Sampling-ADC based Pile-Up Recovery and Digital Signal Processing in the CBELSA/TAPS-Experiment

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

vorgelegt von

Jan Alexander Macht
geb. Schultes
aus
Solingen

Bonn, 14.05.2024

Angefertigt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der
Rheinischen Friedrich-Wilhelms-Universität Bonn

Gutachterin/Betreuerin:	Prof. Dr. Ulrike Thoma
Gutachter:	Prof. Dr. Sebastian Neubert
Tag der Promotion:	05.06.2025
Erscheinungsjahr:	2026

Abstract

The CBELSA/TAPS experiment in Bonn investigates the spectrum and properties of baryon resonances by performing photoproduction experiments off the nucleon to reach a better understanding of quantum chromodynamics in the non-perturbative regime. The main calorimeter of the experiment, the Crystal Barrel, consists of 1320 scintillating CsI(Tl) crystals which are read out by avalanche photodiodes and an FPGA-controlled sampling analog-to-digital converter (SADC) readout system. Using a custom firmware enables access to the sampled waveforms and the extraction of different features from this data. Its main advantages over the previous readout lie in increased rate capabilities, an event-based baseline subtraction and the possibility to detect and correct cases of pile-up.

Significant parts of this work were spent on testing and enhancing feature extraction algorithms, with improvements based upon findings from this thesis. The baseline-finding algorithm was rewritten and improved, further enhancing the energy resolution. Timing capabilities of the constant fraction discriminator were enhanced, surpassing the capabilities of the dedicated current time readout for $E < 8\,\mathrm{MeV}$. In addition, a digital filter was proposed and implemented, which more efficiently suppresses high frequency noise above 1 MHz in the sampled data, improving the performance of the subsequent feature extraction.

A central achievement was the development and refinement of pile-up detection and recovery algorithms, addressing previous algorithmic limitations which led to a considerable amount of unrecognized pile-up in the data. A new detection approach was proposed and implemented, which shows significant improvements and does not exhibit systematic weaknesses anymore. Specific attention was paid to the interplay between the pile-up detection and the several feature extraction algorithms. Some specific causes of pile-up, such as $\mu^{\pm}$-decay at rest, were identified and discussed. Furthermore, the pulse shapes recorded by the new readout setup were carefully examined. Since CsI(Tl) is used, they depend on the type of detected particle. The ensuing complications and opportunities with regards to a potential particle identification, specifically a potential discrimination between p and γ pulses, were investigated and implications for the feature extraction and pile-up detection and recovery process were discussed.

A custom deconvolution method was developed and employed to disentangle the pile-up signals, which was found to perform far better for the waveforms at hand than a standard moving window deconvolution. It is followed by a custom fitting and evaluation routine, which takes care of more complex pile-up topologies. The end result is a robust method which achieves a satisfactory performance for over 99% of cases. Finally, the positive effects on the reconstruction of physical particles from corrected data was discussed as well.

Contents

CHAPTER 1

Introduction

1.1 The Standard Model of Particle Physics

The Standard Model of Particle Physics describes all of the fundamental particles of nature and their interactions[1]. Within this model, all matter is composed of fermions with spin $\frac{1}{2}$ and bosons with spin 1 or 0, which mediate the forces between these particles. With the discovery of the Higgs boson in 2012 [ATL12] [CMS12], which was predicted by the model, its validity was confirmed beyond any reasonable doubt.
The fermions belong to one of two categories: leptons and quarks. While leptons carry integer electric charge (0 or -1) but no color charge, quarks carry fractional electric charge (see Fig. 1.1 and have color charge.
The interactions between these particles are mediated by bosons: The $W^{\pm}$ and Z^0 bosons are the exchange particles of the weak force, to which all particles of the standard model, except the other bosons, are subject to. The electromagnetic interaction is mediated by the photon (γ) and acts on all particles that carry an electric charge (quarks, charged leptons and $W^{\pm}$ bosons). It is described by the field theory of Quantum Electro Dynamics (QED), which is among the most successful and well-tested theories in all of physics. The electromagnetic and the weak force were elegantly explained together by Glashow, Weinberg and Salam in the "Electroweak unification" [Gla61] [Wei67] [Sal68]. The third interaction, the strong force, is discussed in more detail in the following section.

1.1.1 QCD

The strong interaction acts on all particles that carry color charge (quarks and gluons). Similar to the electromagnetic force, the corresponding field theory is known as Quantum Chromo Dynamics (QCD). While the theory bears some similarities to the electromagnetic interaction, there are a few unique properties, which set QCD and QED apart and make the theoretical description more difficult. Table 1.1 summarizes these properties.
One fundamental difference is that gluons, unlike photons, carry color charge and are thus subject to the very same interaction they mediate. This leads to the fact that gluons also interact with each other, which in turn leads to another major difference: the effect of vacuum polarization.

[1] With the exception of gravity.

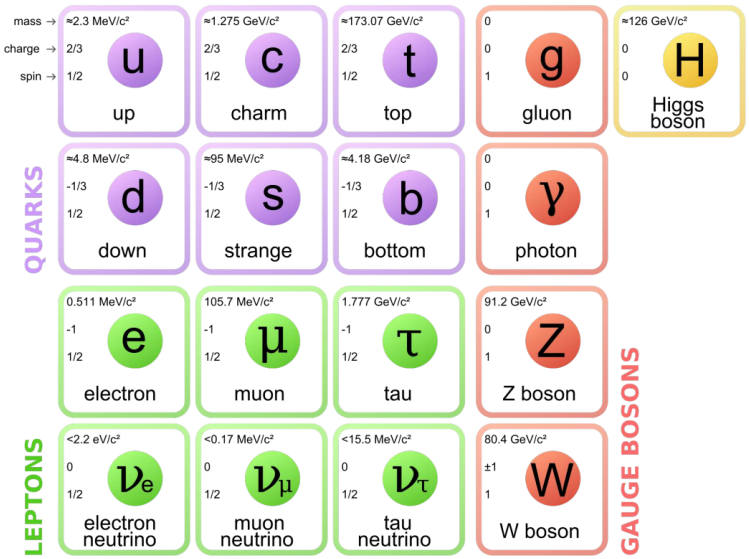


Figure 1.1: Elementary particles of the Standard Model of Particle Physics. Illustration from [Wik24].

In the electromagnetic case, the possibility to form fermion-antifermion pairs from the vacuum leads to an effective screening of a test charge at larger distances. As distance is inversely related to the 4-momentum transfer Q^2 in a given reaction, this explains the observed (slight) increase of the coupling constant $\alpha_{\text{em}}(Q^2)$ seen in Fig. 1.2.
In case of the strong interaction, the same would be expected due to the creation of quark-antiquark-pairs from the vacuum. Virtual gluon-loops, however, have the exact opposite effect[2]; they effectively lead to "anti-screening", i.e. an increase of the coupling constant with decreasing Q^2 or increasing distance (see Fig. 1.3). This somewhat unintuitive behavior implies, that objects carrying a net color charge (like single quarks or gluons) can never be observed experimentally, the so-called "confinement". If one were to attempt to separate e.g. a quark (q) and an antiquark ($\bar{\text{q}}$) of opposite color charge, the force to pull them apart would not decrease, but rather stay constant. By increasing the distance between the quarks, more and more energy is stored in the system, until it is energetically favorable to create another $\text{q}\bar{\text{q}}$ pair from the vacuum, such that two separate $\text{q}\bar{\text{q}}$ pairs would be formed. As a result, only color-neutral bound states of QCD, so-called hadrons[3], are observable. For the same reason, the range of the strong force is effectively limited (because gluons have color charge), even though gluons are in principle massless like the photon. On the opposite end, the decrease of α_{s} with Q^2 results in what is known as "asymptotic freedom": at very high values of Q^2, strongly interacting particles are only weakly bound, allowing perturbative calculations.

[2] The discussion of why exactly this is the case goes beyond the scope of this thesis.
[3] There are two types of hadrons, depending on the baryon number B given by the valence quarks. Hadrons with $B = 1$ are known as baryons, whereas $B = 0$ hadrons are called mesons.

	QED	**QCD**
Charge	Electric charge $(+/-)$	Color charge $(r/g/b$ & $\bar{r}/\bar{g}/\bar{b})$
Symmetry group	U(1)	SU(3)
Mass of boson	massless	massless
Self-interaction	no (γ is uncharged)	yes (g carries color charge)
Range	∞ (drops with $\frac{1}{r^2}$)	limited (confinement)
Coupling constant	$\lim_{Q^2\to 0}(\alpha_{\text{em}}) \approx \frac{1}{137}$ increases with Q^2 (fig. 1.2)	$\alpha_s(M_Z^2) \approx 0.118$ decreases with Q^2 (fig. 1.3)

Table 1.1: Summary of similarities and differences between electromagnetic and strong interaction.

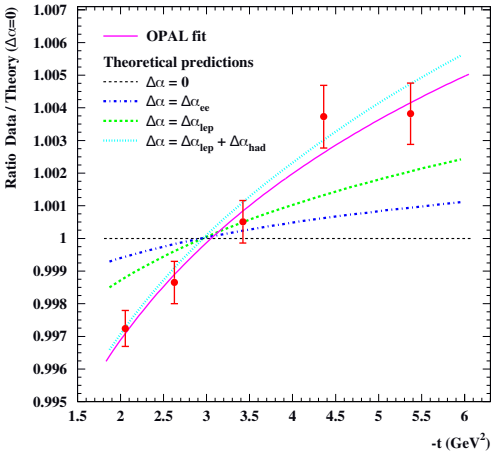


Figure 1.2: Running coupling of α_{em}. Data in red, together with various predictions if only virtual $e^{\pm}$ pairs, only lepton pairs, or lepton and hadron pairs are considered. $-t = Q^2$ on the x-axis. [OPA06]

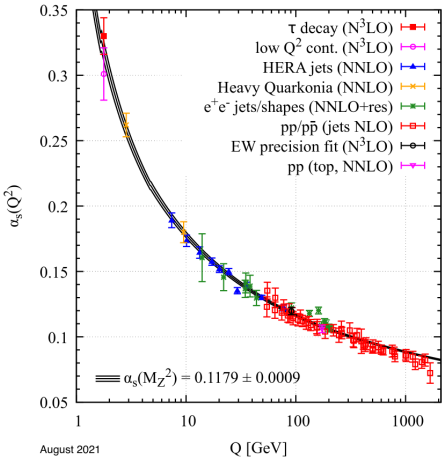


Figure 1.3: Running coupling of α_s, which was calculated from various measurements with different degrees of precision ("LO" = leading order, "NX" = next-to-X). [PDG22]

Bound States The behavior of the coupling constant has huge implications for the calculation of bound states in QCD. Because α_s is close to unity for low Q^2, a perturbative expansion in powers of α_s is not possible, as it would not converge. This makes it difficult to predict the masses and other properties of hadrons.

