiers have correct gain.

An initial analog test should be done with sufficiently large charge injected, so that even pixels with slightly elevated threshold contribute to passing the test. It possibly can be followed by a second test with a charge close to the threshold and more relaxed criteria. An analog scan on the full matrix will take a few minutes on OBELIX, but provides valuable insight in the analog operation.

[5] It is still possible to test a few hand selected values on these registers, but writing arbitrary random numbers is not recommended.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Deterministic TRU test

This test uses the DI module to verify the TRU operation in triggered mode. It configures the TRU with a suitable trigger latency and switches on the DI module, followed by a pattern of `TriggerX` commands. For a given pattern of `TriggerX` commands, the behavior of this test is fully deterministic. Therefore, an exact sequence of DI hits should appear at the chip output. With this test, it is possible to reach all memory cells of S2 inside each TRG, which would otherwise be hard to reach. While the state machines and BCID memories also contain mixed logic 0s and 1s, the address memories inside the S2 cells always contain the DI hit signature and cannot be fully tested.

The longer this test runs, the more coverage it will accumulate. With a trigger rate of 100 kHz at a DI setting of 120 MHz/cm^2, the chip output will transmit about 2.6 MHits/s. Since S2 of OBELIX contains about $32 \cdot 112 = 3584$ storage cells in total (14 336 for S1), a few seconds of run-time will cover all cells multiple times. It is possible that the DAQ software cannot cope with this high number of transmitted hits. In this case a lower trigger rate will decrease the load on the computer while increasing the required runtime linearly. This highly depends on the implementation and amount of optimization of the OBELIX DAQ system.

Test of the TTT module

While the TTT is a feature intended mostly for oVTX, it is necessary to test the correct operation. For this, it is possible to configure the chip for charge injection and inject individual pixels while monitoring the **`TTT output`**. In addition, the FPGA of the BDAQ53 board can emulate the behavior of partner chips to verify the slave and master operation modes of the TTT. Additionally, the synchronization of the 32 MHz **`clock`** can be verified either by analyzing the phase of the **`TTT output`** data stream in the idle pattern (preferred) or by directly accessing the clock over the debug outputs. Even if all columns are injected a few times, this test should only take a few seconds.

Analog scan in triggered mode with DI

This test is a combination of an analog scan with the deterministic TRU test. It configures the TRU for a suitable trigger latency and the analog frontend for operation. Subsequently, the DI module is enabled while injecting charge and sending `TriggerX` commands commands to the chip. In this case, the `TriggerX` commands are targeted to the injected hits. Similarly to the deterministic TRU test, all cells of S2 are tested. In contrast, the hits from the analog injection allow the verification of the address memory and the full functionality of the chip, i.e. this is the test as close to VTX operation as possible. However, since part of the hits, namely the ones originating form the analog injection, go through the analog domain, the strict deterministic behavior is lost for this test. As a result, the received data is not exactly the same from chip to chip but needs statistical analysis like in the analog scan.

TDAC tuning + S-curve (optional)

While the other tests are critical for correct operation, the test of the in-pixel threshold tuning is less important. Tuning all pixels by injecting at the target threshold

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

takes a few minutes. Subsequently it is possible to either run an analog scan slightly above threshold demanding high efficiency or run a full s-curve. Running a full s-curve scan even only for a few charge points will take about an hour on the OBELIX chip. It depends on the degree of automation of the testing if this test is feasible on scale or only performed on selected chips.

Summary

The previous sections proposed a few tests that could be performed on scale. Table 5.29 summarizes the tests along with the chip components tested and an estimate of runtime. Excluding the TDAC tuning and s-curve, the tests take a total of less than 20 minutes. For a full wafer with 32 chips, the total testing time will be about 9 h. This is significant time, but not impossible to handle. Additionally, many types of catastrophic defects such as shorts on supply rails or clock tree will become obvious in the first two tests already, slightly reducing the average time for testing as following tests can be skipped immediately upon detection.

Test	Runtime	Tested components
Power on	30 s	LDOs, shorts on any part of the chip
Register r/w	30 s	SCU, CRU, parts of TXU
Analog scan untrig.	300 s	Pixel matrix, parts of TRU, TXU
Deterministic TRU test	30 s	TRU
TTT module	30 s	TTT
Analog scan trig./DI	600 s	Full chip operation
Total	< 20 min	All critical components

Table 5.29: Summary of the proposed test suite with time estimates.

5.13.2 Applications in Chip Characterization

The DI module can also help to perform certain tests during the initial characterization of the OBELIX chip by artificially increasing the amount of data in the TRU. Two main applications of the DI module in chip testing are presented in the following:

Power Consumption

The power consumption of the OBELIX chip is expected to change with the hit-rate. The main contribution here is the TRU, where more rate translates to more digital activity, increasing the power dissipation. To measure the power consumption at high rate, it is possible to lower the threshold of the pixel amplifiers into the noise floor. This will create a lot of noise-hits in the pixel matrix and is similar to the case of real activity because also the pixel readout is used. However, such measurements with TJ-Monopix2 have proven to be difficult, as even slight variations in temperature vastly change the noise rate. Additionally, the TJ-Monopix2 chip frequently freezes when it is overloaded, complicating the procedure to get a stable noise rate further. This is expected to improve with OBELIX in Triggered mode, however, the error-bars on the rate remain high as it changes at timescales of seconds.

A complementary approach is to use the DI module for this type of measurement. While such a measurement neglects the power consumption in the matrix and pixel readout, it allows a precise control of the equivalent hit-rate in high granularity.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

In particular, this can be used during an X-ray irradiation campaign to measure TID tolerance. In such a campaign, the chip is irradiated and in certain intervals, automated measurements procedures are performed. The DI module enables this, as no user intervention is required and a power measurement procedure can be easily automated for a few different rate equivalents. Since the TID changes the properties of transistors, the power consumption has typically a nontrivial relation with dose, increasing with irradiation up to a peak, followed by gradual reduction in power dissipation above. It is therefore important to characterize the TRU power dissipation at different TID values and only at the final target of the specification.

Test-beam

The second application is during a test-beam campaign. Typical test-beam facilities only provide rates in the kHz range. This is not sufficient to fully test the TRU, as it is designed to operate at much higher rates. While tests with a high rate of actual particles ultimately will need to be performed, they require a lot of preparation. For example, an accelerator providing a sufficient rate is needed and a suitable beam-telescope that also can handle these rates is required.

For this reason, initial tests can use the DI module during a test-beam with the standard, low-rate setup. This does not require any changes to the setup in the hardware or software, except the additional initialization of the DI module by setting the corresponding registers. Since the DI-rate is variable, this allows an easy test of efficiency versus rate within the first test-beam campaigns as soon as the DAQ system runs reliably in Triggered mode.

Since the DI module sends the generated hit-data immediately to the arbiter, this corresponds to hits with zero ToT in terms of timing. This is the worse case, as it maximizes the time-disorder of the hits corresponding to real particles with nonzero ToT. The chance of those hits to be lost is therefore higher and the resulting efficiency measurement is a lower limit on TRU performance.

5.14 Debugging Ports

Various internal signals can be accessed via the debugging port interface. It exposes four channels as CMOS outputs, that can be configured individually via a register configuration, the relevant registers are shown in table 5.30 along with the default values and corresponding default signal selection. Table 5.31 shows all the individual signals that can be selected for each output individually. The first two channels are also available as LVDS outputs, that need to be enabled separately. In TJ-Monopix2, the **HitOr** output is a dedicated LVDS and CMOS output, but for OBELIX, it is available via the debugging interface not present in TJ-Monopix2.

