Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

MASTERTHESIS

Time Resolution of a Fully-Integrated Digital Silicon Photo-Multiplier

submitted by

Stephan Lachnit

2023

Faculty of Mathematics, Informatics and Natural Sciences

Institute of Experimental Physics

Master of Science Physics

Supervisors:

Prof. Dr. Erika Garutti

Dr. Simon Spannagel

Abstract

Silicon Photo-Multipliers (SiPMs) are pixelated semiconductor detectors consisting of SPADs, usually used as single-photon sensitive detectors. SiPMs are often analog devices that require separate digitization. Combining SiPMs with digital readout known from modern pixel sensors offers new possibilities like full hitmap readout, pixel masking, or fast timestamping. Due to their potentially fast timestamping capabilities, digital SiPMs are becoming an interesting contender for 4D-Tracking in High Energy Physics.

At DESY such a digital SiPM was developed. It is a monolithic chip manufactured in a 150 nm CMOS process by LFoundry. The chip has a 32×32 pixel matrix with $76\,\mu m \times 70\,\mu m$ pixels and four on-chip TDCs that allow for timestamps with a resolution of less than 100 ps. It uses the Caribou DAQ system for a fully digital readout.

The chip was tested in the DESY II testbeam facility with a 4 GeV electron beam. In this thesis, the analysis procedure will be described and the results for the timing performance will be discussed. The time resolution was estimated to be (46 ± 5) ps in the center of the SPAD cells, while a slower response was found at the edge of the SPAD cells.

Contents

Chapter 1

Introduction

1.1 Motivation

The European Committee for Future Accelerator (ECFA) published a Detector R&D Roadmap in 2021 [1], that points out the most pressing issues to be resolved for future High Energy Physics (HEP) experiments, such as the ILC [2], CLIC [3] or FCC-ee [4]. The described path forward for vertex and tracking detectors is a technique called 4D tracking, that makes use of a combination of time and position measurements. This requires the development of solid-state sensors with high spatial precision, low material budget, and ultrafast timing.

There are several approaches to building such a detector system. One possible approach is the use of Silicon Photo-Multipliers (SiPMs) in an ultrafast timing layer (less than 100 ps) in conjunction with other Silicon pixel sensors for high spatial precision. SiPMs are composed of an array of Single-Photon Avalanche Diodes (SPADs), which have a charge multiplication layer, leading to large avalanches where the collected charge is not proportional to the energy of the traversing particle anymore. Unlike other pixel designs, SPADs can thus only provide binary information.

SiPMs are already used widely in HEP detectors, mostly in the readout of scintillating material of calorimeters thanks to their ability to detect and count single photons. Typically, these SiPMs are analog devices where spatial information of the fired SPAD cell is not available. However, thanks to commercial CMOS manufacturers, designs of SPAD cells became available that can be integrated into an ASIC (Application-Specific Integrated Circuit). This allows digital SiPMs the possibility to gain access to the position resolution required for 4D tracking while keeping a low material budget thanks to being monolithic. Such a digital SiPM has been designed and tested at DESY in Hamburg, Germany [5].

However, due to their background as analog photon detectors, only a few studies have been carried out testing the performance of SiPMs in charged particle - as opposed to photon - detection (e.g. [6]). Testing sensors for tracking detectors successfully in a testbeam environment, where the sensor is traversed by Minimum Ionizing Particles (MIPs), is the golden standard required before integration in a large HEP detector can begin. Such a test has been performed for DESY's digital SiPM.

In this thesis, a data analysis of DESY's digital SiPM in such a testbeam environment is performed. The main goal is to study the timing performance of the SiPM in MIP detection to test the feasibility of the design as a possible concept for 4D tracking. Besides giving numerical values, a key aspect is to find potential unexpected issues and provide solutions or possible design improvements.

In the remaining part of this chapter, the basic principles of particle and detector physics are summarized. In chapter 2 details about the digital SiPM are given, and chapter 3 describes the concept and the setup of the testbeam measurement evaluated in this thesis. In chapter 4 the analysis chain for the testbeam data is described. The results of the testbeam measurement are given in chapter 5. Lastly, the timing performance of the digital SiPM is evaluated and discussed in chapter 6.

1.2 Particle Interaction with Matter

To measure a particle with a detector, the particle needs to interact with the detector. These interactions with the detector can lead to a change in the energy and trajectory of the particle. Thus understanding how particles interact with matter is essential to design and test particle detectors.
For this thesis, the most important interactions are the energy loss and scattering of electrons and the absorptions of photons in Silicon.

1.2.1 Charged Particles

For heavy charged particles, the most significant energy loss relevant in HEP is the ionization of the material they traverse. This mean energy loss is described by the well-known Bethe formula ([7, 8])

$$-\left\langle \frac{\mathrm{d}E}{\mathrm{d}x} \right\rangle = \frac{4\pi n z^2}{m_\mathrm{e} c^2 \beta^2} \cdot \left(\frac{e^2}{4\pi\varepsilon_0} \right)^2 \cdot \left[\log\left(\frac{2 m_\mathrm{e} c^2 \beta^2}{I \cdot (1-\beta^2)} \right) - \beta^2 - \frac{\delta(\beta\gamma)}{2} \right] \tag{1.1}$$

where n is the electron density in the material, z the charge of traversing particle in multiples of e, m_e the rest mass of the electron, c the speed of light in vacuum, β the relative velocity $\beta = v/c$ of the traversing particle, v the velocity of the traversing particle, e the elementary charge, ε_0 the vacuum permittivity, I the mean excitation energy of the material, $\delta(\beta\gamma)$ the density effect correction and γ the Lorentz factor $\gamma = 1/\sqrt{1-\beta^2}$ of the traversing particle.

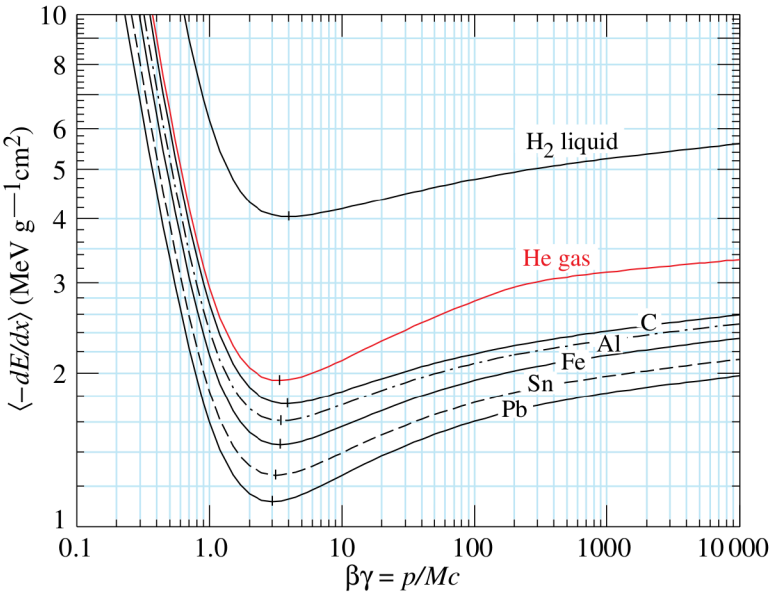


Figure 1.1: Mean energy loss of heavy particles for different materials. Taken from [8].

A plot of the Bethe formula is shown in Fig. 1.1 for different materials. The curves have a minimum at around $\beta\gamma \simeq 3$. This means for every particle there is a certain energy where it loses the least amount of energy when traversing the material. Particles traversing the material with this energy are called Minimum Ionizing Particles (MIPs). Particles that have an energy loss in the same order as MIPs are often called MIPs as well.
For lower momenta the energy loss increases because the interaction time with the material increases. This causes an energy loss that is roughly proportional to β^{-2}. For higher momenta relativistic effects become more significant, resulting in a logarithmic rise of the energy loss. The electric field gets compressed along the longitudinal axis, which leads to a higher interaction strength perpendicular to the direction of the particle momentum.
For the much lighter electrons and positrons, Bremsstrahlung starts to become more relevant. Electrons also experience so-called Fermi pressure, while positrons can annihilate with the electrons in the material. The Berger-Seltzer formula [9] gives a better description for electrons and positrons than the Bethe formula. The overall shape of the Berger-Seltzer formula is not vastly different from Fig. 1.1 however.

1.2.2 Multiple Coulomb Scattering

Besides the energy loss, particles traversing a material also scatter in the material, changing the trajectory of the particle. The dominant scattering in HEP detectors is Coulomb scattering, where the particle gets deflected from the Coulomb field of the nucleus. Since this scattering is a statistical process it can happen multiple times in a material, leading to a net deflection angle (see Fig. 1.2). The root mean square (RMS) of the angular deflection can be described using the Highland equation [10]

$$\Theta_{\text{rms}} = \frac{13.6\,\text{MeV}}{\beta cp}\, z \sqrt{\frac{x}{X_0}} \left(1 + 0.038 \log\left(\frac{x}{X_0}\right)\right) \tag{1.2}$$

where x is the distance traveled in the material and X_0 the radiation length of the material. The other parameters have the same meaning as in Eqn. 1.1.
From Eqn. 1.2 follows that thin sensors result in less angular deflection, which is often advantageous in HEP detectors where the momentum of a particle is reconstructed from its trajectory.

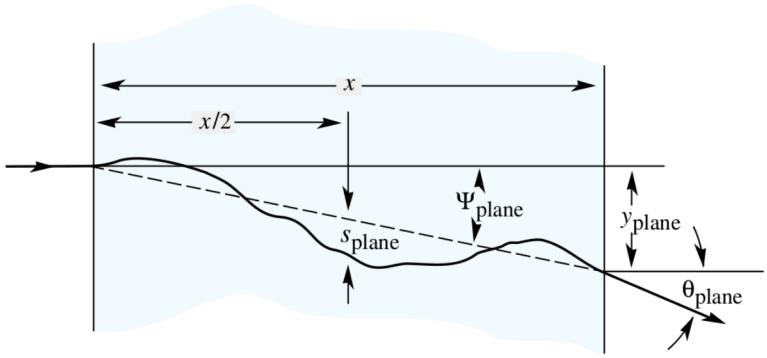


Figure 1.2: Schematic drawing of multiple Coulomb scattering. Taken from [8].

1.2.3 Photons

Photons can interact with matter in various ways. The dominant interaction depends on the material and the energy of the photon. For the purpose of this thesis, only interactions of photons in the visible spectrum with Silicon are relevant. Here the most dominant interaction is the photo-effect, where a photon is absorbed by an atom that then emits an electron.
The photon absorption probability can be described by an exponential decay

$$p_{\text{abs}} \propto \exp\left(-\frac{x}{\lambda}\right) \tag{1.3}$$

where x is the distance traveled in the material and λ the material-specific absorption length. In Fig. 1.3 the absorption length λ of Silicon is shown as a function of the wavelength.

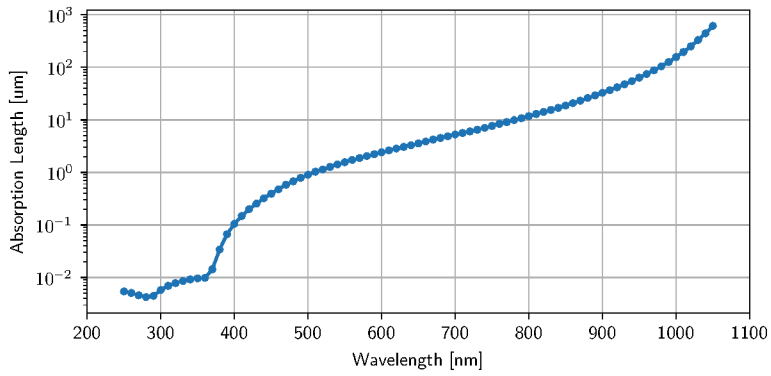


Figure 1.3: Absorption length of photons in Silicon. Data from [11].

1.3 Semiconductor Physics

Semiconductors are described by solid-state physics, and in particular by energy bands of crystal lattices. These energy bands arise from the energy levels of the atoms in the lattice, which are split up due to the influence of neighboring atoms. Some energy levels are so close to each other that they are grouped in energy bands. For example, the valence band of a lattice is the last energy band where all energy states are occupied by electrons at 0 K. The energy band above the valence band is called conduction band. The energy difference between these two bands is called band gap.
The Fermi energy is defined as the highest occupied energy state at 0 K. For insulators, the Fermi energy corresponds to the highest energy state of the valence band and the band gap is so high that electrons in the valence band cannot jump to the conduction band with thermal or external excitation. For conductors, the Fermi energy lies inside the conduction band, which leads to free charge carriers. Semiconductors lie in the middle of these two cases: they have a band gap that is small enough so that electrons can jump to the conduction band from thermal or external excitation. This means that the conductivity of semiconductors increases with temperature. When an electron moves to the conduction band, it leaves a "hole" in the valence band. This hole can be filled from another electron in the valence band or it can recombine with an electron from the conduction band. In semiconductors, holes can be described as charge carriers just like electrons.

1.3.1 Doping

The most commonly used semiconductor is Silicon. Silicon atoms have four valence electrons and form a diamond lattice. The band gap of Silicon is about 1.1 eV, making it a semiconductor at room temperature. However, the charge carrier density in the lattice is relatively low compared to the atomic density, making it effectively act like an insulator.
To increase the charge carrier density, Silicon can be doped. For example, adding some phosphorus atoms to the lattice, which have five valence electrons, results in one loosely bound electron per phosphorus atom in the lattice. This process is called n-doping and phosphorus the (electron) donor.
Similarly, p-doping can be achieved by adding atoms with less than four valence electrons to the lattice, which are then called (electron) acceptors. For example, boron with three valence electrons adds one hole per boron atom as charge carrier to the lattice.
It is worth noting at this point, that while doping might add positive or negative charge carriers into the lattice, the overall space charge is still zero.

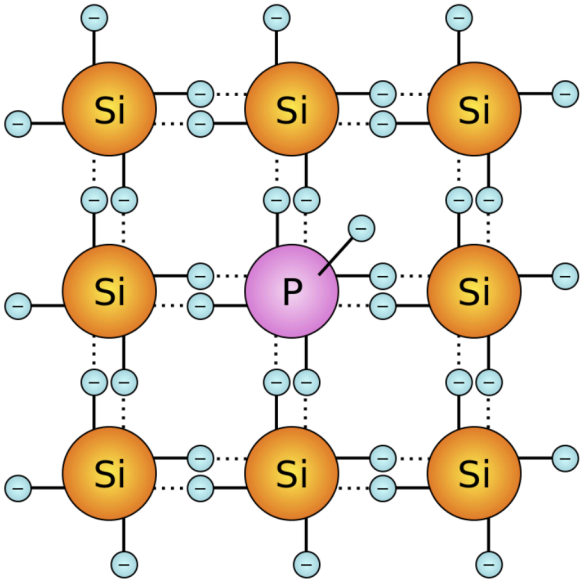


Figure 1.4: Schematic drawing of the lattice in n-doped Silicon. Taken from [12].

1.3.2 pn-Junction

A pn-junction is the junction between two differently doped regions, one doped with donors and the other with acceptors. At the border, electrons from the n-doped region diffuse into the p-doped region and holes from the p-doped region diffuse into the n-doped region. This diffusion current creates a net positive charge in the n-doped region and a net negative charge in the p-doped region.
This charge distribution leads to a potential difference and thus an electric field in the pn-junction. The electric field results in a drift current pointing in the opposite direction of the diffusion current. The field increases until an equilibrium between the diffusion and drift currents is reached.
A region without free charge carriers emerges from this equilibrium, called depletion zone. Any charge carriers in this region drift outwards due to the electric field or recombine with the other charge carriers.

The width of the depletion zone can be increased or decreased by applying a bias voltage between the p-doped and n-doped regions. In reverse bias, the p-doped side is connected to the negative terminal, and the n-doped side to the positive terminal. Here the negative charge carriers in the n-doped region are pulled towards the positive terminal. This leaves behind positive ions on the n-doped side and thus increases the depletion zone. Likewise, holes in the p-doped region are pulled towards the negative terminal and leave behind a negative space charge.
In forward bias, the p-doped side is connected to the positive terminal, and the n-doped side to the negative terminal. The potential difference now leads to a decreased depth of the depletion zone.
A pn-junction operated in reverse bias with a high voltage also has a high electric field. If the electric field is strong enough, avalanche breakdown can start to appear. The minimum bias voltage required for this avalanche breakdown to appear is called breakdown voltage. An avalanche is created when a free electron in the depletion zone, for example created from thermal excitation, is accelerated to such high speeds that it has enough energy to kick an electron from the valence band to the conduction band. Since this electron is also accelerated by the electric field, an avalanche begins. This avalanche continues until the bias voltage drops below a certain voltage.

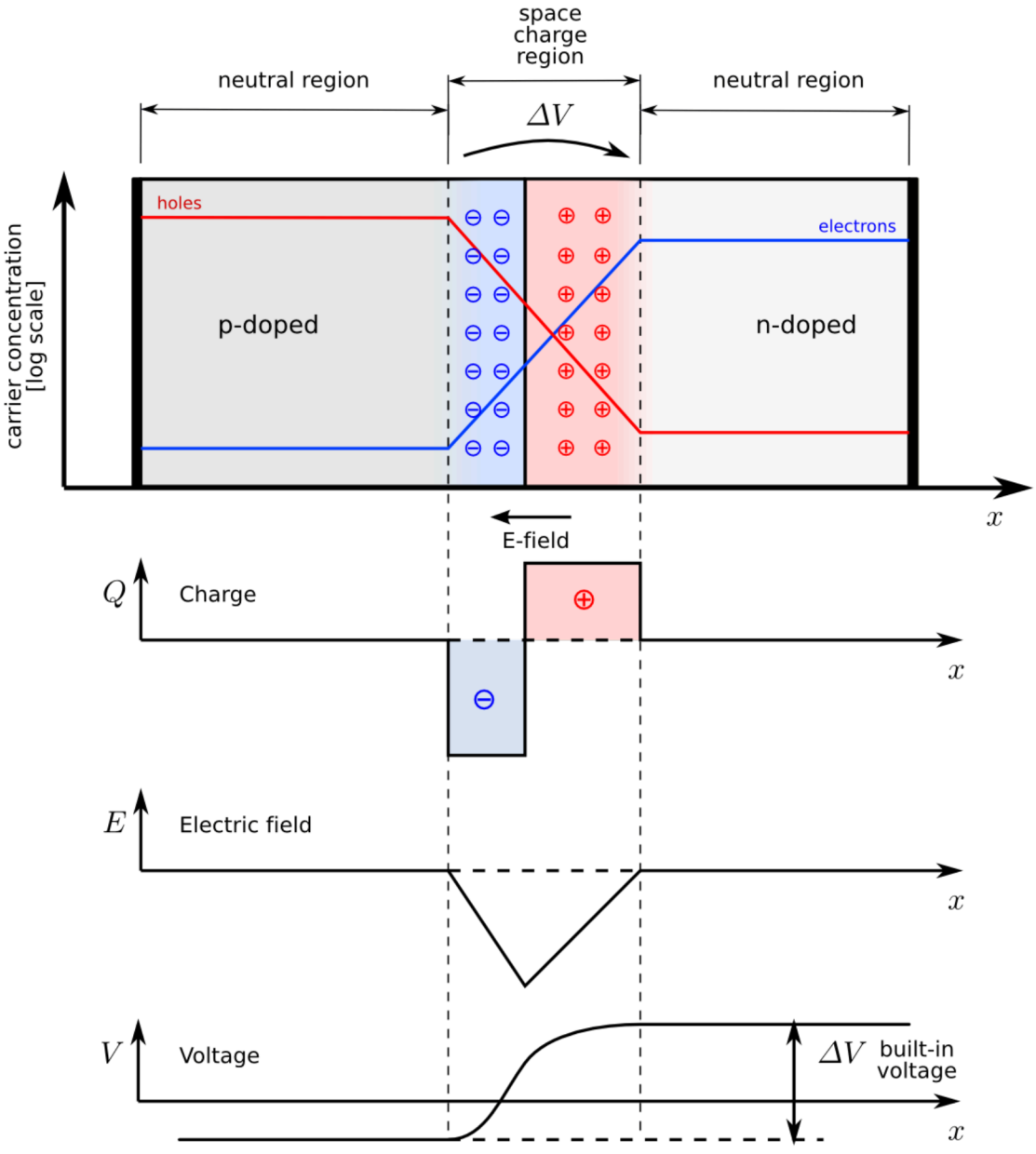


Figure 1.5: Schematic drawing of a pn-juction in equilibrium. Taken from [13].

1.4 Silicon Photo-Multipliers

A Silicon Photo-Multiplier (SiPM) is a semiconductor detector designed to detect single photons. SiPMs are pixelated devices, which means that they have multiple so-called Single-Photon Avalanche Diodes (SPADs) arranged in a grid. To understand the principle of SPADs and SiPMs, one first has to understand the basic principle of "ordinary" Silicon pixel sensors.

1.4.1 Silicon Pixel Sensors

A Silicon pixel sensor is a particle detector using a pn-junction (described in subsection 1.3.2) to detect charged particles with the particle interactions described in subsection 1.2.1. When a charged particle traverses the pn-junction, it loses energy by ionizing the Silicon atoms, which creates free charge carriers. As discussed in subsection 1.3.2, free charge carriers in the depletion zone either recombine or drift in the electric field. If a reverse bias is applied to the pn-junction, the electrons move to the n-doped region with the positive terminal and the holes in the other direction (see Fig. 1.6).

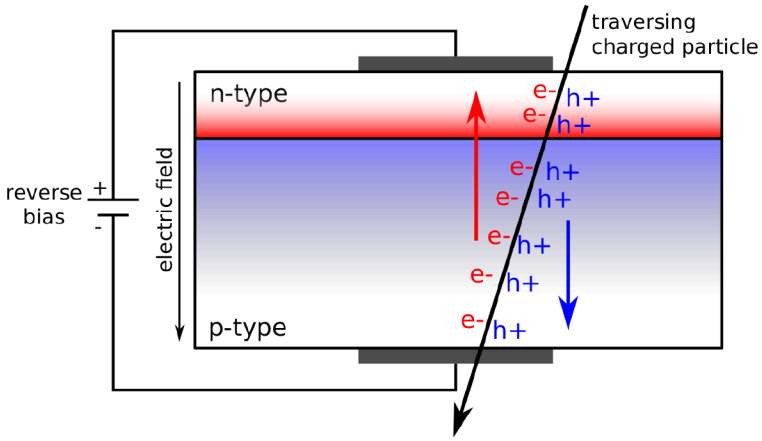


Figure 1.6: Schematic drawing of a pn-junction traversed by a charged particle. Taken from [14].

A particle traversing the sensor can be detected by measuring the current between the positive and negative terminal. Usually, the current created from a traversing particle is so small that it requires amplification before it can be measured.
To complete the Silicon pixel sensor, the pn-junction has to be pixelated. This means that at least one side is biased not along the entire area of the sensor, but by small biasing points forming the center of the pixels. Charge sharing can appear, which means the charge carriers created by a traversing particle are collected in multiple pixels. By measuring the current for every pixel, the position of the traversing particle can be estimated.

1.4.2 Single-Photon Avalanche Diodes

SPADs work similarly to Silicon pixel sensors but implement an amplification layer. An amplification layer is a pn-junction, where the n- and p-doped regions are strongly doped. This is often denoted as n^+- and p^+-doping. The strong doping results in a high electric field in the amplification layer, where charge carriers can start an avalanche if the applied bias voltage is above the breakdown voltage (see subsection 1.3.2 for details).
To detect photons with a SPAD, the photons need to be absorbed within the depletion zone, creating an electron-hole pair. One charge carrier drifts towards the amplification layer, where it starts an avalanche. To stop this avalanche after it started, a quenching circuit is required. Quenching is implemented usually in the form of a resistor in series with the SPAD (see Fig. 1.8a) that reduces the voltage over the cell below the breakdown voltage when there is a large current, i.e. an avalanche.
The gain of a SPAD is typically in the order of $\mathcal{O}(10^6)$, which makes it possible to detect a single electron-hole pair and thus single photons. A SPAD is not able to detect every photon however. One requirement is that the photon is absorbed in or close to the depletion zone. As explained in

subsection 1.2.3, the interaction depth depends on the absorption length of the material, which can be wavelength-dependent. In Silicon particular (see Fig. 1.3), the relevant wavelength lies within the visible spectrum. For example, red photons have an absorption depth of about 3 µm. Photons with a much larger wavelength (for example infrared photons) might not interact within the depletion zone or sensor at all, and photons with a much smaller wavelength might interact too close to the sensor surface. The exact wavelength range where a SPAD is efficient depends on the design of the SPAD.
Given the design of a SPAD, it can also be used to detect charge particles since they create more than one electron-hole pair (eh-pair) when traversing the sensor (for MIPs $\mathcal{O}$(80 eh-pairs per µm) in Silicon). Fig. 1.7 shows a schematic drawing of a SPAD and the interaction with a MIP and photons of different wavelengths.

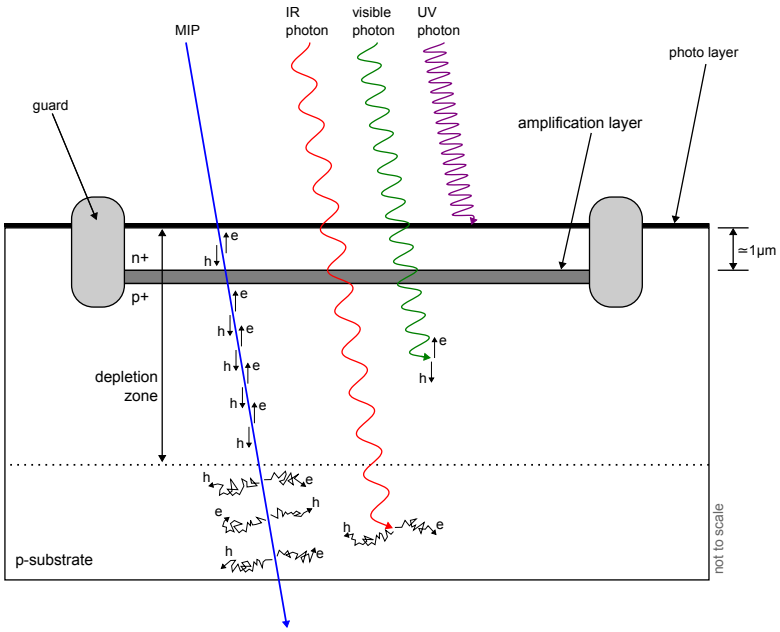


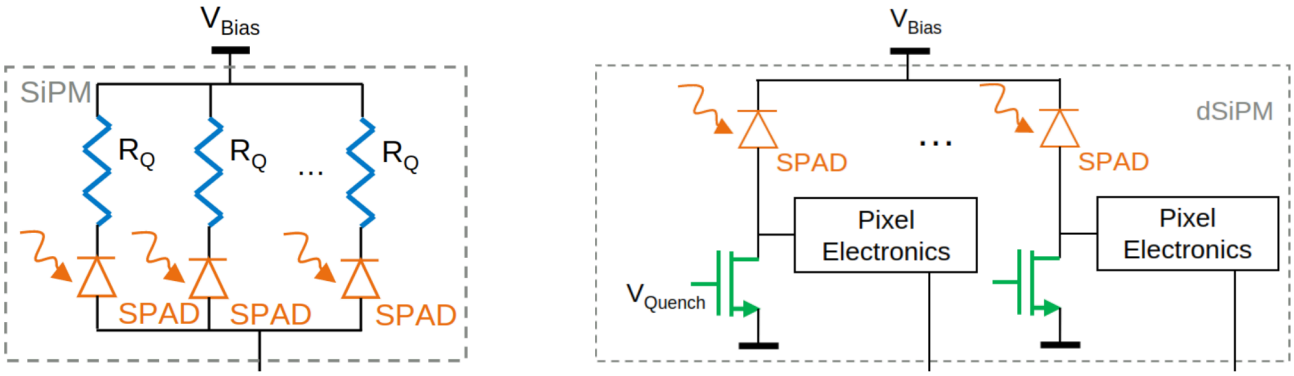
Figure 1.7: Schematic drawing of a SPAD traversed by photons and a MIP.

SPADs require guards to prevent avalanches from spreading to neighboring cells and to prevent breakdown at the edge of the highly-doped pn-junction. This can be implemented with a physical trench in the crystal or a specific doping called Junction Termination Extension (JTE).
This reduces charge sharing between neighboring SPADs. However, optical crosstalk can appear where photons are created from the accelerated charge carriers in the avalanche. These photons can then trigger an avalanche in a neighboring cell.

1.4.3 Analog and Digital SiPMs

SiPMs can be divided into two distinct groups: analog SiPMs and digital SiPMs. In analog SiPMs, all SPADs are connected in parallel. A measurement is conducted by reading the current or the integrated charge. This gives a characteristic integrated charge histogram with peaks for different amounts of fired pixels.[1] In this measurement, analog SiPMs do not provide the information in which SPAD a hit was registered but only the amount of hits.
Digital SiPMs combine the SPAD array directly with a digital readout ASIC (Application-Specific Integrated Circuit). This can either be implemented in a hybrid form where a chip with the SPAD array is bump-bonded to a readout chip, or in a monolithic approach. A monolithic digital SiPM uses a CMOS manufacturing process to combine the SPAD array and readout circuitry on the same chip. Digital SiPMs have the possibility to provide digital information directly. Most notably would be a hitmap to provide spatial information, but timing information is also possible.
In Fig. 1.8 schematics for an analog and a digital SiPM are shown.

[1]The distribution includes more effects than just the number of measured photo-electrons. However, since the focus of this thesis is a digital SiPM, a detailed description of the signal from analog SiPMs will not be given here.



(a) Analog SiPM where all the SPADs are connected in parallel.

(b) Digital SiPM where each SPAD is connected to pixel electronics and quenching is implemented with a transistor.

Figure 1.8: Schematics of an analog and a digital SiPM [15].

1.4.4 Noise and Detection Efficiency

SiPMs have a particular noise pattern, namely "dark counts". These hits can occur via thermal excitation even when no photon or particle is hitting the sensor. An electron might get elevated to the conduction band and travel to the amplification region, which results in an avalanche and thus a hit. The noise performance is measured as Dark Count Rate (DCR), often also as DCR per (active) area. Higher temperature results in more charge carriers being thermally excited, thus increasing the DCR. However, dark counts can also occur when charge carriers are created in the amplification region due to quantum tunneling. This effect increases with higher bias voltages.
Another important performance metric of SiPMs is their Photon Detection Efficiency (PDE). It measures the probability of registering a photon when it hits the SiPM. Due to the effects described in subsection 1.4.2, this quantity depends on the wavelength of the photon. The PDE is comprised of the quantum efficiency, which is the probability for a photon to create an eh-pair, the fill factor for the active area of the SiPM, and the Geiger probability, which is the probability of a charge carrier triggering an avalanche. The PDE is further affected by factors such as the surface reflectivity of the SiPM and the dead time of the sensor. The final PDE of SiPMs highly depends on the SPAD design and is typically only significant in the visible range.
The expected MIP detection efficiency of a SiPM is dominated by the fill factor for the active area of the SiPM since MIPs traversing a SPAD create significantly more than one eh-pair.

1.4.5 Timing Characteristics

SiPMs are known for their fast responses that can reach below $\mathcal{O}(100\,\text{ps})$. Typically the Single Photon Time Resolution (SPTR) is given as Full Width Half Maximum (FWHM) and measured with a fast laser and an optical timing reference. In a study by *Carnesecchi et al. (2022)* [6] the timing performance of SiPMs has been measured with charged particles and a time resolution of 30 ps to 70 ps (standard deviation) was reached with several different SiPMs.
By increasing the overvoltage, which is the voltage over the breakdown voltage, the timing performance of SiPMs can often be improved (e.g. [16, 17]).
An interesting phenomenon in the timing performance of SiPMs is the timing tail (e.g. [6, 16, 17, 18, 19]). This timing tail, also called diffusion tail, is on the slow edge of the timing histogram and commonly explained by deep photon interaction. This means that a photon might get absorbed deep in the SPAD, such that the created electron first needs to diffuse into the depletion zone and then drift to the amplification layer. In a paper by *Nemallapudi et al. (2016)* [16] it was found that the SPTR is higher at the edges of the SPADs, suggesting that this also contributes to the timing tail. It should be mentioned that the timing response can be different from device to device since it depends on the electric field inside the SPAD. Various functions have been used to describe the distribution, such as a q-Gaussian [6] or an exponentially modified Gaussian [16].

Chapter 2

DESY dSiPM

2.1 Chip Details

The chip discussed in this thesis is a monolithic digital SiPM that was designed and tested at DESY [5]. The chip is manufactured in LFoundry's 150 nm CMOS process.
It consists of a 32×32 pixel matrix. The pixels have an area of $69.6\,\mu m \times 76\,\mu m$. Each pixel consists of four SPADs connected in parallel and a quenching circuitry. Compared to a layout with a single SPAD per pixel, this layout allows to increase the fill factor to 30 % at the cost of spatial resolution. A picture of a pixel is shown in Fig. 2.1.
The SPAD type is available in LFoundry's process development kit (PDK). The breakdown voltage was measured to be 18.2 V at 0 °C and the DCR to be 5 kHz per pixel at 1.5 V overvoltage at 0 °C. More details can be found in *Diehl et al. (2023)* [5].
The digital features of the chip include a full hitmap readout with per quadrant timestamping and less than 100 ps resolution (described in section 2.2 and section 2.3). The chip is also capable of masking, meaning that every pixel can be switched off individually.
The chip will be referred to as "dSiPM" in this thesis.