Instead, one has to resort to other methods like so-called constituent quark models or Lattice QCD (LQCD), to only name two. The idea of constituent quark models is to avoid QCD-calculations entirely and to follow a more phenomenological approach, whereas the approach of LQCD is to solve the equations numerically on a discrete space-time lattice. Of particular interest is the study

of the most common hadrons of which most of our everyday matter is composed: the nucleons[4]. Since the nucleon is assumed to consist of 3 valence quarks and a "sea" of quark-antiquark pairs and gluons, it has an associated excitation spectrum, which theoretical models should be able to predict. Depending on their respective isospin, these excited states are referred to as either N^* ($I = \frac{1}{2}$) or Δ^* ($I = \frac{3}{2}$). Fig. 1.4 shows the respective predictions for the excitation spectrum of the nucleon, together with the experimentally observed states that were known at the time[5]. While constituent quark models are arguably still in better agreement with the data, the steady increase in computing power as well as the application of Graphics Processing Units (GPUs) have lead to significant improvements in the field of LQCD over the last years. Nonetheless, it is not yet possible to carry out LQCD-calculations of excitation spectra with the physical quark masses. The discrepancy is usually characterized by the pion mass, and how far it diverges from the physical mass of 140 MeV. In addition, the states are not yet able to decay.
It is immediately obvious, that the agreement between data and theory predictions, be it from quark models or LQCD, is not great. While the number of states and their approximate position seems reasonably well reproduced for the lowest lying states of each J^P, the masses of several states can not be explained within their respective uncertainties. Even more strikingly, the number of predicted states greatly exceed the number of observed states, especially for higher excitation energies, a conundrum which has been dubbed the "missing resonance problem". The question which immediately arises is whether the predictions are wrong or if those missing resonances actually exist but just have not been found yet.

1.2 Baryon Spectroscopy

One way to resolve the discrepancies between theory and experiment is the search for further excitations of the nucleon. In practice however, this is not as straightforward as one might think. An often used analogy is to compare the excitations of the nucleon with the excitations of the hydrogen atom. But in contrast to atomic physics, where the identification of excited states comes down to identifying sharp lines of a particular wavelength, the resonances in hadron physics are usually extremely short-lived and therefore wide and overlapping[6]. Fig. 1.5(a) shows the cross sections for various final states when a proton is excited by a high-energy photon: $\gamma N \left(\rightarrow R^*\right) \rightarrow NX$, where N is a nucleon, R^* is a (potential) resonance and X stands for one or more mesons in the final state. Such processes are generally referred to as *Photoproduction*. It is obvious that apart from a few clearly distinguishable peaks (the first and biggest one being the well known $\Delta(1232)$ resonance), the cross section $\gamma p \rightarrow X$ seems rather featureless.
To make matters worse, the resonances are not only broad and overlapping, but might also (given identical quantum numbers) interfere with each other. The situation improves, when only specific

[4] While proton and neutron are two different particles, they are, in terms of the strong interaction, just different isospin states of one particle: the nucleon.
[5] It is worth noting that a number of additional states have been discovered since, in large parts due to new data from photoproduction (see section 1.2 (Baryon Spectroscopy))
[6] Typical lifetimes are in the order of 10^{-24} s, which (using $\Gamma\tau = \hbar$) corresponds to typical widths around $\approx 10^2$ MeV

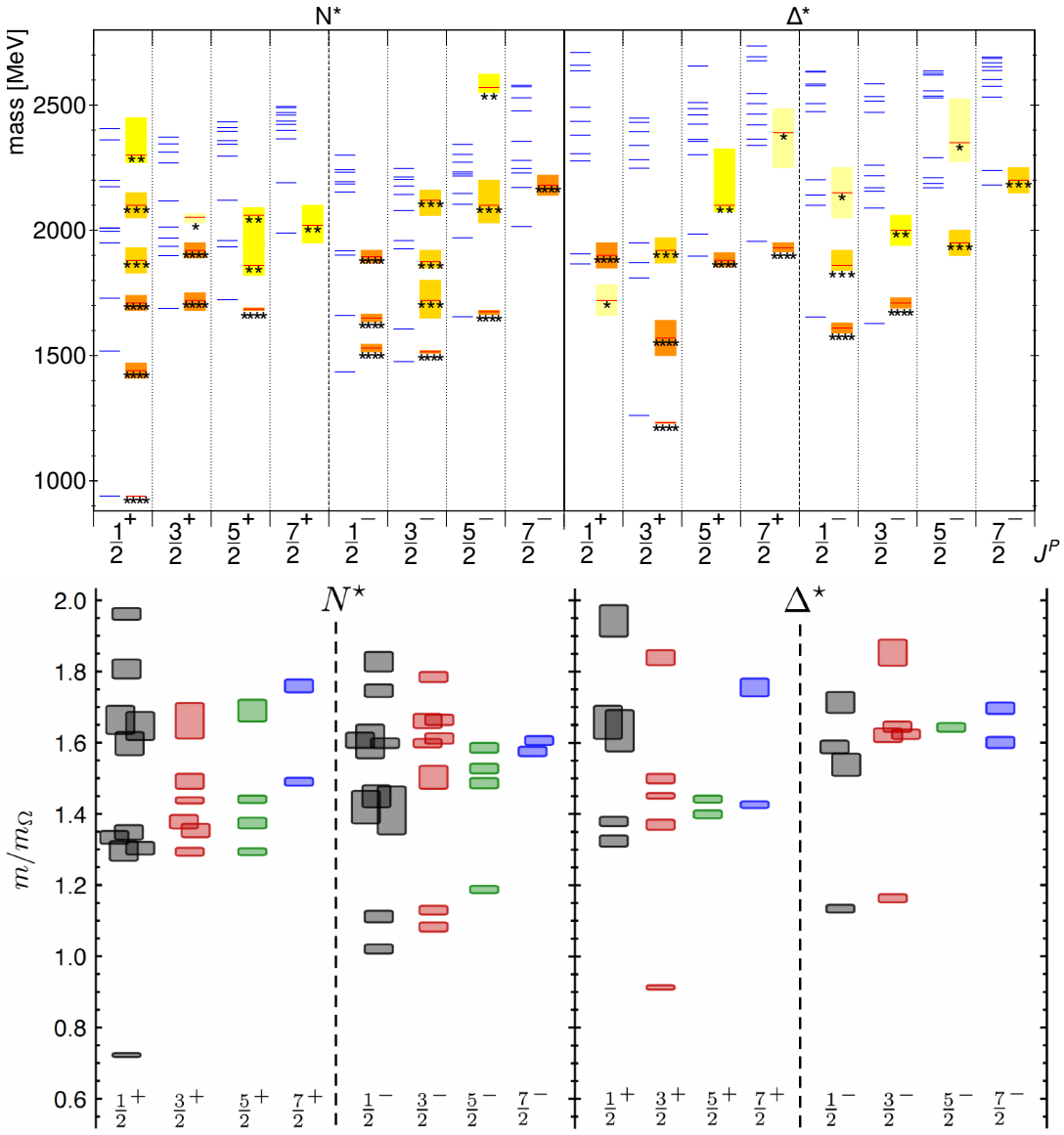
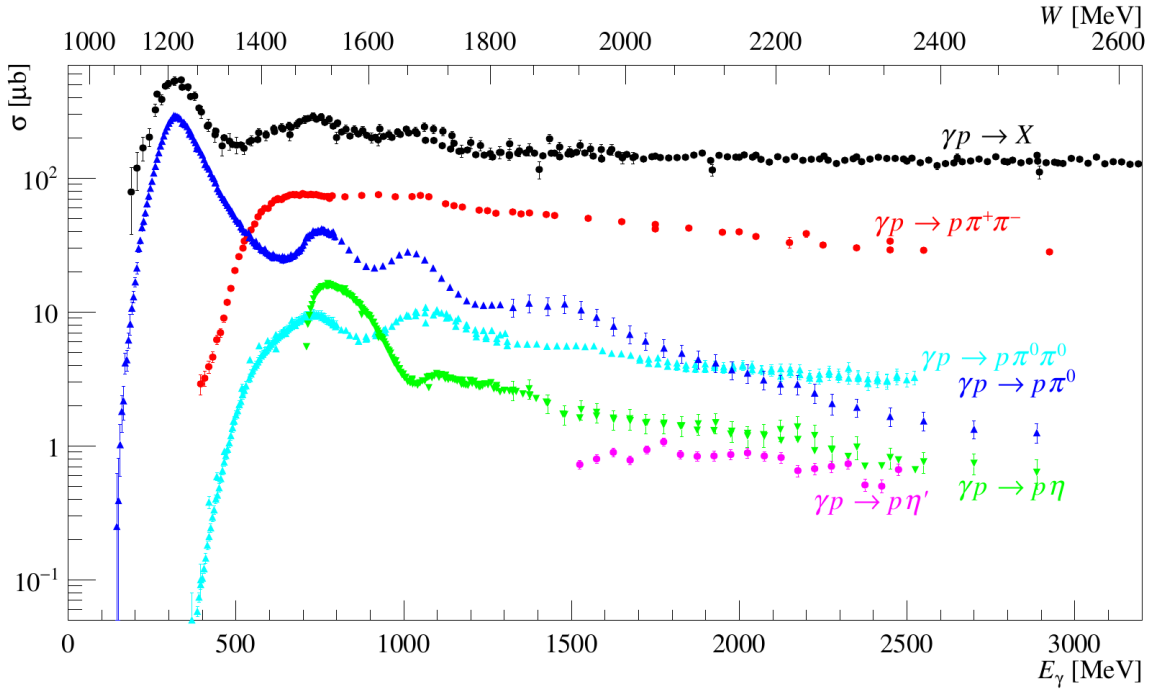


Figure 1.4: Predictions for the excited states of the nucleon. N^* states with $I = \frac{1}{2}$ are shown on the left, $I = \frac{3}{2}$ Δ^* states on the right.
Top: Quarkmodel predictions (recompiled from [RM11]). For each J^P, the experimentally observed states (as listed by the Particle Data Group (PDG) in 2024) are shown on the right (red lines), together with model predictions to the left (blue).
Bottom: LQCD predictions at the unphysical pion mass of $m_\pi = 396\,\text{MeV}$ from [Edw+11]. Masses are expressed in terms of the Ω mass, which has a physical value of 1 672 MeV.