Register	Default		Description
CMOS_DEBUG_SEL0[4]	0xE	sync_locked_out	Signal selector CH0
CMOS_DEBUG_SEL1[4]	0x3	HitOr	Signal selector CH1
CMOS_DEBUG_SEL2[4]	0x6	20 MHz clock	Signal selector CH2
CMOS_DEBUG_SEL3[4]	0xD	CMD_SYNCLOSS_FLAG	Signal selector CH3
LVDS_DEBUG_ENABLE[2]	0x0	Off	Enable LVDS Driver for CH0, CH1

Table 5.30: Registers related to the debugging interface.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

The defaults are chosen to allow debugging of chips that do not respond to commands. The **sync_locked_out** and **CMD_SYNCLOSS_FLAG[1]** signals can be used to determine if the chip successfully locks onto the command stream and if the lock is stable. The **HitOr** output can be used together with a radioactive source to detect weather the analog parts of the chip are functional in principle, and the exposed 20 MHz clock is crucial for all parts of the OBELIX chip.

CMOS_DEBUG_SELx[4]		
Value	Signal	Description
0x0	**tru_set_bcid**	Pulses when the BCID is set
0x1	**bcid_nonzero**	Low when the BCID is 0x000
0x2	**trigger**	Pulses at the delayed BCID of a trigger
0x3	**HitOr**	**HitOr** output
0x4	**injection_pulse**	Output of the injection pulse circuit
0x5	Reserved	Unused, constant 0
0x6	20 MHz clock	Internal 20 MHz clock
0x7	32 MHz clock	Internal 32 MHz clock
0x8	**bcid_current_bin[0]**	Current BCID bit 0 (in binary)
0x9	**bcid_current_bin[1]**	Current BCID bit 1 (in binary)
0xA	**bcid_current_bin[2]**	Current BCID bit 2 (in binary)
0xB	**bcid_current_bin[3]**	Current BCID bit 3 (in binary)
0xC	**bcid_current_bin[4]**	Current BCID bit 4 (in binary)
0xD	**CMD_SYNCLOSS_FLAG[1]**	Flag for nonzero SYNCLOSS counter
0xE	**sync_locked_out**	Flag for successful lock to CMD stream
0xF	Off	Constant logic 0

Table 5.31: Debugging signals for different settings of the **CMOS_DEBUG_SELx[4]** register.

The debugging interface exposes additional signals to debug timing and synchronization. The lower 5 bits of the current BCID counter in binary format can be accessed, as well as a flag indicating the overrun of the BCID counter (**bcid_nonzero**). When a trigger command is sent to the chip, the **trigger** signal pulses exactly at **bcid_current_bin[9]** minus the configured trigger latency. This can be used relative to the **bcid_nonzero** signal to measure which BCID the chip should have triggered on. The **tru_set_bcid** pulses when the BCID is set to a new value via a command. Finally, the generated clocks are exposed to allow the observation of the phase synchronization, either in single-chip or multi-chip configurations.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

5.15 Power Budget

In this chapter, simulations and measurements from TJ-Monopix2 are used estimate the total power consumption of OBELIX.

Analog domain

The analog power consumption is dominated by the `IBIAS[8]` value. As this current is consumed by the amplifier circuit in each pixel, it scales with the total number of pixels and therefore area. The other contribution is a mix of all other bias currents with the static power from the DACs itself. As the DAC modules supply 32 columns each, this contribution scales with the number of columns. The values from measurements in section 4.8 are used here to extrapolate for OBELIX in table 5.32.

Contribution	TJ-Monopix2 Measured	Scaling	OBELIX Extrapolated	OBELIX Simulation
`IBIAS[8]` = 100	159 mA	$896 \cdot 464/512^2$	252 mA	222 mA
Analog other	14 mA	896/512	25 mA	18 mA

Table 5.32: OBELIX power consumption of the analog domain from measurement.

BCID Distribution

With the extracted BCID capacitance from section 4.8 of TJ-Monopix2 of $C_{BCID}^{TJ} = 8.99\,\text{pF}$, the capacitance of the OBELIX BCID lines can be calculated with a scaling factor of 464/512 for the length of the lines. With a clock frequency of $f_{BCID} = 21.2\,\text{MHz}$, a voltage swing of $V_{BCID} = 0.74\,\text{V}$ (see section 5.5) and 896 columns, the current due to BCID distribution I_{BCID} can be calculated as:

$$I_{BCID} = f_{BCID} \cdot C_{BCID}^{OBELIX} \cdot V_{BCID} \cdot n_{columns} = 115\,\text{mA}. \tag{5.4}$$

Digital Periphery

The power consumption of the digital periphery depends on the hit-rate and has been simulated in section 5.9.6. The power consumption of the iVTX (PTD off) with the hit-rate f_{Hit} resulted in:

$$I_{Dig.Periph.}^{iVTX} = 50.8\,\text{mA} + 1.43\,\text{mA} \cdot \text{cm}^2/\text{MHz} \cdot f_{Hit}, \tag{5.5}$$

and

$$I_{Dig.Periph.}^{oVTX} = 86.5\,\text{mA} + 2.46\,\text{mA} \cdot \text{cm}^2/\text{MHz} \cdot f_{Hit}, \tag{5.6}$$

for oVTX with the PTD module enabled. As hitrates, the values from the pessimistic simulation presented in section 2.4.1 of 34.0 MHz/cm^2 for L1 and 1.12 MHz/cm^2 for L3 were used to estimate a supply current of 99 mA and 89 mA for iVTX and oVTX respectively. For the power consumption the average hit-rate rate over hundreds of milliseconds needs to be taken into account. For this reason, the numbers from the VTX simulation are taken directly and not the design maximum rate of 120 MHz/cm^2 that may be reached during top-up injections for only a few milliseconds. In addition, a current of 43 mA is consumed by the LVDS drivers and receivers, as well as for driving the readout of the matrix.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Summary

The total power consumption of the OBELIX chip can be calculated as product of the total current with the input supply voltage of the LDOs. With a total chip area of 5.67 cm^2 the power density is calculated and shown in table 5.33.

	Supply Current	V_{in} 2 V	V_{in} 3 V
iVTX	436 mA	154 mW/cm^2	231 mW/cm^2
oVTX	426 mA	150 mW/cm^2	225 mW/cm^2

Table 5.33: Total OBELIX power consumption.

5.15.1 Digital Periphery: Power Simulation Scaling

To achieve the best accuracy of the power estimation of the digital logic, netlist simulations were performed after the last stage of implementation to obtain the results presented in section 5.9.6. At this stage, the correct capacitance values of signals were available, and the simulation provides toggle rates of each signal at a realistic scenario. The netlist simulations were performed at different hit-rates and with he PTD module disabled and enabled to allow for a more detailed analysis and optimization of the design.

Running an implementation flow of the full chip with 448 DCs takes 20 hours on the server available. The netlist simulation takes additional 55 hours and the power analysis 10 hours. At the prototyping stage, it is unfeasible to run this for each design iteration step. The code was written in a size-agnostic way, allowing an arbitrary number of DCs, as long as the number is divisible by 16. A full run with 32 DCs takes 3 hours for implementation, simulation and power analysis in total. To extrapolate the power P_{448}^{tot} to the target 448 DC version, following approximation was used:

$$P_{448}^{tot} = P_n^{tot} + P_n^{TRU} \cdot (448/n - 1). \tag{5.7}$$

Here, n is the number of DCs used in the particular case, P_n^{tot} is that total power estimate for the implemented design and P_n^{TRU} is only the contribution of the TRU to the total power. Therefore, this formula assumes a linear increase of the TRU power with the number of DCs and no effect on other parts of the chip. Due to the composition of TRU of individual TRGs of the same internal structure, this approximation works. The main source of error is the effort needed to reach timing closure. The larger the design, the more difficult it is to fix all timing problems for the implementation tools, and therefore more power is needed to drive certain signals sufficiently fast. This means the extrapolation is generally a slight underestimate. This mainly affects the clock tree as high fanout, fast signal.