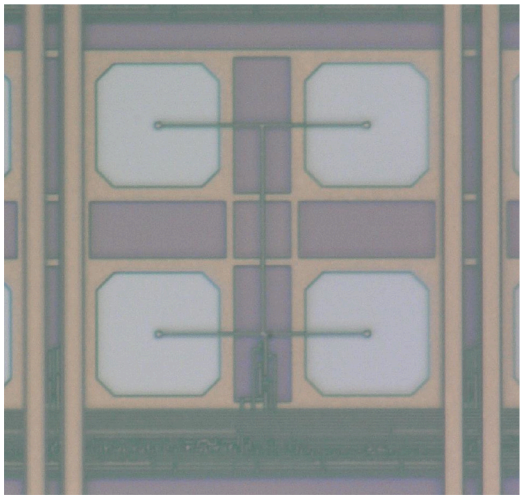


Figure 2.1: Picture of a dSiPM pixel consisting of four SPADs in parallel and corresponding quenching and readout circuitry underneath [5].

2.2 Readout and DAQ

The readout of the chip uses a frame-based scheme. Data is taken and stored in frames running on a 3 MHz clock (bunch clock), resulting in frames with a length of about 333 ns. The chip is divided into four identical quadrants with 16×16 pixels each (shown in Fig. 2.2). Each quadrant contains one 12 bit Time-to-Digital Converter (TDC) with a TDC resolution of less than 100 ps. Details on the TDC are given in section 2.3.
The Caribou system (described in subsection 2.2.1) provides the 3 MHz bunch clock signal and uses it to generate a 408 MHz clock signal, which is used as the system clock in the chip. Both the 3 MHz and 408 MHz clock signals have to be provided externally to the chip by Caribou.

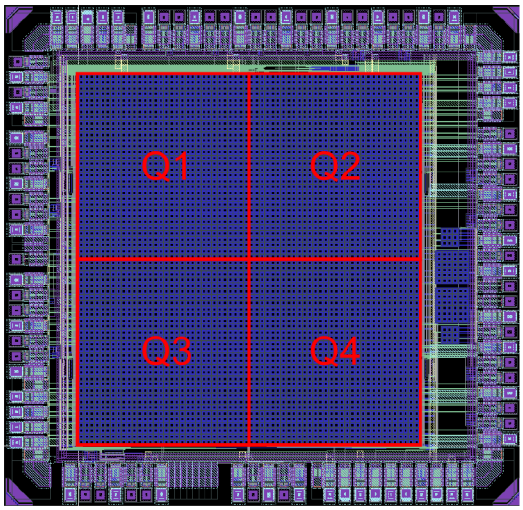


Figure 2.2: ASIC layout of the dSiPM with the 16×16 pixel quadrants [5].

The acquisition window comprises 128 of the 136 system clock cycles building one frame, which is about 314 ns out of the 333 ns. The residual 8 clock cycles, 3 at the beginning of the frame and 5 at the end of the frame, are used for counter resetting and data storing.
If a pixel fires within the acquisition window, a hit flag for the pixel is set in the frame and the timestamp in the corresponding quadrant is stored. If a different pixel in the same frame fires after that, the corresponding hit flag will be set but the corresponding timestamp is not stored. This means that only the first hit sets the timestamp for the quadrant in the frame. The chip has an optional 2 bit mode, in which the hit flag for each pixel is replaced with a 2 bit counter. This allows to count up to three hits per pixel in a single frame.
The chip also provides a fast wired-OR, which can be used for a trigger output if any pixel fires. This allows the chip to be used as a fast trigger, for example in testbeam environments. Additionally, the chip has a four-step validation logic. Each row is connected in pairs to a configurable AND/OR logic gate. These logic gates are then again connected in pairs to another logic gate. This is cascaded in four steps such that there is only one final "valid bit". This can be used in high light conditions (e.g. fiber readout) to filter out noise hits that only consist of single pixel clusters.

2.2.1 Caribou DAQ System

Every detector requires a data acquisition (DAQ) system, whoose task is to convert the collected data of the detectors to digital data that can be stored and analyzed. Additionally, the DAQ system also takes care of the control, configuration, and powering of the detector. The dSiPM relies on the Caribou DAQ system [20], which allows for fast integration of new detector prototypes. The hardware can be split into three components: a System-on-Chip (SoC) board, a Control-and-Readout (CaR) board, and a chipboard. The SoC board consists of an FPGA/CPU combination, that allows for communication with the chip and running the DAQ software. The CaR board hosts various components commonly required to operate a sensor, such as a power supply, a clock generator, and an Analog-to-Digital Converter (ADC). The chipboard hosts the sensor and detector-specific components. The three components can be seen in Fig. 2.3. The chipboard of the dSiPM is encased in an aluminum case to shield the sensor from external light sources and to act as a heat sink [5]. The Caribou DAQ system provides open access to hardware schematics, common firmware components, and the DAQ software Peary [21]. The system has been used in various detector prototypes already such as in the CLIC detector pixel prototypes or the ATLASPix series [20].

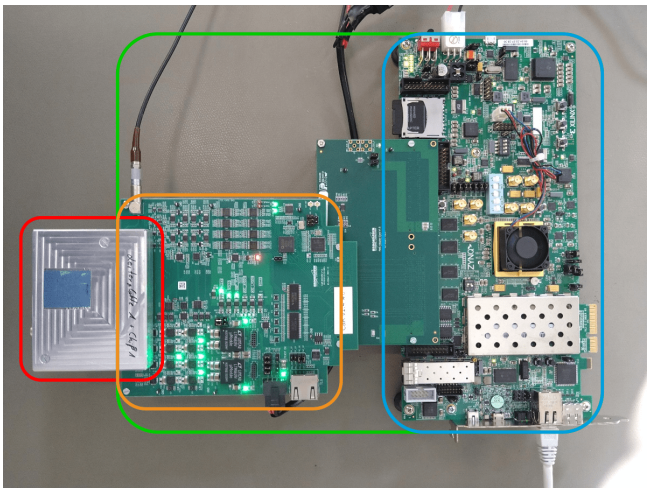


Figure 2.3: Caribou DAQ system (green), composed of the SoC board (blue) and the CaR board (orange), equipped with the dSiPM chipboard in an aluminum housing (red).

2.3 Time-to-Digital Converter

Time-to-Digital Converters (TDCs) are circuits used to convert the time information of an event into a digital data format. There are different types of TDC designs, three of which will be presented here to explain the TDC used in the dSiPM.

2.3.1 Basic Counters

A simple implementation of a TDC is a counter on a clock signal. Every time a new clock cycle begins, the counter increases its value by 1. To measure the time of an event, the counter is stopped and its value is read out. The resolution of such a TDC scales with the frequency of the clock signal.
For example, a counter on a 5 MHz clock signal would have a time binning of 200 ns, which when assuming a binary distribution results in a time resolution of $\frac{200\,\mathrm{ns}}{\sqrt{12}} \simeq 58\,\mathrm{ns}$.[1]
Thus, to have a good time resolution, high clock frequencies are required. However, higher frequencies also require counters that can update their count within the time of one clock cycle. This directly depends on the switching speed of the logic gates and the length of the signal traces in the counter, which is given by the CMOS process used for the TDC.
Due to this limit, many TDCs with high time resolution use a hybrid design utilizing a low-frequency counter with a large range in time and combine it with a fine TDC design like a tapped delay line for higher resolution.
One type of counter is a ripple counter. In a ripple counter, the clock is only fed into the least significant Flip-Flop, and the output of that Flip-Flop is used for the next bit. They are also called asynchronous counters since the bits do not change synchronously. Compared to synchronous counters, ripple counters can be realized with fewer transistors, which is beneficial if a synchronous count is not required.

2.3.2 Tapped Delay Line

A Tapped Delay Line (TDL) is a line of delay elements, where each element introduces a delay τ to its input. To measure time with this circuit, a start signal is injected into the delay line at a known time. When an event occurs, the signal after each delay element is read out. The time of the event relative to the start signal can be calculated by counting how far the start signal propagated through the line of delay elements.
Typically, TDLs are combined with Flip-Flops or D-Latches that are placed after every delay element. A stop signal is then emitted from the event to store the state of the delay line in the Flip-Flops. A circuit diagram of a TDL with D-Latches is given in Fig. 2.4.
The native output of a TDL is thermometer code, which stores the state of every Flip-Flop. In a TDL with 32 delay elements, the thermometer code would be 32 bit wide.
In practical realizations of TDLs, it is nearly impossible to ensure that all delay elements have the same delay τ due to fluctuations in the manufacturing process. A common measure to describe the performance of a TDL-based TDC is the Differential Non-Linearity (DNL). It is defined as

$$\mathrm{DNL}_n = \frac{\tau_n - \tau}{\tau} \tag{2.1}$$

for the nth delay element, where τ_n is the delay length of the nth delay element. It is typically given units of Least Significant Bit (LSB), i.e. as fraction of the design delay length τ.
Since these non-linearities can add up over the length of the delay line, thus another measure is the Integral Non-Linearity (INL). It is defined as

$$\mathrm{INL}_n = \sum_{j=1}^{n-1} \mathrm{DNL}_j \tag{2.2}$$

for the nth delay element.

[1] This follows directly when applying the definition of the variance to a uniform distribution: $\sigma^2 = \int_{-t/2}^{t/2} x^2 \frac{1}{t} \mathrm{d}x = \frac{t^2}{12}$

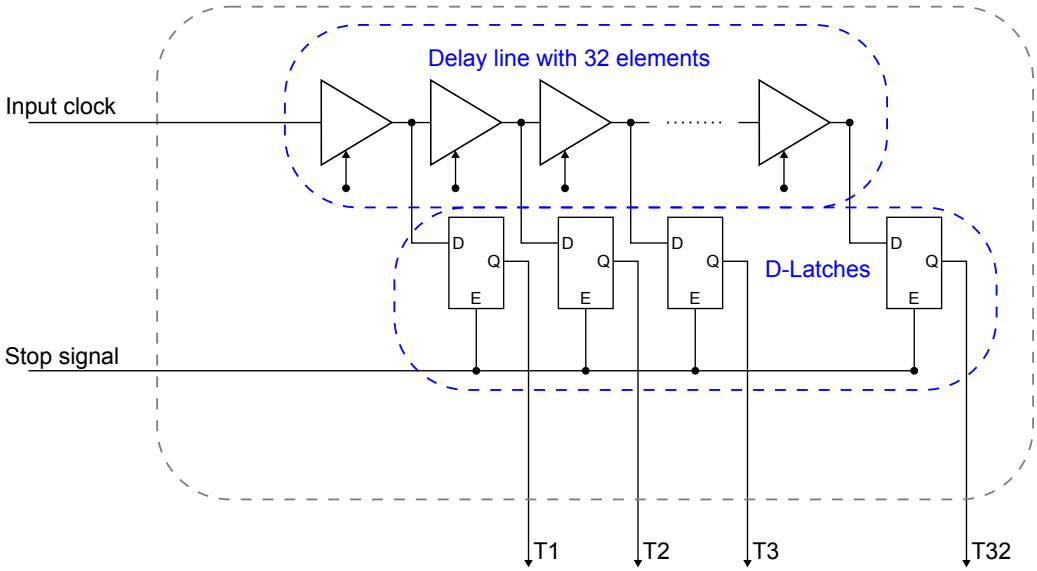


Figure 2.4: Schematic of a 32 bit Tapped Delay Line TDC. The clock enters the delay line from the left. After each delay element the signal is also routed to the data pin D of the D-Latch. The stop signal enters the enable pin E of the D-Latch, which stores the value at D. The stored value can be read from Q.

2.3.3 Delay Locked Loop

Another challenge is ensuring that the delay τ stays constant over long periods of time. Some delay elements like inverters require a bias voltage that controls the delay length of the element. To ensure stable conditions, a Phase Locked Loop (PLL) or Delay Locked Loop (DLL) can be employed.
In a DLL, the output of a periodic (clock) signal after the delay line is compared to the signal before the delay line in a phase detector. In the ideal case, the signal after the delay line is delayed exactly by the period of the clock frequency. The phase detector can detect if the delay signal is either too fast (delayed clock signal appears to be in front of the reference clock signal) or too slow (delayed clock signal appears to be running behind the reference clock signal).
The phase detector is coupled with a charge pump, such that the bias voltage is adjusted to either increase or decrease the delay length of the delay elements. This builds a positive feedback loop that ensures that the length of the delay line stays constant over long periods of time. A circuit diagram of a DLL can be found in Fig. 2.5.
TDCs based on a TDL often employ the concept within a DLL circuit for higher stability.

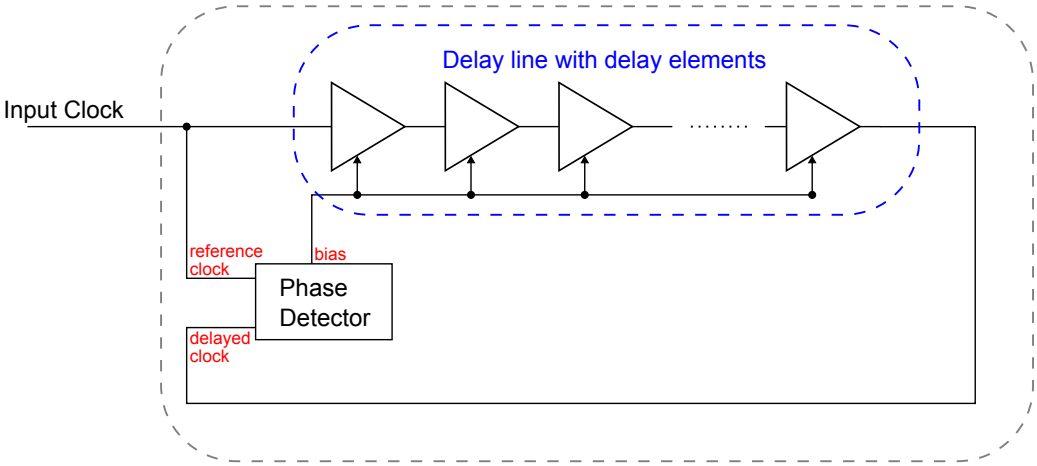


Figure 2.5: Schematic of a Delay Locked Loop. The clock enters the delay line from the left. The clock signal after the delay line is routed to a phase detector, that adjusts the length of the delays if the phase of the delayed clock signal is too early or too late.

2.3.4 dSiPM TDC

The TDC of the dSiPM is a combination of three different TDC types. The base is a counter in the FPGA that counts the number of frames since a start signal, thus running on the 3 MHz bunch clock. The dynamic range is 40 bit, which results in a maximum measurement length of about 100 h.
On the chip, the "coarse TDC" gives the course timestamp within the frame. It is a 7 bit ripple counter running on the 408 MHz system clock, thus having a bin width of about 2.45 ns.
Lastly, the "fine TDC" is a 5 bit TDL with a DLL running on top of the 408 MHz system clock, meaning that it divides each coarse clock cycle into 32 fine bins. The theoretical bin width is 76.6 ps.
Fig. 2.6 shows a sketch of the fine TDC. In addition to the TDL-based TDC presented in Fig. 2.4, the TDC also includes a thermometer code encoder. It compresses the 32 bit thermometer code to 5 bit by just storing the last delay element that was reached.
More detailed schematics are given in *Diehl et al. (2023)* [5].

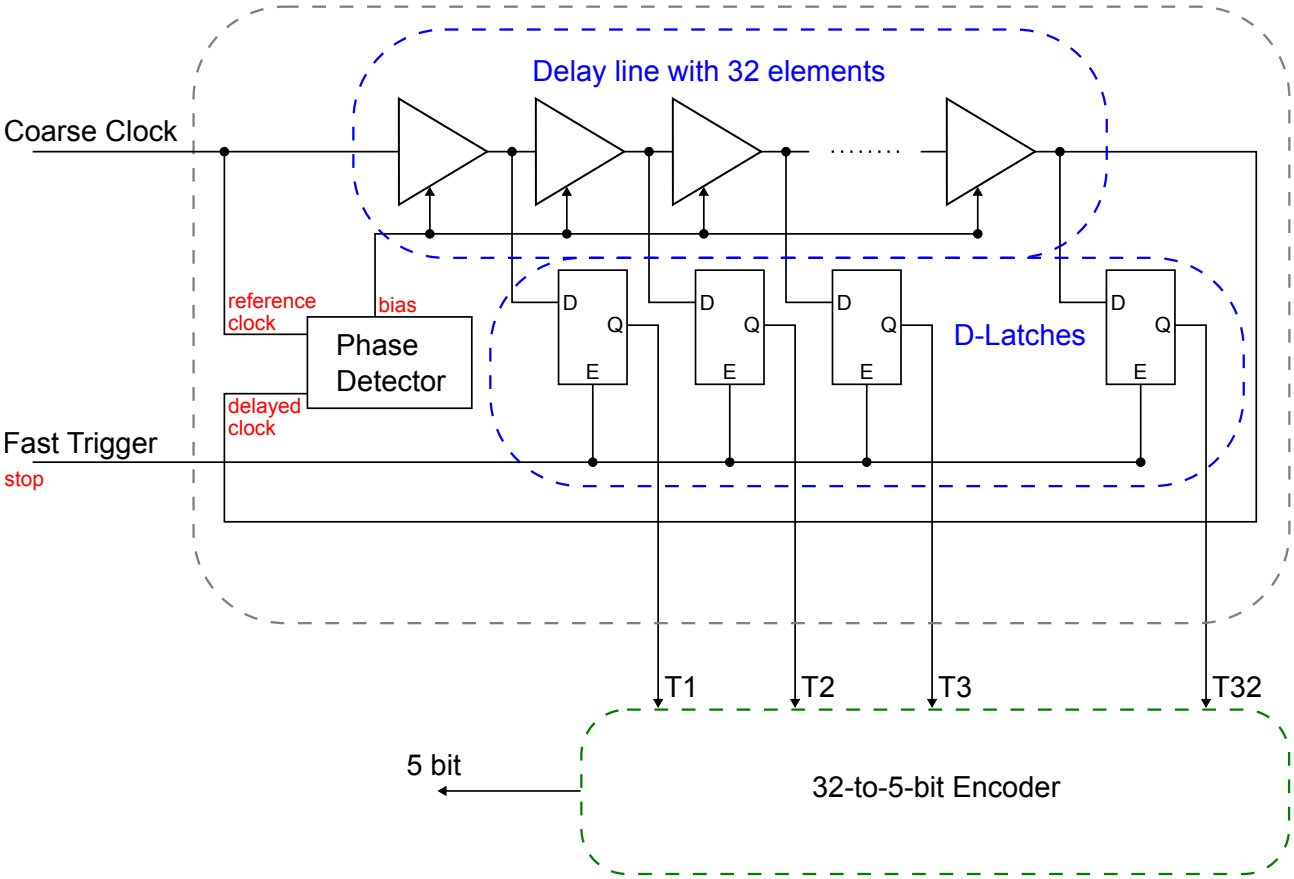


Figure 2.6: Schematic of the fine TDC of the dSiPM. It combines a 32 bit TDL running on a 408 MHz clock with a DLL. The thermometer code of the TDL is encoded to a 5 bit value.

Chapter 3

Testbeam Setup

3.1 Testbeam Concept

A common method to study the characteristics of pixel sensors is a testbeam measurement, where the sensor is placed in a particle beam. The detector that is being tested is referred to as Device-Under-Test (DUT). In a testbeam measurement other, usually well-studied, detectors are used to determine certain properties of each particle passing through the DUT. Examples are the particle type or its trajectory, however the setup can vary between different detector types and testbeam facilities.
Since a pixel sensor is studied in this thesis, testbeam measurements for such a device are assumed for the rest of this chapter. Measurements with a pixel sensor in a testbeam environment fulfill usually at least one of four aspects:

- Testing the Data Aqusition (DAQ) system.
- Measuring the particle detection efficiency of the sensor.
- Measuring the spatial resolution of the sensor.
- Measuring the time resolution of the sensor.

For all these aspects, measuring the particle trajectory precisely is a key requirement.

3.2 dSiPM Testbeam Campaign

In March 2023 a testbeam was conducted at the DESY II testbeam facility [22] with two dSiPMs as DUTs. The DUTs were tested in a 4 GeV electron beam. A chiller provided cooling for the DUTs. Runs were taken with a chiller temperature of −24 °C (about −9 °C on chip[1]) and −14 °C (about −4 °C on chip[1]) for overvoltages ranging from 0.7 V to 3.0 V.
An overview of the setup is given in Fig. 3.1. From the given perspective, the beam is entering the setup on the right side. The different components shown in the figure will be explained in the following sections.
In this thesis, the focus will be the reconstruction of the time resolution from the data acquired at this testbeam campaign. The spatial resolution and efficiency are studied elsewhere [23].

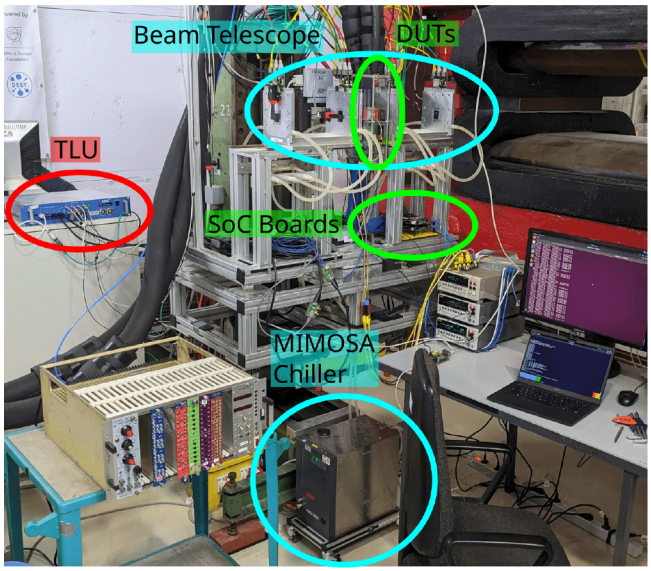


Figure 3.1: Picture of the testbeam setup.

[1]The on-chip temperature diodes were not calibrated.

The breakdown voltage of the dSiPMs was determined using the relative derivate method for each temperature and pixel mask. In this method, an IV-Curve is measured and the relative derivate $\mathrm{d}\ln I/\mathrm{d}V$ is taken. The breakdown voltage V_{BD} is then defined as the peak of the relative derivate.

3.3 Particle Tracking

An important measure for the spatial performance of a DUT are spatial residuals, given via

$$\Delta x = x_{\text{track}} - x_{\text{hit}} \tag{3.1}$$

where x_{track} is the position of a particle passing through the DUT in one axis and x_{hit} the hit position of that particle given by the DUT in the corresponding axis. The width of the spatial residual distribution is a measure of the spatial resolution of the DUT, which is typically given as standard deviation or full width half maximum (FWHM). It is important to note that the distribution also depends on other factors, such as the tracking resolution, which should ideally be significantly smaller than the expected spatial resolution of the DUT.
If the DUT has an ideal spatial resolution determined only by the pixel pitch and x_{track} is known with infinite precision, the spatial residuals would be distributed uniformly in a region given by the pixel pitch. In practice, there are physical effects like charge sharing to consider and the particle position is not known with infinite precision.

To determine the position of the particle when it passes through the DUT the track of the particle needs to be reconstructed. For this, a so-called beam telescope is used. It consists of several other pixel sensors (planes), where usually three planes are placed upstream and three downstream of the DUT. If the particle is detected by more than one telescope plane, a track can be reconstructed from which the position of the particle at the DUT can be calculated.
It is important to keep in mind that the calculated track interception with the DUT has a limited resolution. This tracking resolution depends on the resolution of each telescope plane, the particle type being tracked and its momentum, the material budget of the setup, and the geometry of the telescope, i.e. the distance between the telescope planes and the DUT along the beam direction.
In particular, if the measured particles experience a lot of multiple scattering (see subsection 1.2.2) like low energy electrons, it is important that the material budget is reduced as much as possible. This is because scattering is a stochastic process and thus reduces the precision of the particle track in the reconstruction.

A key aspect of achieving accurate spatial residuals is good alignment. The alignment describes the position of the telescopes and the DUT relative to each other. Mechanical measurement of the positions only has a very limited precision, given that modern pixel sensors have a pixel pitch of less than 100 µm. Thus the alignment is refined in software. This is described more in detail in section 4.5.

In the testbeam from March 2023, the beam telescope was provided by DESY. It is a EUDET-type telescope [24], consisting of six MIMOSA26 telescope planes with a spatial resolution of about 3.5 µm. A hybrid approach for the geometry was chosen, where two planes were placed close to the DUTs on each side and the other two planes far apart from the DUTs. A study of the tracking resolution in this testbeam is shown in Fig. 3.2. The tracking resolution was calculated using the *GBL Track Resolution Calculator* [25] and gives a resolution at the DUT positions of about 3.4 µm for the upstream DUT (dSiPM_0) and 3.7 µm for the downstream DUT (dSiPM_1).

3.4 Triggering

In the last section, the tracking of the particle was described. Up until now, it was assumed that the detectors just store the data when a particle passes through the setup. This process usually does not happen automatically and thus requires a triggering scheme. In short, a trigger is a signal distributed to all detectors to inform them that a particle just crossed the setup and data needs to be stored.
This mainly serves two purposes: on the one hand it ensures that data is only stored when a particle

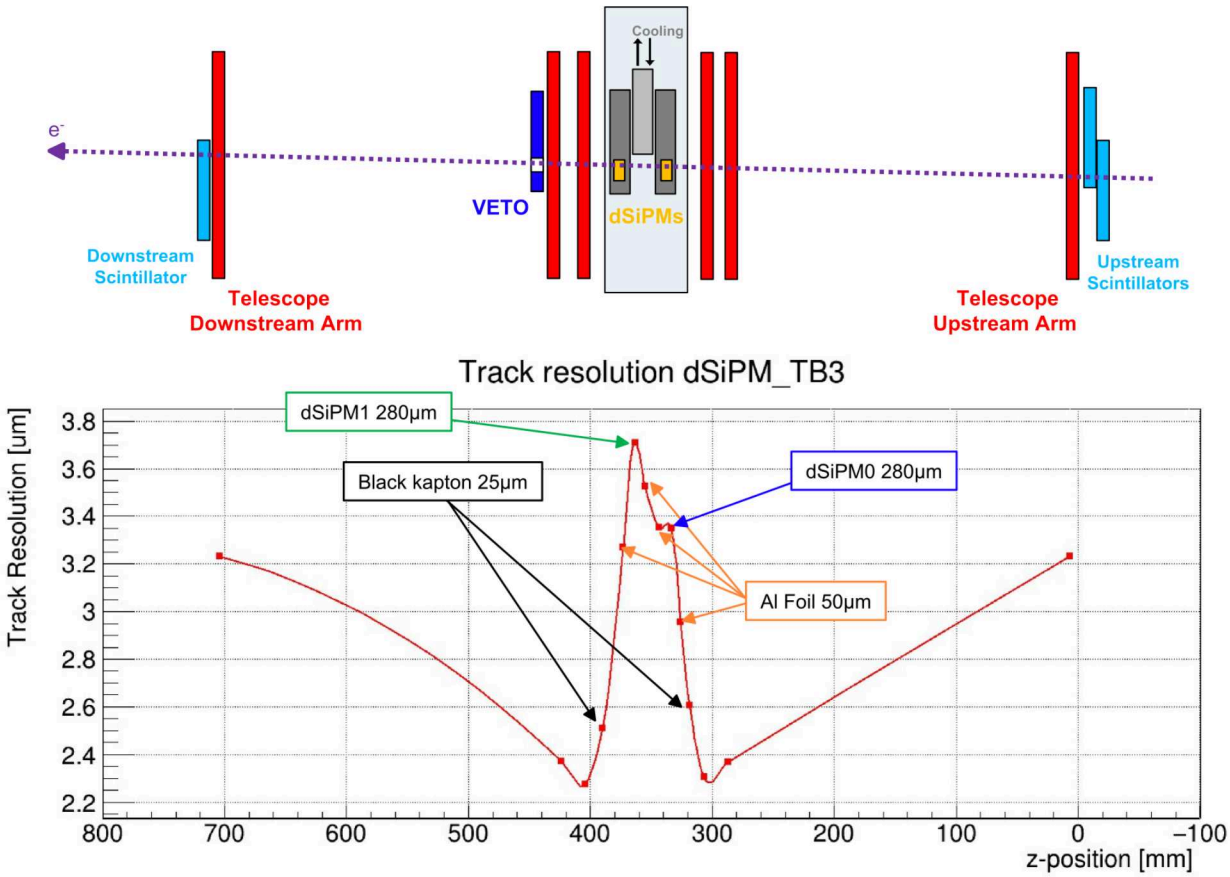


Figure 3.2: Tracking resolution of the the testbeam setup [15].

actually traversed the setup. On the other hand, it also provides synchronization of the data if the triggers are counted and the trigger number is attached to the data.[2]

To provide trigger signals, some particle detector is required. Typically a combination of scintillators is used, but this can also be provided by pixel detectors (e.g. [26]). In most setups, a device providing the trigger directly is not enough. One reason is that some detectors require some time for the readout of the data, in which they are busy and cannot process a second trigger. If the other detectors are not aware that one detector is busy and consequently the trigger is dropped, the trigger numbers go out of sync. A device that can solve this and many other issues related to triggering is called a Trigger Logic Unit (TLU). Its main purpose is to combine and discriminate the trigger inputs and distribute them to the detectors if they are able to receive a trigger signal. A TLU can also provide synchronous clocks to the detectors. This is important when performing timing measurements. If the clocks between the time reference and the DUT are not synchronized, they can drift apart over time.

In this testbeam, a combination of three scintillators and one veto scintillator was used (see Fig. 3.2). A trigger signal is only sent if a particle was detected in three scintillators but not in the veto scintillator, which has a small hole in the middle. This means that only particles that pass through this small hole lead to a trigger signal being sent out.
For the TLU, an AIDA-2020 TLU [27] is used. The DUTs are operated in the synchronous AIDA mode. In this mode, the TLU provides 40 MHz clock signal for time synchronization. At the beginning of a data-taking run, a T0 is distributed to reset the clock counters. A busy signal is sent to the TLU for the duration of the data readout once a trigger is received.

[2]It is worth mentioning that there are many different DAQ schemes, and some of these schemes do not rely on trigger numbers to synchronize.

3.5 Data Acquisition Systems

The main task of a data acquisition (DAQ) system is to ensure that the data collected by the detectors can be stored in a digital format for later analysis. The DAQ system of the dSiPM was described in subsection 2.2.1. However for a testbeam analysis, the different detectors with different DAQ systems need to work together. In this particular testbeam the EUDAQ2 framework [28] was used to control the DAQ systems of the two dSiPMs, the TLU and the telescope.
The most important aspect is to start and stop "data runs" and provide network features to send data from the detector to a storage device. Timing critical aspects such as clock synchronization have to be implemented in hardware and are thus provided by the TLU.

3.6 Timing Measurement

A time reference is required in order to measure the time resolution of a DUT. Similar to spatial residuals given in Eqn. 3.1, time residuals can be defined as

$$\Delta t = t_{\text{ref}} - t_{\text{DUT}} \tag{3.2}$$

where t_{ref} is the timestamp provided by the reference detector for a given particle and t_{DUT} the timestamp of the DUT for that particle. The distribution of the time residuals is a measure of the DUT time resolution. As with the spatial residuals, this distribution also includes the time resolution of the reference.

Ideally, the time reference has a faster expected time resolution than the DUT, however this is not always easily possible. For example, the MIMOSA26 sensors in the EUDET have a rolling shutter with an integration time of 115.2 µs. This is significantly higher than the time resolution of many modern pixel detectors.
In the case of the dSiPM with an expected time resolution of less than 100 ps, acquiring a faster timing reference that is well studied and easy to integrate into testbeam is quite challenging. The solution chosen in this testbeam is taking two dSiPMs and using one as a timing reference, which suffices to demonstrate a time resolution below 100 ps.
Additionally, the AIDA TLU also provides timestamping for the triggers with a bin width of 781.25 ps. However, the time resolution of the timestamps from AIDA TLU depends on the time resolutions of its trigger input. In this case, this is the jitter of the scintillators and the photomultipliers, which is difficult to estimate.

In this thesis, the focus lies on digitized timestamps and not timestamps extracted from waveforms. This means that both the DUT and the reference require synchronized clocks. A reference clock signal for this is provided by the TLU.

In the case of digitized timestamps, the components contributing to the time resolution σ_{DUT} can be described by the following equations:

$$\sigma^2_{\text{DUT}} = \sigma^2_{\text{TDC}} + \sigma^2_{\text{INL}} + \sigma^2_{\text{Delay}} + \sigma^2_{\text{Phys}} \tag{3.3}$$

$$\sigma^2_{\text{TLU}} = \sigma^2_{\text{TDC}} + \sigma^2_{\text{Sci}} \tag{3.4}$$

$$\sigma^2_{\text{TLU-DUT}} = \sigma^2_{\text{DUT}} + \sigma^2_{\text{TLU}} \tag{3.5}$$

$$\sigma_{\text{DUT-DUT}} = \sqrt{2} \cdot \sigma_{\text{DUT}} \tag{3.6}$$

where σ_{TDC} is the binary TDC resolution, σ_{INL} the contribution from INL effects[3] (see subsection 2.3.2), σ_{Delay} the contribution from different delay lengths from the pixels to the TDC, σ_{Phys} the intrinsic time resolution of the DUT, and σ_{Sci} the jitter of the scintillator combination. Note that for Eqn. 3.6 it is assumed that the non-TDC contributions are not significantly smaller than the binary TDC resolution σ_{TDC}, in which case the phase between the two TDCs becomes a significant factor.

[3] It is assumed that INL effects of the TLU are negligible.