final states are considered[7], but it would still be impossible to identify the contributing states from the examination of the cross section alone. In order to unambigously disentangle the contributing resonances, more observables are needed in addition to the (total) cross section. One possibility is to use a polarized beam, a polarized target, or both, and to study the angular distribution of various decay products. From this, a number of so-called *(Double-) Polarization Observables* can be extracted, which encapsulate additional information about the scattering process under investigation. Contributing resonances are then identified in a *Partial Wave Analysis (PWA)*, for which multiple datasets on different final states are brought together. There exist several different models by different groups (e.g. Bonn-Gatchina [AKN+15], MAID [Tia+18], SAID [Wor+12], Jülich-Bonn-Washington [Rön+22], etc.) which aim to predict the measured observables. By incorporating both confirmed and proposed resonances with their respective positions, widths, and coupling strengths as parameters and subsequently fitting these models to the data, properties of (potentially new) resonances are inferred.
In the past, most of the worldwide data on nucleon excitations came pion-nucleon scattering, i.e. processes of the form $\pi^{\pm}N \rightarrow R^{*} \rightarrow NX$. Since this method of excitation is mediated by the strong force, the corresponding cross sections are about three orders of magnitude larger than for photoproduction[8]. As a drawback, only resonances with a significant coupling to πN can be produced this way, which then (preferably) decay back into πN as well. It was proposed, that most of the missing resonances might have a small or non-existent coupling to πN.
In fact, photoproduction experiments like CBELSA/TAPS have led to the discovery of most of the newly established resonances in recent years. In addition to more modern experimental hardware and improved analysis methods, this mostly results from the fact that photoproduction is able to complement the existing data with resonances with a particularly weak coupling to πN (provided their coupling to γN is not equally small). The table in Fig. 1.5(b) summarizes the recent progress that has been made.

[7] As a very simple example, the final state $p\eta$ is a pure $I = \frac{1}{2}$ state (the η meson has isospin 0 and acts as an "isospin filter"), so it can never come from the decay of an $I = \frac{3}{2}$ Δ-state, but must come from the decay of an N^{*} resonance with $I = \frac{1}{2}$

[8] At the $\Delta(1232)$ resonance, $\sigma_{\text{total}}(\pi^{+}p) \approx 200\,\text{mb}$, compared to $\approx 0.54\,\text{mb}$ for γp.



(a)

Particle	J^P	Rating	Particle	J^P	Rating
p	$1/2^+$	****	Δ (1232)	$3/2^+$	****
n	$1/2^+$	****	Δ (1600)	$3/2^+$	****
N (1440)	$1/2^+$	****	Δ (1620)	$1/2^-$	****
N (1520)	$3/2^-$	****	Δ (1700)	$3/2^-$	****
N (1535)	$1/2^-$	****	Δ (1750)	$1/2^+$	*
N (1650)	$1/2^-$	****	Δ (1900)	$1/2^-$	***
N (1675)	$5/2^-$	****	Δ (1905)	$5/2^+$	****
N (1680)	$5/2^+$	****	Δ (1910)	$1/2^+$	****
N (1700)	$3/2^-$	***	Δ (1920)	$3/2^+$	***
N (1710)	$1/2^+$	****	Δ (1930)	$5/2^-$	***
N (1720)	$3/2^+$	****	Δ (1940)	$3/2^-$	***
N (1860)	$5/2^+$	**	Δ (1950)	$7/2^+$	****
N (1875)	$3/2^-$	***	Δ (2000)	$5/2^+$	**
N (1880)	$1/2^+$	***	Δ (2150)	$1/2^-$	*
N (1895)	$1/2^-$	****	Δ (2200)	$7/2^-$	***
N (1900)	$3/2^+$	****	Δ (2300)	$9/2^+$	**
N (1990)	$7/2^+$	**	Δ (2350)	$5/2^-$	*
N (2000)	$5/2^+$	**	Δ (2390)	$7/2^+$	*
N (2040)	$3/2^+$	*	Δ (2400)	$9/2^-$	**
N (2060)	$5/2^-$	***	Δ (2420)	$11/2^+$	****
N (2100)	$1/2^+$	***	Δ (2750)	$13/2^-$	**
N (2120)	$3/2^-$	***	Δ (2950)	$15/2^+$	**
N (2190)	$7/2^-$	****			
N (2220)	$9/2^+$	****			
N (2250)	$9/2^-$	****			
N (2300)	$1/2^+$	**			
N (2570)	$5/2^-$	**			
N (2600)	$11/2^-$	***			
N (2700)	$13/2^+$	**			

(b)

Figure 1.5: Left: Photoproduction cross sections for different final states. Data is compiled from several different measurements.
Right: N^* and Δ^* resonances as listed by the PDG in 2020. Listing as of 2002 in black and changes since then in red. Rating ranges from $****$ (existence is certain) to $*$ (evidence of existence is poor). Not all states shown on the right contribute equally to the cross sections on the left. Plot and table taken from [TAW22]

1.3 Interactions of Particles with Matter

Any attempt to detect a particle ultimately relies on its interactions with matter from which the detector is built of. While there are particle detectors that focus on the detection of only weakly interacting particles (neutrinos, WIMPs[9] etc.), most particle detectors are built to detect electromagnetically interacting particles. As such, one has to distinguish mainly between (heavier) charged particles on the one hand and photons and electrons/positrons on the other hand. The detection of neutrons and other neutral hadrons is an entirely different topic, as it relies on hadronic interactions, and will not be discussed here. The following discussions largely follow the review from the PDG [PDG22].

1.3.1 Interactions of Heavy Charged Particles

When traversing a medium, heavy[10] charged particles of intermediate energy lose most of their energy by scattering off the atomic electrons[11] (ionization or excitation). The mean energy loss per pathlength $\frac{dE}{dx}$ is well described by the Bethe equation [Bet30]. Some corrections have to be made for for very small energies (not relevant for this thesis), as the particle energy approaches the binding energy of electrons in the atomic shell. For very high energies, radiative effects like bremsstrahlung dominate.

[9] Weakly Interacting Massive Particles (WIMPs): hypothetical candidate particles for dark matter (unconfirmed).
[10] heavier than an electron
[11] (elastic) scattering of atomic nuclei occurs as well, but the respective energy loss is insignificant in comparison due to the large mass of the recoiling particle.

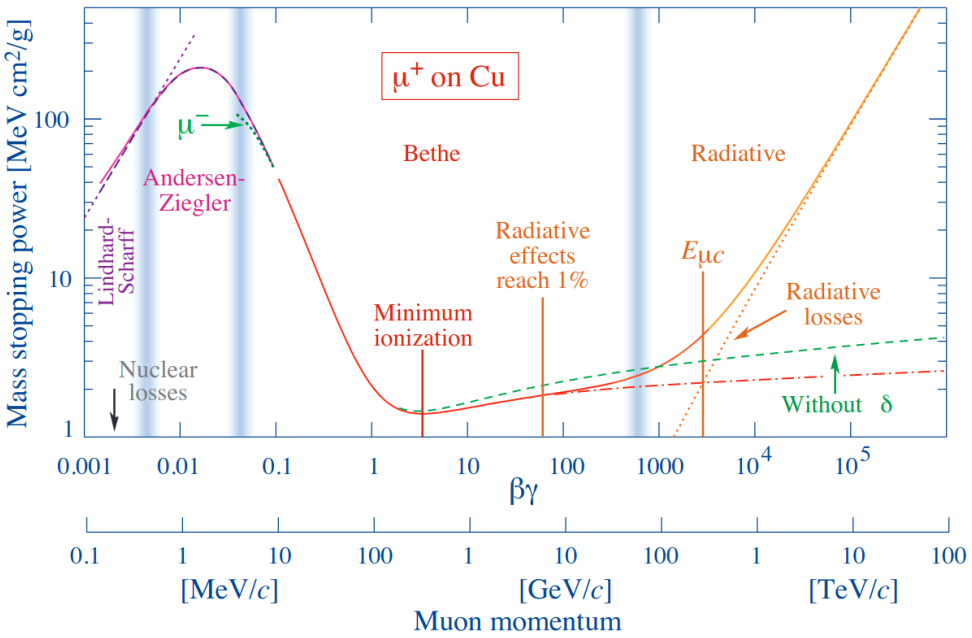


Figure 1.6: Mean energy loss for muons in copper. [PDG22]

Ionization Fig. 1.6 shows $\frac{dE}{dx}$ for muons over a wide range of momenta. Muons were chosen as model particles because they only interact electromagnetically[12], but the same arguments would hold for e.g. charged pions or protons if hadronic interactions are neglected. The behavior of electrons on the other hand is insufficiently modeled by Fig. 1.6. While they lose energy by the same processes in general, the cross sections have to be modified because of their light mass and the fact that they are quantum-mechanically indistinguishable from the electrons they collide with.
A notable feature of this plot is the region around $\beta\gamma = 3$, where the energy loss reaches a broad minimum. Particles in this energy range are referred to as Minimum Ionizing Particles (MIPs). Relaxing this definition a bit, this name is also often used for particles with larger momenta, since the energy loss only increases logarithmically with $\beta\gamma$.
A further characteristic of the ionization process is the steep rise for lower momenta ($0.1 < \beta\gamma < 1$). This entails that once a particles momentum drops below $\beta\gamma \approx 1$, it will lose more and more energy per unit length as its momentum decreases. Accordingly, a large portion of a particles energy will be deposited in a small volume concentrated at the end of the track, the so-called "Bragg peak".
While the concept of a "mean energy loss" is very useful, it is worth noting that it does not encapsulate the full nature of the interactions. Most notably, the mean is in general higher than the most probable energy loss, since rare events with high energy transfer (e.g. δ-electrons[13]) can substantially skew the mean. This effect is most pronounced for thin absorbers, due to the limited number of interactions in total. For thick absorbers, the most probable energy loss approaches the mean due to the central limit theorem[14].

Bremsstrahlung Bremsstrahlung is one of the fundamental interactions by which highly relativistic charged particles lose energy in a medium. Loosely speaking, deceleration in a Coulomb potential (i.e. of an atomic nucleus in the material), which also absorbs the recoil from the electron[15] leads to the emission of a photon. The quantum mechanical process was first described by Bethe and Heitler [BH34]. The cross section scales with the square of the nuclear charge Z^2 and is further (approximately) proportional to $\frac{1}{m^2}$. At moderate energies, bremsstrahlung is therefore almost exclusively limited to electrons and positrons, since the next lightest particle, the muon, is already around 200 times as heavy.