A particularly timing critical object is the PTD, and a slight positive slope is visible in the power estimate graph in figure 5.35 only when the PTD module is enabled. The slope is independent of the hit-rate, as expected for a contribution of the clock tree. In general, it is visible that the extrapolated power is mostly independent on the particular number of DCs used, which qualifies this method of power estimation.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

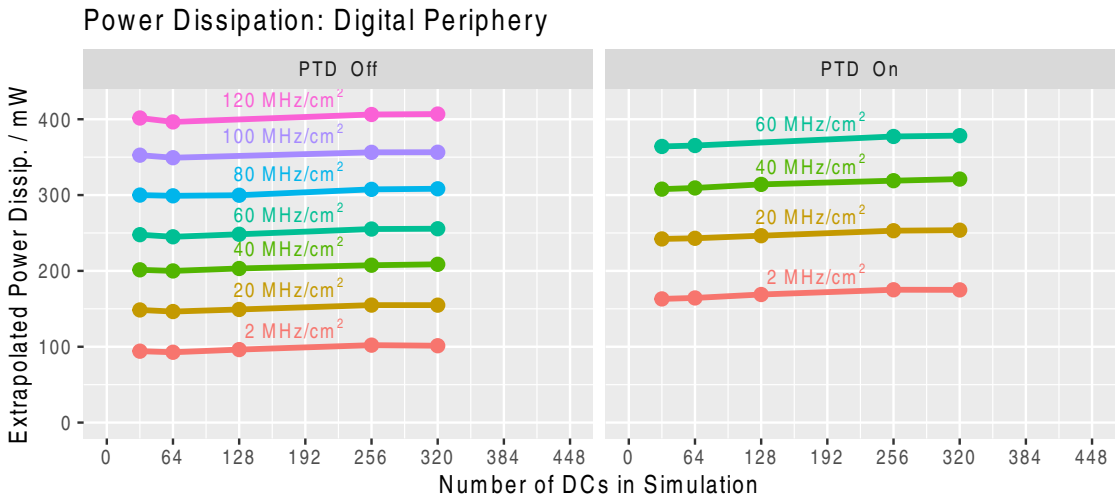


Figure 5.35: Power dissipation of the digital periphery for implementations with different number of DCs.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

Chapter 6

Conclusion

The Belle II VTX upgrade planned for 2032 will improve performance while increasing tolerance to beam background due to the fully pixelated design. Beam background is expected to increase to unprecedented levels when the luminosity of the SuperKEKB collider is increased to $6 \times 10^{35}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$ as planned.

The small collection electrode DMAPS chip OBELIX is tailored to the VTX upgrade and brings a variety of features for integration into Belle II. It is based on the TJ-Monopix2 chip, particularly inheriting the pixel matrix with only slight modifications, while the periphery of the OBELIX chip is heavily adapted for integration into Belle II. The whole OBELIX design was presented in technical detail with a special focus on the newly implemented features.

In particular, a trigger memory algorithm has been developed and implemented by the author for compatibility with the Belle II Level 1 trigger system, allowing a latency of 10 µs at hit-rates up to $120\,\mathrm{MHz/cm}^2$. Performance has been evaluated with realistic data patterns.

A precision timing module allows for increased timing resolution of the outer VTX layers to 3 ns, proposed, conceptualized and implemented by the author. This allows for the determination of the T0 event time of Belle II by VTX while enabling rejection of off-time tracks caused by beam background. Timing measurements on the TJ-Monopix2 were presented to show feasibility and propose the correction methods required for achieving this resolution.

The OBELIX chip also allows contribution to the Level 1 trigger system, enabling a track trigger after the VTX upgrade. This module, proposed and implemented by the author, allows the transmission of low-granularity hit-data at a low latency in parallel to normal chip operation. The expected occupancies and the time reconstruction for a track trigger have been presented and simulated. A time resolution below 10 ns is possible as required by the Level 1 trigger. Further simulations are needed to estimate trigger efficiency and fake rate of the proposed design.

A variety of measurements of the TJ-Monopix2 chip have been performed by the author and the whole upgrade group and results have been presented. The suitability of the pixel matrix implementation at the VTX design fluence of $5 \times 10^{14}\,\mathrm{n_{eq}/cm}^2$ has been shown. A high particle detection efficiency was reached even above room temperature to guide the development of the VTX cooling system. A variety of timing measurements were performed and analyzed by the author to obtain insight into the performance and tune the simulations of the OBELIX design.

While not completed, the OBELIX development is reaching the final phase. Verification and physical implementation is progressing, and the chip is planned to go into production towards the end of 2025.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Bibliography

[1] O. Brüning et al. “LHC Design Report”. In: CERN Yellow Reports: Monographs (2004). DOI: 10.5170/CERN-2004-003-V-1.

[2] The ATLAS Collaboration. *Expected Performance of the ATLAS Experiment - Detector, Trigger and Physics*. 2009. URL: https://arxiv.org/abs/0901.0512.

[3] The CMS Collaboration. “The CMS experiment at the CERN LHC”. In: *Journal of Instrumentation* 3.08 (Aug. 2008), S08004. DOI: 10.1088/1748-0221/3/08/S08004.

[4] T. Abe et al. “Belle II Technical Design Report”. In: (Nov. 2010). URL: https://arxiv.org/abs/1011.0352.

[5] M. Benedikt et al. *Future Circular Collider Feasibility Study Report Volume 1*. Tech. rep. Geneva: CERN, 2025. DOI: 10.17181/CERN.9DKX.TDH9.

[6] A. Kazunori et al. “SuperKEKB collider”. In: *Nucl. Instrum. Meth. A* 907 (2018), pp. 188–199. ISSN: 0168-9002. DOI: https://doi.org/10.1016/j.nima.2018.08.017.

[7] A. Abashian et al. “The Belle detector”. In: *Nucl. Instrum. Meth. A* 479.1 (2002). Detectors for Asymmetric B-factories, pp. 117–232. ISSN: 0168-9002. DOI: https://doi.org/10.1016/S0168-9002(01)02013-7.

[8] S. Kurokawa et al. “Overview of the KEKB accelerators”. In: *Nucl. Instrum. Meth. A* 499.1 (2003). KEK-B: The KEK B-factory, pp. 1–7. ISSN: 0168-9002. DOI: https://doi.org/10.1016/S0168-9002(02)01771-0.

[9] H. Aihara et al. *The Belle II Detector Upgrades Framework Conceptual Design Report*. 2024. DOI: https://doi.org/10.48550/arXiv.2406.19421.

[10] A. Natochii et al. “Measured and projected beam backgrounds in the Belle II experiment at the SuperKEKB collider”. In: *Nucl. Instrum. Meth. A* 1055 (2023), p. 168550. ISSN: 0168-9002. DOI: https://doi.org/10.1016/j.nima.2023.168550.

[11] M.J. French et al. “Design and results from the APV25, a deep sub-micron CMOS front-end chip for the CMS tracker”. In: *Nucl. Instrum. Meth. A* 466.2 (2001), pp. 359–365. ISSN: 0168-9002. DOI: https://doi.org/10.1016/S0168-9002(01)00589-7.

[12] P. Ahlburg et al. “The new and complete Belle II DEPFET pixel detector: Commissioning and previous operational experience”. In: *Nucl. Instrum. Meth. A* 1068 (2024), p. 169763. ISSN: 0168-9002. DOI: https://doi.org/10.1016/j.nima.2024.169763.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

[13] A. Gabrielli et al. “The silicon vertex detector of the Belle II experiment”. In: *Nucl. Instrum. Meth. A* 1064 (2024), p. 169401. ISSN: 0168-9002. DOI: https://doi.org/10.1016/j.nima.2024.169401.

[14] L Andricek et al. “Initial requirements for an upgraded VXD in Belle II”. In: *Belle II VXD open workshop CERN-Geneva*. 2019. URL: https://indico.cern.ch/event/810687/attachments/1827433/3072784/BELLE2-NOTE-TE-2019-011.pdf.