Chapter 4

Analysis Chain

4.1 Overview

In the previous chapter, the hardware side of a testbeam was discussed. In this chapter, the analysis side of a testbeam is discussed. The Corryvreckan testbeam data reconstruction framework [29, 30] is used to perform the various analysis tasks. It uses a modular concept with several steps in the analysis chain. An overview of this analysis chain is shown in Fig. 4.1.

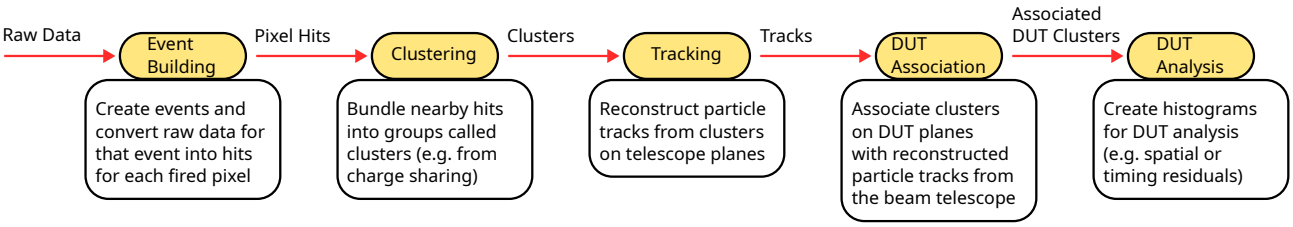


Figure 4.1: Overview of the analysis chain in Corryvreckan.

The first step in every analysis chain is the event building from the raw data, which turns the raw data into events with pixel hits of the detectors. The next step is clustering, which groups nearby pixel hits into groups called clusters, that correspond to particle hits. With these clusters, reconstruction of the particle tracks can be performed using the telescope planes. After tracking, the DUT data of the event can be searched for a cluster that could belong to the particle track. Finally, an analysis of various performance metrics can be performed, such as efficiency, spatial resolution, and time resolution.
Selection criteria can be applied at every step. These selection criteria can impact the results of the final performance and thus need to be selected carefully. This is especially important for efficiency studies. Too harsh selection criteria can reduce statistics and remove good events, too lax selection criteria might include too much noise or random combinations of hits to form tracks.

4.2 Data Decoding

One important part of the event building is reading the raw data and converting it to a data structure from which events can be created. This is detector-specific and can be highly non-trivial depending on the readout scheme.

In the testbeam studied in this thesis, data was recorded with the EUDAQ2 framework [28] (see section 3.5). It defines data structures for raw events, in which the raw data is stored during data taking, and standard events, which include the fired pixels and timestamps for one or more planes. The conversion between these two data structures is implemented in a EUDAQ event converter.
For the dSiPMs, a raw event includes the three frames collected for one trigger. Peary, a DAQ software framework for the Caribou DAQ system [21], is used to convert the raw frames into a usable data structure. It includes the list with all pixel hits and the clock counter values for every quadrant, which

can be used to assign every pixel hits a timestamp.
Before the hits can be passed on to Corryvreckan, invalid data needs to be filtered out. This is something that should only be performed after it is understood which data is invalid and why. In this testbeam, invalid data was found, which is described in detail in subsection 5.1.1. As part of this thesis data filtering of the observed invalid data has been implemented in the EUDAQ2 event converter.[1]

In addition to the filtering of invalid data, the correction of differential non-linearities in the TDC needs to be applied in the event converter as well since Corryvreckan only receives the final timestamps for every pixel and not the counter values. The DNL correction was implemented as part of this thesis.[2] Timestamps are calculated via

$$t_{\text{quadrant}} = \frac{\mathrm{c}_{\text{frame}} - 1}{3\,\text{MHz}} + \frac{\mathrm{c}_{\text{dead}}}{408\,\text{MHz}} + \frac{\mathrm{c}_{\text{coarse}} - 1}{408\,\text{MHz}} + \sum_{j=1}^{\mathrm{c}_{\text{fine}}-1} \tau_j + \frac{\tau_{\mathrm{c}_{\text{fine}}}}{2} \tag{4.1}$$

where $\mathrm{c}_{\text{frame}}$ is the value of the bunch counter, c_{dead} the amount of clock cycles of the dead time at the beginning of a frame (three clock cycles), $\mathrm{c}_{\text{coarse}}$ the counter value of the coarse TDC, c_{fine} the counter value of the fine TDC, and τ_n the DNL-corrected bin width of the nth fine TDC bin. Since the three counters start counting from one instead of zero, a counter value of zero is converted to the highest possible counter value plus one. Additionally, one has to be subtracted from $\mathrm{c}_{\text{frame}}$ and $\mathrm{c}_{\text{coarse}}$ to calculate to correct timestamp. The last term in Eqn. 4.1 corresponds to placing the timestamp in the middle of the measured fine TDC bin, which is a natural approach in delay line based TDCs (see subsection 2.3.2).

4.3 Event Building

While Corryvreckan has a flexible algorithm allowing for different event-building schemes, event-building is done purely with triggers for this testbeam. Here, an event is defined by a trigger of the TLU. For all subsequent detectors added to the chain, data is selected such that the trigger number matches the one from the TLU. The data of each detector is read by Corryvreckan's `EventLoaderEUDAQ2` module. For the dSiPMs only the first two of the three frames are read since it was found that the third event only contains noise hits.

An alternative to this concept would be to load the data of one dSiPM, defining an event for every frame. Next, the TLU would add a trigger number to the event in which its timestamp would belong. The telescope data would be added based on a matching trigger number. In principle, this has the advantage of including the dead time in the analysis for a more precise determination of the efficiency. One drawback however is that the frames of the dSiPMs do not necessarily overlap exactly since the phase of their frame clocks is not synchronized. Another one is the large integration time of the telescope. Sometimes more than one track is found for one event and as a result, it is not possible to determine which track belongs to which frame. In conclusion, it was decided that the trigger-based approach is sufficient for the timing study in this thesis.

Due to the measurement artifacts described in subsection 5.1.2, a filter that can skip events based on the trigger number had to be added. This has been added to the `FilterEvents` module in Corryvreckan.[3]

[1] Implemented in EUDAQ2 PR#715.
[2] Implemented in EUDAQ2 PR#718.
[3] Implemented in Corryvreckan MR!629.

4.4 Track Reconstruction Chain

4.4.1 Clustering

The first step in the tracking chain is the clustering of nearby pixel hits on a plane. The rationale behind this is that a single particle might create a hit in more than one pixel, for example due to charge sharing (see subsection 1.4.1) or in the case of SiPMs optical crosstalk (see subsection 1.4.2). For this thesis, Corryvreckan's `ClusteringSpatial` module is used, which only takes into account spatial information. Corryvreckan also includes a module that includes timing information in the clustering, only grouping pixels with a close timestamp. However, the MIMOSA26 telescope planes do not provide per-pixel timing information and for the dSiPM the limited amount of TDCs on the chip limits the advantages of this approach.
The only cut applied at this step is to ignore pixel hits outside of a defined region of interest (ROI) on the telescope planes to reduce the number of tracks that would never pass through the DUTs. The ROIs were chosen to be slightly larger than the active DUT area projected on each plane.

4.4.2 Tracking

The actual particle tracking is performed with Corryvreckan's `Tracking4D` module. It starts by creating straight lines between all clusters in the first and last plane. Planes in between are added if there they have a cluster within a given search radius. Then the track fit is performed, for which the General Broken Lines (GBL) algorithm [31] algorithm was chosen. GBL tracks allow scattering (see subsection 1.2.2) at and between planes, which gives a better description than straight lines for 4 GeV electrons.
The spatial search radius was chosen to be 200 mm. Although this is rather large compared to the pixel pitch to allow for larger scattering angles, the tracks are filtered by their χ^2/n_{dof} later in the analysis anyway.
The tracking module only takes clusters on telescope planes into account when performing the track fit. This avoids biasing the track when performing spatial studies on the DUTs. Spatial residuals where the corresponding clusters have been used during the tracking are thus called biased residuals, and spatial residuals where the corresponding clusters have not been used during the tracking unbiased residuals. The module is also configured to require a cluster on all six telescope planes for each track.

4.4.3 DUT Association

After a track is created from the clusters of the telescope planes, the cluster on the DUT associated with this track can be searched. This is performed with Corryvreckan's `DUTAssociation` module. It searches for a cluster on the DUT within a given search window to the fitted track intersect, which was chosen to be one pixel, i.e. 69.6 µm in X and 76 µm in Y.

4.5 Alignment

The tracking chain described in section 4.4 assumes that the position (and rotation) of the telescope and DUT planes in the X-Y plane is known with high precision compared to their respective spatial resolution. In the case of the MIMOSA26 chips, measuring the position manually with a resolution of less than 3.5 µm is quite challenging. Thus the position is determined in software using a process called alignment.
Aligning the planes is a multi-step process, refining the position in every step. The first step of this process is usually prealignment, where planes are aligned with correlations relative to a reference frame. Correlations are the difference between cluster positions on one plane compared to the cluster positions on another plane that are within one event. In the Corryvreckan's `Prealignment` module, the correlations in X and Y for all possible clusters on the reference plane are stored in histograms. If the planes are aligned with each other, the distribution should be centered around zero. If not, one plane can be shifted relative to the reference plane such that this is the case.

To get a better estimate of the position of the planes as well as their rotation, tracking as described in section 4.4 can be used. After the tracking is performed, the alignment of the telescope is optimized by minimizing the sum of the χ^2 values of all track fits. This means that in every fit iteration, the planes' alignment is changed according to the fit parameters, all tracks are refitted to the clusters of that track, and the χ^2 values of all fitted tracks are summed up. This method is implemented in Corryvreckan's `AlignmentTrackChi2` module using ROOT [32, 33] for minimization.
This method was used to align the telescope planes. Multiple iterations with different selection criteria were used using data from the run with the highest statistics (run 5545).
The alignment of the DUTs has to be performed differently to avoid biasing the alignment of the telescope planes. This is performed with Corryvreckan's `AlignmentDUTResidual` module. The module requires an aligned telescope and a tracking chain, after which the unbiased residuals of the DUTs are minimized.
The method was run for multiple iterations and performed for every run for maximum precision. However in long runs, it was found that the alignment of the DUT planes was still insufficient. It was suspected that the cause might be a misalignment over time. To test this, an alignment was performed using only a small portion of the data from the beginning of the run, which was then used to perform an analysis of the entire run. In Fig. 4.2 the unbiased residuals in Y of one plane from this analysis is shown (red data points). It can be seen that the residuals shift over the duration of a long run by almost 15 µm, confirming the hypothesis. Possible causes are mechanical changes in the position stage due to temperature changes in the testbeam hall.

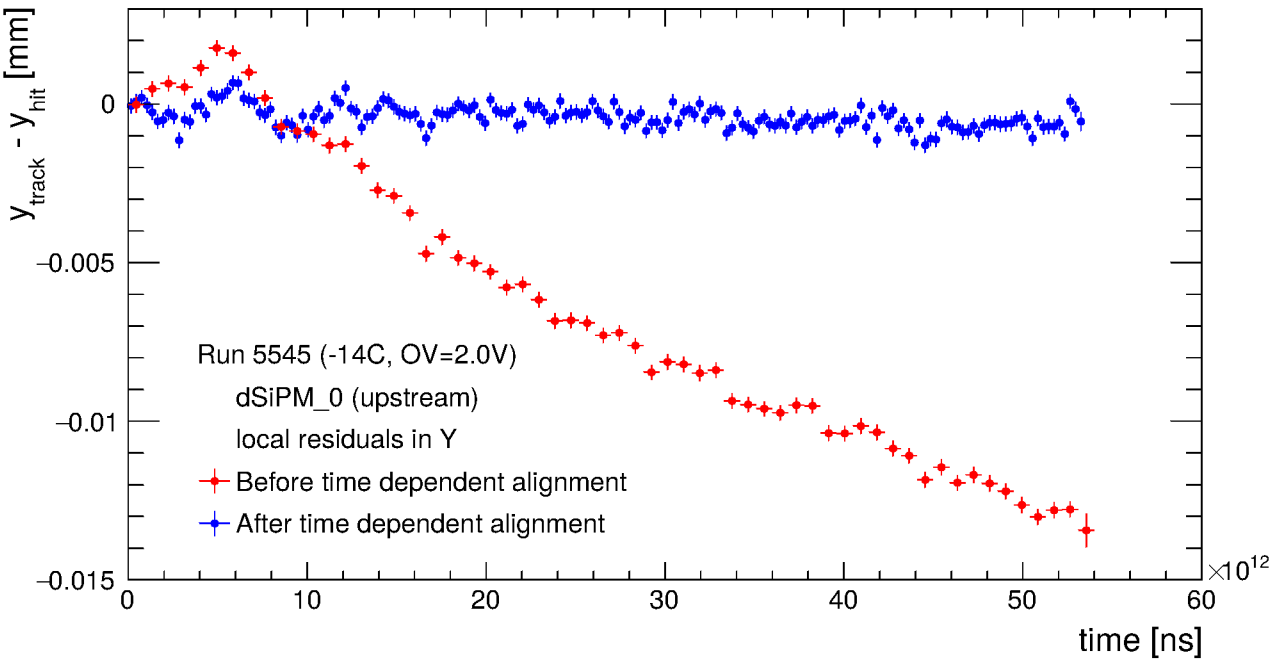


Figure 4.2: Alignment over time with and without time dependent alignment (run 5545).

To mitigate this, time-dependent alignment[4] (TDA) can be used. In TDA an alignment is performed using only some data in the beginning of the run. Then, the residuals over time in X and Y are fitted with a polynomial. This polynomial is then added to the static alignment. In Fig. 4.2, the blue data points show the unbiased residuals after TDA. They are within 1 µm, which is significantly better than before.
One possible pitfall with TDA is to overfit the residual distribution. If the polynomial with a large order and fine binning is used, the fit might artificially improve or even degrade the spatial resolution. Thus TDA was only applied in long runs that had suspicious residual distributions with a 4th order polynomial in X and a 6th order polynomial in Y.

[4]See Corryvreckan MR!542 for more details.

The Z positions of the planes (both telescope and DUT) have been measured relative to the reference plane by hand using a ruler. No further alignment for the Z positions has been performed.
Another alignment step independent of the spatial alignment is the alignment in time. This means that the time offset between the DUTs and the TLU needs to be found, which was performed using time correlations.

4.6 DUT Analysis

The last step in the analysis chain is performing the actual analysis of the DUTs. This step can vary depending on the goals of the study but usually includes spatial resolution studies and efficiency studies. However, both these things will be studied in a future paper [23] and thus will not be discussed in detail here. In particular, the study of the efficiency is not trivially possible due to measurement artifacts described in subsection 5.1.1. Instead, the analysis chain used in this thesis focuses on the timing analysis of the DUTs.

4.6.1 Timing Analysis

At the time when the testbeam was performed, Corryvreckan had no built-in possibility to calculate time residuals (Eqn. 3.2) between DUTs. Calculating the time residuals between a DUT and the track timestamp provided by the TLU is relatively simple since an associated cluster on the DUT is always associated to a track with a timestamp. This is not true for the time residuals between two DUTs, since one DUT might be inefficient. As part of this thesis, a new `AnalysisTiming` module has been implemented.[5] It allows for timing studies between two arbitrary planes by looping over all tracks and checking if a cluster is available on both detectors. It uses the timestamp of the closest associated cluster for DUTs and the timestamp of the cluster used in the reconstruction of the track for other detectors to calculate the time residual. The module implements a track selection discarding tracks with high χ^2/n_{dof}. Additionally, the module provides various 2D histograms to study the timing performance on a sensor and pixel scale. An important aspect of the implementation was to make this module flexible and not specific to the dSiPM detector, such that it can be used with other sensors in the future.

For the purpose of this thesis, the output data of this module has been exported to a `TTree` in order to conduct specific binning and in-pixel tests. A track selection criterion of $\chi^2/n_{\text{dof}} < 5$ was chosen.

[5] Implemented in Corryvreckan MR!633 and MR!640.

Chapter 5

Testbeam Measurements

5.1 Measurement Artifacts

5.1.1 Unphysical Timestamps

An interesting observation was made during the first analysis of the testbeam data. In Fig. 5.1, the raw time residual is plotted over time. In this context, raw time residuals are extracted by reading the raw data of the DUT and the TLU for every trigger and subtracting the TLU timestamp from the DUT timestamp in the event where at least one hit is available. If multiple hits are available the first frame and/or last quadrant with a hit is preferred. The spatial information is not used.

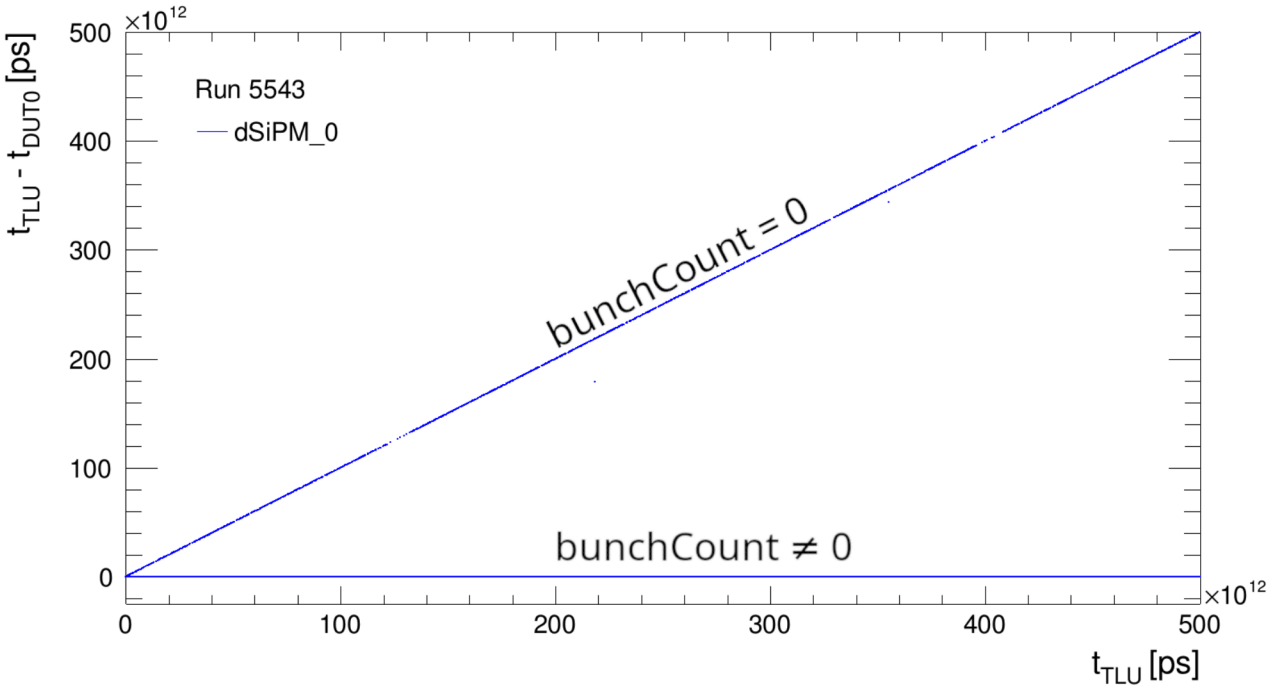


Figure 5.1: Raw time residuals over time with zero timestamps (run 5543).[1]

The time residual is expected to always be in a certain window defined by the time resolution of the setup. However, two lines can be seen in Fig. 5.1, one is constant as expected and one increases linearly with slope 1. On further inspection, it was found that for entries in the linearly increasing line, the bunch counter is at 0. This explains the slope of 1.

[1] At $t_{\mathrm{TLU}} = 125\,\mathrm{s}$ and $t_{\mathrm{TLU}} = 405\,\mathrm{s}$ small gaps can be seen in the linearly increasing line that are also visible when zooming in the constant line. This gap is where the DESY II beam is filled into the PETRA III beam, which happens about every 5 min.

Interestingly, the spatial information of the frame with these zero timestamps is not compromised. Further investigation indicates that hits arriving during the dead time between the frames are the cause for the zero timestamps. This can be caused by real particles, but also by noise hits.

Since hits with a zero timestamp cannot be used for studying the timing performance of the DUT, they are filtered out in the analysis. Additionally, this also means that an efficiency study is not easily possible with accurate timestamps. For example, it can happen that an efficient hit was registered, but a noise hit in the dead time results in a zero timestamp, leading to the frame with the efficient hit being discarded. It was decided that efficiency studies go beyond the scope of this thesis.

5.1.2 Electrical Interference

Another interesting observation was made after the events with zero timestamps had been filtered out. The time residuals between the DUTs showed multiple sharp peaks, which is not expected. It has been found that the timestamps of the DUTs have jumps at certain times. In Fig. 5.2 the raw time residuals over time are shown, excluding DUT timestamps with a zero timestamp.

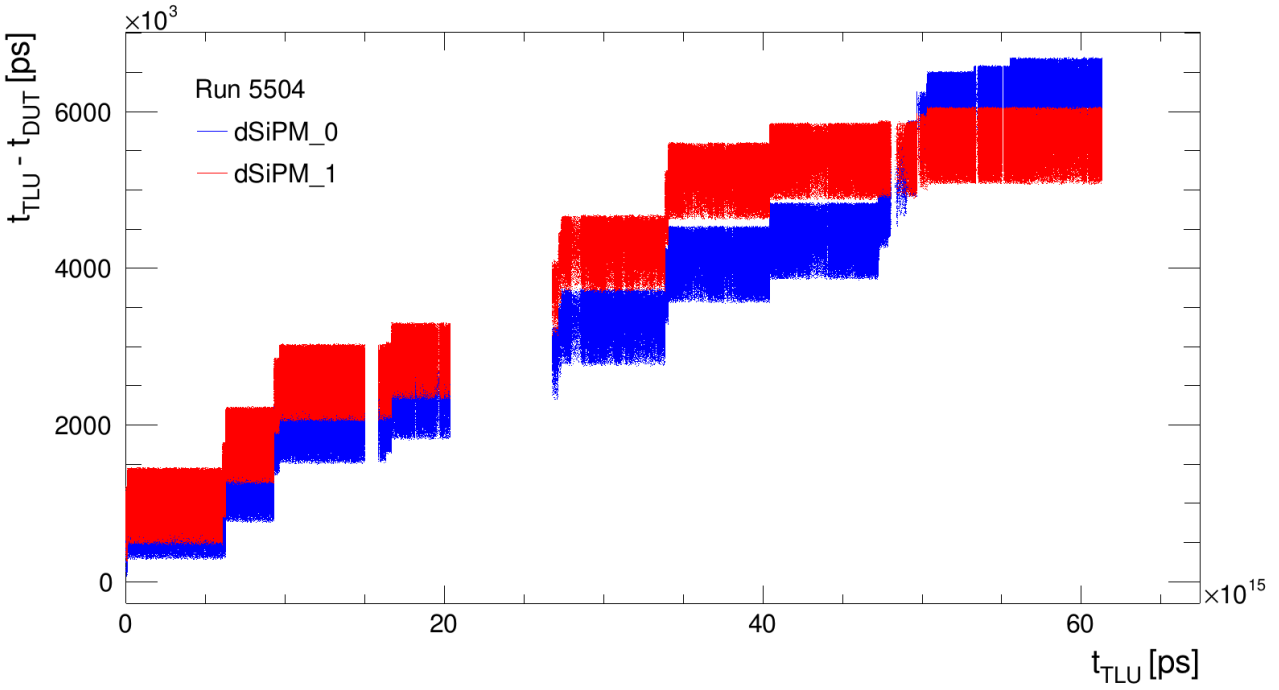


Figure 5.2: Raw time residuals over time without zero timestamps (run 5504).

The time residuals in Fig. 5.2 have several interesting properties. The overall offset of the time residuals should be a constant given by the length of the trigger cable and the delay to register the T0 signal. This offset is supposed to be constant over the entirety of the run though since the clocks of the DUTs are synchronized with the TLU clock. The width of the time residual band is about 1 µs, which is the expected width when picking a timestamp from the three frames.
At first sight, the jumps in the distribution seem to happen at the same point in time and are roughly in the same order of magnitude. The jumps also only ever go in one direction, namely the clocks of the DUTs are lagging behind the TLU clock. The gaps in the data are caused by DESY II not providing particles to testbeam. This was expected in run 5504 since it was taken on a Wednesday, which is maintenance day at DESY II.

The first idea for the cause of the jumps was a decoding error of the timestamps, given that the jumps happen at the same time. The code of the decoder was investigated, but no bug in the decoding was found. Another hint that the jumps are not caused by a decoding issue is the magnitude of the jumps, which should be identical in both DUTs. This is not the case when inspecting Fig. 5.2 carefully and can be seen even more clearly in Fig. 5.3, where jumps are only observed in one DUT in the same run.

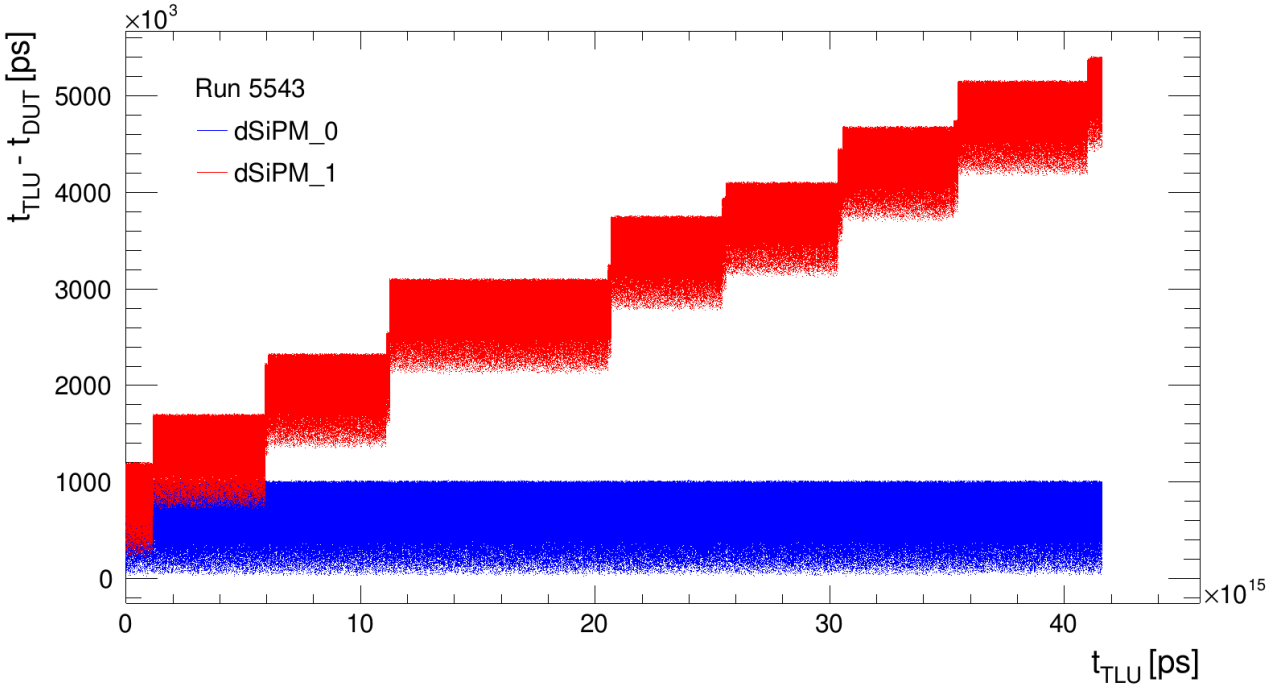


Figure 5.3: Raw time residuals over time without zero timestamps (run 5543).

Another possible cause that has been investigated is possible sticky bits in the DUT. A sticky bit is a bit in the digital representation of the timestamp that is permanently either 1 or 0 in all or most timestamps. Histograms with the bins of the clocks of each TDC have not shown any sticky bits however. This was further confirmed by investigating the time difference between the DUT timestamps and the trigger timestamp provided by the FPGA. Fig. 5.4 shows this as a function of time. For the FPGA timestamp, the bunch counter is used to highlight the 3 MHz frame structure of the DUT. The fact that no jumps between the DUT and FPGA exist, which have separate TDC circuits running on the same clock signal, indicates that the issue might be outside of our DAQ system. The next suspicious device to be investigated was the TLU. In particular, it was thought that the clock signal from the TLU might include artifacts that disturb the DAQ system.

A spare TLU was available to test if the issue is reproducible. In a first test in the lab, random triggers were sent to one DUT in a long run over a weekend. In this scenario, timestamps in the DUT are provided by noise hits only. However, no jumps have been found in the run, hinting that the issue had to be introduced somewhere between the TLU and DUT by something that is in the testbeam area but not in the lab. This rendered the chiller of the MIMOSA telescope a suspect as it made problems in a different testbeam before.

The week after the testbeam was maintenance week, which made it possible to go back to the testbeam area and perform further tests. The same TLU used in the testbeam provided random triggers to the DUTs. In addition, the chiller of the MIMOSA telescope was turned on and off while noting down the trigger number when a click was heard from the chiller, probably from a relay switching. It was observed that turning the chiller on or off caused a jump in the time residuals exactly at the trigger number when the click was heard. A test with a probe of an oscilloscope hanging in the air close to the chiller showed a large electrical interference. The chiller problem was handed over to the DESY testbeam crew and is currently being investigated.

It was concluded that the jumps were caused by EMI (ElectroMagnetic Interference), which allows devising a migitation strategy. Since the jumps are caused at essentially random times with random magnitudes, any attempt to correct it has the potential to bias the data. Instead, a region in the data was chosen where no jumps are observed in either DUT. Unfortunately, this partially reduces the available statistics significantly. A list of runs with the selected amount of triggers is given in Tab. 5.1.

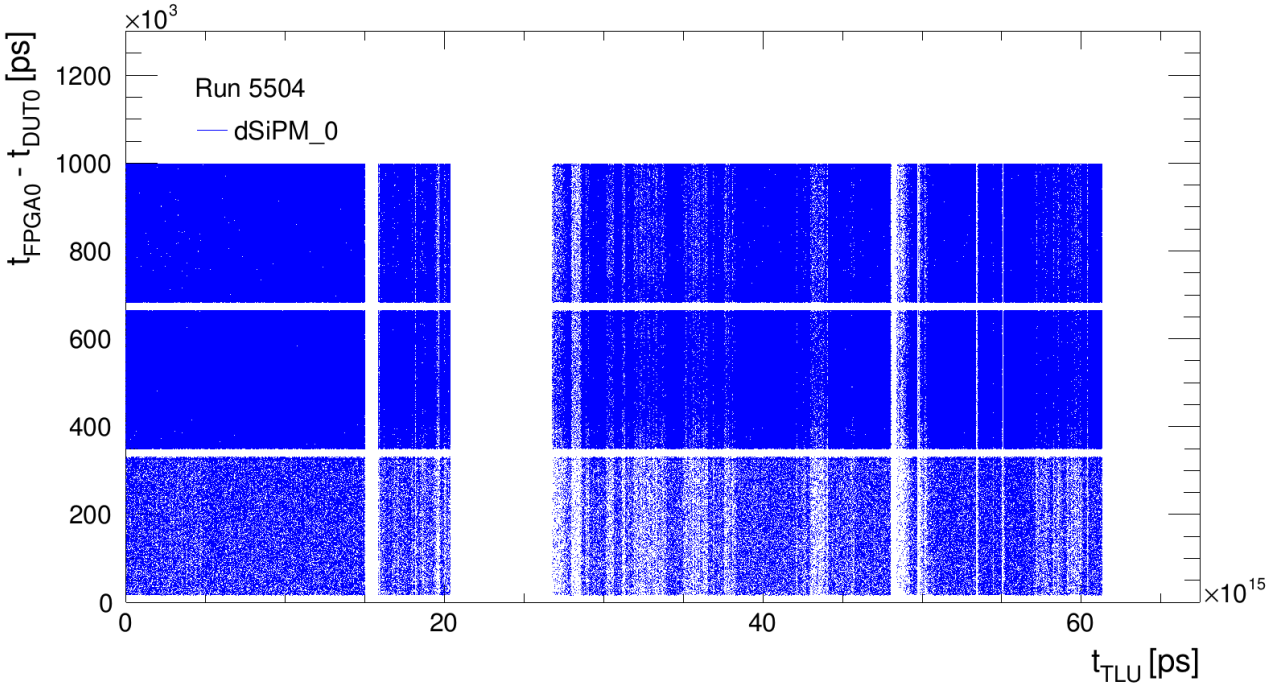


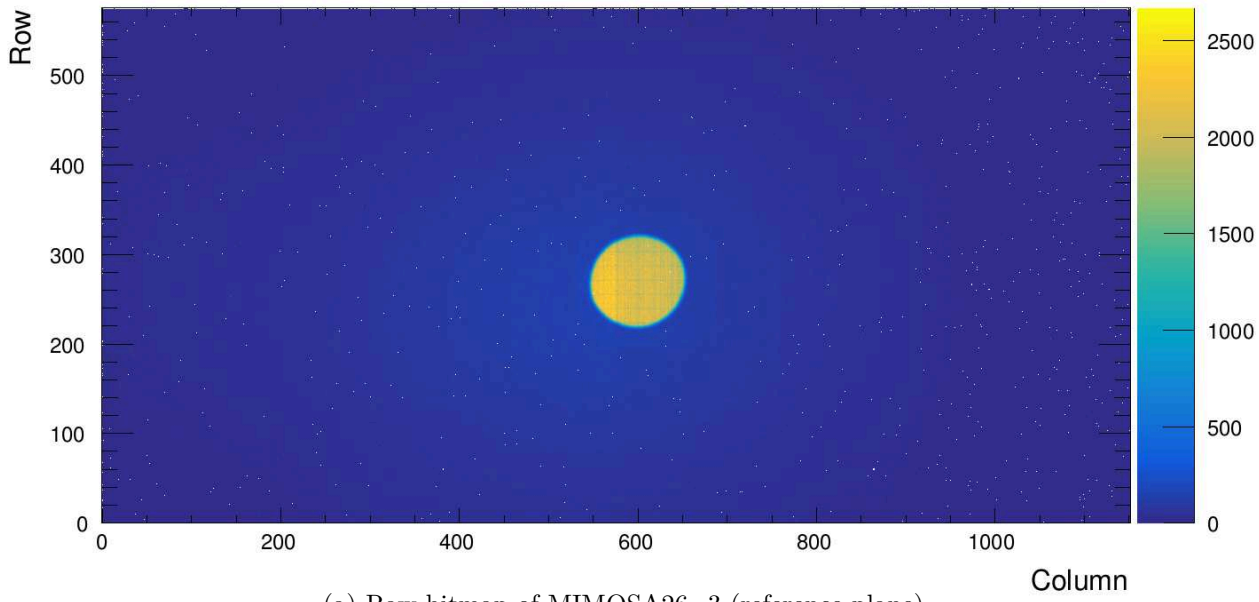
Figure 5.4: Raw time difference between DUT and FPGA frame counter over time (run 5504).