Cherenkov Radiation Another effect that charged particles exhibit when traversing a medium is Cherenkov radiation. It is emitted when the speed of the particle exceeds the speed of light in the medium $c_{\text{medium}} = \frac{c}{n}$, with n being the refractive index.
Energy losses by Cherenkov radiation are generally insignificant compared to the other processes; the resulting emission of light can, however, be used for detection purposes.

[12] The weak interaction is completely irrelevant at these energies

[13] Secondary electrons, knocked from the atomic shells of the material, which possess enough energy to ionize further atoms on their own.

[14] The distribution of energy losses for each individual particle approaches a gaussian, where the mean equals the Most Probable Value (MPV)

[15] The fundamental process $e^- \rightarrow e^- + \gamma$ would violate energy-momentum-conservation without the presence of another particle to absorb some of the momentum

Nuclear Interactions Strongly interacting particles, in particular protons and neutrons, might be subject to interactions with the nuclei of the target material. Potential reactions include the ejection of one or more protons, neutrons, or heavier charged particles from the struck nucleus. This might alter the nature of energy deposition, or prevent part of the total energy from being detected entirely, especially if uncharged neutrons were produced. Also see section 6.5.3.4 for further discussions of such effects.

1.3.2 Interactions of Photons

Once a photon enters matter, there are several different effects which can occur. Fig. 1.7 shows their respective cross sections in dependence of the photon's energy for $E \geq 10\,\mathrm{eV}$. In order of relevant energy range, these are the following processes:

- Photoelectric effect: The photon is absorbed by an electron bound in the atom, which gets ejected from the coulomb potential of the nucleus.
- Rayleigh scattering: The photon scatters from the entire atom (or molecule).
- Compton scattering: The photon scatters from an electron to which it imparts part of its energy. The photon changes its wavelength to accommodate for the lost energy.
- Pair production: Only possible if $E_\gamma > 2m_\mathrm{e}$. The photon creates a lepton-antilepton[16] pair from the vacuum. The fundamental vertex of this process exhibits crossing symmetry to bremsstrahlung. Again, 4-momentum conservation requires that this process also can only occur in the vicinity of another particle to absorb the excess recoil. This is (in most cases) the atomic nucleus. Like bremsstrahlung, the cross section also scales with Z^2 of the absorber material.
- Photonuclear interactions: The photon is absorbed by the nucleus, which in turn emits other particles or disintegrates. The most prominent form of photonuclear excitation is known as the Giant Dipole Resonance (GDR).
- Hadronic interactions: The photon is absorbed by a single nucleon and excites it. The excited baryon then decays to other hadrons, which in turn might be unstable (also see "photoproduction" in section 1.2). At typical accelerator-energies, these interactions are not a relevant process for energy loss, since the cross section is much lower than for e.g. pair production[17].

1.3.3 Electromagnetic Shower

Above a certain energy, pair production is the dominant interaction of photons in matter. Likewise, electrons and positrons of sufficient energy lose most of their energy via bremsstrahlung. Since pair production produces $\mathrm{e}^+/\mathrm{e}^-$ pairs from an initial γ, and bremsstrahlung produces a γ in

[16] Most likely electrons, but also muons or taus are possible if enough energy is available.

[17] At the well-known $\Delta(1232)$-resonance, the hadronic γp cross section reaches its maximum at $\approx 0.54\,\mathrm{mb}$ (see Fig. 1.5(a)). Even for the lightest element, hydrogen, the pair production cross section at the same energy is already much bigger ($\approx 19\,\mathrm{mb}$ [Ber+10]). For other materials, the difference will be even larger, due to the Z^2-dependence.

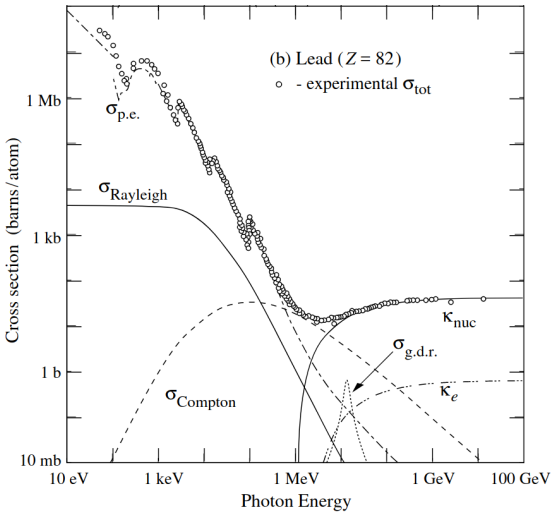


Figure 1.7: Cross section of photons in lead for different processes [PDG22]. p.e.=photoelectric effect; κ=pair production; g.d.r.=giant dipole resonance

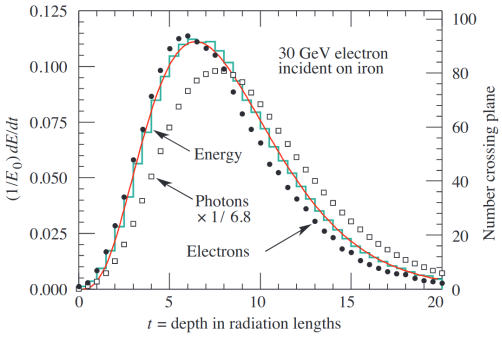


Figure 1.8: Simulation of an electromagnetic shower [PDG22]. The fractional energy loss per radiation length is shown in green, the red line is a fit to the histogram. The axis on the right hand side denotes the respective number of electrons (filled circles) and photons (open squares; scaled down) with an energy above 1.5 MeV crossing an imaginary plane perpendicular to the initial direction after penetration depth τ.

addition to the original lepton, it is easy to see how a "shower" of electrons, positrons and photons will emerge from an initial high-energy particle; regardless of whether the cascade was started by a photon or an electron (or positron). This phenomenon is known as an electromagnetic shower. It is characterized by several material-specific parameters:

The *radiation length* X_0 is defined as the mean distance after which a high-energy electron has lost all but $\frac{1}{e}$ of its energy due to bremsstrahlung. It is the appropriate scale to describe longitudinal shower development.

At some point, the energy of the particles in the shower will be too low to create e^+/e^- pairs or produce bremsstrahlung, respectively. Instead, the leptons will deposit their energy by ionization, and the lower energetic photons might liberate further electrons from the material via Compton scattering. The energy at which an electron loses as much energy to bremsstrahlung as it does to ionization is known as the *critical energy* E_c. There also exists another popular definition by Rossi, which states that E_c^{Rossi} is the energy at which the energy loss due to ionization over X_0 equals the electron energy itself. Both definitions yield different, albeit comparable values. Once a particle's energy drops below E_c, it will not significantly contribute further to the shower development. Conversely, particles with $E > E_c$ will not contribute much to the actual deposition of energy in the material and instead produce more shower particles.

Fig. 1.8 clearly illustrates this point, most of the energy deposition happens several radiation lengths deep into the material. At the same time, the number of electrons falls off more quickly than both the number of photons and the total energy deposition, meaning that an increasing fraction of the shower energy is carried by the photons.

While X_0 describes the longitudinal shower development, the transverse profile of a shower is

characterized by the *Molière radius* R_M, which is given by this simple equation:

$$R_M = X_0 \frac{E_s}{E_c^{\text{Rossi}}} \quad \left(E_s = \sqrt{\frac{4\pi}{\alpha}} m_e c^2 \approx 21.2\,\text{MeV} \right)$$

On average, 90% of a shower's energy deposition will happen within a cylindrical volume with a radius of R_M around the direction of the incident particle.

1.4 Properties of CsI(Tl)

Thallium-doped cesium iodide is a widely used material for inorganic scintillators. It was and is currently used by several nuclear and particle physics experiment around the world, including e.g. the Belle-II experiment at KEK in Japan, the BaBar experimant at SLAC, the CLEO experiment at CESR (both in the USA; decommissioned), the R^3B experiment at FAIR in Germany, or the BES III experiment at the BEPC in China, to only name a few. It is also used in the main calorimeter of the CBELSA/TAPS experiment in Bonn (the Crystal Barrel detector), which is why it is crucial to understand its properties for this thesis.
Its main advantages are the high light yield of ≈ 50 000 photons per MeV[18] and the high density of 4.53 g cm^{-3} [Ake+92]. In comparison to other widely used materials like NaI(Tl), its lower hygroscopicity is another advantage. The radiation length X_0 of 1.86 cm is also rather short[19], which facilitates compact detector setups. The critical energy (Rossi definition) is 11.17 MeV and the Molière radius 3.8 cm (properties taken from [Gro+22] and [Ake+92]).
While undoped CsI can also be used as a scintillator, the light output increases significantly[20] by adding a small amount of another element - most commonly Thallium (Tl) - when growing the crystal. Due to their similar chemical properties, Tl^+ atoms replace Cs^+ atoms within the otherwise regular lattice (see Fig. 1.9). The main purpose of adding this "activators" is to make the scintillation process more efficient and (by shifting the wavelength of the emitted radiation) prevent re-absorption of the scintillation light. More on the details of this processes will be explained in the next section.

Despite its prolific use in various experiments, the exact mechanism of scintillation is still not completely understood. However, after an initial phase of research in the 1950s and 60s ([MM61], [SJW58], [GM63]) several advancements have been made in recent years ([Gri+14], [Lu+17], [Syn+08], [Zaz01]) leading towards a more complete understanding of the relevant physics and the peculiarities of CsI(Tl) in particular. The following section is meant to give a brief overview over the relevant mechanisms with special emphasis on effects that are of interest for the application in electromagnetic calorimeters.

[18] Knoll: 65000 photons per MeV [Kno00]; Saint-Gobain (crystal manufacturing company): 54000 γ/MeV [Sai22]; Aker et al: 45000 γ/MeV [Ake+92]. The exact value depends on the amount of doping, temperature, the cutoff energy for photons to be counted, and other factors.