[15] T.V. Dong et al. “Calibration and alignment of the Belle II central drift chamber”. In: *Nucl. Instrum. Meth. A* 930 (2019), pp. 132–141. ISSN: 0168-9002. DOI: https://doi.org/10.1016/j.nima.2019.03.072.

[16] Y. Iwasaki et al. “Level 1 trigger system for the Belle II experiment”. In: (2010), pp. 1–9. DOI: 10.1109/RTC.2010.5750454.

[17] H. Nakayama. “Overview of BELLE II injection experience and issues”. In: *KEK-CERN mini workshop* Presentation (2025). URL: https://indico.cern.ch/event/1542123/contributions/6490283/attachments/3062505/5416221/20250508_KEK-CERN_injection.pdf.

[18] H. Bae et al. “A New Scheme for the Injection Veto System of the Belle II Experiment”. In: (2024), pp. 1–1. DOI: 10.1109/NSS/MIC/RTSD57108.2024.10658627.

[19] A. Piwinski. *The Touschek Effect in Strong Focusing Storage Rings*. 1999. URL: https://arxiv.org/abs/physics/9903034.

[20] Tower Semiconductor. *imaging processes*. 2024. URL: https://towersemi.com/technology/cmos_image_sensor/ (visited on 03/30/2024).

[21] S. Bettarini et al. “The DMAPS upgrade of the Belle II vertex detector”. In: *JINST* 19.02 (2024), p. C02060. DOI: 10.1088/1748-0221/19/02/C02060.

[22] B. Abelev et al. “Technical Design Report for the Upgrade of the ALICE Inner Tracking System”. In: *Journal of Physics G: Nuclear and Particle Physics* 41.8 (July 2014), p. 087002. DOI: 10.1088/0954-3899/41/8/087002.

[23] H. Kolanoski et al. *Particle Detectors: Fundamentals and Applications*. Oxford University Press, June 2020. ISBN: 9780198858362. DOI: 10.1093/oso/9780198858362.001.0001.

[24] S. M. Sze et al. *Physics of Semiconductor Devices, 3rd ed.* John Wiley and Sons, Inc., 2007. ISBN: 9783540283324.

[25] D. Schroder. *Semiconductor Material and Device Characterisazion*. John Wiley and Sons, Inc., 2005. ISBN: 9780471749097.

[26] F. Hartmann. *Evolution of Silicon Sensor Technology in Particle Physics*. Springer, 2024. ISBN: 978-3-031-59720-6. DOI: https://doi.org/10.1007/978-3-031-59720-6.

[27] C. Patrignani et al. “Particle Data Group”. In: *Chin. Phys. C, 40, 100001 (2016) and 2017 update* (2017). URL: https://pdg.lbl.gov/2017/html/authors_2017.html.

[28] M. Moll. “Radiation Damage in Silicon Particle Detectors”. Universität Hamburg, 1999.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

[29] *The Phase-2 Upgrade of the CMS Tracker*. Tech. rep. Geneva: CERN, 2017. DOI: 10.17181/CERN.QZ28.FLHW. URL: https://cds.cern.ch/record/2272264.

[30] *Technical Design Report for the ATLAS Inner Tracker Pixel Detector*. Tech. rep. Geneva: CERN, 2017. DOI: 10.17181/CERN.FOZZ.ZP3Q. URL: https://cds.cern.ch/record/2285585.

[31] Institut Pluridisciplinaire Hubert Curien IN2P3-CNRS / UdS Strasbourg. *MIMOSA26 User Manual*. 2011. URL: https://indico.cern.ch/event/231527/contributions/488098/attachments/384366/534629/M26_UserManual_light.pdf.

[32] J. Baudot et al. "First test results Of MIMOSA-26, a fast CMOS sensor with integrated zero suppression and digitized output". In: Dec. 2009, pp. 1169–1173. DOI: 10.1109/NSSMIC.2009.5402399.

[33] P. Sieberer. " Monolithic active pixel sensors for high rate tracking detectors". TU Wien, 2023. DOI: https://doi.org/10.34726/hss.2023.86681.

[34] C. Bespin. "Characterization of the TJ-Monopix2 Depleted Monolithic Active Pixel Sensor for High-Energy Physics Experiments". Rheinische Friedrich-Wilhelms-Universität Bonn, 2024. DOI: https://doi.org/10.48565/bonndoc-427.

[35] K. Moustakas. "Design and Development of Depleted Monolithic Active Pixel Sensors with Small Collection Electrode for High-Radiation Applications". Rheinische Friedrich-Wilhelms-Universität Bonn, 2021. URL: https://hdl.handle.net/20.500.11811/9315.

[36] R. Diener et al. "The DESY II test beam facility". In: *Nucl. Instrum. Meth. A* 922 (2019), pp. 265–286. ISSN: 0168-9002. DOI: https://doi.org/10.1016/j.nima.2018.11.133.

[37] F. Ulrich-Pur et al. "Commissioning of low particle flux for proton beams at MedAustron". In: *Nucl. Instrum. Meth. A* 1010 (2021), p. 165570. ISSN: 0168-9002. DOI: https://doi.org/10.1016/j.nima.2021.165570.

[38] EBG MedAustron. *IR1 Instructions for Use*. Version 11. 2024.

[39] L. Huth et al. "TelePix2: Full scale fast region of interest trigger and timing for the EUDET-style telescopes at the DESY II test beam facility". In: *Nucl. Instrum. Meth. A* 1080 (2025), p. 170720. ISSN: 0168-9002. DOI: https://doi.org/10.1016/j.nima.2025.170720.

[40] Y. Liu et al. "EUDAQ2—A flexible data acquisition software framework for common test beams". In: *Journal of Instrumentation* 14.10 (Oct. 2019), P10033–P10033. ISSN: 1748-0221. DOI: 10.1088/1748-0221/14/10/p10033.

[41] P. Baesso et al. "The AIDA-2020 TLU: a flexible trigger logic unit for test beam facilities". In: *Journal of Instrumentation* 14.09 (Sept. 2019), P09019. DOI: 10.1088/1748-0221/14/09/P09019.

[42] J. Kröger et al. "Efficient Analysis of Test-beam Data with the Corryvreckan Framework". In: *JPS Conf. Proc.* 34 (2021), p. 010024. DOI: 10.7566/JPSCP.34.010024.

[43] CYRCé. *Irradiation facility*. 2024. URL: https://cyrce.fr/en/irradiation-2/ (visited on 03/30/2024).

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

[44] RARiS. *Research Center for Accelerator and Radioisotope Science, Tohoku University*. 2025. URL: https://www.raris.tohoku.ac.jp/eng/ (visited on 07/08/2025).

[45] M. Mager. *The Telescope Optimiser V1.0*. https://mmager.web.cern.ch/telescope/tracking.html. Accessed: 2025-06-30.

[46] *Basil - a modular data acquisition system and system testing framework in Python and Verilog*. https://github.com/SiLab-Bonn/basil. Accessed: 2024-12-05.

[47] M. Babeluk et al. "The DMAPS upgrade of the Belle II vertex detector". In: *Nucl. Instrum. Meth. A* 1064 (2024), p. 169428. DOI: 10.1016/j.nima.2024.169428.

[48] L. Schall et al. "Cross-talk of a large-scale depleted monolithic active pixel sensor (DMAPS) in 180nm CMOS technology". In: *NIMA* 1064 (2024), p. 169381. DOI: 10.1016/j.nima.2024.169381.

[49] D. E. Norgaad. *ELECTRONIC VOLTAGE REGULATOR*. US Patent 2,219,195. Oct. 1940. URL: http://www.google.it/patents/US4741207.

[50] *Integrated Circuit and System Design. Power and Timing Modeling, Optimization and Simulation - 16th International Workshop, PATMOS 2006, Proceedings*. 2006. DOI: https://doi.org/10.1007/11847083.

[51] Maurice Garcia-Sciveres. *The RD53A Integrated Circuit*. Tech. rep. Geneva: CERN, 2017. URL: https://cds.cern.ch/record/2287593.