Run	T_{chiller}	V_{OV}	Total Triggers	Lower Selection	Upper Selection	Fraction
5504	−24 °C	1.50 V	3.1×10^6	8×10^3	439×10^3	14 %
5506	−24 °C	2.00 V	4.0×10^6	2.05×10^6	3.71×10^6	42 %
5507	−24 °C	2.50 V	3.5×10^6	2.74×10^6	3.455×10^6	20 %
5508	−24 °C	3.00 V	3.0×10^6	765×10^3	1.745×10^6	33 %
5509	−24 °C	0.70 V	5.0×10^6	3.49×10^6	4.705×10^6	24 %
5512	−24 °C	0.90 V	3.2×10^6	585×10^3	3.2×10^6	82 %
5513	−24 °C	1.25 V	3.1×10^6	385×10^3	1.065×10^6	22 %
5514	−24 °C	1.75 V	3.0×10^6	465×10^3	1.185×10^6	22 %
5515	−24 °C	2.25 V	3.9×10^6	2.6×10^6	3.26×10^6	17 %
5516	−24 °C	2.75 V	3.9×10^6	720×10^3	1.405×10^6	18 %
5519	−14 °C	0.70 V	3.0×10^6	0	612×10^3	20 %
5520	−14 °C	1.00 V	4.0×10^6	0	880×10^3	22 %
5521	−14 °C	1.50 V	4.3×10^6	2.895×10^6	4.175×10^6	30 %
5524	−14 °C	2.50 V	2.7×10^6	2.1×10^6	2.7×10^6	22 %
5526	−14 °C	3.00 V	4.3×10^6	345×10^3	2.33×10^6	46 %
5534	−14 °C	0.80 V	4.4×10^6	668×10^3	1.935×10^6	29 %
5537	−14 °C	1.25 V	1.7×10^6	0	1.48×10^6	87 %
5539	−14 °C	1.75 V	4.3×10^6	2.205×10^6	2.715×10^6	12 %
5541	−14 °C	1.25 V	4.2×10^6	1.86×10^6	2.895×10^6	25 %
5543	−14 °C	2.75 V	4.5×10^6	1.23×10^6	2.245×10^6	23 %
5545	−14 °C	2.00 V	5.8×10^6	0	5.8×10^6	100 %

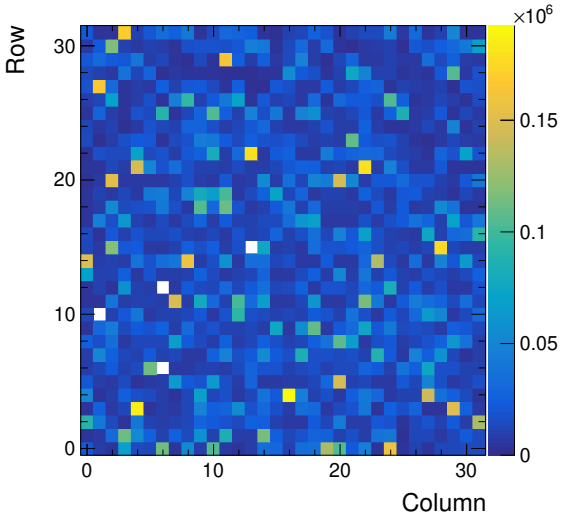
Table 5.1: List of runs with DUT chiller temperature T_{chiller}, overvoltage V_{OV}, total amount of triggers in the run, the selected trigger region in the run to avoid time jumps and the fraction of triggers in the selection. The breakdown voltage V_{BD} was determined using the relative derivate method.

5.2 Hitmaps

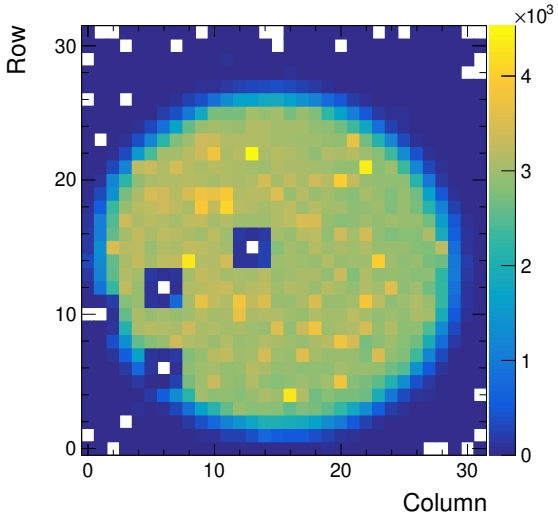
Hitmaps are 2D maps that show how many times each pixel was hit. Hitmaps are an important first check to ensure the setup is working as intended. In Fig. 5.5 various hitmaps are presented.
Fig. 5.5a show the raw hitmap of the reference plane for one run. A "raw" hitmap is created from the data without further analysis. As can be seen, most hits are centered around a small circle. This is the region selected by the scintillator+veto combination, in particular by the hole of the veto. Hits outside of this region can be caused by noise hits or a second particle that traverses the sensor in the relatively long integration time of the MIMOSA26 planes.
Fig. 5.5b show the raw hitmap of dSiPM_1. A few pixels have no entries because they have been masked. Most of the hits in this hitmap are noise hits. This becomes clear when looking at Fig. 5.5c, where the hitmap of associated hits is shown. As explained in subsection 4.4.3, this means that all hits in Fig. 5.5c are associated to a particle track reconstructed from the telescope. In this hitmap, the hole of the veto can also be seen.



(a) Raw hitmap of MIMOSA26_3 (reference plane).



(b) Raw hitmap of dSiPM_1.



(c) Hitmap with associated hits of dSiPM_1.[2]

Figure 5.5: Hitmaps of a telescope plane and a DUT (run 5545).

[2]The low-entry pixels around a masked pixel are an artifact from the specific Corryvreckan module, which excludes hits around a masked pixel. However this cut has not been used for the timing analysis.

5.3 Spatial Residuals

Inspecting spatial residuals (Eqn. 3.1) of a DUT is an essential check for its alignment. Examples of spatial residual distributions are shown in Fig. 5.6 and Fig. 5.7. The double peak structure seen in the distributions corresponds to the four SPADs per pixel layout described in section 2.1, for which a mathematical approximation is given in Appendix A. The deeper valley in Fig. 5.6 compared to Fig. 5.7 corresponds to a better tracking at the position of dSiPM_0 than at the position of dSiPM_1, which is in agreement with the expected tracking resolution at the DUT positions (see Fig. 3.2). While detailed results will be published elsewhere [23], the distributions show a pixel-scale spatial resolution.

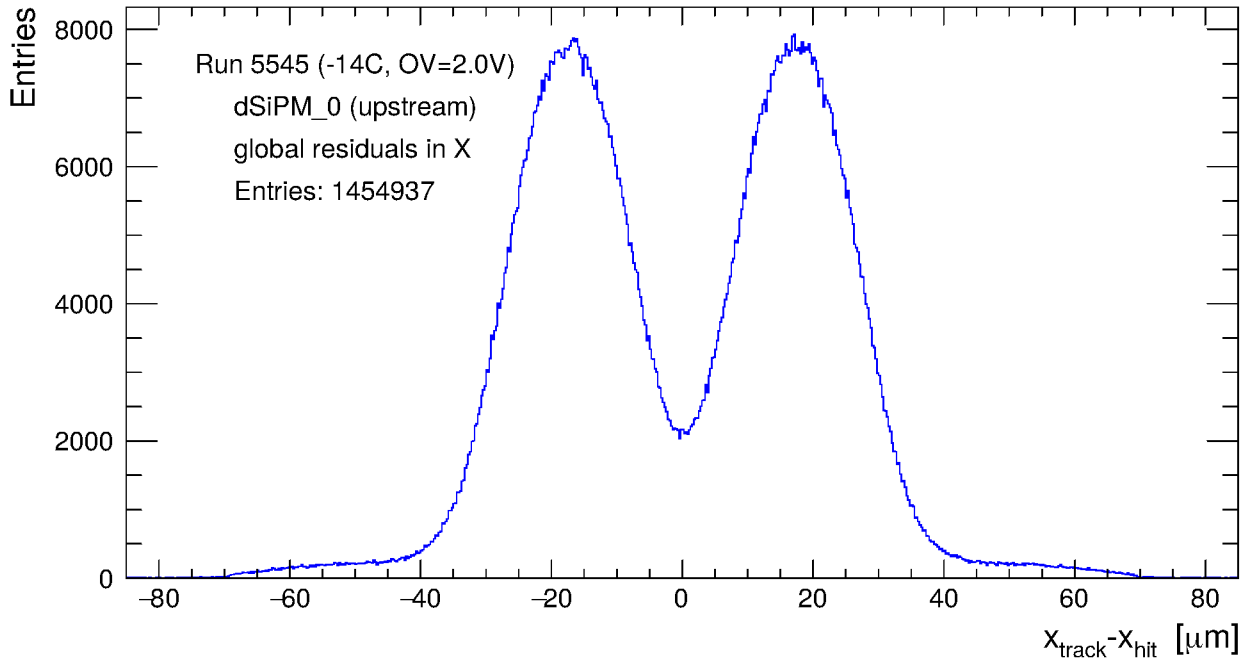


Figure 5.6: Spatial residuals in X of dSiPM_0 in run 5545.

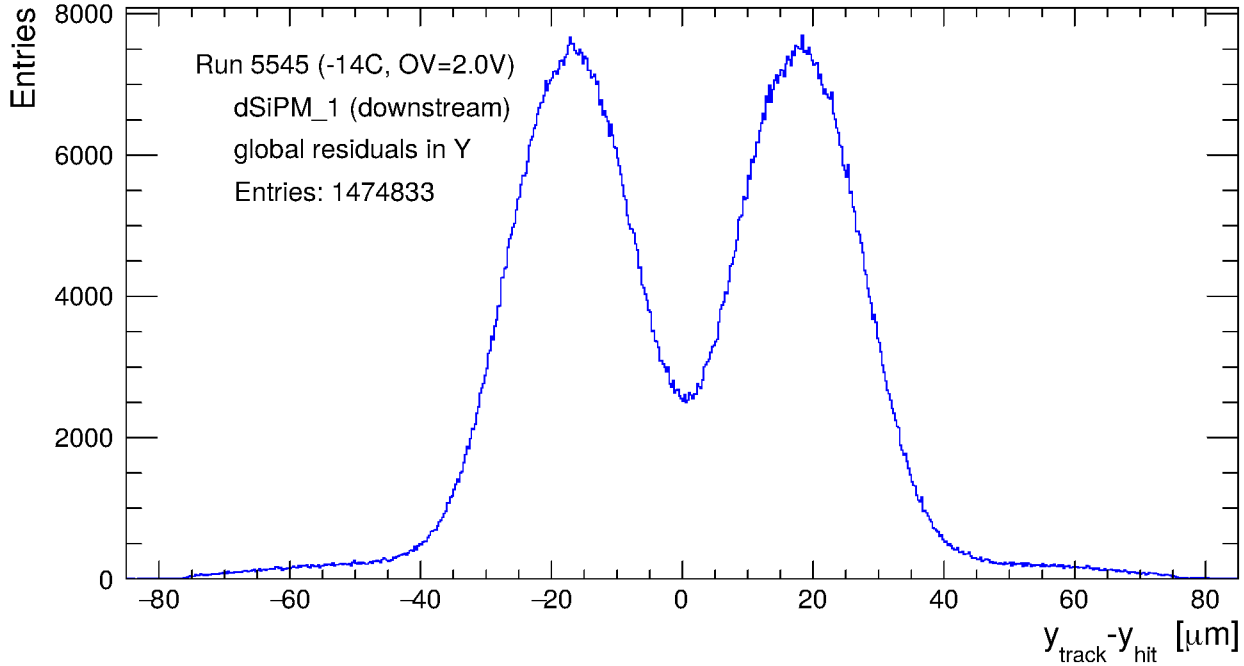


Figure 5.7: Spatial residuals in Y of dSiPM_1 in run 5545.

5.4 Differential and Integral Non-Linearity

To achieve the optimal timing performance of the chip, the extent of DNL and INL effects (see subsection 2.3.2) have to be understood and corrected if necessary. The equation used to calculate DNL corrected timestamps (given in Eqn. 4.1) requires that the bin width τ_n for all fine TDC bins is known, which is just the delay length of the delay element.
One observation with TDL-based TDCs is that if the TDC is sampled randomly, the probability of measuring a certain bin is proportional to the delay length. For example, in a 2 bit TDL-based TDC where bin 0 has twice the delay length compared to bin 1, a random measurement would return bin 0 with a probability of $^2/_3$. Consequently, it is possible to evaluate the individual delay lengths by sampling a TDL-based TDC randomly.
This approach was chosen to calculate the delay length, with the assumption that the particles and noise hits arrive at random times. Both assumptions are simplifications, for one the particles are not provided at random times from DESY II, but it is assumed that there is no structure relevant for timescales of less than one 408 MHz clock cycle ($\sim$2.5 ns). For noise timestamps, it is known that due to the readout system always taking the first timestamp of all hits in the quadrant (see section 2.2) there is a preference for smaller timestamps. However, at the temperatures where the chip operated in the testbeam, this structure is not significant on a 2.5 ns scale due to the low DCR (see Fig. 5.17 and Appendix B).
In Fig. 5.8, a histogram is shown with the number of entries for each fine TDC bin of one particular TDC. One observation is that bins larger than bin 27 have no entries. This effect was known already before testbeam and exists because all delay elements in the TDC have a too large delay. This causes the incoming clock signal to never reach bins above 27 before a new clock signal enters the delay line. Since the TDC counter starts at bin 1, this means that the dynamic range is reduced from 5 bit to effectively $\sim$4.8 bit for this particular TDC.
Another observation is that the number of entries in the bins is not equal and their difference is statistically significant. The differences seen are caused by DNL as described before. The last bin is a special case, which can be seen in the significantly smaller amount of entries. For this bin, the number of entries is not proportional to the real delay length since a new clock signal enters the delay line while the old clock signal is still in this delay element. It will be called "cutoff bin" in the following.

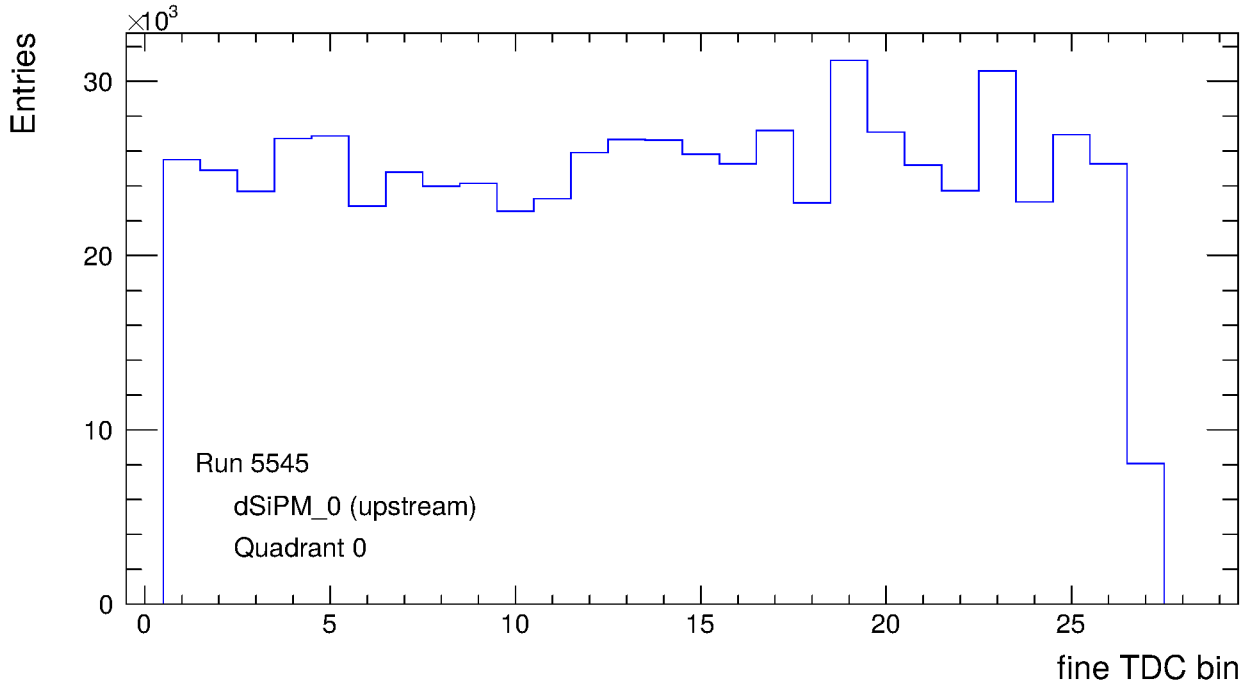


Figure 5.8: Histogram showing how often each fine TDC bin has been measured in the TDC for quadrant 0 of dSiPM_0 (run 5545).

To evaluate the bin width (delay lengths) of each bin, the following approach is used:

1. Determine the cutoff bin n_{cutoff}. It is defined as the last bin with with more than 5 % entries compared to the maximum bin to exclude rare cases where another bin is reached.
2. Calculate the mean number E_{mean} of entries excluding the cutoff bin:

$$E_{\text{mean}} = \frac{1}{n_{\text{cutoff}} - 1} \sum_{j=1}^{n_{\text{cutoff}}-1} E_j \tag{5.1}$$

   The cutoff bin is excluded in the mean since it is not possible to calculate its true delay length, but only the effective bin width.

3. Calculate fraction f_n between bin entries and mean for all bins:

$$f_n = E_n \;/\; E_{\text{mean}} \tag{5.2}$$

4. Calculate conversion factor C between bin entries and bin width:

$$C = (408\,\text{MHz})^{-1} \;/\; \left(E_{\text{mean}} \cdot \sum_{j=1}^{n_{\text{cutoff}}} f_j \right) \tag{5.3}$$

   Here the insight is used that the number of bin entries is proportional to the probability to measure that bin. Thus the sum of all entries is proportional to the length of one clock cycle.

5. Calculate bin width for all bin width using fraction and mean bin width:

$$\tau_{\text{mean}} = C \cdot E_{\text{mean}} \tag{5.4}$$

$$\tau_n = f_n \cdot \tau_{\text{mean}} \tag{5.5}$$

It is possible to simplify the final result of the calculation, but this approach was given to highlight the derivation of the calculation. Error propagation is done by assuming that the error on the number of entries E_j in a bin is the statistical error $\sqrt{E_j}$.

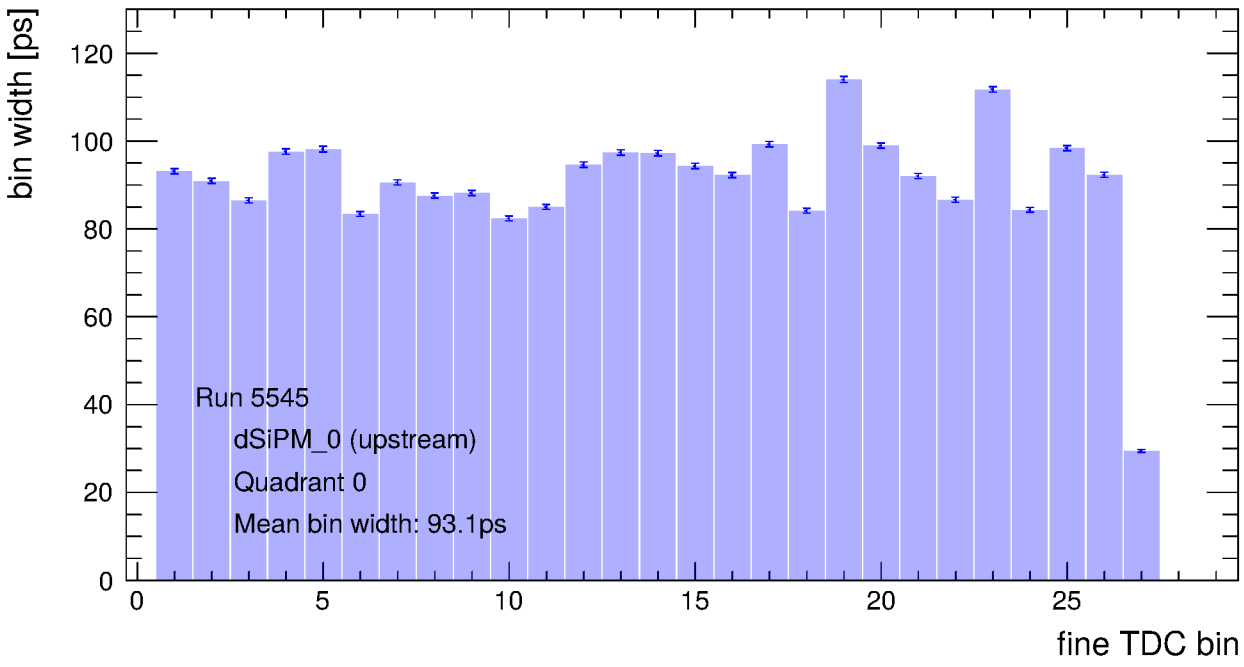


Figure 5.9: Calculated fine TDC bin widths in the TDC for quadrant 0 of dSiPM_0 (run 5545).

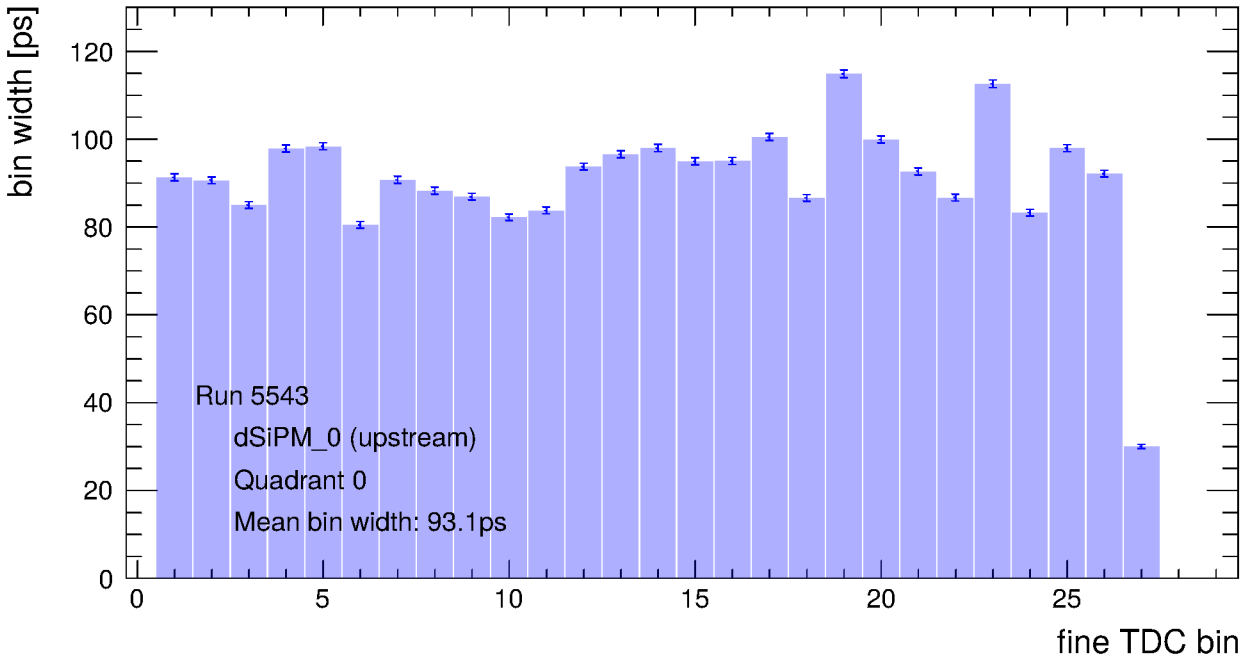


Figure 5.10: Calculated fine TDC bin widths in the TDC for quadrant 0 of dSiPM_0 (run 5543).

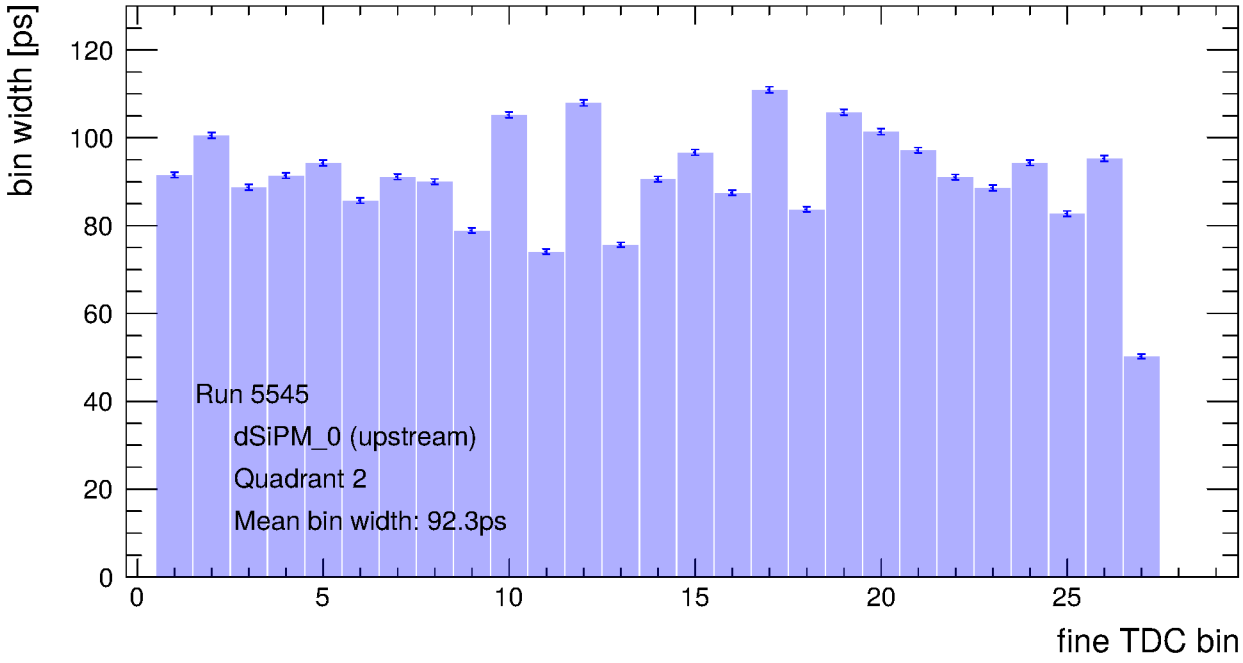


Figure 5.11: Calculated fine TDC bin widths in the TDC for quadrant 2 of dSiPM_0 (run 5545).

In Fig. 5.9, the calculated bin widths corresponding to the bin entries given in Fig. 5.8 is shown. The calculated mean bin width τ_{mean} excluding the cutoff bin is (93.14 ± 0.11) ps. Due to the large number of entries, the errors are quite small and show that the differences in bin entries are not just statistical fluctuation. To further confirm this, Fig. 5.10 shows the calculated bin widths for the same TDC using data from a different run. Besides minor statistical fluctuations, the pattern is identical. Since run 5545 has the highest statistics, the DNL correction in the analysis uses the bin widths calculated from this run.
It is important to keep in mind that every TDC can have different bin widths since DNL is caused by process variations. In Fig. 5.11, the calculated bin widths for a different TDC on the same chip are shown. The pattern is visibly different.

The DNL and INL measures described in subsection 2.3.2 are used to quantify the impact of the bin width fluctuations. In Fig. 5.12 and Fig. 5.13, the DNL and INL corresponding to the bin widths given in Fig. 5.9 are shown. Here, the design bin width $(408\,\text{MHz})^{-1} \,/\, 2^5 \simeq 77\,\text{ps}$ is used for τ in Eqn. 2.1. However as can be seen in Fig. 5.9, the mean bin width τ_{mean} is quite far from the design value and systematically higher. If this systematic offset is not accounted for, the INL in the last bin results in a value over 5 LSB, which is a deviation over 380 ps.
An effective measure against this deviation (without deploying DNL correction) is using the mean bin width τ_{mean} instead of the design value as the fine TDC bin width during the timestamp decoding. To evaluate the deviation of this correction, the DNL and INL relative to this mean (labeled rDNL and rINL) are useful quantities:

$$\text{rDNL}_n = \frac{\tau_n - \tau_{\text{mean}}}{\tau_{\text{mean}}} \tag{5.6}$$

$$\text{rINL}_n = \sum_{j=1}^{n-1} \text{rDNL}_j \tag{5.7}$$

Consequently, the unit in which the data is represented is not LSB anymore. The corresponding unit for rDNL and rINL will be labeled as effective LSB, or eLSB for short.
Fig. 5.14 shows the rDNL corresponding to the DNL shown in Fig. 5.12 relative to the mean bin width of 93.1 ps. The maximum rDNL is 0.22 eLSB, which is about 20.5 ps.
While DNL and rDNL are interesting quantities for ASIC designers, on the analysis side INL and rINL are more relevant as they quantify the deviation if the timestamps are not corrected for non-linearities. In Fig. 5.15, the corresponding rINL is shown. Compared to the INL shown in Fig. 5.13, the rINL does not rise continuously anymore but comes back to around 0 eLSB. The maximum rINL is 0.44 eLSB, which is about 41 ps. The largest uncertainty is about 0.05 eLSB, which is less than 5 ps.
While the 41 ps maximum deviation is much less compared to the over 380 ps in the completely uncorrected scenario, the value is still relatively high compared to the binary TDC resolution of about 27 ps (22 ps with design values).

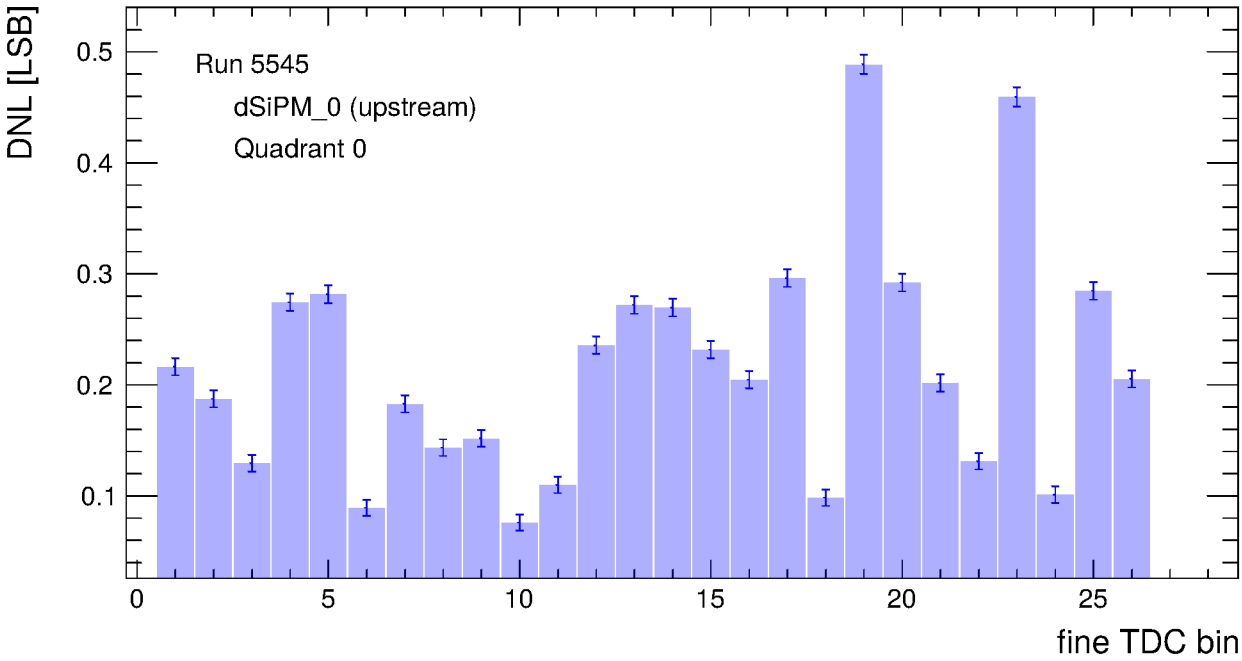


Figure 5.12: Differential Non-Linearity (DNL) in the TDC for quadrant 0 of dSiPM_0, given in units of LSB, where 1 LSB $\simeq$ 77 ps (run 5545).