[19] Compared to e.g. 2.59 cm for NaI or 2.03 cm for BaF_2

[20] The light yield of pure CsI is only 3200 photons per MeV at room temperature [Ams+02]

1.4.1 Scintillation Mechanism in CsI(Tl)

From the standpoint of an experimentalist, the main purpose of a scintillator is to convert the energy of an incoming particle to an amount of light proportional to the deposited energy. Following the treatise in [Lec20] and [Ben14], the scintillation process can be subdivided into 5 different phases:

1. **Energy conversion:** Fundamentally, by whatever process the initial particle has lost its energy in the material, the last step of the energy-deposition cascade will always involve the ionization of atoms in the material. Since CsI(Tl) as well as other solid inorganic scintillators are crystalline materials forming a regular lattice, the physical processes are most easily understood in terms of the well known band model for semiconductors: Within this model, the process of ionization is equivalent to raising an electron form the valence band to the conduction band, thereby leaving a hole[21] in the valence band. A schematic depiction of the band structure can be found in Fig. 1.10. Also note the position of the activator states (resulting from the doping) in the band gap. Direct excitation of the activator states is also possible, but due to the generally relatively low concentration of doping, it does not contribute significantly.

2. **Thermalization:** The created charge carriers lose their respective kinetic energy by interactions with the lattice. Most importantly, by scattering with so called "phonons"[22]. At the end of this phase, the electrons are at the bottom of the conduction band, whereas the holes are at the top of the valence band.

3. **Localization:** After losing most of their kinetic energy, the charge carriers are free to drift within the crystal. There exist several processes by which charge carriers can get localized or "trapped" at specific positions within the lattice:
   - Trapping of e^- by a Tl^+ atom: $e^- + Tl^+ \rightarrow Tl_0$
     This trapping happens very fast within a few ps [Lu+17].
   - Trapping of h^+ by a Tl^+ atom: $h^+ + Tl^+ \rightarrow Tl^{2+}$
     Happens not quite as fast as e^- trapping due to lower mobility of the holes.

[21] In solid-state physics, these "holes" are usually treated as quasi-particles of positive charge.
[22] Quasi-particle describing quantized vibrations ("sound waves") within a lattice

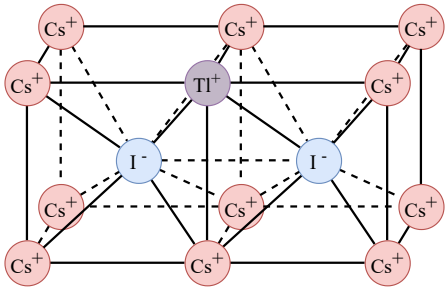


Figure 1.9: Lattice structure of CsI(Tl). The structure is bcc (body centered cubic). [own illustration]

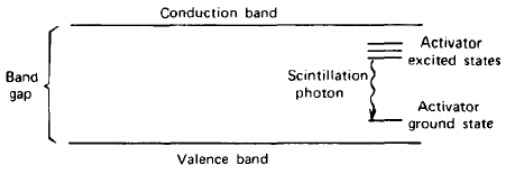


Figure 1.10: Energy band structure of a doped inorganic scintillator [Kno00]. Activator states are situated in the band gap.

- Self Trapped Hole (STH): $h^+ + 2I^- \rightarrow I_2^-$ (also known as "V_k-Center")
  The holes can be trapped in an energetically favorable configuration "in between" two I^- atoms to form a I_2^- quasi-molecule within the lattice (see Fig. 1.9)
- Direct formation of excitons, leads to Self Trapped Excitons (STEs): $e^- + h^+ \rightarrow ex^0$
  Free electrons and holes can form a metastable bound state, which is also energetically favorable. This requires that both charge carriers come close together with low kinetic energy and will mostly happen for very low activator concentrations where the probability of trapping the charge carriers by other means is low.
- Deep trapping of e^- or h^+ in lattice defects:
  Lattice defects, for example the absence of a Cs^+ or I^- atom from its designated position results in a strong local electric field which attracts charge carriers of the respective opposite charge and traps them very deeply.

4. **Transfer:** In order to produce measurable scintillation light, the charge carriers must travel to the activator sites (Tl^+) to form an excited state (Tl^{+*}). During the de-excitation of this excited state, the charge carriers finally recombine with their respective counterparts and measurable light is produced. After the localization phase, almost all charge carriers will be located at specific spots within the lattice. Several mechanisms are possible, by which opposite charge carriers could recombine:

   (a) Trapped electron (Tl_0) plus STH $\rightarrow$ Tl^{+*}. This process generally happens fast, especially because the formation of self-trapped holes is most efficient (energetically favorable) [Ben14] in the vicinity of activator sites. The corresponding process of a trapped hole (Tl^{2+}) recombining with a free electron is also possible, but rare, since the trapping of free electrons happens so quickly.

   (b) Trapped electron (Tl_0) plus trapped hole (Tl^{2+}). Since it is highly unlikely to have two activator centers in direct proximity, either one of the charge carriers has to be released from its trap and drift to the corresponding other site in order to form a Tl^{+*} excited state. The most common mechanism of "untrapping" is via thermal excitation.

   (c) In case an STE was formed, the direct reaction STE plus $Tl^+ \rightarrow Tl^{+*}$ is also possible. However, as stated above, this is only really relevant for very low activator concentrations and usually contributes less than 5% to the overall light output [Lu+17]

5. **Radiative relaxation:** The Tl^{+*} states de-excite under emission of light of a characteristic wavelength. New research [Zaz01] seems to indicate, that it is actually not the relaxation of excited activator states themselves, but the de-excitation of self-trapped excitons in the vicinity of the activator centers which gives rise to most of the characteristic radiation[23]. Either way, for the purpose of this overview, the exact relaxation mechanism is not important. In both cases, the eventual recombination of electrons and holes at or near an activator site gives rise to the light output which is measured.

[23] Details also depend on the specific configuration; i.e. how close exactly the exciton is to the activator centers.

1.4.2 Light Output of CsI(Tl)

While earlier measurements have reported the existence of two distinct components in the scintillation light, a "fast" component with an exponential decay time $\tau_f \approx 0.7\,\mu s$ and a slower one with $\tau_s \approx 3.3\,\mu s$ [Val+93] (at room temperature), recent measurements with high precision indicated the existence of another, even slower, "tail" component with $\tau_s \approx 16\,\mu s$ [Syn+08]. There is, however, no consensus over the exact number of decay modes or their respective decay times. Other sources provide somewhat different values[24] for the first two components. Apart from the methodological difficulty of fitting a superposition of an (a priori unknown) number of exponential decay curves, different experimental conditions, crystal manufacturers, or doping concentrations might account for some of the discrepancies. Nonetheless, the existence of *at least* two different components with decay times around 750 ns and 3 µs seems certain.
According to Lu et al. [Lu+17], the existence of these different components seems to be a result of the different transfer mechanism described above. While the fast component is assumed to come from process 4(a) (and, to some degree from 4(c)[25]), i.e. the prompt recombination of self-trapped holes with electrons trapped at the activator sites, both the "slow" and the "tail" component must come from process 4(b). Since this process requires (thermal) untrapping of a charge carrier, it is obvious why this process takes longer on average. This is supported by the fact, that τ_s decreases/increases significantly, when the crystal is heated/cooled [Val+93].
It is theorized, that due to their higher mobility, the electrons tend to spread out further radially from the initial track than the holes. In a simplified view, this gives rise to two different regions. One close to the initial track where both electrons and holes are trapped at different sites, but within the same cylindrical volume around the track. The other region is further away from the track, where mostly electrons are trapped and almost no holes. Within the first region, a single untrapping event will most likely be sufficient to bring an untrapped electron to a trapped hole at an activator site (or vice versa) and facilitate the radiative recombination ($\rightarrow$ slow component). In the outer region however, there exists a high probability that an untrapped charge carrier (most likely an electron) can get re-trapped at another site site and has to be untrapped again (potentially several times) before eventually reaching its respective partner for recombination. The same happens in the inner region, once a substantial number of charge carries is already depleted (leaving many unpaired holes). This gives rise to the tail component.

1.4.3 Dependence on Ionization Density

It is well known since the late 1950s [SJW58], that the light output from CsI(Tl) depends on the ionization density $\rho_I = \frac{N}{V}$ (where N denotes the number of ionized charge carriers) of the initial particle. If most energy is lost via ionization, then ρ_I is, in first order approximation, proportional to the differential energy loss $\frac{dE}{dx}$ and hence to the stopping power[26] $\frac{dE}{dx}\frac{1}{\rho}$, which (apart from a particles energy) depends on its mass and charge.
In particular, a larger proportion of the scintillation light seems to be deposited via the fast

[24] e.g. [Gas+13]: $\tau_f = 0.87\,\mu s$; $\tau_s = 2.5\,\mu s$, [Ake+92]: $\tau_f = 0.9\,\mu s$; $\tau_s = 7\,\mu s$.

[25] The exact role and importance of the direct formation of excitons depends on various factors, such as activator concentration, temperature, ionization density and others. A detailed discussion of this topic would go beyond the scope of this overview.

[26] Depending on the context, "stopping power" might also refer to $\frac{dE}{dx}$ (i.e. without division by the density).

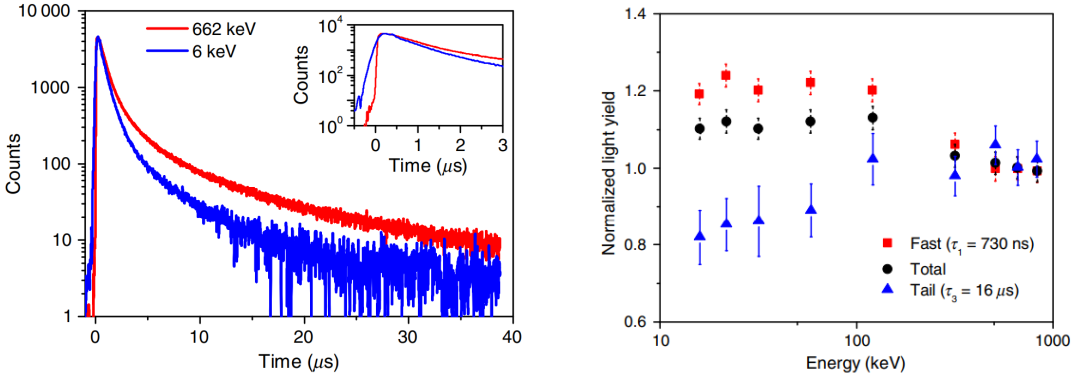


Figure 1.11: Left: Measured amplitude of scintillation light versus time for electrons of different energies. Right: Extracted relative contributions of different components to the overall light yield. Figures taken from [Lu+17], originally by [Syn+08].

component for particles of high ionization density, whereas particles of low ionization density seem to produce more slow scintillation light[27]. This implies, that CsI(Tl) might be used for the purpose of Particle Identification (PID). In recent years, Bendel et al. [BGH+13] have obtained impressive results in the identification of photons versus protons based on careful analysis of the scintillation light from CsI(Tl). They were also able to differentiate between protons that were stopped in the material and those which punched through the detector due to the characteristic area of high ionization density in the Bragg peak (see section 1.3).
The reason for this behavior lies again in the transfer processes. The higher the ionization density, the more free charge carriers will be produced within a small volume around the initial track, creating a vast supply for the fast process. On the contrary, for low ionization densities, the initial concentration of charge carriers will be low, so that the fast process becomes more unlikely. Instead, it becomes more likely that the charge carriers will have to untrap at least once from their first trapping site in order to be able to recombine. Fig. 1.11 illustrates the relative contributions of the three components for electrons of several energies. For this energy range, a lower energy corresponds to higher ionization density. It can be clearly seen, that the fast component's contribution declines, while contributions from the slow (not shown) and tail components increase for higher energies.