[52] Y. Zhao. "Radiation hardened design of CMOS pixel sensor for the micro-vertex detector of the CBM experiment". PhD thesis. Université de Strasbourg, 2021. URL: https://theses.hal.science/tel-03284188v1.

[53] B. Arnoldi-Meadows et al. "Results from single event effect tests with MIMOSIS-1". In: *Journal of Instrumentation* 18.04 (2023). DOI: 10.1088/1748-0221/18/04/C04002.

[54] I. Perić et al. "The FEI3 readout chip for the ATLAS pixel detector". In: *NIMA* 565.1 (2006), pp. 178–187. ISSN: 0168-9002. DOI: https://doi.org/10.1016/j.nima.2006.05.032.

[55] A. X. Widmer and P. A. Franaszek. "A DC-Balanced, Partitioned-Block, 8B/10B Transmission Code". In: *IBM Journal of Research and Development* 27.5 (1983), pp. 440–451. DOI: 10.1147/rd.275.0440.

[56] M. Babeluk et al. "The OBELIX chip for the Belle II VTX upgrade". In: *Nucl. Instrum. Meth. A* 1067 (2024), p. 169659. DOI: 10.1016/j.nima.2024.169659.

[57] S. Huhn et al. *Design for Testability, Debug and Reliability: Next Generation Measures Using Formal Techniques*. Springer International Publishing AG, 2021. ISBN: 3030692086.

[58] L. Wang et al. *VLSI test principles and architectures: design for testability*. Elsevier Morgan Kaufmann Publishers, 2006. ISBN: 128096684X.

[59] M. Babeluk et al. "OBELIX: A monolithic pixel sensor with triggered readout for the Belle II upgrade". In: *Nucl. Instrum. Meth. A* 1080 (2025), p. 170654. DOI: 10.1016/j.nima.2025.170654.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

[60] Z. Zulfikar et al. “FPGA-based Design System for a Two-Segment Fibonacci LFSR Random Number Generator”. In: *International Journal of Electrical and Computer Engineering* 7.4 (2017), pp. 1882–1891. DOI: 10.11591/IJECE.V7I4.PP1882-1891.

[61] M. Standke et al. “RD53B Wafer Testing for the ATLAS ITk Pixel Detector”. In: 2374.1 (2022), p. 012087. DOI: https://doi.org/10.1088/1742-6596/2374/1/012087.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

Glossary

ADC Analog-to-Digital Converter. 85, 89

BCID Initially: Bunch Crossing ID, for Belle II it is just a timestamp. 48, 68, 70, 85, 91, 93, 94, 97, 98, 101, 102, 105, 108, 109, 112, 116, 119, 120, 124, 130, 134, 135, 139

CDC Central Drift Chamber, gaseous central tracking detector of Belle II.. 14, 16, 17, 20, 24, 123

CMOS Complementary metal–oxide–semiconductor, in this case: single ended chip input or output with logic levels corresponding to supply voltage (0 and 1.8 V for logic 0 and 1 respectively). 49, 91, 138

CMOS Complementary metal–oxide–semiconductor, circuit technology used in nearly all current microchip technologies. 131

CRU Control Unit, responsible for command processing and global configuration.. 85, 95

D-FF D-Flipflop: Each rising clock edge, the D-input is stored and becomes the new output. 86, 94, 116, 125, 129, 131

DAC Digital-to-Analog Converter. 47, 51, 89

DC Double Column, two columns that share one readout BUS and logically form one unit.. 102, 105, 107, 141

DDR Doube Data rate: making use of rising and falling clock edges to double the frequency. 92, 100, 122, 124

DMAPS Depleted Monolithic Active Pixel Sensor. 46, 84

DUT Device Under Test. 54, 56

EoC End of Column Controller, circuit for reading out pixel data from the matrix.. 85, 102, 107, 108

FIFO First-in First-out memory, acting like a queue. 106, 108, 114, 118, 133

HER High energy ring, storage ring for electrons in Belle II.. 10, 18

IR Interaction region, volume in the center of Belle II where collisions are occuring.. 11, 12, 17, 18, 20, 23

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

LDO Low Dropout Regulator, voltage regulators for LV supply.. 86, 89, 91

Le-BCID Leading edge BCID code, start of the ToT period. 48, 106, 108, 109, 118, 122, 130, 133

LER Low energy ring, storage ring for positrons in Belle II.. 10, 18

LSB Least significant bit. 112

LVDS Low Voltage Differential Signaling, a hardware standard for high speed differential digital transmission between chips. 49, 92, 122, 125, 138

MSB Most significant bit. 91

PTD Peripheral Time to Digital Converter, precision timing module and part of each TRG.. 112, 118, 141

PVT Process, Voltage and Temperature variations, the main causes for variations of timing in digital circuits or performance in analog circuits. 50, 65

PXD Pixel Detector, innermost part of the current Belle II vertex detector.. 13, 17, 123

S-Box Substitution Box, a term commonly used in cryptography for a bijective function, typically rearranging and possibly inverting individual bits to scramble common patterns. 133

SCU Sync Clock Unit, responsible for command deserialization and clock division.. 85, 93

SEU Single Event Upset, an inversion (bit-flip) of a signal due to an ionizing particla in the digital logic. 129

SVD Silicon Vertex Detector, outer part of the current Belle II vertex detector.. 13, 17, 20, 21, 24, 123

Te-BCID Trailing edge BCID code, end of the ToT period. 48, 106, 122, 133

TLU Trigger Logic Unit, device to synchronize detectors in a testbeam scenario. [41]. 57

ToT Time over Threshold. 48, 51, 54, 64, 66, 68, 100, 105, 106, 108, 111, 113, 115, 121, 124, 135, 138

TRG Trigger Group, the TRU is composed of 112 TRGs with one two-stage memory block each.. 105, 107, 112, 141

TRU Trigger Unit, the data processing and trigger part of the OBELIX chip, composed of TRGs.. 85, 102, 105, 111, 121, 141

TTT Track Trigger Transmission, module for contribution to the Belle II L1 Trigger.. 86, 123

TXU Transmission Unit, packs data to frames and serializes it.. 85, 105, 107, 118, 121, 122