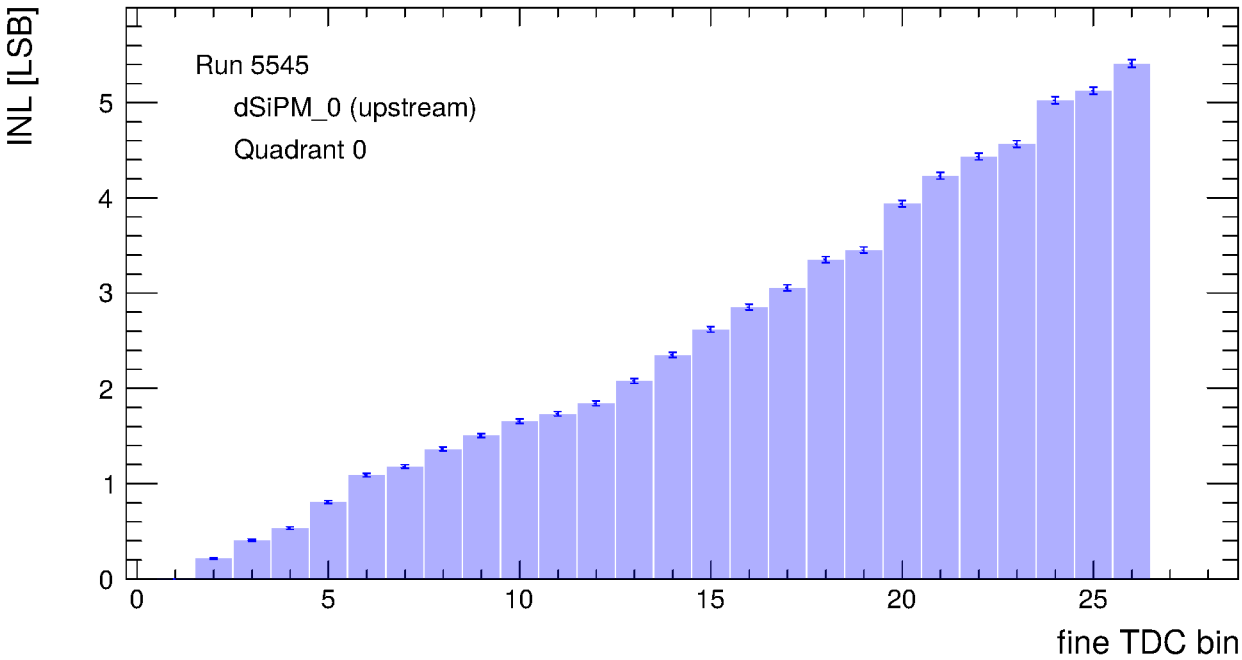


Figure 5.13: Integral Non-Linearity (INL) in the TDC for quadrant 0 of dSiPM_0, given in units of LSB, where 1 LSB $\simeq$ 77 ps (run 5545).

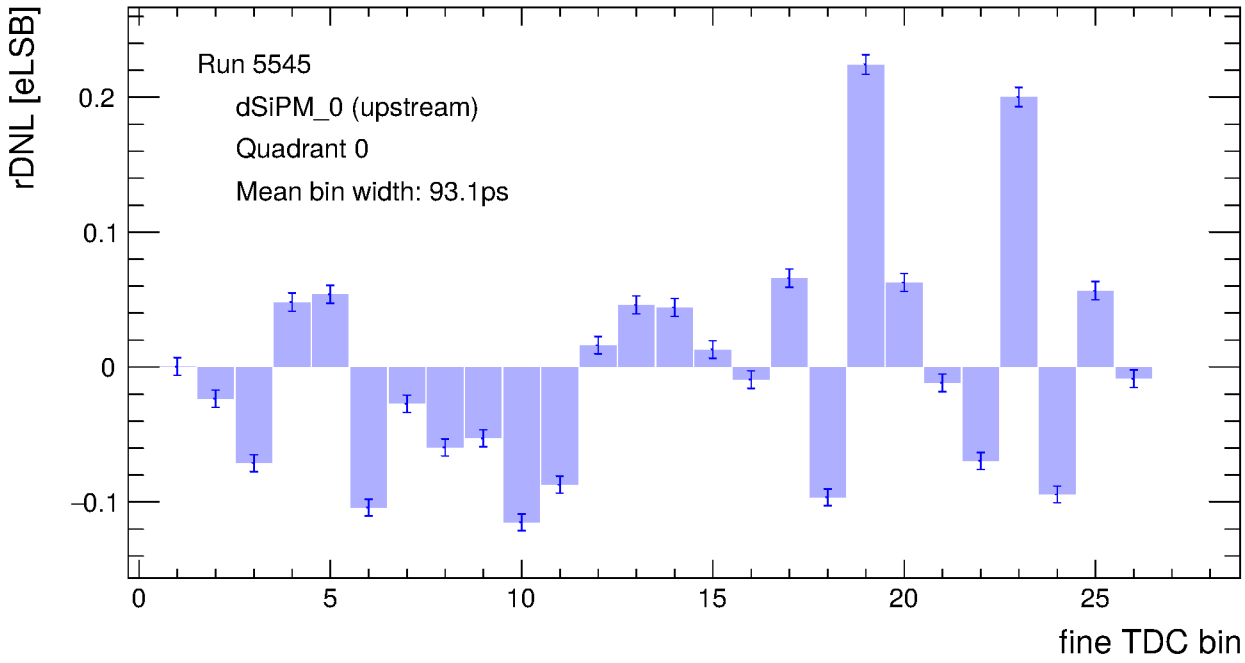


Figure 5.14: Differential Non-Linearity (DNL) relative to the mean bin width (93.1 ps) instead of the design value (77 ps) in the TDC for quadrant 0 of dSiPM_0, given in units of eLSB, where 1 eLSB $\simeq$ 93.1 ps (run 5545).

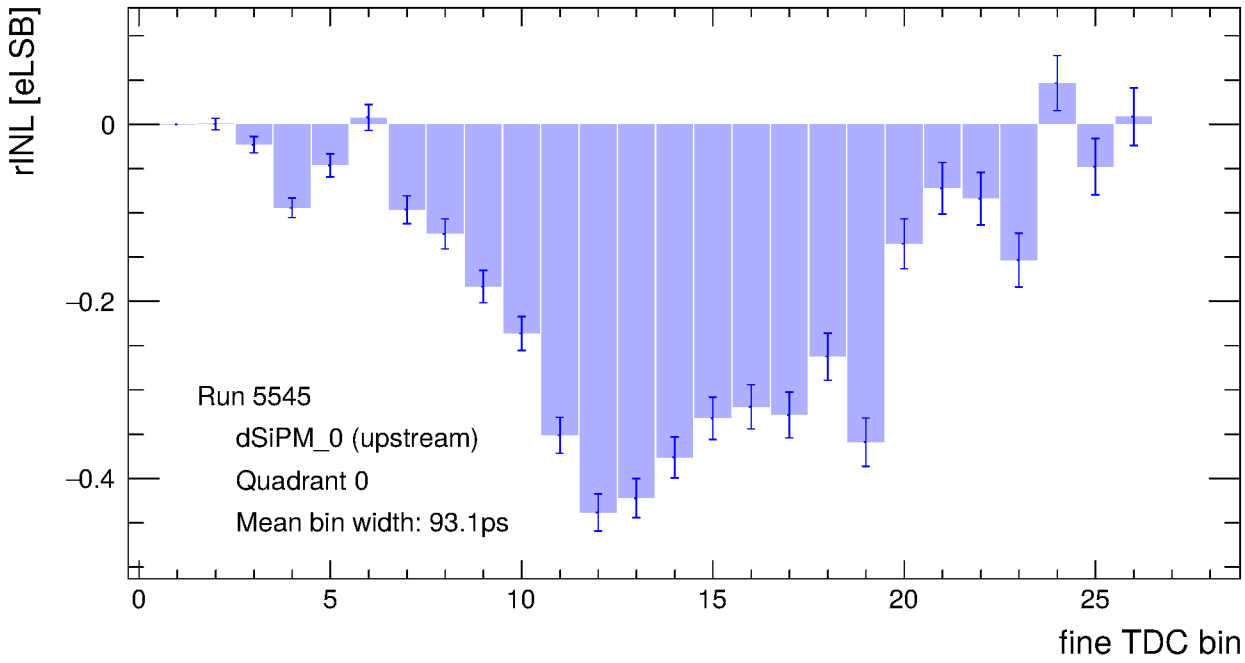


Figure 5.15: Integral Non-Linearity (INL) relative to the mean bin width (93.1 ps) instead of the design value (77 ps) in the TDC for quadrant 0 of dSiPM_0, given in units of eLSB, where 1 eLSB $\simeq$ 93.1 ps (run 5545).

5.5 Time Residuals between TLU and DUT

In this section, a first look is taken at the time residuals between the TLU and DUT (Eqn. 3.2). Although the time resolution of the TLU is expected to be significantly larger than the one from the dSiPM, the distribution can provide an upper limit, which smooths out binning effects of the dSiPM. In Fig. 5.16, the time residual between the TLU and DUT0 is shown. The FWHM of the distribution is about 782 ps, and assuming a gaussian core[3] gives $\sigma_{\text{TLU-DUT}} \simeq 332\,\text{ps}$. The overall time resolution of this distribution can be broken down into various contributions (see also Eqn. 3.5):

$$\sigma^2_{\text{TLU-DUT}} = \sigma^2_{\text{Sci}} + \sigma^2_{\text{TLU}} + \sigma^2_{\text{Delay}} + \sigma^2_{\text{TDC}} + \sigma^2_{\text{INL}} + \sigma^2_{\text{SPAD}} \tag{5.8}$$

Estimations of each component, if known, are given in Tab. 5.2. The estimation of σ_{Delay} has been made using early measurements with a laser setup, however these measurements are still a work in progress and were not precise enough to allow for a good correction.
To estimate σ_{Sci}, it is assumed that the only other major contributions to $\sigma_{\text{TLU-DUT}}$ are σ_{TLU} and σ_{Delay}. Assuming a gaussian core for the TLU-DUT time residual distribution[3] gives

$$\sigma_{\text{Sci}} = \sqrt{\sigma^2_{\text{TLU-DUT}} - \sigma^2_{\text{TLU}} - \sigma^2_{\text{Delay}}} = \sqrt{\left(\frac{\text{FWHM}_{\text{TLU-DUT}}}{2\sqrt{2\log 2}}\right)^2 - \sigma^2_{\text{TLU}} - \sigma^2_{\text{Delay}}} \simeq 230\,\text{ps} \tag{5.9}$$

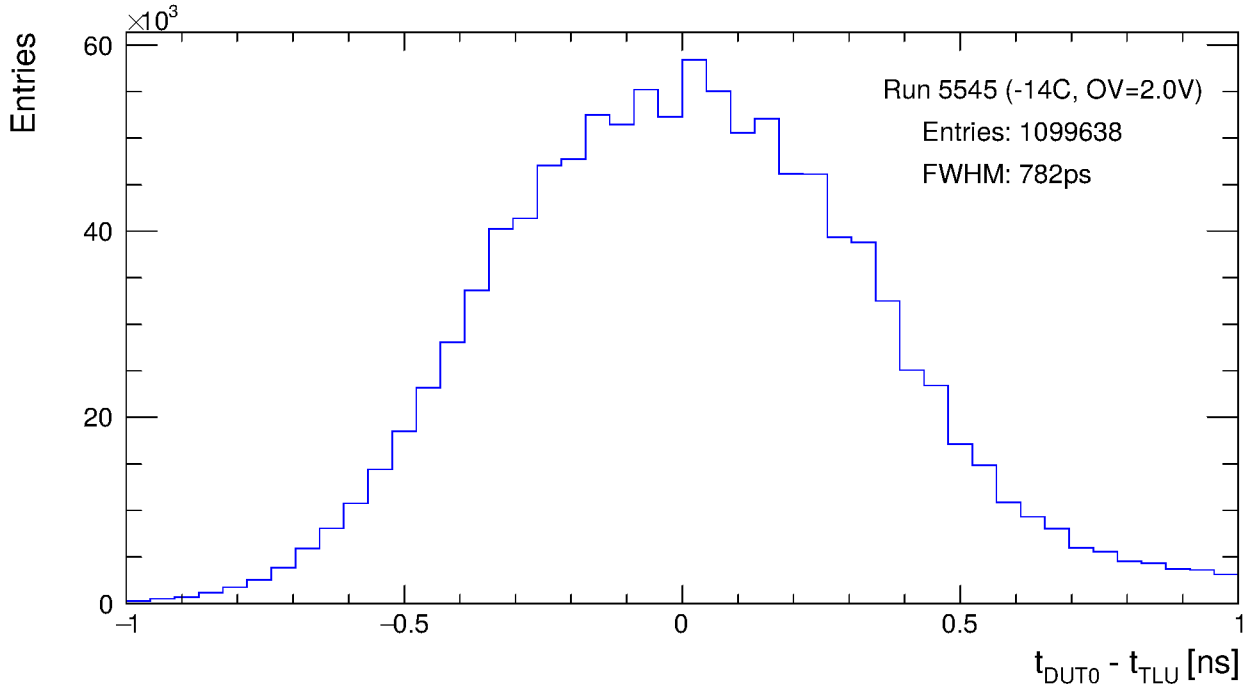


Figure 5.16: TLU-DUT time residual distribution (run 5545).

Component	Estimation	Description
σ_{Sci}	$\sim 230\,\text{ps}$	Jitter from scintillators and veto logic
σ_{TLU}	$\sim 226\,\text{ps}$	TLU time resolution assuming binary resolution with 781.25 ps bins
σ_{Delay}	$\sim 70\,\text{ps}$	Signal delay variations from pixel to TDC
σ_{TDC}	$\sim 27\,\text{ps}$	TDC resolution assuming binary resolution with 95 ps average binning
σ_{INL}	$< 5\,\text{ps}$	INL uncertainty from DNL corrections (see section 5.4)
σ_{SPAD}	unknown	Jitter from physical processes in the SPAD

Table 5.2: Estimations of the contributions to the TLU-DUT time residual.

[3]This assumption is very approximate. A more precise description of the core includes a convolution of a "box" function that describes the bin width of the TLU and a Gaussian for the dSiPM. This "Gaussian box" function is given by $\text{gBox}(t) = \frac{N}{2}\left(\text{erf}\left(\frac{t+b}{\sqrt{2}\sigma}\right) - \text{erf}\left(\frac{t-b}{\sqrt{2}\sigma}\right)\right)$, where $2b$ is the TLU bin width. A fit with this function gives a similar value for σ.

5.6 Time Residuals between DUTs

In this section, a first look at the time residual distribution between the two DUTs is taken, while detailed studies will be conducted in chapter 6. Compared to the time residuals with the TLU in section 5.5, components with a large contribution to the time residual distribution vanish or become significantly smaller. In particular, the jitter of the trigger system contained in σ_{Sci} and σ_{TLU} vanishes entirely. Also, due to the relative alignment of the DUTs with an offset of less than 3 pixels in x and y, delay effects contained in σ_{Delay} are reduced.
The overall time resolution can be broken down into uncorrelated contributions as before:

$$\sigma^2_{\text{DUT-DUT}} = \sigma^2_{\text{Delay}} + \sigma^2_{\text{TDC}} + \sigma^2_{\text{Phase}} + \sigma^2_{\text{INL}} + \sigma^2_{\text{SPAD}} \tag{5.10}$$

Compared to Eqn. 5.8, the only new component is σ_{Phase}. It describes the contribution to the time resolution between two TDCs that use the same clock signal for synchronization. It depends on the phase offset between the two clocks with respect to the synchronization signal. For this thesis it is assumed that σ_{Phase} is not significant in the sum for $\sigma^2_{\text{DUT-DUT}}$.
A new lower estimation σ_{Delay} was not performed, since the early measurements of the delays are not understood well enough to give a sufficiently precise value.

Noise

Fig. 5.17 shows the time residual distribution in a time window of 600 ns. A background structure with a length of two frames can be seen. This background comes from noise events with uncorrelated timestamps. There are two kinds of noise events for this case, in one of them the particle was not detected on one DUT but there was a dark hit in the association region. In the second case, the particle was detected by the DUT but there was a dark hit in the same frame and quadrant before the particle hit, which already set the timestamp for that quadrant in the frame (see section 2.2). In Appendix B an approximate derivation of the expected distribution is given.

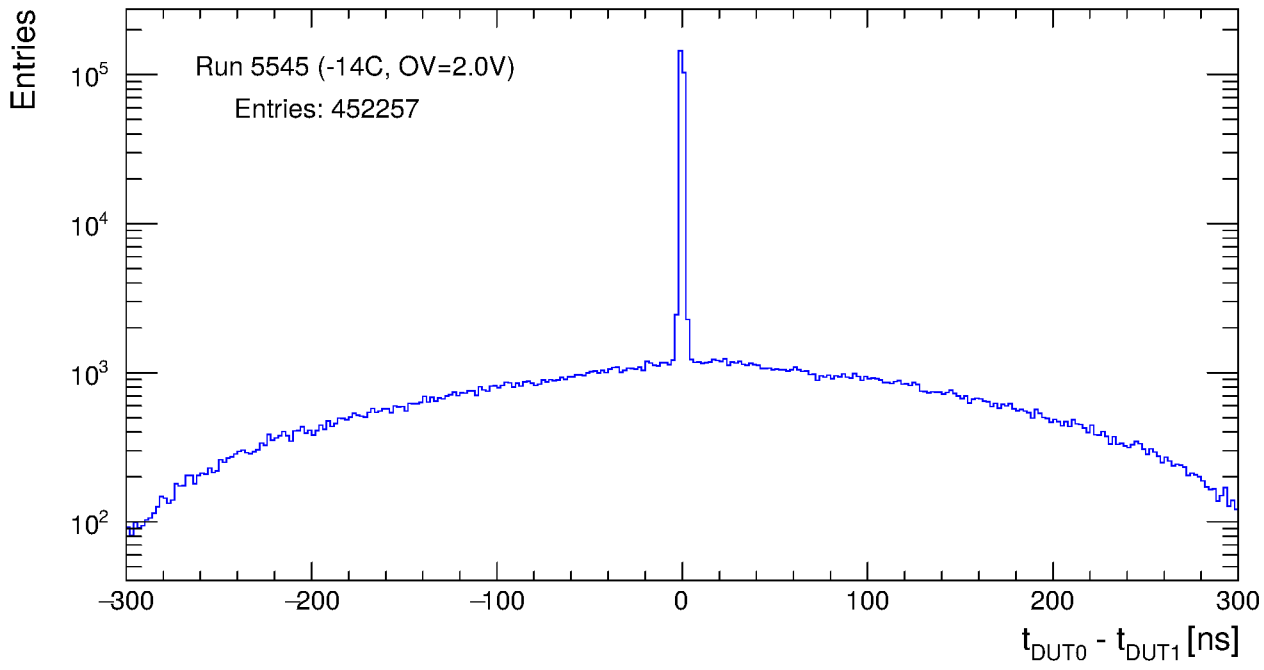


Figure 5.17: DUT-DUT time residual distribution in frame range (run 5545).

Timing Tails

The next interesting feature appears at $\mathcal{O}(10\,\text{ns})$. In Fig. 5.18, the time residual distribution is shown with a logarithmic y-axis. A $\sim 2\,\text{ns}$ tail is visible on each side of the peak. These timing tails will be discussed in detail in section 6.1.

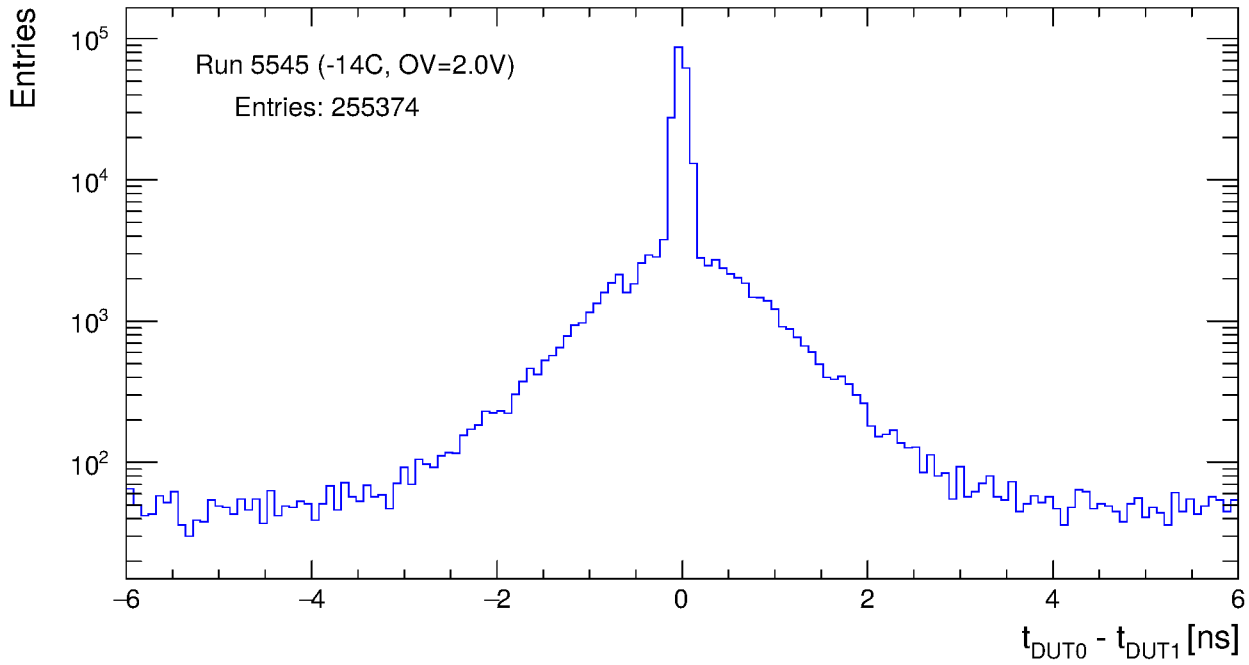


Figure 5.18: DUT-DUT time residual distribution in timing tail range (run 5545).

Peak

Fig. 5.19 shows the time residual between the two DUTs without DNL correction. The fine TDC bin size was set to a fixed value of 92.1 ps for all bins in the event decoder (see section 4.2), which is the average bin size considering all TDCs. The binning of the histogram has a bin size of $1/4 \cdot 92.1\,\text{ps} \simeq 23\,\text{ps}$ to highlight the binning effects. As can be seen in the histogram, the core of the peak only consists of about three 92.1 ps (and thus fine TDC) bins. This shows that σ_{SPAD} is not significantly larger than the TDC resolution σ_{TDC}.

Fig. 5.20 shows the time residual now with DNL correction. The binning of the histogram also has a bin size of $1/4 \cdot 92.1\,\text{ps} \simeq 23\,\text{ps}$. Compared to Fig. 5.19, the peak now consists of more bins than before. This is explained by the fact that the difference between two timestamps is not a fixed number of the TDC bin widths anymore, but a combination of two TDC bins with different bin widths. This indicates that INL effects can have a positive effect on determining the time resolution more precisely if they are corrected. A disadvantage of corrected INL effects is however that picking a suitable binning is not trivial anymore.

Both histograms show that determining the time resolution is not trivial. In Fig. 5.19 the small number of bins in the peak prevents using the FWHM in a meaningful way, and a fit is also not really possible. Fig. 5.20 on the other hand has the problem that it is difficult to find an appropriate binning for the histogram. In section 6.2 a procedure to determine the time resolution is discussed.

Further, while the contribution of the phase σ_{Phase} is assumed to be negligible in $\sigma_{\text{DUT-DUT}}$, the phase between the TDCs of the two DUTs can impact the bin structure of time residual histograms.

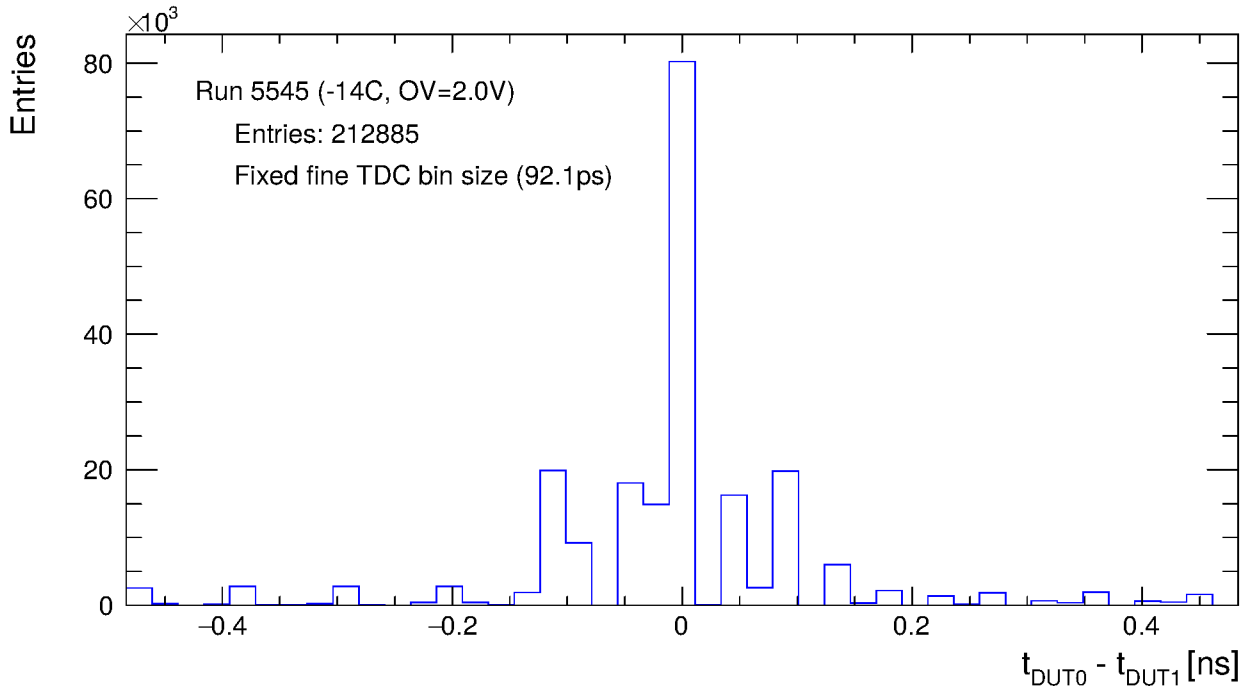


Figure 5.19: DUT-DUT time residual distribution in peak range with fixed fine TDC bin size of 92.1 ps during the event decoding (run 5545).

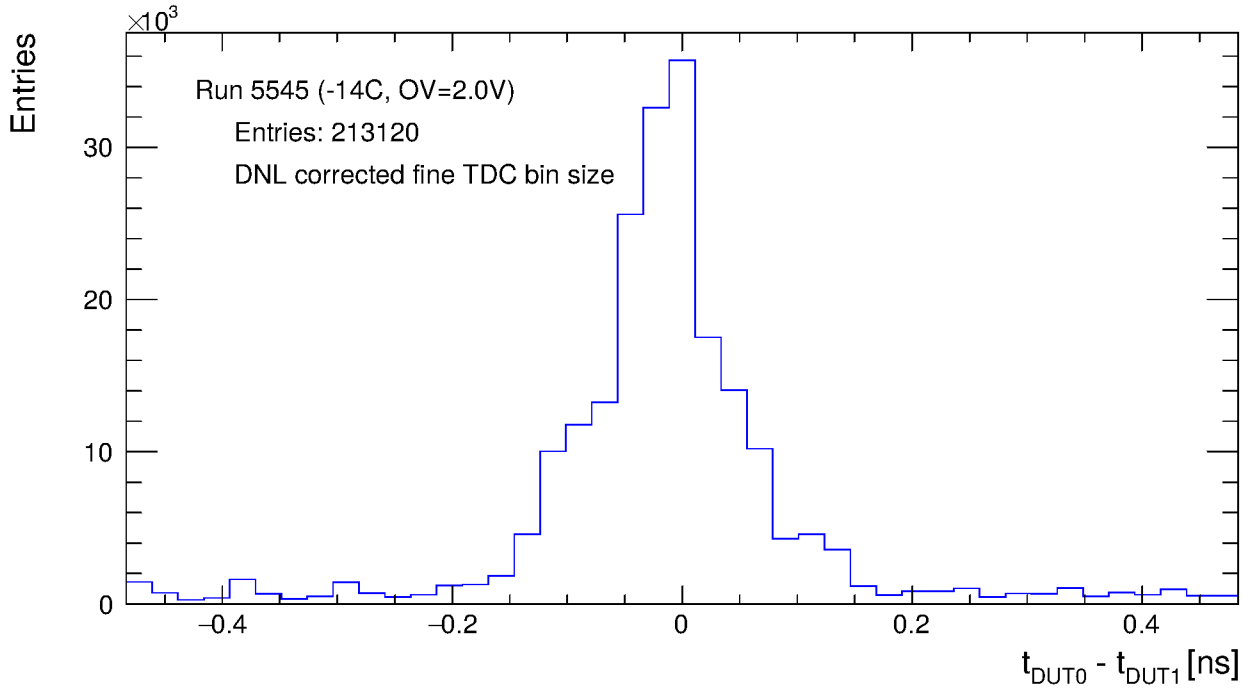


Figure 5.20: DUT-DUT time residual distribution in peak range with DNL correction applied to the fine TDC bins during the event decoding (run 5545).

Chapter 6

Timing Performance

6.1 Timing Tails and SPAD Edge Effects

In this section, the timing tails in the TLU-DUT time residual distribution (see section 5.5) and the DUT-DUT time residual distribution (see section 5.6) will be discussed.

To test whether the two tails are correlated or not, the DUT-DUT time residual distribution can be plotted only for entries where the corresponding timestamp is in the tail of TLU-DUT time residual distribution. In Fig. 6.1, the TLU-DUT distribution used for this test is shown. In Fig. 6.2, the corresponding DUT-DUT distribution is plotted in four different ways: the distribution without any cut applied, one with a cut on $1\,\text{ns} < t_{\text{DUT1}} - t_{\text{TLU}} < 5\,\text{ns}$, one with a cut on $1\,\text{ns} < t_{\text{DUT0}} - t_{\text{TLU}} < 5\,\text{ns}$, and one with a cut on $t_{\text{DUT0}} - t_{\text{TLU}} > 700\,\text{ps}$ and $t_{\text{DUT1}} - t_{\text{TLU}} > 700\,\text{ps}$.

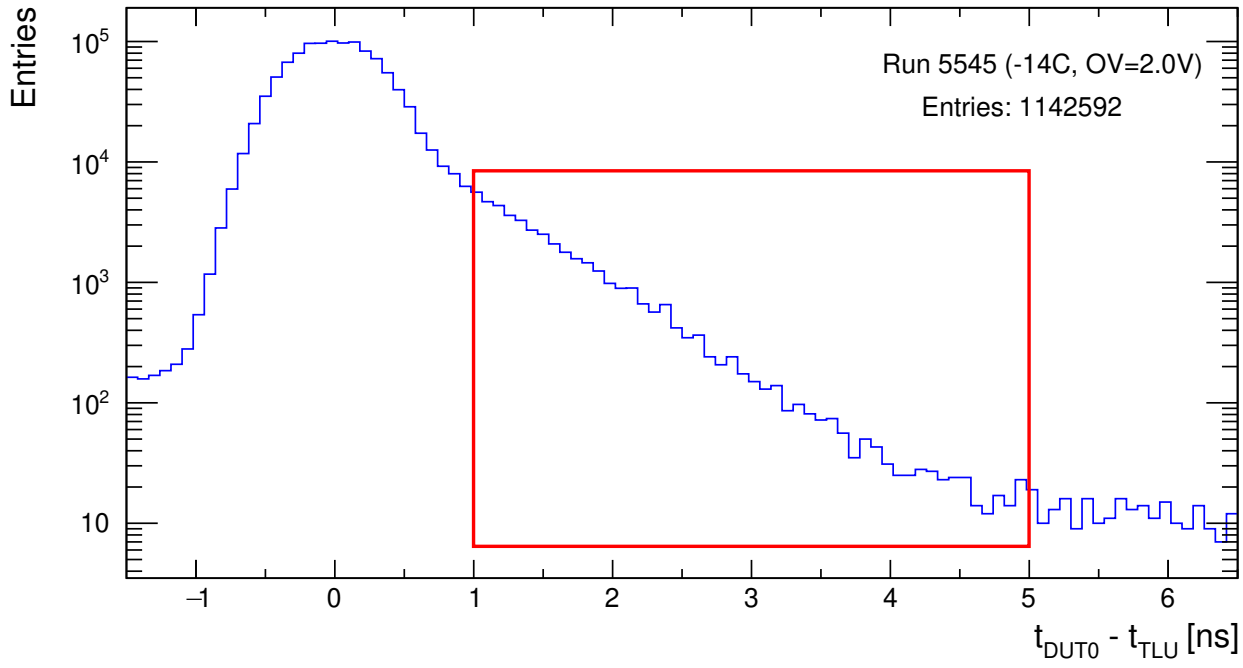


Figure 6.1: TLU-DUT time residual with timing tail cut visualized ($1\,\text{ns} < t_{\text{DUT0}} - t_{\text{TLU}} < 5\,\text{ns}$).

The histograms in Fig. 6.2 indicate a clear correlation between the tails in the DUT-DUT time residual distribution and the TLU-DUT time residual distribution. Because the left and right tails can be separated by selecting an explicit cut on either DUT, the tails are an effect of a single DUT and not an artifact of using two DUTs.
It should be mentioned that a simple cut in the TLU-DUT time residual distribution can not separate

the tail from the edge of the peak cleanly and thus only gives limited information for the inner side of the tail. Additionally, the histograms in Fig. 6.2 include entries where both DUT0 and DUT1 are slow, giving a small peak around zero.