1.4.4 Non-Linearities

It is also well established, that the total light output decreases for low as well as for very high ionization densities [MM61] [GM63] (see Fig. 1.12). There are several effects at play which can lead to such non-proportionality in the light output, when CsI(Tl) is used to measure energy:

- Charge carriers that are trapped by defects in the crystal or other "deep" traps tend to stay trapped for long periods of time and are practically lost for the production of scintillation

[27] The opposite is true for (among others) BaF_2. The "slow" component of BaF_2 has a similar decay time and similar physics to the fast component in CsI(Tl), whereas the ultra-fast component of BaF_2 ($\tau \approx 600$ ps) is due to an entirely different process know as cross-luminescence [Die+15].

light. Since the density of defects in the crystal volume is constant, the relative impact of this is bigger for low $\frac{dE}{dx}$, i.e. when the number of charge carriers per volume is small to begin with. From Fig. 1.12 it seems that this applies to stopping powers $\frac{dE}{dx}\left(\times\frac{1}{\rho}\right)$ below $\approx 10\,\mathrm{keV\,cm^2\,mg^{-1}}$, which corresponds to $\frac{dE}{dx} \approx 50\,\mathrm{MeV\,cm^{-1}}$ in CsI.

- For very large stopping powers (above $\approx 100\,\mathrm{keV\,cm^2\,mg^{-1}}$ or $\approx 500\,\mathrm{MeV\,cm^{-1}}$ in CsI), the light yield drops as well, a phenomenon known as "quenching". In general, quenching can be attributed to non-radiative processes of de-excitation which become more probable with growing density of charge carriers. Without going into too much detail ([Mos+12] gives a good overview on the topic), one such process would be the dissociation of an exciton involving another exciton[28] in an Auger-like process: One exciton recombines and transfers its energy to the other one, which in turn dissociates into free charge-carriers again, thus decreasing the overall number of charge carriers available for radiative recombination.

- An important detail when talking about "light yield", is the length of the integration window: In any practical application, the scintillation light[29] has to be integrated in order to get the deposited energy. Ideally, this integration would take infinitely long to get all the energy. In practice, some finite integration length must be used. Typical choices for CsI(Tl) range somewhere between 4 and 10 µs. It is obvious, that such integration windows will most likely capture the bulk of the fast component, but only part of the slow, let alone the tail component of the signal. Therefore, the measured integral per deposited energy might also decrease for smaller ionization densities, depending on the integration window.

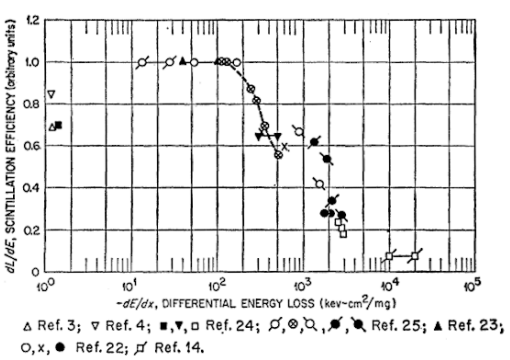
With regards to the particles relevant at CBELSA/TAPS, the average ionization density of an electromagnetic shower can be considered to be quite low, since most of the energy deposition happens below the critical energy of 11.17 MeV, so in an energy range where the electrons and positrons are close to the minimum of their $\frac{dE}{dx}$-curve[30] (see Fig. 1.13). For cosmic muons of sufficiently high energy and other weakly ionizing particles, the ionization density will be similarly low and therefore a similar scintillation characteristic and a comparable light yield can be expected. For protons, the behavior depends on their energy. For kinetic energies above $\approx 200\,\mathrm{MeV}$ the ionization density is again similar, but for protons of low kinetic energy (especially below $\approx 25\,\mathrm{MeV}$), the ionization density becomes big enough such that the light yield will be larger than for the other particles discussed. By how much exactly seems hard to tell, since there does not seem to be a comprehensive data set anywhere in the literature and the actually measured light yield depends on several factors besides the scintillator material itself. As a ballpark figure, [Ben14] reported a factor of 1.27 by which the light yield of protons[31] exceeds the one from photons of similar energies. The same non-linearity is also visible in data from the CBELSA/TAPS experiment; see Fig. 6.13.

[28] For very high ionization densities (or very low doping concentrations), the number of activator centers can be locally insufficient to quickly trap all charge carriers. In that case, the formation of excitons becomes a very important mechanism.

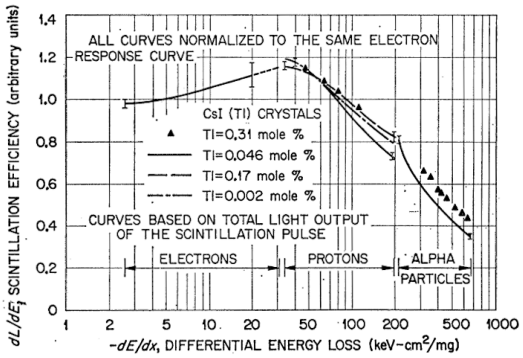
[29] More precisely: the electrical signal generated by the scintillation light in a photomultiplier, APD, etc.

[30] Even though $\frac{dE}{dx}$ increases again for decreasing particle energies, the minimum region between ≈ 0.1 and $\approx 10\,\mathrm{MeV}$ already covers almost all of the particles total energy to be deposited.

[31] The maximum kinetic energy of protons in the experiment was 16 MeV.



(a) Compilation of different results by [MM61].



(b) Measurements by [GM63].

Figure 1.12: Light yield non-proportionality in CsI(Tl) from two different publications.
Gwin and Murray [GM63] (r.h.s) used an integration time of $\tau = 7\,\mu s$ (extrapolated to ∞), whereas the integration times for the various datapoints compiled in [MM61] were left unspecified.

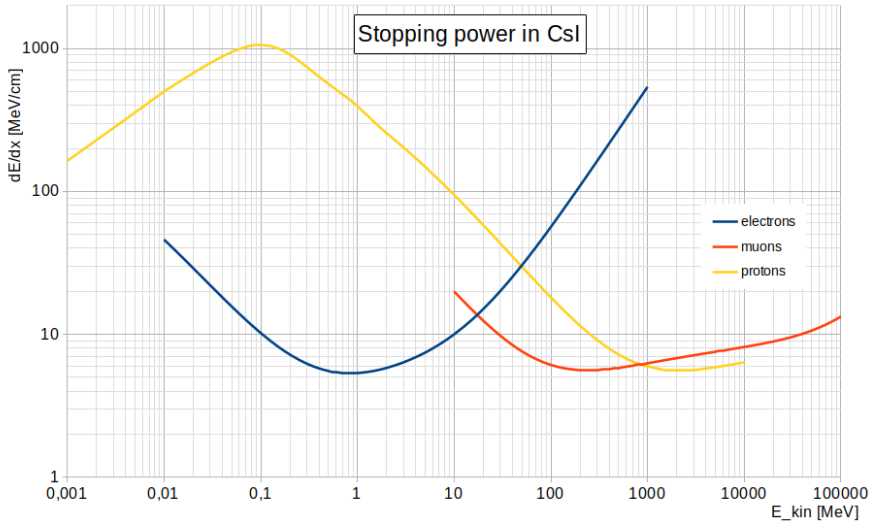


Figure 1.13: Stopping power (differential energy loss $\frac{dE}{dx}$) of several particles in CsI (values for CsI(Tl) would be almost identical, due to the small amount of doping atoms). Values taken from PDG for muons [Gro+22] and NIST [Ber+05] for electrons and protons.

CHAPTER 2

The CBELSA/TAPS-Experiment

Situated at the Electron Stretcher Accelerator (ELSA) facility in Bonn, the CBELSA/TAPS-Experiment studies baryonic resonances produced in photoproduction reactions. The experiment features a polarized beam as well as a polarized target to enable the extraction of not only differential cross sections, but also single and double polarization observables. The ELSA beam energy of up to 3.2 GeV gives it a maximum center of mass energy of 2.62 GeV[1], which is ideal for studies of the light baryon excitation spectrum. Due to its electromagnetic calorimeters, the CBELSA/TAPS experiment is especially well suited to study final states with neutral mesons, decaying into photons[2]. In terms of charged and strange final states, it is complemented by the BGO-OD experiment located at the same accelerator facility.