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

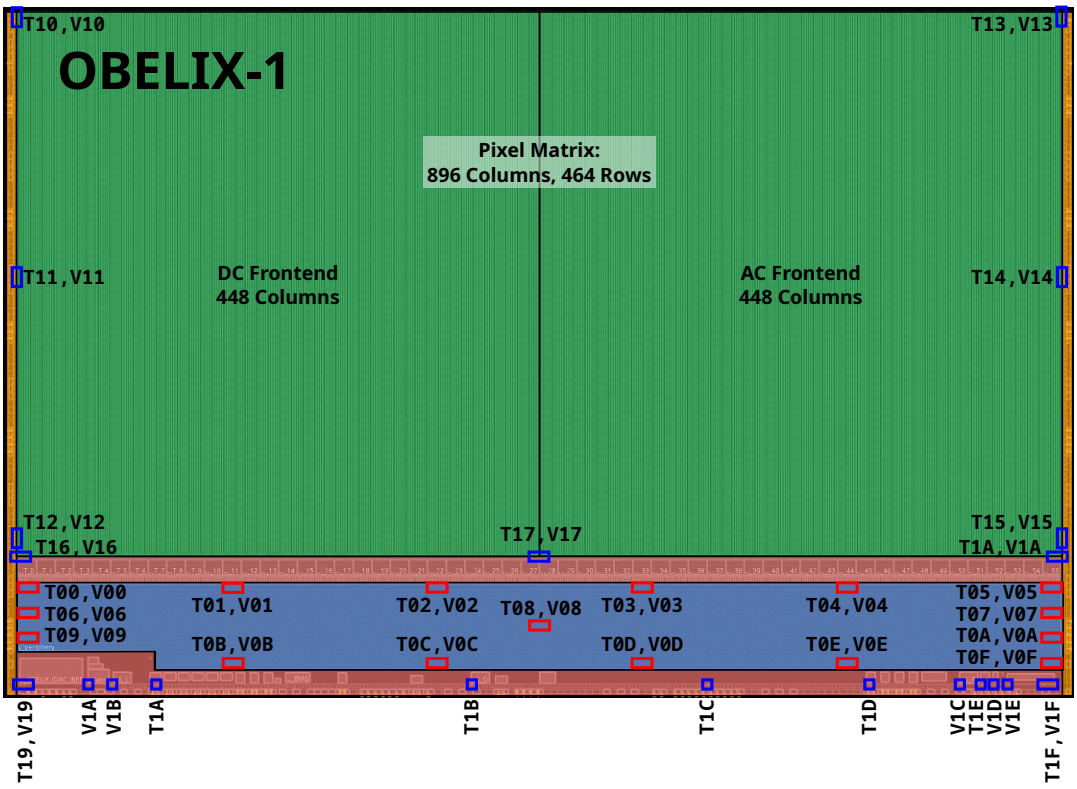
Appendix A

OBELIX Cheatsheet

Clock signal	Rounded	Real value	Period
Belle II RF		508.887 MHz	1.965 ns
Fast clock	160 MHz	169.7 MHz	5.89 ns
Slow clock	20 MHz	21.2 MHz	47.1 ns
TXU clock	32 MHz	33.9 MHz	29.5 ns

Hitrate:	120 MHz/cm^2
Trigger rate:	3 * 50 kHz
Triggered BCIDs:	0.7%
Output:	3.8 MHits/s
TXU Bandwidth:	$>$ 4.8 MHits/s
TXU Utilization:	80%

	TJ-Monopix2	OBELIX
Pixels	512 × 512	896 × 464
Pixel pitch	33.04 × 33.04 µm^2	33.04 × 33.04 µm^2
Active area	16.92 × 16.92 mm^2	29.60 × 15.33 mm^2
Active area	2.86 cm^2	4.54 cm^2
Die area	20.20 × 20.60 mm^2	30.17 × 18.81 mm^2



Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Appendix B

Auxiliary Plots

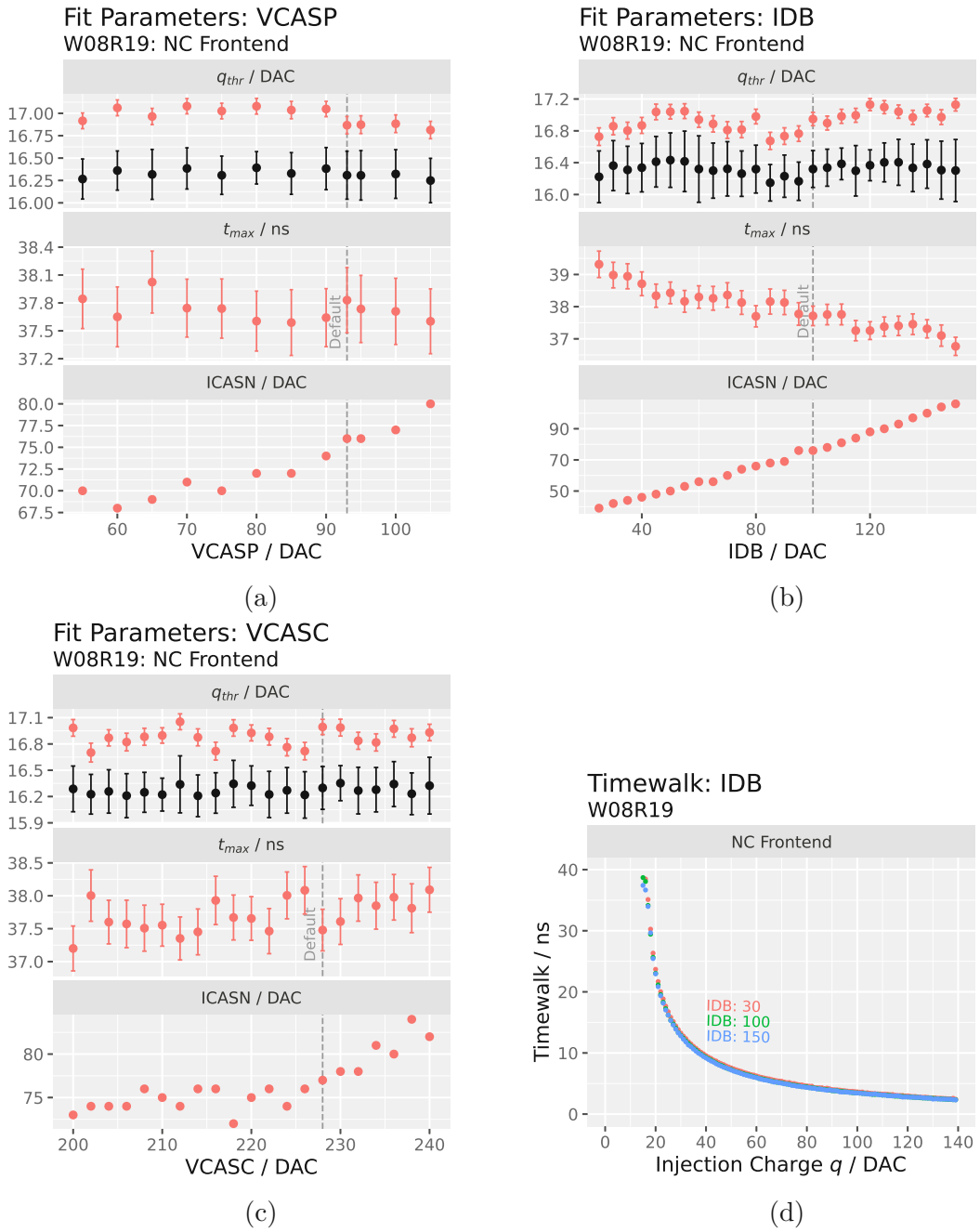


Figure B.1: Timewalk parameters for **VCASP** (a), **IDB** (b and d) and **VCASP** (c).

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Appendix C

Additional OBELIX Material

Payload	Encoded data		Payload	Encoded data	
0x00	0x6A	0b01101010	0x10	0xA6	0b10100110
0x01	0x6C	0b01101100	0x11	0xA9	0b10101001
0x02	0x71	0b01110001	0x12	0x59	0b01011001
0x03	0x72	0b01110010	0x13	0xAC	0b10101100
0x04	0x74	0b01110100	0x14	0xB1	0b10110001
0x05	0x8B	0b10001011	0x15	0xB2	0b10110010
0x06	0x8D	0b10001101	0x16	0xB4	0b10110100
0x07	0x8E	0b10001110	0x17	0xC3	0b11000011
0x08	0x93	0b10010011	0x18	0xC5	0b11000101
0x09	0x95	0b10010101	0x19	0xC6	0b11000110
0x0A	0x96	0b10010110	0x1A	0xC9	0b11001001
0x0B	0x99	0b10011001	0x1B	0xCA	0b11001010
0x0C	0x9A	0b10011010	0x1C	0xCC	0b11001100
0x0D	0x9C	0b10011100	0x1D	0xD1	0b11010001
0x0E	0xA3	0b10100011	0x1E	0xD2	0b11010010
0x0F	0xA5	0b10100101	0x1F	0xD4	0b11010100

Table C.1: Table of 5b/8b encoding for RD53 protocol.