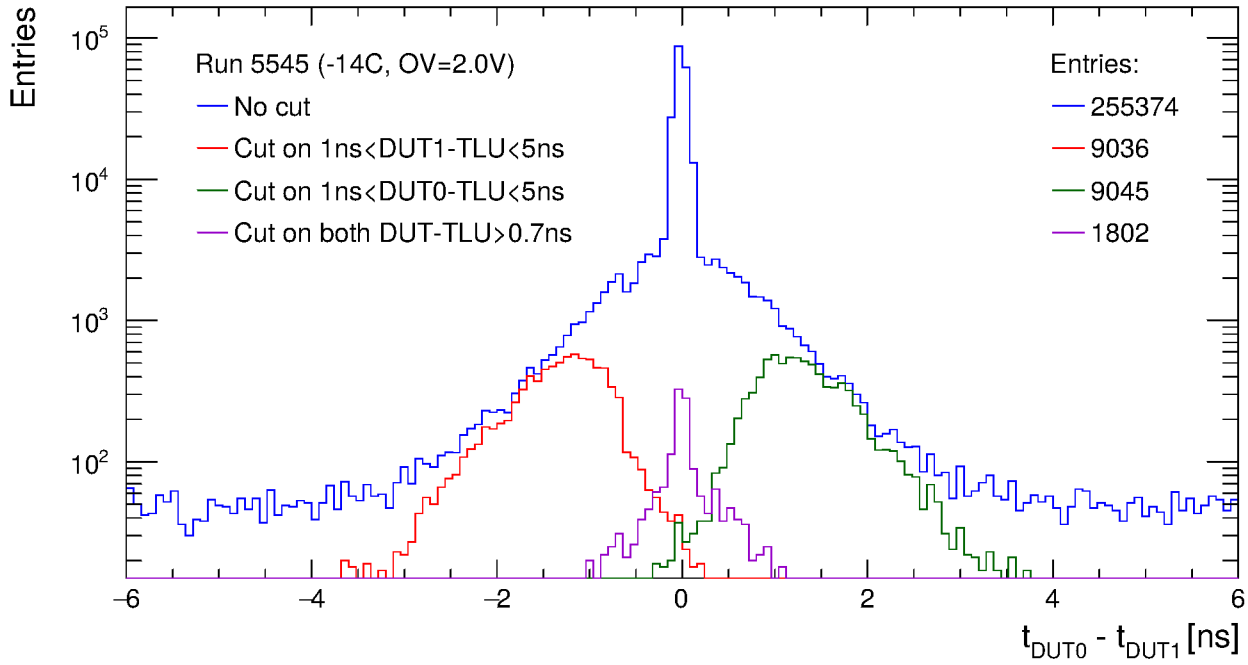
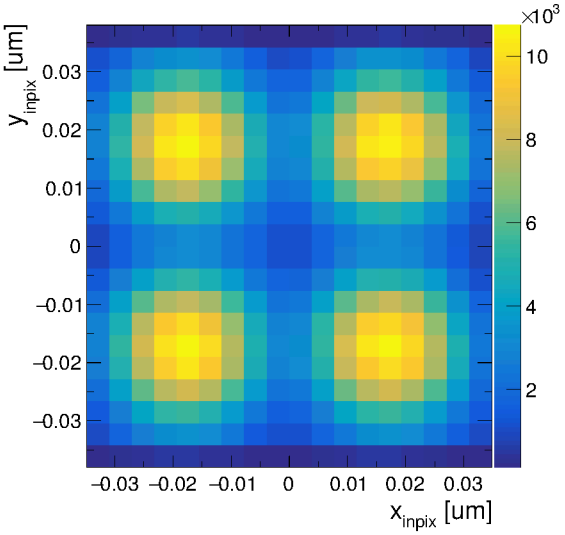
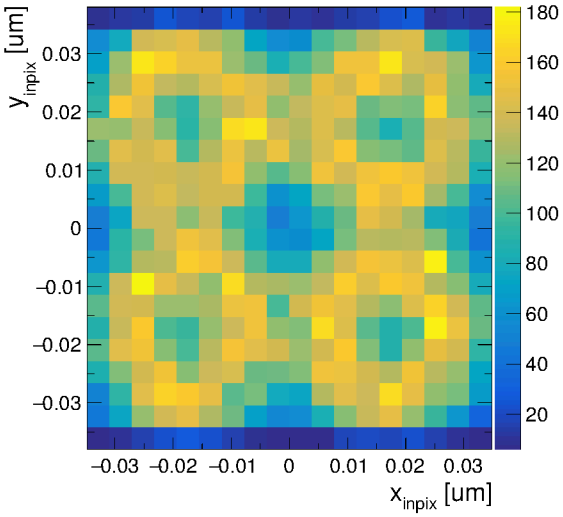


Figure 6.2: DUT-DUT time residual with cuts on timing tails in TLU-DUT distributions.

To test the origin of the tail, an in-pixel study of the time residuals has been performed using reconstructed track positions. In Fig. 6.3a, a histogram is shown filled with the track position of all associated tracks that have a valid timestamp (see subsection 5.1.1) and track χ^2 of smaller than 5. In Fig. 6.3b, the same histogram is shown with a cut applied on the TLU-DUT time residuals of $1\,\text{ns} < t_{\text{DUT0}} - t_{\text{TLU}} < 5\,\text{ns}$. The cut corresponds to the timing tail region shown in Fig. 6.1. Since the tracking resolution of the testbeam setup is estimated to about 4 µm (see Fig. 3.2), the binning of the in-pixel histograms is chosen to be 3.87 µm in x and 3.80 µm in y.



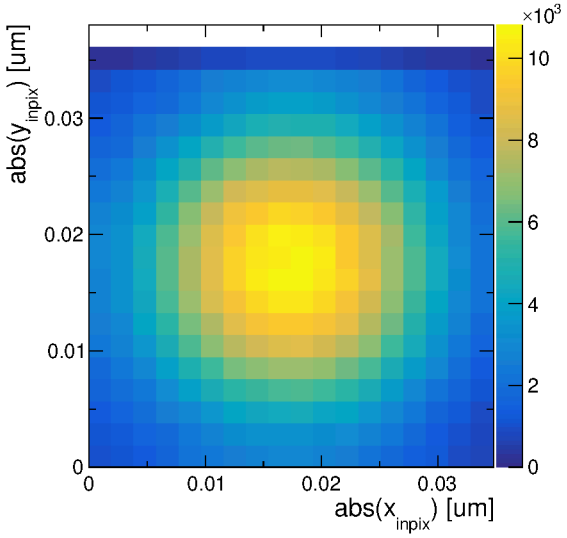
(a) In-pixel hitmap without cut.



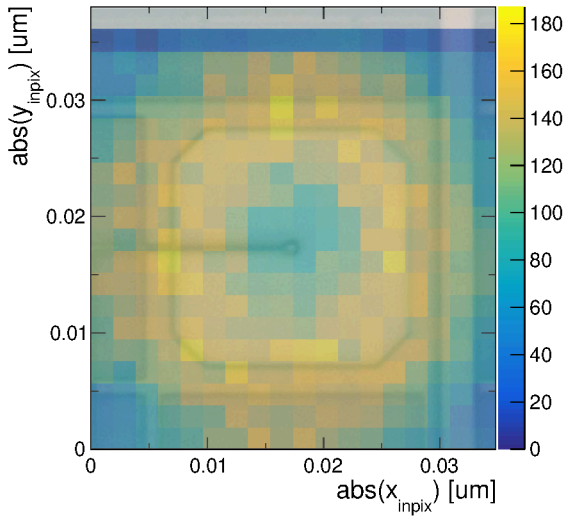
(b) In-pixel hitmap with cut on timing tail ($1\,\text{ns} < t_{\text{DUT0}} - t_{\text{TLU}} < 5\,\text{ns}$).

Figure 6.3: In-pixel hitmaps (axes to scale).

In Fig. 6.3a four areas with a high number of entries can be seen. This correlates to the four SPADs per pixel structure (see section 2.1), mainly smeared by the tracking resolution (see also section 5.3). When a particle crosses the sensor outside of a SPAD and as a result does not trigger an avalanche, the DUT is not efficient in these cases. Thus, most tracks that have an associated hit, pass through the SPAD.
In Fig. 6.3b, the area where the most tracks are associated is not the SPAD center but the area around it. This difference in entries compared to Fig. 6.4a leads to the conclusion that some particles that cross the edge of a SPAD result in a slower response of the DUT than most particles crossing the center of a SPAD.
Above and below the four SPADs a low entry region is found in both Fig. 6.3a and Fig. 6.3b. This corresponds to the area where the pixel electronics are placed, which is passive area.
Since the histograms are centered around zero, they can be folded along the center of the x and y axis. This results in an in-SPAD histogram, which is shown in Fig. 6.4 overlayed to scale with a picture of a SPAD from Fig. 2.1.



(a) In-SPAD hitmap without cut.



(b) In-SPAD hitmap with cut on timing tail ($1\,\text{ns} < t_{\text{DUT0}} - t_{\text{TLU}} < 5\,\text{ns}$).

Figure 6.4: In-SPAD hitmaps (axes to scale) overlayed with a picutre of a SPAD.

Overall, the results imply that the timing tail observed in the measurements is similar to the effect of the diffusion tail described in subsection 1.4.5. In particular, the spatial distribution of this tail within the SPAD as seen in Fig. 6.4b shows a similar behavior as the SPTR map at SPAD level shown in the paper by *Nemallapudi et al. (2016)* [16, Fig. 11].
However, when performing such comparisons it is important to keep in mind that a photon response of a SPAD is fundamentally different from a charged particle time response since a photon only creates a single eh-pair. Diffusion from depth effects can thus also appear in the middle of the SPAD for photons (see Fig. 1.7), while MIPs create so many eh-pairs that there are essentially no depth effects (except when the particle traverses at a large angle, which is not the case in this testbeam). Additionally, the SPAD design can impact edge effects.
In order to test the hypothesis that the timing tails are caused by diffusion, further studies are needed. For example, a simulation could be performed if the electric field of the SPAD design would be available.

6.2 Evaluation of the Time Resolution

The time resolution of a SiPM is typically given as FWHM of the peak. However, this approach is not feasible in this measurement for two reasons.
Firstly, the TDC resolution of the chip is in the order of the expected intrinsic SPAD time resolution σ_{SPAD}. This results in a peak that consists of only a few bins (typically not more than three before the tail begins), which limits the accuracy of a FWHM significantly.
Secondly, the DNL correction is applied in the analysis chain. This results in a binning that is extremely uneven. A test of determining the FWHM as a function of the binning showed that it can change quite drastically and quickly becomes difficult to evaluate at high binning (low entry bin between maximum and "true" half maxima).
This also makes a typical binned fit of the peak unfeasible, since the fit highly depends on the binning of the histogram. In theory, an optimal binning can be found since the bin widths of all eight TDCs are known, but this results in a binning that exceeds 10^4 bins and also requires proper reweighting of each bin. In this thesis, an unbinned maximum likelihood (ML) fit was chosen instead, which solves the binning problem to some degree. The maximum likelihood estimator is given as

$$L(\theta) = \prod_{j=1}^{N} f(y_j \,|\, \theta) \tag{6.1}$$

where f is the probability density function (PDF) of the function to be fitted, y_j the measurements, and θ the fit parameters. In short, the maximum likelihood $L(\theta)$ gives the probability for the measurement given the parameters θ, which is to be maximized by varying the fit parameters.
For better numerical performance, the negative log-likelihood (NLL) estimator

$$-\ln L(\theta) = -\sum_{j=1}^{N} f(y_j \,|\, \theta) \tag{6.2}$$

is used. Minimizing the NLL estimator is equivalent to maximizing $L(\theta)$.
RooFit [34] (which is part of ROOT [32, 33]) was chosen to perform the unbinned ML fits.

6.2.1 DUT-DUT Fit Model

The DUT-DUT time residual can be split into five components:

- Noise background
- Fast response in DUT0 and DUT1
- Fast response in DUT0 but slow response in DUT1
- Fast response in DUT1 but slow response in DUT0
- Slow response in DUT0 and DUT1

The noise background is modeled as a uniform distribution, which is sufficient in the close region around the peak. The `RooUniform` class is used for the noise PDF.
A motivation for this choice is provided in Appendix B.

The fast response for a single DUT can be modeled as a normal distribution with standard deviation σ. To describe the difference between two DUTs with a fast response, the two distributions need to be convolved. In this case, this results in a normal distribution with a standard deviation of $\sqrt{2}\,\sigma$. The `RooGaussian` class is used for this component.

The slow response of a single DUT will be modeled by an exponential distribution. Consequently, the component describing one DUT with a fast response and the other one with a slow response is a convolution of a normal distribution with an exponential distribution. The resulting distribution is

called an exponentially modified Gaussian distribution [35]. This distribution was used to fit the time response of a SPAD in *Nemallapudi et al. (2016)* [16, Eqn. 2.2]. It is given by:

$$f_{\text{fast,slow}}(t \,|\, \sigma,\, \lambda,\, t_{\text{tail}}) = \frac{\lambda}{2} \exp\left(\lambda\,(\mu + t_{\text{tail}} - t) + \frac{1}{2}\lambda^2\sigma^2\right) \operatorname{erfc}\left(\frac{\mu + t_{\text{tail}} - t + \lambda\sigma^2}{\sqrt{2}\sigma}\right) \tag{6.3}$$

The parameters of the distribution are the standard deviation σ of the fast response, the fall rate of the tail λ, and a time offset for the exponential tail t_{tail}. The model was implemented in RooFit as part of this thesis.

The component where both DUTs exhibit a slow response is described by a convolution of two exponential distributions (see Appendix C). The resulting distribution is given by:

$$f_{\text{slow,slow}}(t \,|\, \sigma,\, \lambda,\, t_{\text{tail}}) = \frac{\lambda}{2}\ \exp(-\lambda\,|t|) \tag{6.4}$$

Since this PDF is not available in RooFit at the time of writing, it was implemented as part of this thesis.

To get the final PDF for the noise and all signal components, the different PDFs need to be added with corresponding probability fractions. For the signal PDF, the probability fraction p_{peak} gives the probability for a DUT to exhibit a fast response. The combined signal PDF is then given by the sum:

$$f_{\text{signal}} = p_{\text{peak}}^2 \cdot f_{\text{fast,fast}} + p_{\text{peak}} \cdot \left(1 - p_{\text{peak}}\right) \cdot \left(f_{\text{fast,slow}} + f_{\text{slow,fast}}\right) + \left(1 - p_{\text{peak}}\right)^2 f_{\text{slow,slow}} \tag{6.5}$$

The final PDF for signal and noise is then given by:

$$f_{\text{signal+noise}} = (1 - p_{\text{noise}}) \cdot f_{\text{signal}} + p_{\text{noise}} \cdot f_{\text{noise}} \tag{6.6}$$

An overview of the fit parameters with an initial guess and the parameter limits is given in Tab. 6.1.

Parameter	Description	Initial Guess	Parameter Limits
t_{mean}	Mean offset of the distribution	mean value	$-0.2\,\text{ns}$ to $0.2\,\text{ns}$
σ	Standard deviation of the fast response	$45\,\text{ps}$	$30\,\text{ps}$ to $70\,\text{ps}$
λ	Exponential fall rate of the tail	$1.58\,\text{ns}^{-1}$	$1.48\,\text{ns}^{-1}$ to $1.68\,\text{ns}^{-1}$
t_{tail}	Time offset of the tail reponse	$40\,\text{ps}$	$-0.1\,\text{ns}$ to $0.3\,\text{ns}$
p_{peak}	Probability for fast response per DUT	0.85	0 to 1
p_{noise}	Probability for entry in noise background	$V_{\text{OV}} \cdot 0.01\,\text{V}^{-1}$	0 to 1

Table 6.1: Fit parameters in the DUT-DUT model.

It should be noted, that the model does not take into account the binning structure of the TDCs. While an unbinned fit is unaffected by the chosen binning of a histogram, it does not solve the problem of uneven binning in the measurement itself.

6.2.2 DUT-DUT Fit Results

The unbinned ML fits are performed in the range from −6 ns to 6 ns to include the noise floor in the fit. All figures shown in this subsection use a binning density of 150 bins over 12 ns.

Fig. 6.5 shows the fit result for run 5504 with a log scale in y to highlight the fit in the tail region. The dotted purple line in the figure shows the component where both DUTs exhibit a fast reaction on top of the noise floor (dotted blue line). It dominates the signal in the peak region. The dotted green line shows the component where DUT0 exhibits a fast response while DUT1 exhibits a slow response, and the dotted orange line shows the components where both DUTs exhibit a slow response. The dotted red line gives the combination of all components with at least one slow response.

In Fig. 6.6, the same fit is given on a linear scale around the peak. Overall the fitted distribution is in good agreement with the data. χ^2/n_{dof} evaluates to 2.4 in Fig. 6.5, which is the best value of all the analyzed runs. However, it should be noted that this value is highly dependent on the chosen binning and can vary by an order of magnitude. This is an effect of the unevenness in the measured data and drastically reduces the usefulness of the absolute value of the χ^2 test for this measurement. A useful χ^2 can probably be only achieved when using the reweighted binning mentioned before, however it was chosen to not follow that approach in this thesis. For the rest of this thesis, the χ^2/n_{dof} values of the fits will be given but they will not be further discussed.

An interesting feature that can be observed in Fig. 6.5 is the component where both DUTs are slow. While it does not contribute significantly to the signal, it shows similarities to the histogram in Fig. 6.2 where the slow component on both DUTs has been selected.

The parameter σ can be interpreted as the time resolution of the DUT. The fitted value of 51 ps lies within the expected range for the time resolution. An interesting value is the peak fraction p_{peak}. The value of 0.85 indicates that in about 15 % of the events a slow response in the DUT can be expected in a scenario with evenly distributed single particle tracks.
In the next section, a deeper analysis of all fit parameters will be given.

For comparison, the fits for run 5545, which is the run with the highest statistics, and run 5508, which has the highest overvoltage, are also shown. Run 5545 is shown in Fig. 6.11 ($\chi^2/n_{\text{dof}} = 13$) and Fig. 6.12, run 5508 in Fig. 6.13 ($\chi^2/n_{\text{dof}} = 3.7$) and Fig. 6.14. In both runs the fit is in good agreement with the data.

However, the fit did not work well for all runs. In particular, the fit of run 5512 gives a large χ^2/n_{dof} value of 99. The reason is quickly identified by looking at the fit in the peak region in Fig. 6.8.

To get an insight into why there is a drop in the peak in Fig. 6.8, it helps to look at the data with a finer binning. In Fig. 6.9, this is shown for run 5512 and for comparison also in Fig. 6.10 for run 5504, which has the best χ^2/n_{dof}.
In Fig. 6.9, there are four distinct peaks with a separation of about 90 ps, which is about the mean bin widths for the TDCs. This behavior is less pronounced in Fig. 6.10.
It is suspected that this difference is caused by different phases of the two clocks. In Fig. 6.9 the phases would be in sync. This leads to hits providing a timestamp from the same fine TDC bin in both DUTs. Entries between the peaks can still occur for example due to the DNL correction, but on average hits would be separated by the mean bin width. In Fig. 6.10, the phases would not be in sync but have some relative phase, which results in a higher probability for hits to land in two different fine TDC bins. Whether this effect would be pronounced or not depends on the inherent physical time resolution of the DUTs.
This theory has not been verified though and requires further studies. It is also unclear if this effect would even have an impact on the unbinned fitting procedure. If this turns out to be false, then there is something else at work in run 5512. In any case, the fact that the fit model does not describe the discrete nature of the measurement shows an inherent weakness of it.

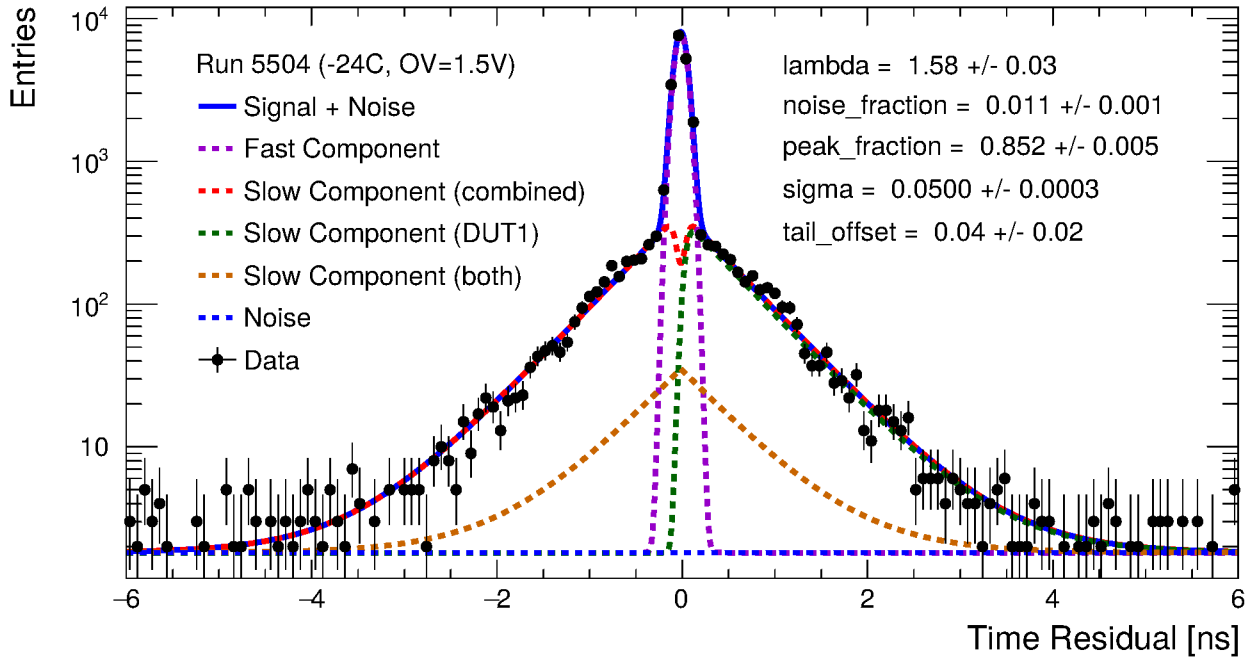


Figure 6.5: Fit for DUT-DUT time residual in log scale (run 5504). The fit for this run has the best χ^2/n_{dof} value of all fits with a value of 2.4.

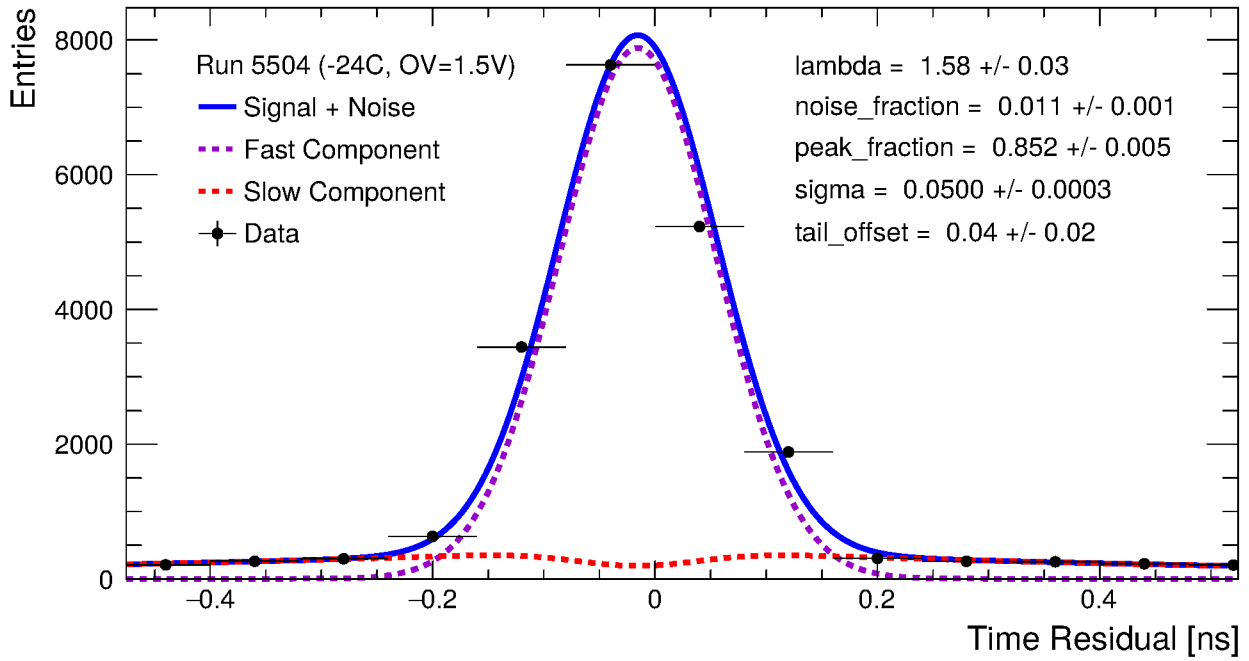


Figure 6.6: Fit for DUT-DUT time residual in peak region (run 5504). The fit for this run has the best χ^2/n_{dof} value of all fits with a value of 2.4.

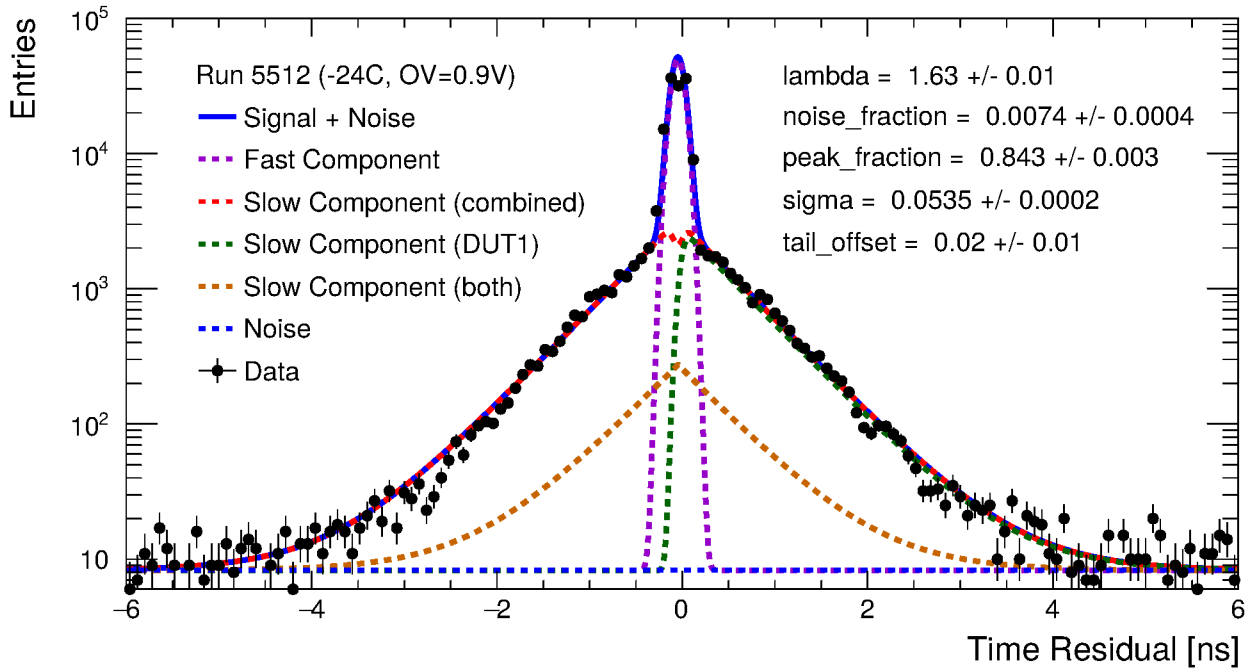


Figure 6.7: Fit for DUT-DUT time residual in log scale (run 5512). The fit for this run has the worst χ^2/n_{dof} value of all fits with a value of 99.

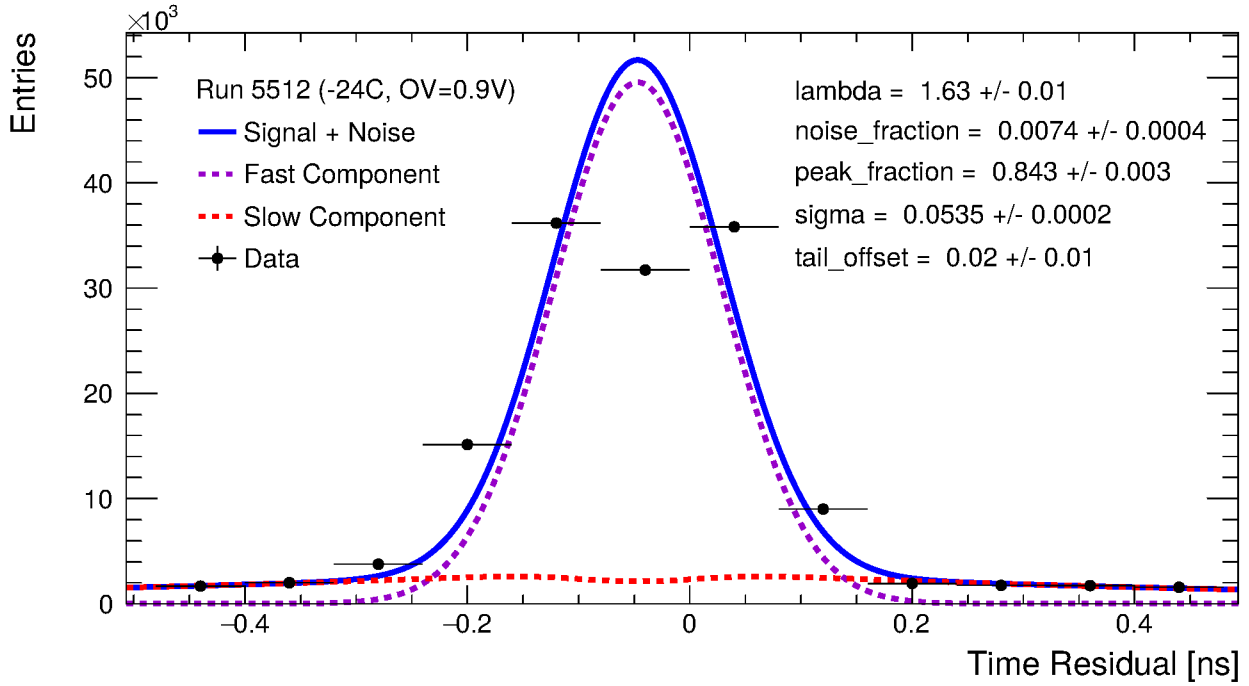


Figure 6.8: Fit for DUT-DUT time residual in peak region (run 5512). The fit for this run has the worst χ^2/n_{dof} value of all fits with a value of 99. The main factor for this is the drop in the middle of the peak.

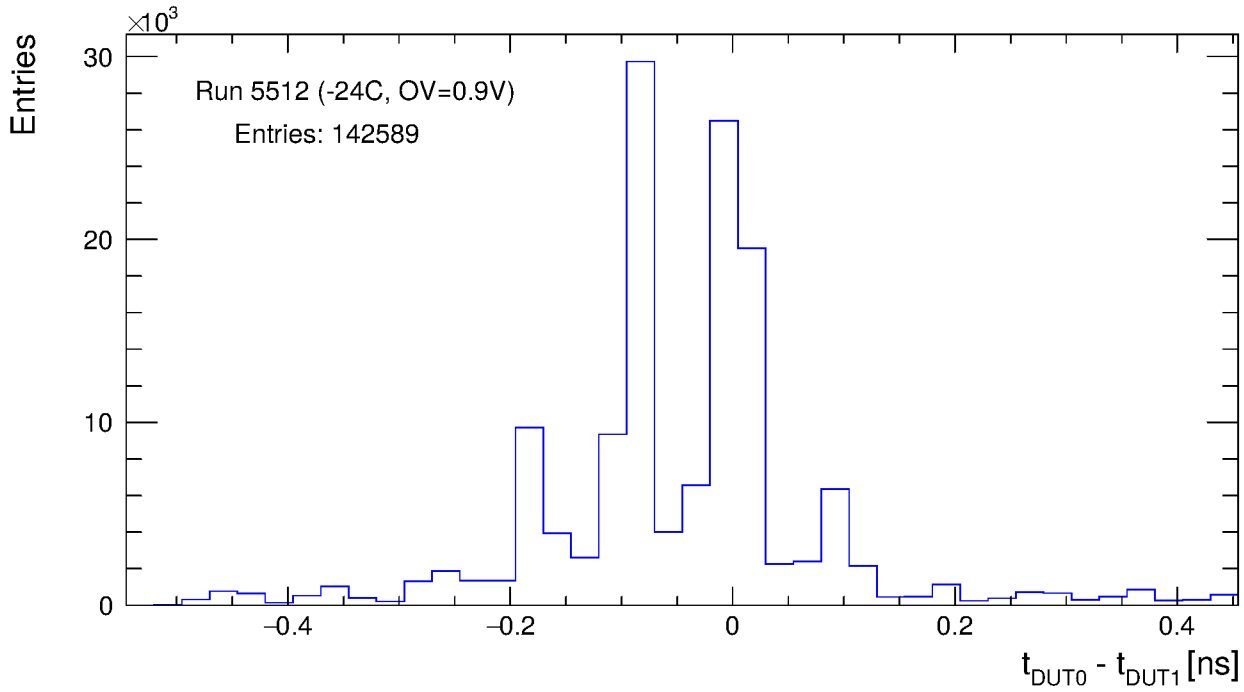


Figure 6.9: DUT-DUT time residual in peak region (run 5512). A clear separation between fine TDC bins can be seen.