2.1 Overview

Fig. 2.1 shows an overview of the experiment. The individual components will be explained in detail in the following sections, after a short synopsis of their respective purpose is provided:
The **ELSA** accelerator delivers a beam of electrons, which can be polarized. The electron beam enters the **Goniometer** tank, wherein it impinges on one of several radiator targets. By the process of bremsstrahlung, the electrons are used to create the beam of photons needed for the photoproduction experiment. Depending on the type of radiator, the resulting photon beam may also be polarized. The produced photons are energy-tagged[3] in the so called **Tagging System**, before they hit the **Target** which is surrounded by an **Inner Detector** for charged particle tracking. Both sit in the center of the experiment's main detector, the **Crystal Barrel** electromagnetic calorimeter. The tracking system is complemented by the **Forward Veto** plates, which cover part of the forward direction. The Crystal Barrel covers all but the most forward and backwards directions of the full solid angle, but since CBELSA/TAPS is a fixed target experiment, particles tend to be emitted under small angles to the initial beam direction due to the forward boost in the lab frame. These angles are covered by a **Gas Cherenkov** detector, which is used

[1] Assuming a proton at rest as the target particle. The e^- beam is converted to a γ beam beforehand (see section 2.3)
[2] The π^0 has a 98.8% Branching Ratio (BR) to 2γ, the η meson has a BR of 39.4% to 2γ and 32.7% to $3\pi^0$ ($\rightarrow 6\gamma$ in most cases). ω (8.35% to $\pi^0\gamma$) and η'(22.4% to $\pi^0\pi^0\eta$) have significant photonic final states as well [PDG22].
[3] Technically: the scattered electrons (details below)

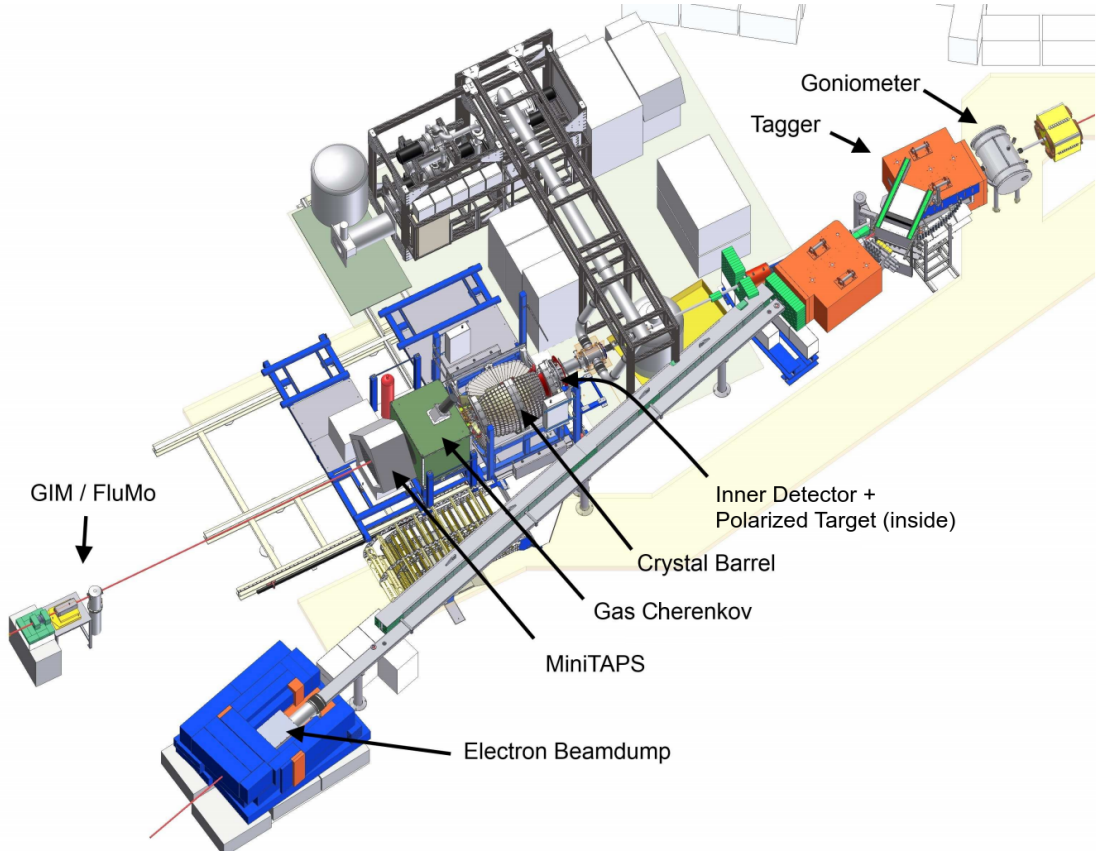


Figure 2.1: Setup of the experiment [Wal07]. The electron beam is coming from the top right, is converted to a photon beam in the goniometer and impinges on a target situated in the middle of the Crystal Barrel detector.

for the suppression of electromagnetic background reactions, and the **MiniTAPS** calorimeter. Additional detectors (**GIM** and **FluMo**) for the flux determination are located even further down the beamline.

2.2 The ELSA Accelerator

The ELSA facility provides an electron beam of up to 3.2 GeV, which can be polarized. It consists of several parts which successively accelerate the electrons to higher and higher energies.
The polarized electron source is capable of delivering a beam current of up to 100 mA and a polarization degree of 83% [Hil00]. The electrons are accelerated in LINear ACcelerator (LINAC) 2[4] to 26 MeV.
Afterwards, the electrons are injected into the booster synchrotron, where they are further accelerated before injection into the final stretcher ring; typically to an energy of 1.2 GeV. The stretcher ring is used to accumulate electrons from the previous stages on the one hand, while on

[4] LINAC 1 was used for unpolarized electrons, but has been decomissioned. A new thermal electron gun was installed at LINAC 2 to provide unpolarized electrons.

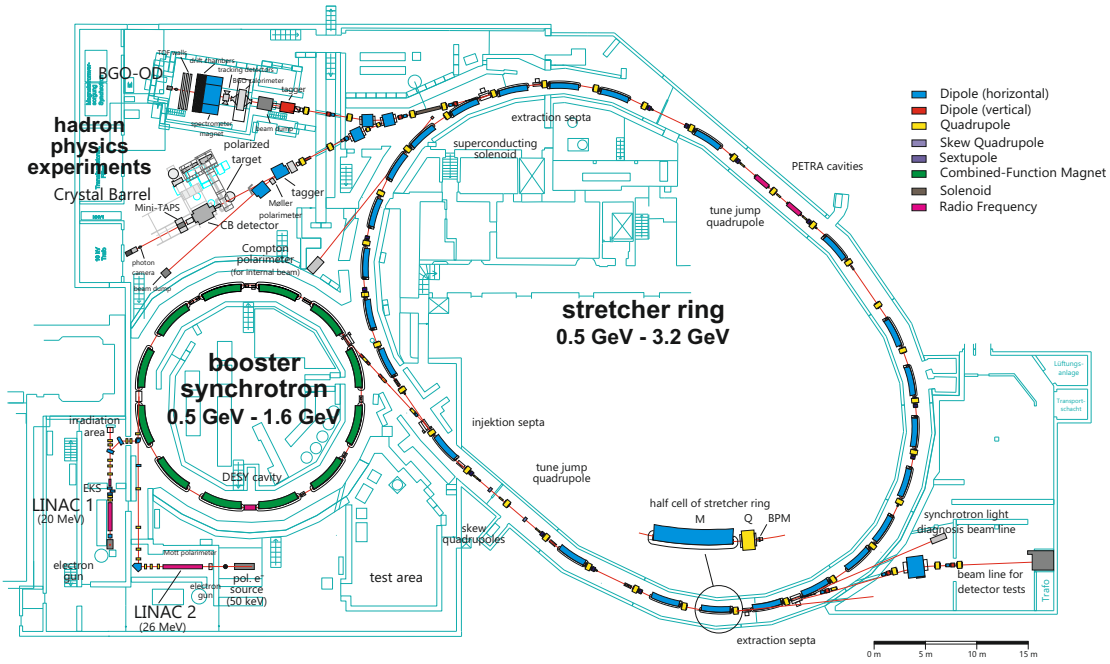


Figure 2.2: Overview of the ELSA setup: After acceleration in a linear accelerator, the electrons are transferred to a smaller booster synchrotron, followed by the main stretcher ring in which they are stored and accelerated to their final energy. The two hadron physics experiments BGO-OD and CBELSA/TAPS are situated in the upper left corner of the graphic. [FH22].

the other hand providing the final acceleration to energies up to 3.2 GeV. Finally, the beam can be extracted to one of two hadron physics experiments[5] and a separate area for detector tests. In its usual mode of operation, a single extraction cycle takes roughly 8 seconds, during which the electrons are accumulated and accelerated in the stretcher ring and then gradually extracted to provide a quasi-continuous[6] beam over the course of several seconds.

2.3 Tagging System

2.3.1 Goniometer

A five-axis goniometer with different radiator targets mounted in a wheel allows to select the appropriate target for the given application (see Fig. 2.3). Unpolarized photons are created by using unpolarized electrons and an amorphous radiator. Polarized photons (circular or linear/transverse) can be produced as well:

Circular polarization By using longitudinally polarized electrons and an amorphous radiator (copper foils of varying thickness[7]), the polarization of the electrons can be transferred to the

[5] The other one being the BGO-OD experiment.

[6] Due to the acceleration method, the beam is, in fact, bunched. Bunches are ≈ 120 ps wide and 2 ns apart [Hil00].

[7] 12, 50, 150 and 300 µm.

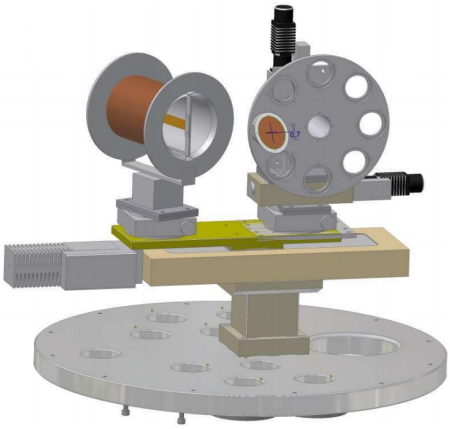


Figure 2.3: Different radiator targets. Møller polarimeter (left) and goniometer wheel (right) [Wal07].

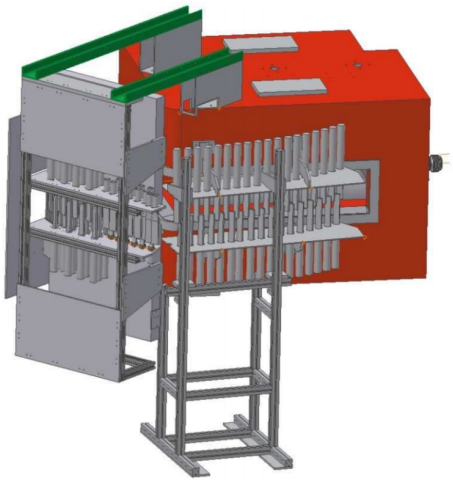


Figure 2.4: The tagging system consisting of dipole magnet (red), scintillating bars and fibers [Wal07].

photons. Due to the helicity transfer mechanism, the resulting polarization degree approaches zero for low momentum transfers and the initial polarization degree of the electron beam for complete momentum transfer [OM59] (see Fig. 2.5). The initial polarization degree is measured with a Møller-polarimeter (depicted to the left of the goniometer wheel). It consists of a thin foil inside of a coil, which produces a magnetic field to polarize the electrons in the foil. Because electron-electron scattering (Møller scattering) is spin-dependent, an asymmetry occurs in the resulting distribution of scattered electrons. By measuring the scattered electrons with dedicated detectors, the polarization degree of the beam can be deduced.