`MON_VDAC_MUX[3]` and `OVR_VDAC_MUX[3]`	
Value	Selection
0	Disabled
1	`VPULSE` Dynamic injection pulse
2	`VCASC`
3	`VCASP`
4	`VCLIP`
5	`VRESET_HV`
6	`VRESET`

Table C.2: Possible selections of the `MON_VDAC_MUX[3]` and `OVR_VDAC_MUX[3]` register.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

MON_I_MUX[4] and OVR_IDAC_MUX[4]	
Value	Mode
0	Disabled
1	IBIAS
2	ITHR
3	ICASN
4	IDB
5	ITUNE
6	IDEL
7	IRAM
8	ICOMP
9	I_LDOA_L: Current through left analog LDO
10	I_LDOA_R: Current through right analog LDO
11	I_LDOD: Current through digital LDO
12	I_LDO_VPC: Current through VPC LDO

Table C.3: Possible selections of the MON_I_MUX[4] and OVR_IDAC_MUX[3] (only 8 and below) register.

C.0.1 Timewalk Measurement with PTD

In principle, the PTD module can be directly used to measure timewalk during a regular s-curve inside the chip without external components. With TJ-Monopix2, an injection pattern was used that enabled every fourth row in one column together for simultaneous injection. Subsequently, the next column was enabled for injection for the same row. This repeats until all columns were injected, at which point the row enabled for injection was shifted to the next by one pixel and again all four residue-classes of rows are scanned. This pattern is minimal in terms of commands required to be sent to the chip while scaling the runtime linearly with number of enabled columns and constant with number of enabled rows.

Since one PTD counter is available per 8 columns, this pattern is not suited for a timewalk scan. Instead, a 90° rotated pattern is required, injecting every 8th column in one row at once. While this pattern requires the same number of commands (except for the factor of two for the 8 instead of 4 pixel spacing), the runtime scales linearly with the number of rows enabled and is constant for any number of columns.

Since quick s-curve scans are routinely executed over a small number of columns on all rows in the typical lab environment, the rotated injection pattern would slow this down significantly. Usually, all rows are injected since this inhibits the systematic errors from variations along the row due to process but also voltage drop and temperature gradients. These effects also exist along the column axis in similar magnitude, but usually concern groups of 32 columns together as these groups share common DACs for the bias voltages and currents. Therefore, a separate scan for timewalk is probably useful implementing the rotated injection pattern, while the regular s-curve keeps the old injection pattern without the timewalk feature.

C.0.2 Register Table

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

Name	Address	Offset	Size	Default	Type	Description
IBIAS	0	0	8	100	hamming	DAC for frontend tuning
ITHR	0	8	8	64	hamming	DAC for frontend tuning
ICASN	1	0	8	0	hamming	DAC for frontend tuning
IDB	1	8	8	100	hamming	DAC for frontend tuning (Discriminator)
ITUNE	2	0	8	53	hamming	DAC for setting in-pixel TDAC LSB width
ICOMP	2	8	8	80	hamming	DAC for readout bit-comperators
IDEL	3	0	8	88	hamming	DAC for BCID delay compensation circuit
IRAM	3	8	8	50	hamming	DAC for in-pixel SRAM cell bias
VRESET_N	4	0	8	143	hamming	DAC for frontend baseline (Normal Cascode Frontend)
VRESET_HV	4	8	8	143	hamming	DAC for frontend baseline (HV Cascode Frontend)
VCASP	5	0	8	93	hamming	DAC for frontend tuning
VCASC	5	8	8	228	hamming	DAC for frontend tuning for cascode flavors
VCLIP	6	0	8	255	hamming	DAC for ToT clipping
VPULSE	6	8	8	255	hamming	DAC for charge injection level
IR2R	7	0	8	128	hamming	DAC For the Injection rail to rail buffer biasing
LDO_VDDAL	9	0	8	170	hamming	Tuning for Left Ananlog Voltage regulator
LDO_VDDAR	9	8	8	170	hamming	Tuning for Right Ananlog Voltage regulator
LDO_VDDD	10	0	8	170	hamming	Tuning Digital Voltage regulator
LDO_VPC	10	8	8	170	hamming	Tuning VPC Voltage regulator
MON_I_MUX	11	0	4	0	hamming	Select channel for monitoring of currents (IDACs and LDOs)
OVR_IDAC_MUX	11	4	4	0	hamming	Select channel for overdriving an IDAC
MON_VDAC_MUX	11	8	3	0	hamming	Select channel for monitoring of VDAC outputs
OVR_VDAC_MUX	11	11	3	0	hamming	Select channel for overdriving a VDAC
MON_VPULSE	11	14	1	0	hamming	Enable monitoring of dynamic pulse voltage
OVR_VPULSE	11	15	1	0	hamming	Enable overdrive of dynamic pulse voltage
MON_TEMP_MUX	12	0	5	0	hamming	Select Temperature sensor channel
MON_VDD_MUX	12	5	6	0	hamming	Select Supply Voltage monitoring channel
ADC_MUX	13	0	5	0	hamming	Main Multiplexer for the ADC
ADC_EXT_CAL	13	5	1	0	hamming	Enable external calibration for ADC
ADC_EN	13	6	1	0	hamming	ADC Enable
ADC_START	13	7	1	0	hamming	ADC Start conversion
ADC_IREF	13	8	8	128	hamming	ADC reference current selector
ADC_DONE	14	0	1	0	input	ADC conversion finished
ADC_RESULT	14	1	15	0	input	ADC result value
CLEAR_EN_SET_BCID	16	0	1	0	hamming	If set, a CLEAR command will set the current-BCID-counter to the value in BCID_INITIAL_VALUE register
CLEAR_EN_SYNC_TTT	16	1	1	0	hamming	If set, a CLEAR command will reset the clock generation of the TTT and TXU clock, this may temperarily disrupt data transmission
CLEAR_EN_RESET_MATRIX	16	2	1	0	hamming	If set, a CLEAR command will reset the matrix
CLEAR_EN_RESET_TRU	16	3	1	0	hamming	If set, a CLEAR command will reset the trigger unit
CLEAR_EN_RESET_TXU	16	4	1	0	hamming	If set, a CLEAR command will reset the transmission unit including the Tx FIFO
CLEAR_EN_RESET_TRIG_ID	16	5	1	0	hamming	If set, a CLEAR command will reset the (mostly) internal trigger id, this will disrupt data in memory at that time
CLEAR_EN_RESET_CHIP	16	6	1	0	hamming	If set, a CLEAR command will soft-reset the whole chip
RESET_PIXEL_MATRIX	16	7	1	0	hamming	Reset the Pixel Matrix, select DCs via EN_RO_RST_CONF
CMOS_DEBUG_SEL0	18	0	4	14	hamming	Select signal for Debug line 0, Default: sync_lock
CMOS_DEBUG_SEL1	18	4	4	3	hamming	Select signal for Debug line 1, Default: HitOr output
CMOS_DEBUG_SEL2	18	8	4	6	hamming	Select signal for Debug line 2, Default: 20MHz output
CMOS_DEBUG_SEL3	18	12	4	13	hamming	Select signal for Debug line 3, Default: SYNC_LOSS_FLAG