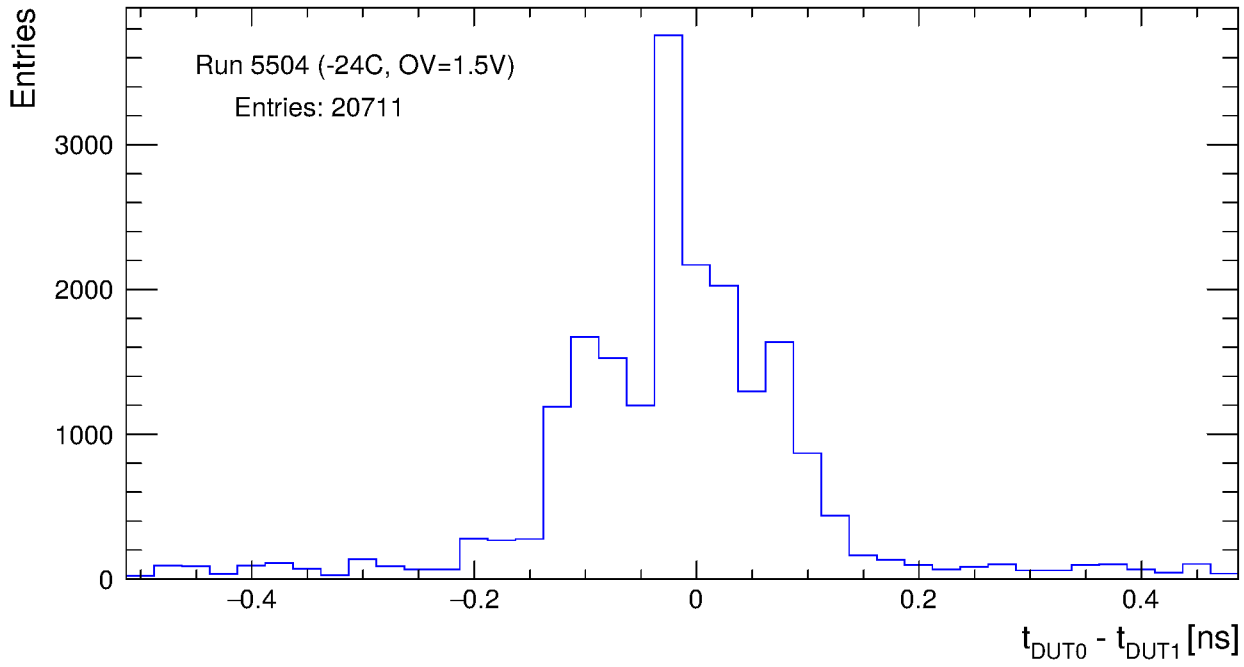


Figure 6.10: DUT-DUT time residual in peak region (run 5504). No clear separation between fine TDC bins can be seen.

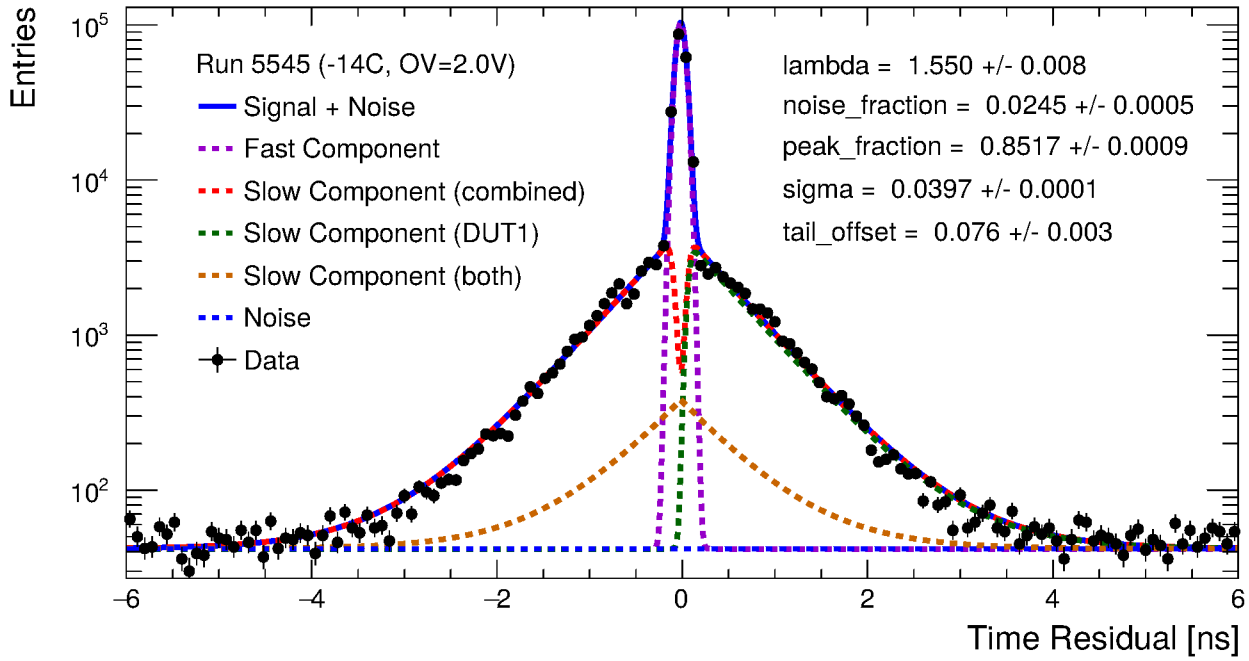


Figure 6.11: Fit for DUT-DUT time residual in log scale (run 5545). This run has the usable highest statistics without measurement artifacts.

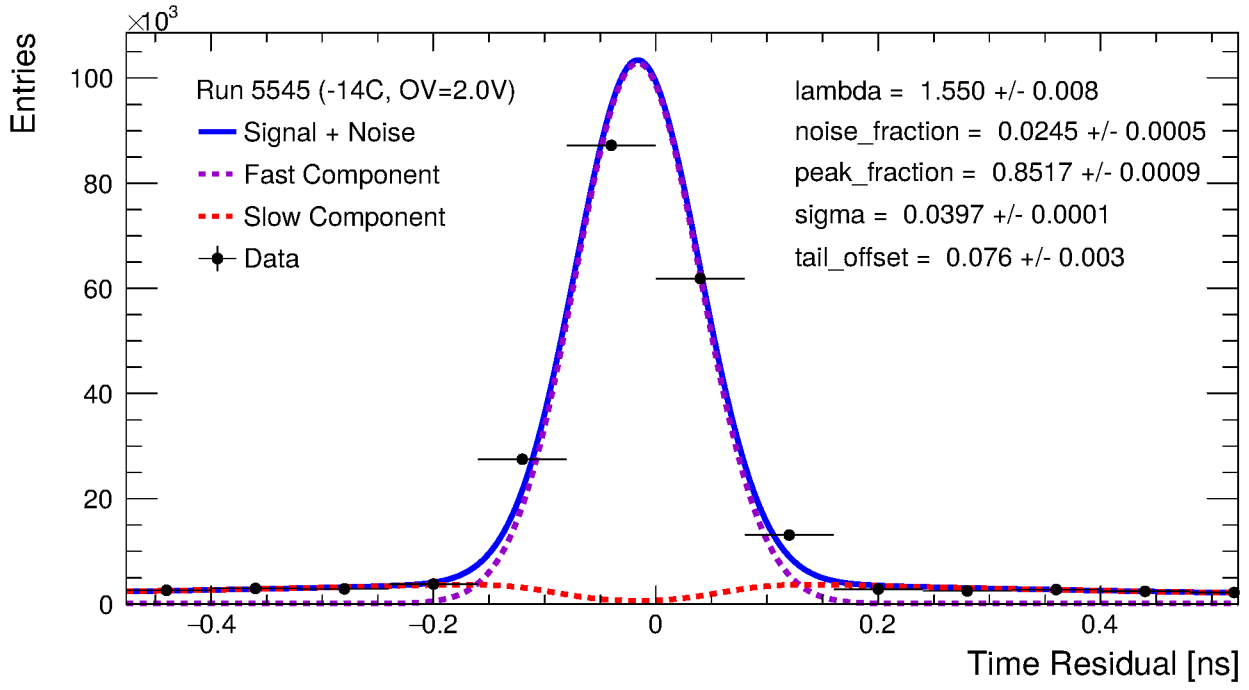


Figure 6.12: Fit for DUT-DUT time residual in peak region (run 5545). This run has the usable highest statistics without measurement artifacts.

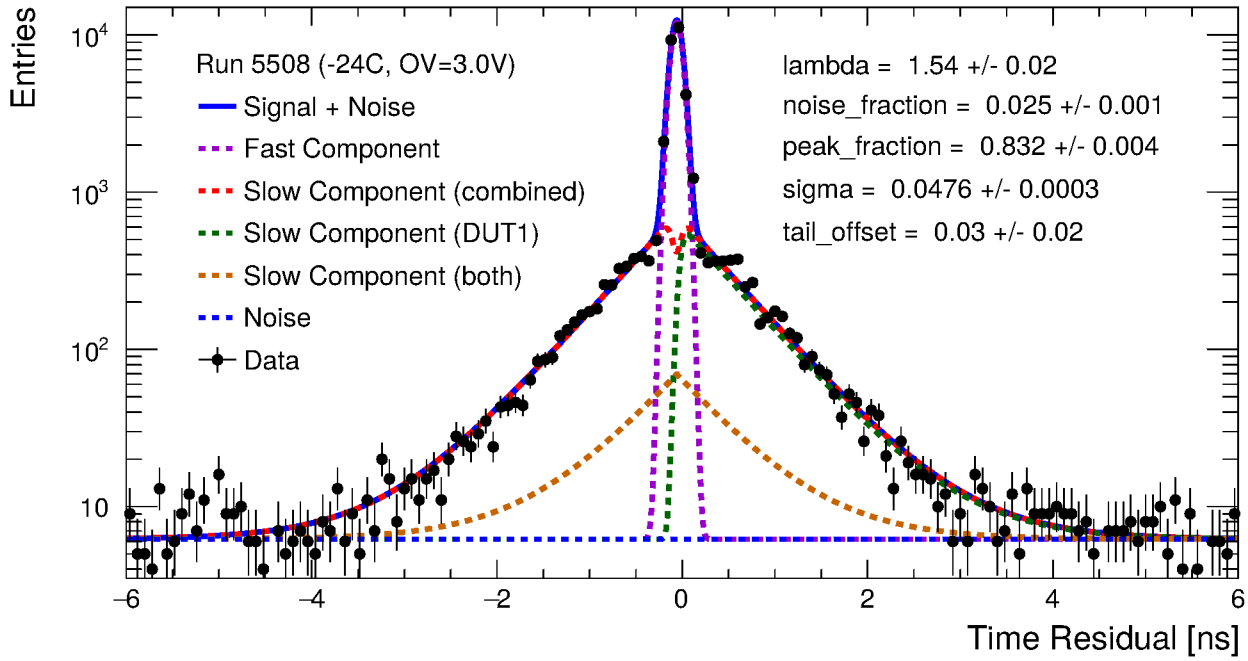


Figure 6.13: Fit for DUT-DUT time residual in log scale (run 5508). This run has the highest overvoltage applied during the testbeam campagin ($V_{OV} = 3\,\mathrm{V}$).

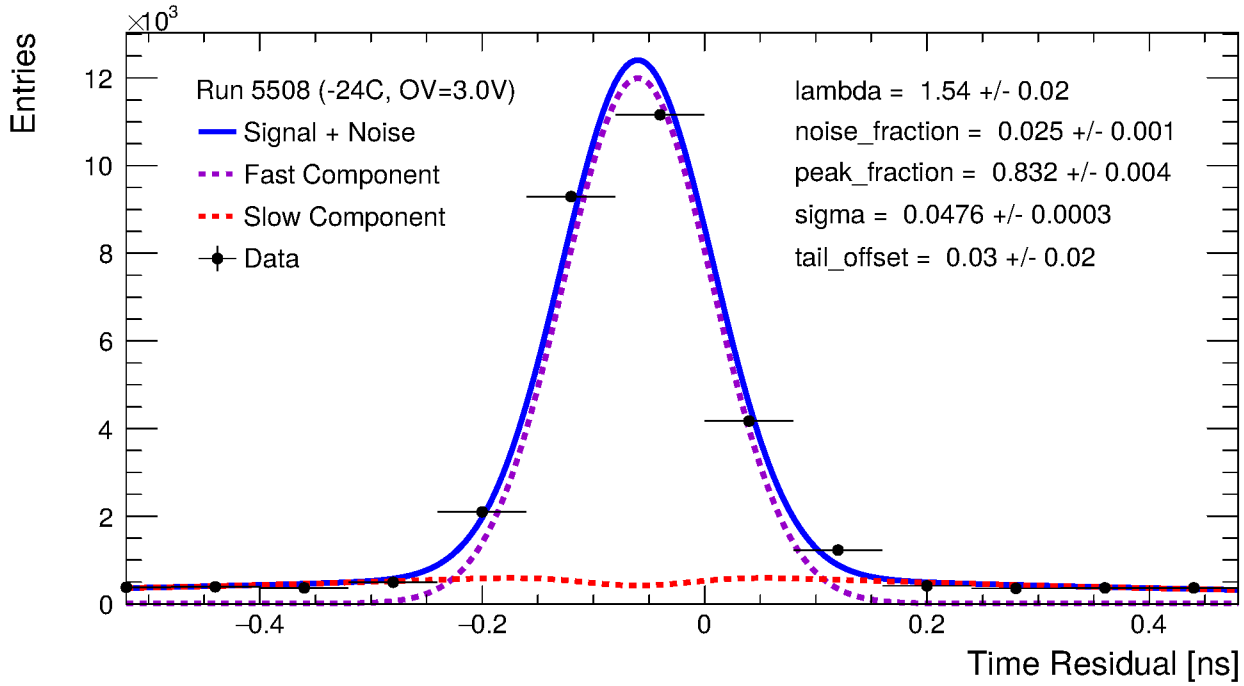


Figure 6.14: Fit for DUT-DUT time residual in peak region (run 5508). This run has the highest overvoltage applied during the testbeam campagin ($V_{OV} = 3\,\mathrm{V}$).

6.3 Timing Performance as a Function of the Overvoltage

Since the testbeam data is available with different overvoltages applied to the DUTs at two different operating temperatures, the performance of the chip can be inspected as a function of the overvoltage and temperature.[1] In thesis the time resolution and the features of the timing tail are of interest.
In the plots for the fit parameters as a function of the overvoltage shown in this section all fit parameters are plotted with an error bar in y. The error of the overvoltage is not studied in this thesis.
In Fig. 6.15, the time resolution σ is plotted as a function of the overvoltage. No clear trend is visible at either temperatures, the values for σ lie between 37 ps and 54 ps. The minuscule error bars of σ indicate that the fit model fails to evaluate the error of the parameter properly. Due to timing constraints it was decided not to investigate this issue within this thesis. As a consequence the overall time resolution independent of the temperature and overvoltage is evaluated using the average and standard deviation of all fitted values for σ, resulting in a time resolution of (46 ± 5) ps.

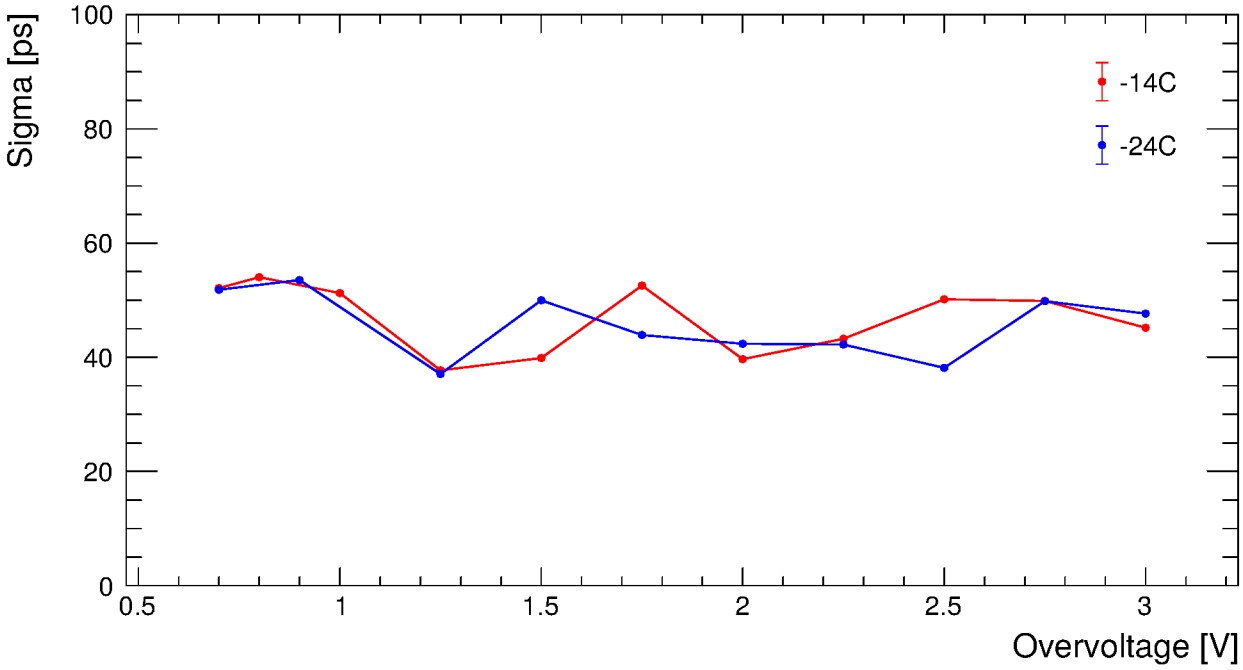


Figure 6.15: Time resolution σ as a function of the overvoltage.[1]

The parameters determining the shape of the timing tail, λ and t_{tail}, are shown in Fig. 6.16 and Fig. 6.17. The peak fraction p_{peak} determining the scale relative to the peak is shown in Fig. 6.18. Similar to the time resolution, none of three parameters show a clear trend for different overvoltages or temperatures.
The peak fraction p_{peak} in Fig. 6.18 is about 0.85 in all cases. As already mentioned in subsection 6.2.2, this value means that 85 % of the entries correspond to the fast response of about 46 ps, while 15 % have a slow repsonse. The magnitude of this factor was not investigated as part of this thesis, however it might be possible to deduce this value from the SPAD design.
It should be noted that while for σ, λ, t_{tail} and p_{peak} no clear trend was found, this does not mean these are independent of the overvoltage or temperature. This is in particular important for the time resolution σ, since it is not equal to the intrinsic time resolution of the SPAD but also includes the TDC resolution and pixel delays as explained in section 5.6. Both contributions are independent of the temperature and overvoltage.
In Fig. 6.19, the noise fraction p_{noise} is shown. It is the only parameter that shows a clear trend for different overvoltages and temperatures: the noise fraction is higher at higher temperatures and increases (linear) with overvoltage. This is the expected behavior of SPADs (see subsection 1.4.4).

[1]The temperature values shown in the legends of the plots are the chiller temperatures and not the chip temperatures since the on-chip temperature diodes were not calibrated.

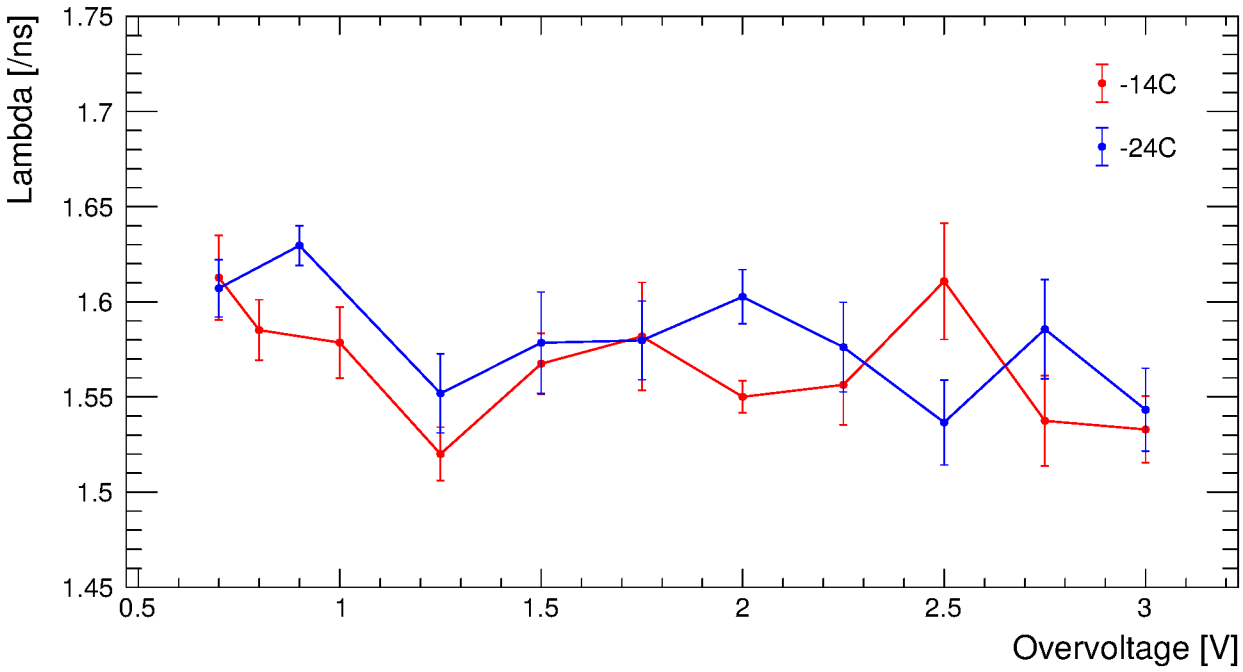


Figure 6.16: Exponential timing tail fall rate λ as a function of the overvoltage.
No clear trend for different overvoltages or temperatures can be seen.[1]

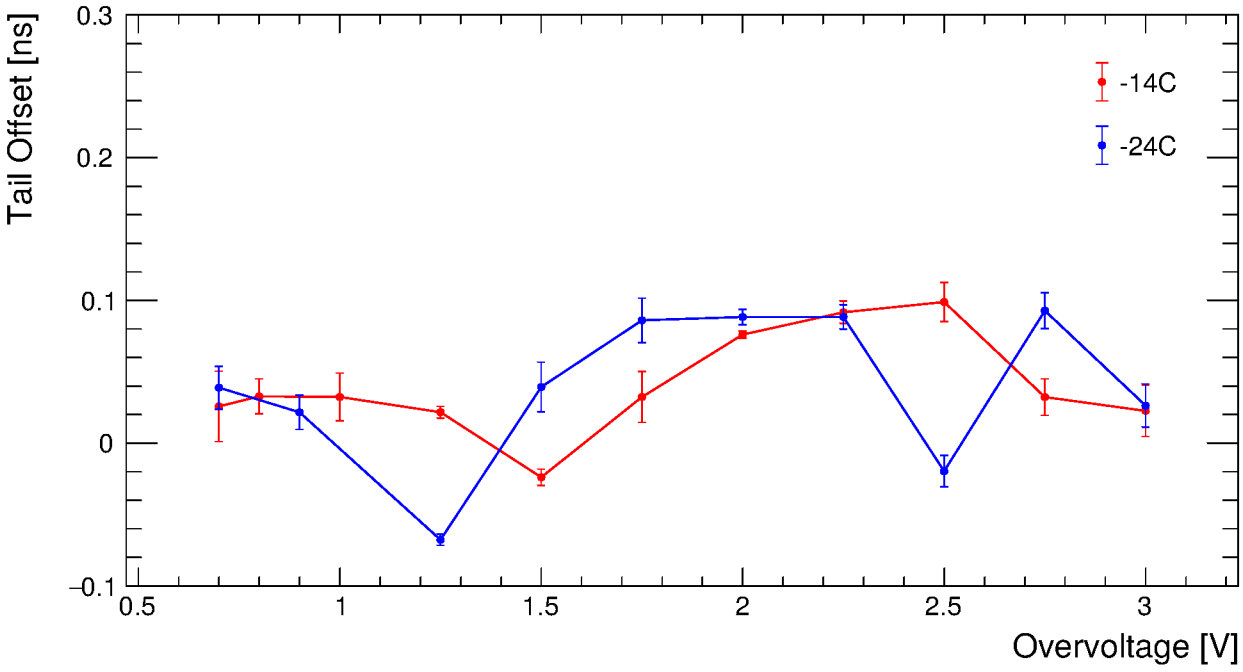


Figure 6.17: Timing tail offset t_{tail} as a function of the overvoltage.
No clear trend for different overvoltages or temperatures can be seen.[1]

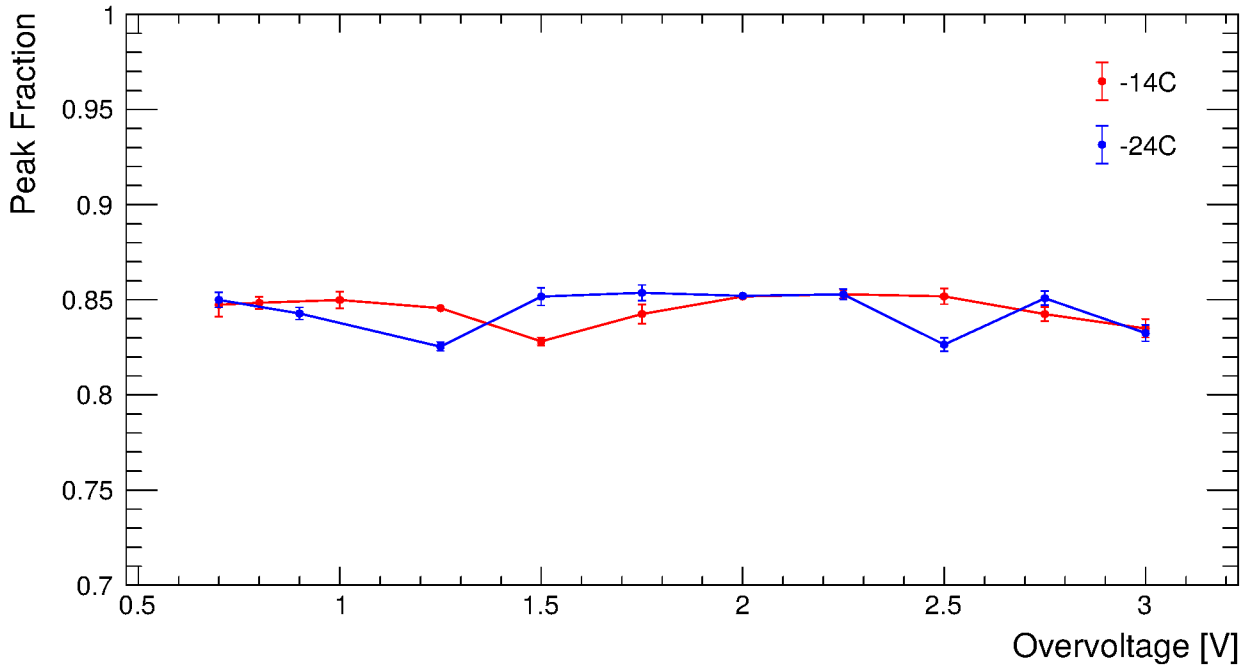


Figure 6.18: Peak fraction p_{peak} as a function of the overvoltage.
No clear trend for different overvoltages or temperatures can be seen.[1]

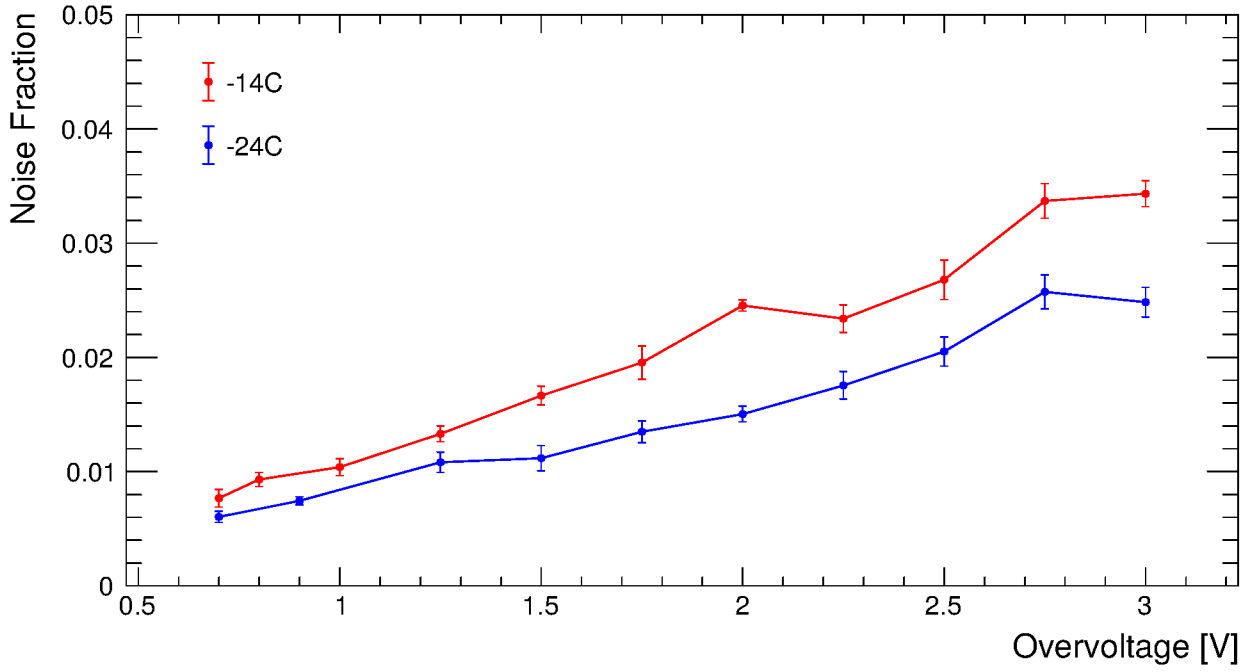


Figure 6.19: Noise fraction p_{noise} as a function of the overvoltage.
As expected, the noise fraction increases with higher overvoltage (via quantum tunneling) and higher temperature (via thermal excitation).[1]

Chapter 7

Conclusion

7.1 Summary

A novel digital monolithic SiPM (dSiPM) developed at DESY has been tested in a 4 GeV electron beam conducted at the DESY II testbeam facility. Two dSiPMs have been used to conduct a coincidence measurement with the goal of studying the timing performance of the dSiPM. The chip was fully integrated into the testbeam setup, requiring no external measurement devices such as oscilloscopes.[1]

Two unexpected measurement artifacts have been found, one arising from an issue in the chip (see subsection 5.1.1) and the other one arising from electrical interference from the testbeam telescope chiller (see subsection 5.1.2). A procedure was developed to filter these measurement artifacts.

The testbeam analysis was performed with Corryvreckan (see chapter 4), for which a new module was developed to study the timing performance between two DUTs (see subsection 4.6.1). A concept to correct the non-linearities in the TDC of the dSiPM has been developed and implemented in the EUDAQ converter to improve the accuracy of the timestamps (see section 5.4 and section 5.6).

In a first comparison with the timestamps provided by the TLU using a scintillator+veto combination, an upper limit of 332 ps was found for the time resolution of the dSiPM (see section 5.5). This limit is dominated by the jitter of the scintillator+veto combination and the time resolution of the TLU.

In an analysis of the time residuals between the two DUTs, a slow timing tail has been found (see section 5.6). It was further found that this timing tail depends on the position of the particle crossing through the chip (section 6.1). In particular, it was found that entries causing a slow response are more common in the SPAD edges, hinting that this effect might be by slow diffusion outside of the amplification region (see also subsection 1.4.5).

To evaluate the timing performance quantitatively, a model was developed including describing both the fast response and the slow response of both DUTs (see subsection 6.2.1). The model has been used to fit time residual distributions for runs with different overvoltages to study the timing performance as a function of the overvoltage (see section 6.3). The fit model was not able to show any significant behavior except for the noise component, which increases with temperatures and overvoltage as expected. This could be caused by the fit model itself since it does not describe the discrete nature of the measurement (see subsection 6.2.2). However, it could also hint that the time resolution is not limited by the SPAD but by factors that do not depend on the overvoltage within the uncertainty of the measurement, such as the TDC resolution or the relative pixel delays between the two devices.

As a final result, the time resolution of the dSiPM was estimated as (46 ± 5) ps in 85 % of the reconstructed events. Overall, the presented studies are a good first step in the characterization of digital SiPMs for application in fast timing tracking detectors.

[1] As far as published literature goes, this measurement has not been performed before. The most similar measurement tested analog SiPMs in charged particle detection [6], where a timing performance of 30 ps to 70 ps has been found.

7.2 Outlook

The work in this thesis is a good step, but further measurements are necessary to show that digital SiPMs are viable for application in fast timing detectors. In this section, an outlook is given for potential future studies.

An ongoing project is the study of the timing performance with a fast (picosecond) laser. This allows for various studies, for example if the spot size is small enough the in-SPAD timing performance could be studied to verify the results found in this thesis. Another possibility would be to study the time delay between a hit and the registered timestamp in the TDC for each pixel. This could potentially be used to correct the pixel delays in testbeam data, further improving the time resolution. A setup for this study is already available.