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

Name	Address	Offset	Size	Default	Type	Description
TRU_LATENCY	21	0	9	0	hamming	Latency for the trigger unit
TRU_RESET_BCID	21	9	1	0	hamming	Reset BCID to value of BCID_INITIAL_VALUE register, use after changing the Latency. Can be kept high to stop BCID counter at BCID_INITIAL_VALUE
TRU_TRIGGER_MODE	21	10	2	0	hamming	Trigger mode
TRU_ENABLE_EMPTY_TRIGGER	21	13	1	0	hamming	If set, an empty frame is created when there is no hit in a trigger
TRU_STOP_EOC	21	14	1	0	hamming	Disable EoC Readout (Complete the current readout cycle when bit is set) but keep TRU running
TRU_SLOW_BCID	21	15	1	0	hamming	Slow down the BCID clock speed by a factor of two (10 MHz)
BCID_INITIAL_VALUE	22	0	9	0	hamming	If BCID is reset, this is the start value
PTD_ENABLE	22	9	1	0	hamming	Enable the PTD precision timing counter (Peripheral Time to Digital), will consume a lot of power
TXU_AUX_SELECT	22	10	2	3	hamming	In each frame, there are 5 bits of free data per hit, 0: trigger id, 1: error flags, 3: first hit in frame contains trigger Id, subsequent hits (if any) error flags
TTT_EN_MASTER	22	12	1	0	hamming	Enable TTT Master/Standalone mode. LVDS Receivers need to be enabled separately!
TTT_EN_SLAVE	22	13	1	0	hamming	Enable TTT Slave mode, automatically enables required LVDS Drivers
TTT_PIXELSIZE	22	14	2	0	hamming	Macropixel size selection for TTT, 0: 8 Macropixels, 1: 4 Macropixels (1 Slave), 2: 2 Macropixels (Up to 3 Slaves)
EOC_T_START	23	0	4	0	hamming	Additional clock cycles between Token rising and Freeze rising edge
EOC_T_FREEZE	23	4	4	0	hamming	Additional clock cycles between Freeze rising and Read rising edge
EOC_T_PRECHARGE	23	8	4	0	hamming	Additional clock cycles between Read rising and Read falling edge
EOC_T_READ	23	12	4	0	hamming	Additional clock cycles between Read falling and optional Read rising edge
DIGITAL_INJ_EN	24	0	1	0	hamming	Enable the digital injection module, reset BCID after enabling
DIGITAL_INJ_RUN	24	1	1	0	hamming	Inject hits with the digital injection module
DIGITAL_INJ_HITRATE	24	8	7	0	hamming	Select hitrate for digital injection, in 4.1 MHz/cm2 units
DIGITAL_INJ_SEED_SELECT	25	0	7	0	hamming	Select trigger group for setting the seed, 0x7F for all (needs to be enabled) - writing to this register writes the RNG seed
DIGITAL_INJ_SEED	25	7	7	0	hamming	Digital injection RNG seed value (needs to be enabled) - writing to this register writes the RNG seed
CMD_LOCKLOSS_FLAG	27	0	1		input	Command decoder lock loss flag - not sure what this is supposed to do
CMD_SYNCLOSS_FLAG	27	1	1		input	Loss of synchronization flag, BCIDs cannot be trusted anymore when this is set
CMD_ERROR_FLAG	27	2	1		input	Command decoder error flag
CMD_BITFLIP_WARNING_FLAG	27	3	1		input	Command decoder bitflip warning flag
CMD_BITFLIP_ERROR_FLAG	27	4	1		input	Command decoder bitflip error flag
CHIP_ID_READ	27	5	5		input	Read back chip_id input
PULSE_START_CONFIG_READ	27	10	6		input	Read back pulse start config
PULSE_STOP_CONFIG_READ	28	0	14		input	Read back pulse stop config
CMD_LOCKLOSS_CNT	29	0	16		input	Command decoder lock loss counter
CMD_SYNCLOSS_CNT	30	0	16		input	Counter of sync loss events
CMD_ERROR_CNT	31	0	16		input	Command decoder error counter
CMD_BITFLIP_WARNING_CNT	32	0	16		input	Command decoder bitflip warning counter
CMD_BITFLIP_ERROR_CNT	33	0	16		input	Command decoder bitflip error counter
CMD_LOCKLOSS_CNT_RESET	34	0	1	0	hamming	Reset lock loss counter
CMD_SYNCLOSS_CNT_RESET	34	1	1	0	hamming	Reset sync loss counter
CMD_ERROR_CNT_RESET	34	2	1	0	hamming	Reset CMD error counter
CMD_BITFLIP_WARNING_CNT_RESET	34	3	1	0	hamming	Reset Bitflip warning counter
CMD_BITFLIP_ERROR_CNT_RESE	34	4	1	0	hamming	Reset bitflip error counter
CMD_CHSYNC_THR_HIGH_CONF	34	6	5	4	hamming	SCU sync threshold high
CMD_CHSYNC_THR_LOW_CONF	34	11	5	31	hamming	SCU sync threshold low

TU WIEN Bibliothek Your knowledge hub
Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

Name	Address	Offset	Size	Default	Type	Description
LVDS_TX_ENABLE	36	0	1	1	hamming	Tx LVDS Enable
LVDS_CLKOUT_ENABLE	36	1	1	1	hamming	Clock Output LVDS Enable
LVDS_TTT_ENABLE	36	2	1	1	hamming	TTT LVDS Enable
LVDS_HITOR_ENABLE	36	3	1	0	hamming	HitOr LVDS Enable
LVDS_TTT_MASTER_RX_ENABLE	36	4	6	0	hamming	Enable for the TTT-Master LVDS inputs (in case fewer Slaves than possible are connected)
LVDS_EXT_PULSE_ENABLE	36	11	1	0	hamming	Enable input for external pulse
LVDS_DEBUG_ENABLE	36	12	2	0	hamming	Enable LVDS drivers for debugging outputs
LVDS_TX_EN_HBRIDGE	37	1	5	28	hamming	Tx/TTT/HitOr LVDS ???
LVDS_TX_EN_CMFB	37	6	5	16	hamming	Tx/TTT/HitOr LVDS ???
LVDS_TX_SET_IBCMFB	37	11	4	7	hamming	Tx/TTT/HitOr LVDS ???
LVDS_TX_EN_PRE	38	0	16	0	hamming	Tx/TTT/HitOr LVDS Preemphasis
LVDS_TX_SET_IVNH	39	0	4	13	hamming	Tx/TTT/HitOr LVDS ???
LVDS_TX_SET_IVNL	39	4	4	7	hamming	Tx/TTT/HitOr LVDS ???
LVDS_TX_SET_IVPH	39	8	4	2	hamming	Tx/TTT/HitOr LVDS ???
LVDS_TX_SET_IVPL	39	12	4	13	hamming	Tx/TTT/HitOr LVDS ???
LVDS_TTTSLAVE_EN_HBRIDGE	40	1	5	28	hamming	TTT-Slave Output LVDS ???
LVDS_TTTSLAVE_EN_CMFB	40	6	5	16	hamming	TTT-Slave Output LVDS ???
LVDS_TTTSLAVE_SET_IBCMFB	40	11	4	7	hamming	TTT-Slave Output LVDS ???
LVDS_TTTSLAVE_EN_PRE	41	0	16	0	hamming	TTT-Slave Output LVDS ???
LVDS_TTTSLAVE_SET_IVNH	42	0	4	13	hamming	TTT-Slave Output LVDS ???
LVDS_TTTSLAVE_SET_IVNL	42	4	4	7	hamming	TTT-Slave Output LVDS ???
LVDS_TTTSLAVE_SET_IVPH	42	8	4	2	hamming	TTT-Slave Output LVDS ???
LVDS_TTTSLAVE_SET_IVPL	42	12	4	13	hamming	TTT-Slave Output LVDS ???
PIXEL_CONF_ROW_SELECT	44	0	9	0	normal	Pixel portal row selector
PIXEL_CONF_WRITE_SPEED	44	9	4	3	normal	Number of clock cycles to write pixel config, 3 corresponds to TJ-Monopix2 timing
PIXEL_CONF_COL_GROUP_SELECT	45	0	8	0	normal	Pixel portal column group select
PIXEL_CONF_PORTAL_DATA	46	0	12	0	normal	Pixel portal - Config for 4 pixels, write to this register automatically cofigures selected 4 pixels
EN_HITOR_COL	48	0	896	'1	normal	HitOr column enable
EN_HITOR_ROW	104	0	464	'1	normal	HitOr row enable
EN_INJ_COL	133	0	896	'0	normal	Injection column enable
EN_INJ_ROW	189	0	464	'0	normal	Injection row enable
EN_RO_CNF	218	0	448	'1	normal	Enable the readout of each DC individually
EN_BCID_CONF	246	0	448	'1	normal	Enable of double column BCID distribution
EN_RO_RST_CONF	274	0	448	'1	normal	Enable of double column readout reset, only hornored when RESET_PIXEL_MATRIX, ResetB and Clear always reset all DCs

Table C.4: Registers of the OBELIX chip.