Sensors for 4D tracking require MIP detection efficiencies of 100 % [1], which is significantly higher than is currently achieved by the dSiPM. Thus overcoming this challenge is a necessary step for the application of digital SiPMs in fast timing detectors. A promising path forward is the combination of a dSiPM with a thin LYSO in front of the sensor. In this approach the LYSO produces many photons when a MIP passes through it, resulting in a theoretical efficiency of almost 100 %, while only reducing the spatial resolution by a small amount. The diffusion tail found in this thesis is less relevant there since many photons create hits in multiple pixels. A prototype for this already exists in the lab but has not been used in a testbeam yet. The analysis from this thesis will be extensively used for the analysis of the new prototype.

Another possibility is to study the chip in a testbeam with a higher spatial resolution to study the in-SPAD effects more precisely. For example, a high energy proton beam from PS or SPS could be used. Further, the statistics of testbeam data for timing could be increased by avoiding the interference from the MIMOSA chiller. The dSiPM could also be used in conjunction with an LGAD to cross-check the time resolution with a different device.

Lastly, to reduce the noise of the chip a redesign with a different SPAD design is possible, although costly. In the last years, new SPAD designs in imaging-specific CMOS process nodes became available that have higher fill factors and lower DCRs. The results presented in this thesis also allow fine-tuning the well-working TDC design for an even better timing resolution.

The project is working on a paper that includes some of the results presented in this thesis.

Appendix A

Shape of the Spatial Residual Distribution

The DUT association search area can be described via:

$$\left(\frac{x}{c_x}\right)^2 + \left(\frac{y}{c_y}\right)^2 \leq 1 \tag{A.1}$$

where c_x and c_y are the respective search radii. Assuming a constant noise probability N_ρ, we can construct the following noise PDF for the spatial residuals:

$$N(x,y) = N_\rho \cdot \begin{cases} 1 & \text{if } \left(\frac{x}{c_x}\right)^2 + \left(\frac{y}{c_y}\right)^2 \leq 1 \end{cases} \tag{A.2}$$

The signal PDF for the spatial residuals can be modeled by summing over the shape functions B of the four SPADs in a pixel:

$$S(x,y) = \sum B\left(x \pm d_x, y \pm d_y\right) \tag{A.3}$$

where d_x and d_y are the distances between the center of the SPAD and the origin in x and y. For simplicity $B(x,y)$ can be assumed to just be a 2D box function for now that is 1 if $|x| \leq b_x \wedge |y| \leq b_y$, where b_x and b_y are the half widths of the SPAD box respectively.
In Fig. A.1, the noise PDF is illustrated in red, and the signal PDF in blue.

Let us define

$$E(x,y) = N(x,y) + S(x,y) \tag{A.4}$$

which gives a PDF for the expected spatial residual distribution. Note that this includes some simplifications and ignores correlations between noise hits and signal hits, which can indeed happen. To get the spatial residual distribution in x, we perform a projection of these functions. This gives

$$E_x(x) = \int_{-c_y\sqrt{1-(\frac{x}{c_x})^2}}^{c_y\sqrt{1-(\frac{x}{c_x})^2}} \mathrm{d}y \; E(x,y) \tag{A.5}$$

$$= \int_{-c_y\sqrt{1-(\frac{x}{c_x})^2}}^{c_y\sqrt{1-(\frac{x}{c_x})^2}} \mathrm{d}y \; \left(N_\rho + \sum B\left(x \pm d_x, y \pm d_y\right)\right) \tag{A.6}$$

The projection of the shape functions B when assuming a simple 2D box function is just a one-dimensional box function B_x. Thus we get

$$E_x(x) = \int_{-c_y\sqrt{1-(\frac{x}{c_x})^2}}^{c_y\sqrt{1-(\frac{x}{c_x})^2}} \mathrm{d}y \; N_\rho + 2\sum B_x\left(x \pm d_x\right) \tag{A.7}$$

$$= 2c_y\sqrt{1-\left(\frac{x}{c_x}\right)^2}\, N_\rho + 2\sum B_x\left(x \pm d_x\right) \tag{A.8}$$

The assumption that B is a perfect box function is flawed in two major ways. For one, the real spatial residual distribution includes Gaussian smearing from the tracking resolution. Secondly, assuming that the SPAD itself has a perfect box response is unrealistic, since the detection efficiency is decreasing at the edges due to the doping profile.
A simplification to solve this issue is to just smear the projected box function B_x with a Gaussian. The result is a "Gaussian box function"

$$\mathrm{gBox}(x) = \frac{1}{2}\left(\mathrm{erf}\left(\frac{x+b}{\sqrt{2}\,\sigma}\right) - \mathrm{erf}\left(\frac{x-b}{\sqrt{2}\,\sigma}\right)\right) \tag{A.9}$$

The projection of the expected spatial residual distribution with a Gaussian box function for B_x is plotted in black in Fig. A.1 (note that the function is not normalized). The dashed line reflects the noise component.

While the function can be used to get a rough description, it uses many simplifications. As mentioned before, assuming a square signal distribution is a simplification, and applying smoothing after the projection to simplify the calculation is less precise. Further, it assumes a homogeneous noise distribution, which is not strictly correct either. The existence of neighboring pixels is also ignored, as well as any reconstruction effects such as clustering.

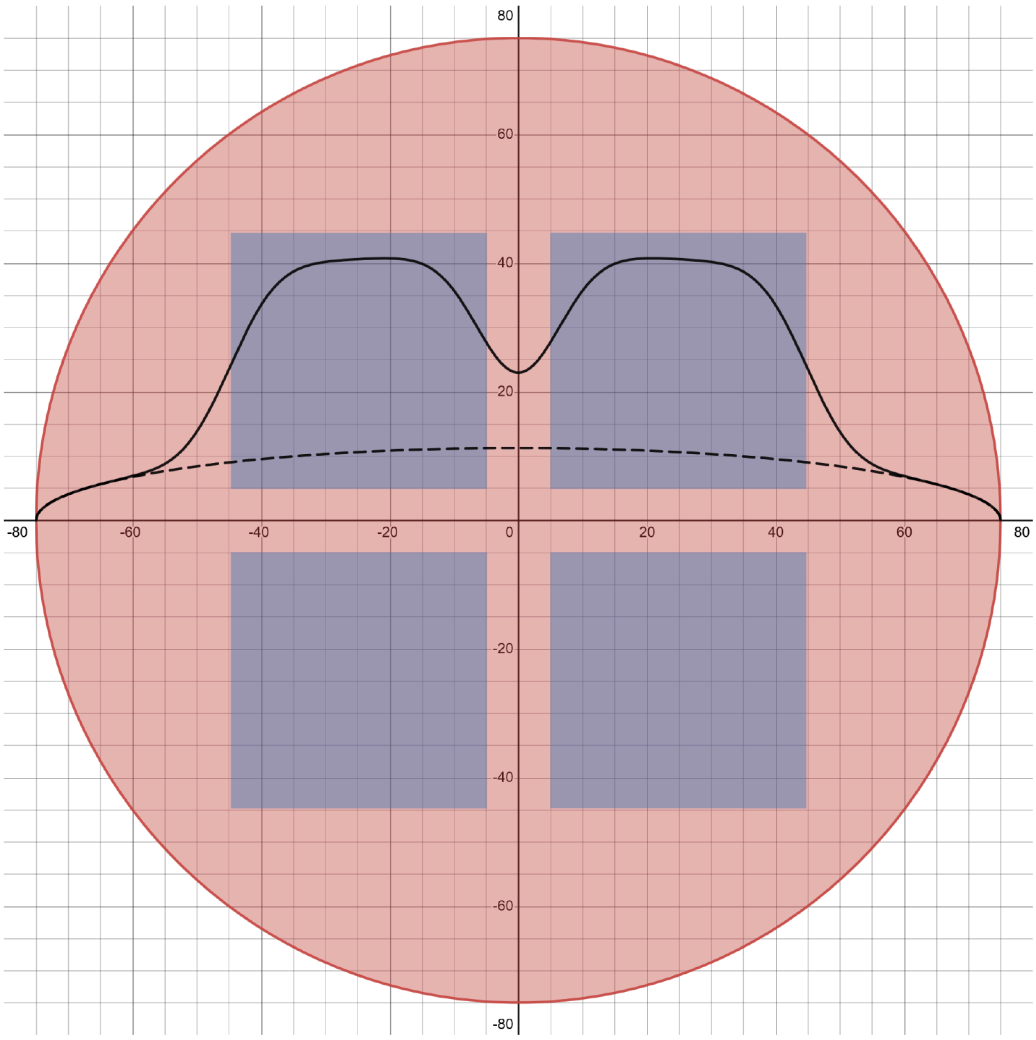


Figure A.1: Illustration of the 2D spatial residual distribution for the noise (red) and signal (blue) and its projection along x in black (with some Gaussian smoothing for the signal component). The dashed line is the noise component. Pixel dimensions and noise roughly to scale.

Appendix B

Noise Background in the Time Residual Distribution

Let's assume that within a time Δt there is a probability Δp to get a noise hit, with the limit $\frac{\Delta t}{\Delta p} = \tau$ for $\Delta t \to 0$. Assuming we only have one TDC for the entire chip, we can construct the probability p_n to get a noise hit with a timestamp of $t_n = n \cdot \Delta t$. The probability p_1 to get a noise hit between t_0 and t_1 is just Δp. For p_2, we require that there was no hit between t_0 and t_1, thus the probability is $(1 - \Delta p) \cdot \Delta p$. We can continue the pattern:

$$p_1 = \Delta p \tag{B.1}$$
$$p_2 = \Delta p \cdot (1 - \Delta p) \tag{B.2}$$
$$p_2 = \Delta p \cdot (1 - \Delta p)^2 \tag{B.3}$$
$$\dots$$
$$p_n = \Delta p \cdot (1 - \Delta p)^{n-1} \tag{B.4}$$

We can form the PDF P by replacing Δp with $\frac{\Delta t}{\tau}$ let $\Delta t \to 0$:

$$p(t_n) = \frac{\Delta t}{\tau} \cdot \left(1 - \frac{\Delta t}{\tau}\right)^{\frac{t_n}{\Delta t} - 1} \tag{B.5}$$
$$P(t) = \lim_{\Delta t \to 0} \frac{p(t)}{\Delta t} \tag{B.6}$$
$$= \frac{1}{\tau} \exp\left(-\frac{t}{\tau}\right) \tag{B.7}$$

Let's now assume that there is a particle coming at time t_p within the length of the frame t_f ($0 \leq t_p \leq t_f$) and our detector has an efficiency ε. We can now create the noise PDF within a frame where a particle passes:

$$P_N\left(t, t_p\right) = \frac{1}{\tau} \exp\left(-\frac{t}{\tau}\right) \cdot \begin{cases} 1 & \text{if } 0 \leq t < t_p \\ 1 - \varepsilon & \text{if } t_p \leq t \leq t_f \end{cases} \tag{B.8}$$

We are now interested in the time residual PDF where we correlate a noise hit with a particle. For this we first need to consider the PDF that gives the time of the particle within the frame, which is just a uniform distribution:

$$P_p\left(t_p\right) = \begin{cases} \frac{1}{t_f} & \text{if } 0 \leq t_p \leq t_f \end{cases} \tag{B.9}$$

To now get the desired time residual distribution P_{tr}, let's introduce a new variable $\tilde{t} = t - t_p$. The time residual distribution can be found by convolution:

$$P_{\text{tr}}\left(\tilde{t}\right) = \int_{-\infty}^{+\infty} \mathrm{d}t_p \; P_p\left(t_p\right) \cdot P_N\left(\tilde{t} + t_p, t_p\right) \tag{B.10}$$
$$= \frac{1}{\tau \cdot t_f} \int_0^{t_f} \mathrm{d}t_p \; \exp\left(-\frac{\tilde{t} + t_p}{\tau}\right) \cdot \begin{cases} 1 & \text{if } 0 \leq \tilde{t} + t_p < t_p \\ 1 - \varepsilon & \text{if } t_p \leq \tilde{t} + t_p \leq t_f \end{cases} \tag{B.11}$$

We can split the two cases, the first case gives $\tilde{t} < 0$, and the second $\tilde{t} \geq 0$.

Case $\tilde{t} < 0$: The constrait further gives $t_p \geq -\tilde{t} = |\tilde{t}|$. This restricts the integration limits:

$$P_{\text{tr}}\left(\tilde{t}\right) = \frac{1}{\tau \cdot t_f} \int_{-\tilde{t}}^{t_f} \mathrm{d}t_p \ \exp\left(-\frac{\tilde{t}+t_p}{\tau}\right) \tag{B.12}$$

$$= \frac{1}{\tau \cdot t_f} \exp\left(-\frac{\tilde{t}}{\tau}\right) \int_{-\tilde{t}}^{t_f} \mathrm{d}t_p \ \exp\left(-\frac{t_p}{\tau}\right) \tag{B.13}$$

$$= \frac{1}{\tau \cdot t_f} \exp\left(-\frac{\tilde{t}}{\tau}\right) \left[-\tau \cdot \exp\left(-\frac{t_p}{\tau}\right)\right]_{t_p=-\tilde{t}}^{t_p=t_f} \tag{B.14}$$

$$= -\frac{1}{t_f} \exp\left(-\frac{\tilde{t}}{\tau}\right) \left(\exp\left(-\frac{t_f}{\tau}\right) - \exp\left(-\frac{\tilde{t}}{\tau}\right)\right) \tag{B.15}$$

$$= \frac{1}{t_f} \left(1 - \exp\left(-\frac{\tilde{t}+t_f}{\tau}\right)\right) \tag{B.16}$$

Case $\tilde{t} \geq 0$: The constraint further gives $t_p \leq t_f - \tilde{t}$. Similarly to before:

$$P_{\text{tr}}\left(\tilde{t}\right) = \frac{1-\varepsilon}{\tau \cdot t_f} \int_{0}^{t_f-\tilde{t}} \mathrm{d}t_p \ \exp\left(-\frac{\tilde{t}+t_p}{\tau}\right) \tag{B.17}$$

$$= -\frac{1-\varepsilon}{t_f} \exp\left(-\frac{\tilde{t}}{\tau}\right) \left(\exp\left(-\frac{t_f-\tilde{t}}{\tau}\right) - 1\right) \tag{B.18}$$

$$= \frac{1-\varepsilon}{t_f} \left(\exp\left(-\frac{\tilde{t}}{\tau}\right) - \exp\left(-\frac{t_f}{\tau}\right)\right) \tag{B.19}$$

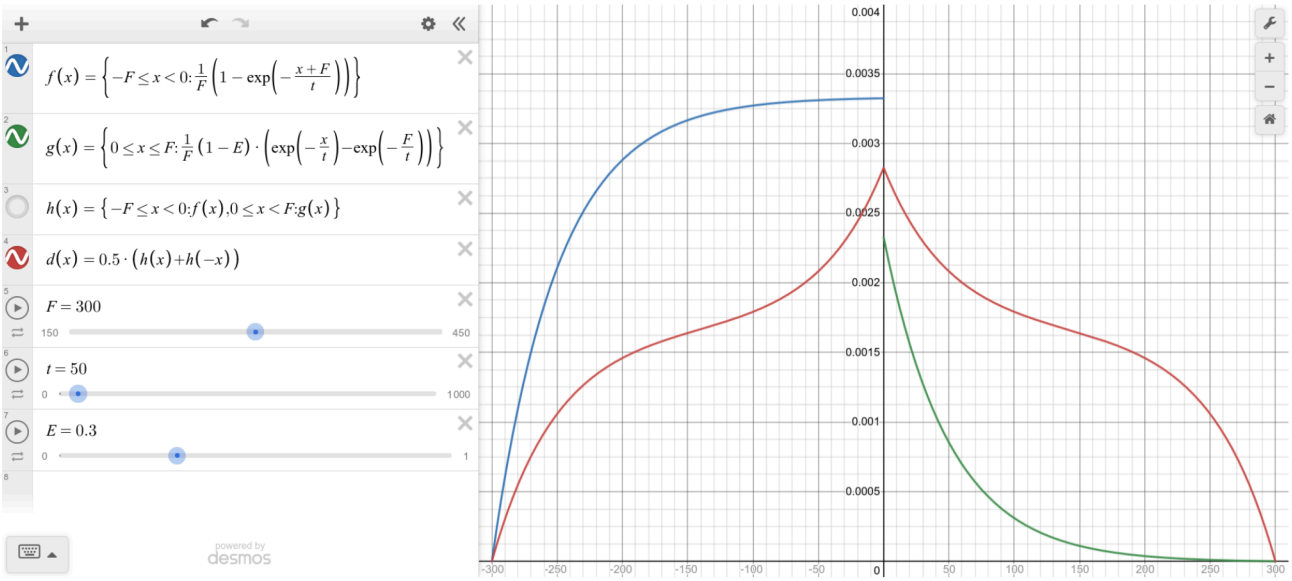
Combination: The PDF can be written for both cases as:

$$P_{\text{tr}}\left(\tilde{t}\right) = \frac{1}{t_f} \cdot \begin{cases} 1 - \exp\left(-\frac{\tilde{t}+t_f}{\tau}\right) & \text{if } -t_f \leq \tilde{t} < 0 \\ (1-\varepsilon)\left(\exp\left(-\frac{\tilde{t}}{\tau}\right) - \exp\left(-\frac{t_f}{\tau}\right)\right) & \text{if } 0 \leq \tilde{t} \leq t_f \end{cases} \tag{B.20}$$

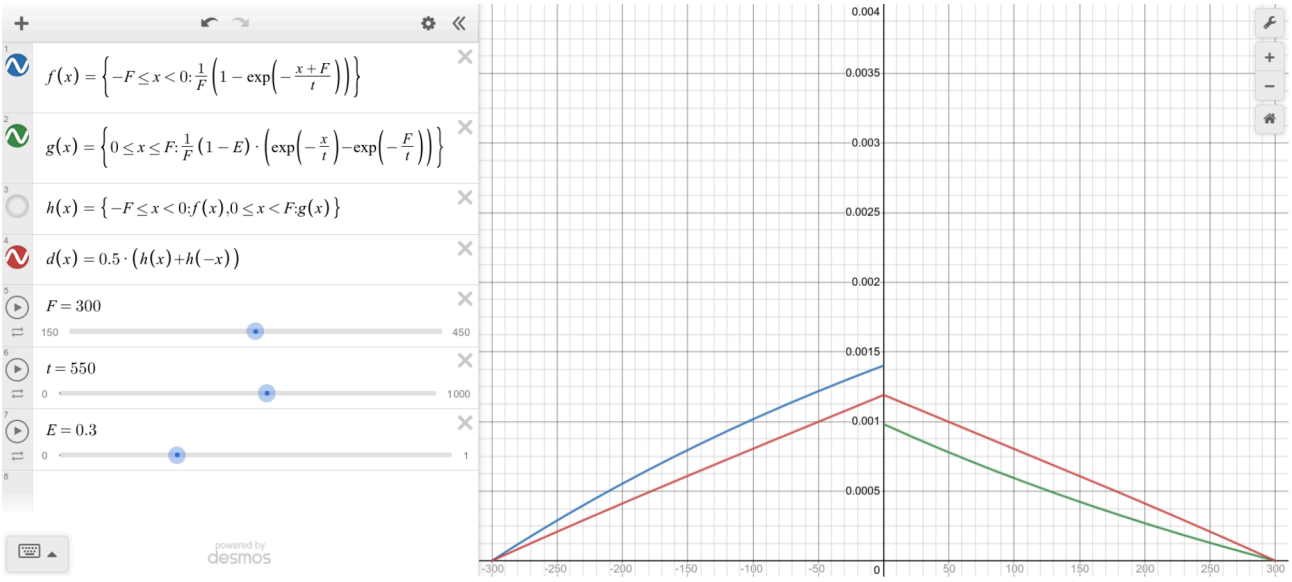
This gives the expected noise background in the TLU-DUT time residual distribution. However some simplifications were made, namely that there is only one frame and no dead time, that there is only one TDC, and that the TDC has perfect timestamping. In Fig. B.1 the function is plotted with two different values for τ.

Intepretation Let's take a look at Fig. B.1b, where $\tau = 550\,\text{ns}$ relative to a frame length of 300 ns. The detection efficiency ε is set to 0.3. In blue the function is shown before the particle arrival, and in green after the particle arrival.
The drop between the two parts is determined by the efficiency, if it is set to 1 then there is no jump. This jump can be seen partially in Fig. 6.1 (smearing by the TLU time resolution and overlapping with the non-noise components).
The left and right parts appear roughly linear. Since τ is larger than the frame length, we don't expect many noise hits per frame. To get an entry of $-300\,\text{ns}$, there must be a noise hit right at the beginning of the frame and the particle must arrive right at the end of the frame. As we get closer to zero, we get more possibilities for the position of the noise and particle timestamp in the frame. Similarly, the distribution decreases after the particle's arrival. Overall the distribution is described by a convolution of two "box" functions (uniform distributions).
Now let's look at Fig. B.1a. As we can see, the linear behavior is gone. Since τ is much smaller than t_f, we expect many noise hits in a frame and thus a timestamp at the beginning of the frame. Thus the slope quickly decreases since many of the new possibilities have a low probability. The part after the particle is dominated by the sharp exponential distribution given by pure noise.

DUT-DUT noise background For the DUT-DUT case, we can simply assume that both DUTs follow the same noise distribution, and thus mirror one distribution and add it. In Fig. B.1 these distributions are given by the red graph. Note that this is only true for the case where only one DUT has a noise hit. However, the probability that both DUTs have a noise hit within the association window is relatively low.
Since τ in our case is close to the 550 ns case in Fig. B.1b than to the 50 ns case in Fig. B.1a, we can assume that the distribution is roughly flat in a small range (less than 10 ns) around zero.



(a) Noise background with $\tau = 50$ ns.



(b) Noise background with $\tau = 550$ ns.

Figure B.1: Noise background in the time residual distribution. In blue the function for the TLU-DUT before the particle arrival, in green the function after the particle arrival. In red the distribution is mirrored along the y axis and averaged with itself to give the noise background in the DUT-DUT time residual distribution.

Appendix C

Derivation of $f_{\text{slow,slow}}$

Start with the exponential distribution with an offset t_0

$$f(t) = \lambda \exp\left(-\lambda\left(t - t_0\right)\right) \Theta(t - t_0) \tag{C.1}$$

where Θ is the Heaviside function. The distribution mirrored around $t = 0$ is then given by

$$g(t) = \lambda \exp(-\lambda\left(-t - t_0\right)) \Theta(-t - t_0) \tag{C.2}$$

$f_{\text{slow,slow}}$ is the convolution of the two functions, thus given by

$$f_{\text{slow,slow}}(t) = (f * g)(t) = \int_{-\infty}^{+\infty} f(\tau)\ g(t - \tau)\ \mathrm{d}\tau \tag{C.3}$$

$$= \int_{-\infty}^{+\infty} \lambda^2 \exp(-\lambda\left(\tau - t_0\right)) \exp(-\lambda\left(-t + \tau - t_0\right)) \Theta(\tau - t_0)\ \Theta(-t + \tau - t_0)\ \mathrm{d}\tau \tag{C.4}$$

Subsituting $\hat{\tau} = \tau - t_0$ leaves the integration boundaries unchanged.

$$f_{\text{slow,slow}}(t) = \int_{-\infty}^{+\infty} \lambda^2 \exp(-\lambda\hat{\tau}) \exp(-\lambda\left(-t + \hat{\tau}\right)) \Theta(\hat{\tau})\ \Theta(-t + \hat{\tau})\ \mathrm{d}\hat{\tau} \tag{C.5}$$

$$= \lambda^2 \int_{-\infty}^{+\infty} \exp(-\lambda\left(2\hat{\tau} - t\right)) \Theta(\hat{\tau})\ \Theta(\hat{\tau} - t)\ \mathrm{d}\hat{\tau} \tag{C.6}$$

Evaluating the Heaviside terms gives $\hat{\tau} \geq 0$ and $\hat{\tau} \geq t$, which results in a new lower limit of the integral.

$$f_{\text{slow,slow}}(t) = \lambda^2 \int_{\min(0,t)}^{+\infty} \exp(-\lambda\left(2\hat{\tau} - t\right)) \mathrm{d}\hat{\tau} \tag{C.7}$$

$$= \lambda^2 \frac{1}{-2\lambda} \left[\exp(-\lambda\left(2\hat{\tau} - t\right))\right]_{\hat{\tau}=\min(0,t)}^{\hat{\tau}\to+\infty} \tag{C.8}$$

$$= -\frac{\lambda}{2} \left(0 - \exp(-\lambda\left(2\min(0,t) - t\right))\right) \tag{C.9}$$

$$t > 0: \tag{C.10}$$

$$= \frac{\lambda}{2} \exp(-\lambda\left(2t - t\right)) = \frac{\lambda}{2} \exp(-\lambda t) \tag{C.11}$$

$$t \leq 0: \tag{C.12}$$

$$= \frac{\lambda}{2} \exp(-\lambda\left(0 - t\right)) = \frac{\lambda}{2} \exp(-\lambda\left|t\right|) \tag{C.13}$$

Thus showing that $f_{\text{slow,slow}}$ is independent of t_0 and symmetric in t.

Bibliography

[1] ECFA Detector R&D Roadmap Process Group. *The 2021 ECFA detector research and development roadmap*. Tech. rep. Geneva, 2020. DOI: 10.17181/CERN.XDPL.W2EX.

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

[3] M Aicheler et al. *A Multi-TeV Linear Collider Based on CLIC Technology: CLIC Conceptual Design Report*. Tech. rep. Geneva, 2012. DOI: 10.5170/CERN-2012-007.

[4] A. Abada et al. „FCC-ee: The Lepton Collider." In: *The European Physical Journal Special Topics* 228.2 (2019), pp. 261–623. ISSN: 1951-6401. DOI: 10.1140/epjst/e2019-900045-4.

[5] Inge Diehl et al. „Monolithic MHz-frame rate digital SiPM-IC with sub-100 ps precision and 70 um pixel pitch." 2023. arXiv: 2311.13220 [physics.ins-det].

[6] F. Carnesecchi et al. „Direct detection of charged particles with SiPMs." In: *Journal of Instrumentation* 17.06 (2022-06), P06007. DOI: 10.1088/1748-0221/17/06/P06007.

[7] H. Bethe and J. Ashkin. In: *Experimental nuclear physics*. Ed. by Emilio Segrè. 1953.

[8] M. Tanabashi et al. „Review of Particle Physics." In: *Phys. Rev. D* 98.3 (2018), p. 030001. ISSN: 1550-7998. DOI: 10.1103/PhysRevD.98.030001.

[9] Stephen M. Seltzer and Martin J. Berger. „Improved procedure for calculating the collision stopping power of elements and compounds for electrons and positrons." In: *The International Journal of Applied Radiation and Isotopes* 35.7 (1984), pp. 665–676. ISSN: 0020-708X. DOI: 10.1016/0020-708X(84)90113-3.

[10] Virgil L. Highland. „Some practical remarks on multiple scattering." In: *Nuclear Instruments and Methods* 129.2 (1975), pp. 497–499. ISSN: 0029-554X. DOI: 10.1016/0029-554X(75)90743-0.

[11] Martin A. Green and Mark J. Keevers. „Optical properties of intrinsic silicon at 300 K." In: *Progress in Photovoltaics: Research and Applications* 3.3 (1995), pp. 189–192. DOI: 10.1002/pip.4670030303.

[12] Wikimedia Commons. *Schema - n-dotiertes Silicium*. 2006. URL: https://commons.wikimedia.org/w/index.php?title=File:Schema_-_n-dotiertes_Silicium.svg&oldid=657675145 (visited on 2023-08-31).

[13] Wikimedia Commons. *pn junction equilibrium graphs*. 2007. URL: https://commons.wikimedia.org/w/index.php?title=File:Pn-junction-equilibrium-graphs.png&oldid=451423606 (visited on 2021-07-13).

[14] Jens Kröger. „Characterisation of a High-Voltage Monolithic Active Pixel Sensor Prototype for Future Collider Detectors." Dissertation. Heidelberg University, 2021. DOI: 10.11588/heidok.00030657.

[15] Gianpiero Vignola. Private communication.

[16] M.V. Nemallapudi et al. „Single photon time resolution of state of the art SiPMs." In: *Journal of Instrumentation* 11.10 (2016-10), P10016. DOI: 10.1088/1748-0221/11/10/P10016.

[17] F. Zappa et al. „Principles and features of single-photon avalanche diode arrays." In: *Sensors and Actuators A: Physical* 140.1 (2007), pp. 103–112. ISSN: 0924-4247. DOI: 10.1016/j.sna.2007.06.021.

[18] Feiyang Sun et al. „A Simple Analytic Modeling Method for SPAD Timing Jitter Prediction." In: *IEEE Journal of the Electron Devices Society* 7 (2019), pp. 261–267. ISSN: 2168-6734. DOI: 10.1109/JEDS.2019.2895151.

[19] A. Gulinatti et al. „A physically based model for evaluating the photon detection efficiency and the temporal response of SPAD detectors.“ In: *Journal of Modern Optics* 58.3-4 (2011), pp. 210–224. DOI: 10.1080/09500340.2010.536590.

[20] Tomas Vanat. „Caribou – A versatile data acquisition system.“ In: *PoS* TWEPP2019 (2020-04), p. 100. DOI: 10.22323/1.370.0100.

[21] Simon Spannagel, Adrian Fiergolski, and Tomas Vanat. *Peary - A DAQ framework for the Caribou DAQ System*. URL: https://cern.ch/peary-caribou.

[22] R. Diener et al. „The DESY II test beam facility.“ In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 922 (2019), pp. 265–286. ISSN: 0168-9002. DOI: 10.1016/j.nima.2018.11.133.

[23] Finn Feindt et al. „Test Beam Characterization of a Digital Silicon Photomultiplier.“ Unpublished.

[24] Hendrik Jansen et al. „Performance of the EUDET-type beam telescopes.“ In: *EPJ Techniques and Instrumentation* 3.1 (2016-10), p. 7. ISSN: 2195-7045. DOI: 10.1140/epjti/s40485-016-0033-2.

[25] Simon Spannagel and Hendrik Jansen. *GBL Track Resolution Calculator v2.0*. 2016. DOI: 10.5281/zenodo.48795. URL: https://github.com/simonspa/resolution-simulator.

[26] Heiko Augustin et al. „TelePix – A fast region of interest trigger and timing layer for the EUDET Telescopes.“ In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 1048 (2023), p. 167947. ISSN: 0168-9002. DOI: 10.1016/j.nima.2022.167947. arXiv: 2212.10248 [physics.ins-det].

[27] P. Baesso, D. Cussans, and J. Goldstein. „The AIDA-2020 TLU: a flexible trigger logic unit for test beam facilities.“ In: *Journal of Instrumentation* 14.09 (2019-09), P09019. DOI: 10.1088/1748-0221/14/09/P09019.

[28] Y. Liu et al. „EUDAQ2 - A flexible data acquisition software framework for common test beams.“ In: *Journal of Instrumentation* 14.10 (2019-10), P10033. DOI: 10.1088/1748-0221/14/10/P10033.

[29] D. Dannheim et al. „Corryvreckan: a modular 4D track reconstruction and analysis software for test beam data.“ In: *Journal of Instrumentation* 16.03 (2021-03), P03008. DOI: 10.1088/1748-0221/16/03/P03008.

[30] Morag Williams et al. *Corryvreckan - A Modular 4D Track Reconstruction and Analysis Software for Test Beam Data*. DOI: 10.5281/zenodo.4384170. URL: https://cern.ch/corryvreckan.

[31] Claus Kleinwort. „General broken lines as advanced track fitting method.“ In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 673 (2012), pp. 107–110. ISSN: 0168-9002. DOI: 10.1016/j.nima.2012.01.024. arXiv: 1201.4320 [physics.ins-det].

[32] Rene Brun and Fons Rademakers. „ROOT — An object oriented data analysis framework.“ In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 389.1 (1997). New Computing Techniques in Physics Research V, pp. 81–86. ISSN: 0168-9002. DOI: 10.1016/S0168-9002(97)00048-X.

[33] Fons Rademakers et al. *ROOT: An open-source data analysis framework used by high energy physics and others*. DOI: 10.5281/zenodo.848818. URL: https://root.cern/.

[34] Wouter Verkerke and David Kirkby. *The RooFit toolkit for data modeling*. 2003. arXiv: physics/0306116 [physics.data-an].

[35] *Exponentially modified Gaussian distribution*. URL: https://en.wikipedia.org/w/index.php?title=Exponentially_modified_Gaussian_distribution&oldid=1158261356 (visited on 2023-07-27).

Acknowledgments

I would like to thank Prof. Dr. Erika Garutti and Dr. Simon Spannagel for giving me the great opportunity to write my thesis about such an amazing device.

The measurements leading to these results have been performed at the Test Beam Facility at DESY Hamburg (Germany), a member of the Helmholtz Association (HGF).

To Gianpiero: thank you so much for helping me when I was confused, frustrated or on the wrong path. Your input was invaluable for this thesis.

I acknowledge Gianpiero Vignola for running a significant part of the alignment, providing me the pixel delay data of the laser measurement and helping to analyze the chiller problem. I acknowledge the help of everyone leading to the successful testbeam, namely Daniil Rastorguev, Finn Feindt, Frauke Poblotzki, Gianpiero Vignola, Inge Diehl, Simon Spannagel and Tomas Vanat.

I want to thank Finn Feindt, Gianpiero Vignola, Inge Diehl, Simon Spannagel and Tomas Vanat and in particular for helping me understand various aspect of the chip, the TDC and the DAQ system. I would also like to thank Adrian Heckert for helping us with the chiller problem.

Thank you Finn, Gianpiero and Inge for reviewing my thesis.

My thanks also go to all other members of the Tangerine and dSiPM groups at DESY for hosting me and made me having a great time.