Coherent Bremsstrahlung The energy spectrum of the bremsstrahlung from an amorphous radiator follows (in first order approximation) a E_γ^{-1} distribution (see red line in Fig. 2.6. When the scattering target forms a regular lattice, the recoil $\vec{q}$ from the electron can be absorbed by the whole lattice instead of single atomic nuclei. Provided that the Laue condition

$$\vec{q} = n \cdot \vec{g}$$

(where $n \in \mathbb{N}$ and $\vec{g}$ is a reciprocal lattice vector) is fulfilled, this leads to the emission of a photon with a fixed polarization. If the lattice was not regular, the different polarizations from different lattice orientations would effectively cancel out and lead to an unpolarized beam.
A diamond crystal provides such a regular lattice. By carefully aligning the crystal, it can be achieved that only a single recoil vector fulfills the Laue condition. The coherent (polarized) part of the resulting energy spectrum are shown in green in Fig. 2.6. It shows a steady increase up to a certain point, after which the spectrum drops steeply. At this point, the so-called “coherent edge”, the recoil momentum drops out of the kinematically allowed region (see e.g. [Har17] for a more detailed summary). Its position in energy is given by the crystals orientation relative to

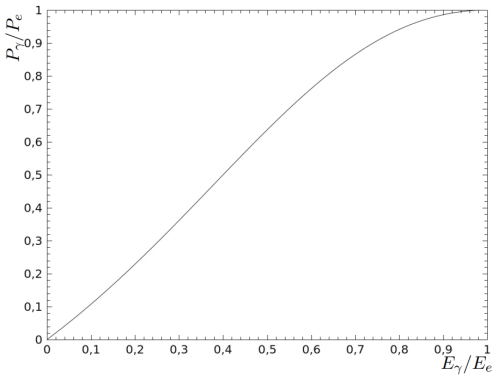


Figure 2.5: Polarization transfer in the production of circularly polarized photons [Thi12].

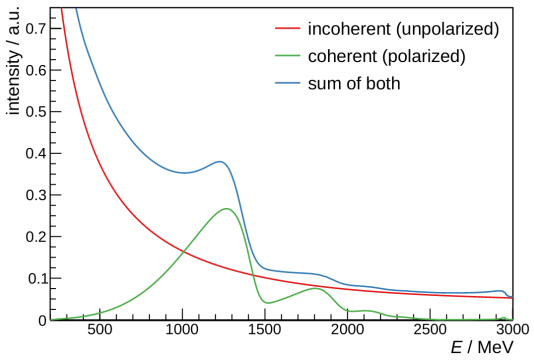


Figure 2.6: Intensity spectrum with the coherent edge at $E_\gamma = 1.2\,\text{GeV}$ and $E_{\text{beam}} = 3.2\,\text{GeV}$ [Har17].

the electron beam and can be changed by tilting the crystal. While the first edge corresponding to $n = 1$ is the most pronounced, further edges are visible at higher energies which correspond to higher values of n. The polarization degree of the photon beam is finally calculated from the measured sum of coherent and incoherent part (blue line) and the incoherent part, which is approximated by a separate measurement with a copper radiator.

2.3.2 Tagging Hodoscope

In order to investigate photoproduction reactions, the initial state needs to be known. While the target particles are assumed to be at rest, the energy of the projectile photons is determined by measuring the momentum and hence the energy of the electrons after the bremsstrahlung process in a dipole spectrometer. The magnetic field is chosen such that the majority of electrons[8] which did not undergo any reaction are deflected by a fixed angle of 9° and are further directed into the electron beam dump. These electrons are of no further use for the experiment.
If an electron emitted a bremsstrahlung photon, its energy will be lower and, accordingly, the angle of deflection in the magnetic field will be larger. The angle of deflection and hence the momentum of the deflected electron can be determined by using an array of overlapping scintillating bars and fibers (see Fig. 2.4). The energy of the produced photon can then be calculated using the simple relation:

$$E_\gamma = E^{e^-}_{\textbf{beam}} - E^{e^-}_{\textbf{tagger}} = E^{e^-}_{\textbf{beam}} - \sqrt{p^2_{\textbf{tagger}} + m_e^2} \tag{2.1}$$

Both bars and fibers are plastic scintillators, read out by Photo Multiplier Tubes (PMTs). The PMT signals are digitized using the same Time-to-Digital Converters (TDCs) as the Crystal Barrel (CB) calorimeter (see section 2.8.1), albeit with a different firmware [Bie16], to provide precise timestamps for each of the produced photons. There are 96 bars with widths between 1.4 to 5 cm, which, in order to suppress background, are positioned such that electrons are detected

[8] In order to ensure a low probability of multiple bremsstrahlung processes by a single electron, the radiator has to be chosen sufficiently thin, resulting in a low probability of interactions overall. In case of multiple scattering, the photon energy would not be directly related to the energy of the remaining electron.

in at least two adjacent bars. In addition, 480 scintillating fibers (diameter 2 mm) cover the high-energy part of the hodoscope. These fibers are much narrower than the bars, increasing the energy resolution for low-energy photons[9]. The bars cover an angular region corresponding to 2.1–82.5% of the electron beam energy, whereas the fibers cover a slightly different region of 16.6–88.1%. Details can be found in [For10].
Before the so produced photons reach the target, the photon beam is collimated. The collimator consists of 6 adjacent cylinders with an outer diameter of 20 mm and a length of 4 cm each. Every cylinder has a small hole with a diameter of 4 mm running through its center. The collimators are made of tungsten, a material with very high density ($\rho = 19.25\,\mathrm{g\,cm^{-3}}$) and proton number ($Z = 74$), which makes it a very efficient absorber for photons. Any charged particles produced by interactions of the photons with the collimator (e.g. e^+/e^- pairs) are deflected away from the beamline by a second dipole magnet which is placed after the collimator.

2.4 Targets

The experiment can be run with a variety of different targets. To study reactions off unpolarized protons, a liquid hydrogen target cell was developed [Ham09]. It can also be filled with liquid deuterium, to study reactions off neutrons as well. However, in order to extract beam-target double polarization observables, a polarized target is needed: the Bonn Frozen Spin Target [Bra+99].

2.4.1 The Polarized Target

The target material is 1-Butanol (C_4H_9OH), which contains several free[10] protons that can be polarized. The target cell itself is a 2 cm × 2 cm cylinder located within a ^{3}He/^{4}He dilution cryostat, which is capable of providing temperatures below 100 mK. By using the method of Dynamic Nuclear Polarization (DNP) [CM97], a polarization degree of over 85% can be achieved at a temperature of 300 mK. The polarization requires a strong magnetic field of 2.5 T, which is not possible to achieve at the position of the target right inside the main detector. The target is therefore polarized outside of the detector and then inserted[11], while a small internal coil provides a weaker magnetic holding field of ≈ 600 mT to maintain the polarization for as long as possible. Additionally, the temperature is further reduced down to 60 mK to reduce the depolarization. Fig. 2.8 shows that the resulting relaxation time can exceed 200h, which allows for large stretches of continuous data taking. Only if the polarization has dropped too far, the target is pulled out of the detector and is repolarized. Reactions of polarized neutrons can be studied as well by using deuterated butanol (D-butanol) instead of 1-butanol.
Butanol not only contains hydrogen, but other atoms (4 C and 1 O atom per molecule) as well, which do not have spin and are therefore unpolarizable. In terms of double polarization analysis, reactions off nuclei in these atoms constitute a large background. Fortunately, these background contributions can be accounted for by performing measurements with a carbon foam target of similar dimensions and density within the same cryostat.

[9] High electron momenta correspond to low photon energies and vice versa.
[10] In this context, "free" means “not bound in other nuclei”, i.e. the 10 hydrogen nuclei.
[11] For mechanical reasons, the targets position is, in fact, fixed, while the detector slides over the target.

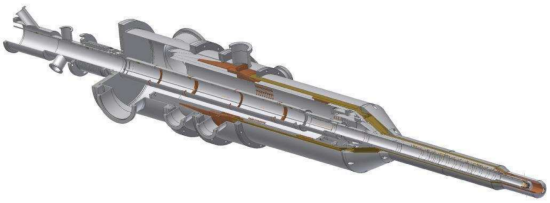


Figure 2.7: Bonn Frozen Spin Target: The actual target cell is located in the tip on the right [Wal07].

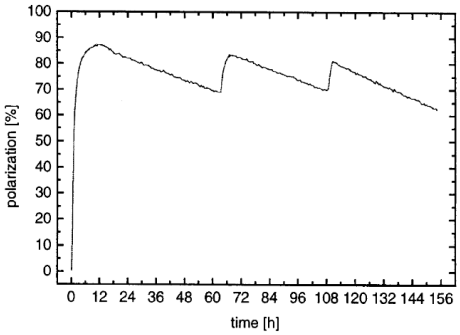


Figure 2.8: Proton polarization during three typical frozen-spin cycles [Bra+99].

For the duration of this thesis, the Bonn Frozen Spin Target was under maintenance [Run+16]. In cooperation with the colleagues from Mainz, the cryostat/cooling system of the Mainz-Dubna Frozen Spin Target was used instead. It is usually installed at the MAMI[12] accelerator facility in Mainz. Both targets are very similar in a variety of ways[13], such that the combination of the cryostat with components from Bonn was possible and data taking could proceed (almost) as normal. With a holding field of 450 mT and a temperature of 25 mK, polarization degrees around 75% can be reached [Tho+13].

2.5 Charge-sensitive Detectors

2.5.1 Inner Detector

The target is surrounded by an array of 513 scintillating fibers in three layers, comprising the inner detector [Suf+05]. Fig. 2.9 illustrates how the geometry is matched to the geometry of the target (see Fig. 2.7). While the fibers in the outer layer are parallel to the beam axis, the other two layers are wound helically half way around, with opposing directions of rotation. The fibers are read out on the backside by mulit-anode PMTs connected via light guides, while the forward end of each fiber is reflective. The signals are digitized using multihit TDCs. By demanding coincidences between hits in at least two of the three layers, this allows for a precise reconstruction of where exactly a charged particle went through the detector. Since the scintillating fibers are only sensitive to charged particles, the inner detector is also used to discriminate between charged and uncharged particles. The detector covers a polar angle range from 21° to 167°, although this range slightly changed when the target from Mainz was installed. Due to geometric constraints, the inner detector had to be shifted slightly to the front, such that the angular coverage is now 13.7° to 155°.

[12] **MA**inz **MI**crotron

[13] The target from Bonn was used at MAMI in 1998 and 2003 for measurements of the Gerasimov-Drell-Hearn GDH sumrule [Ahr+06].

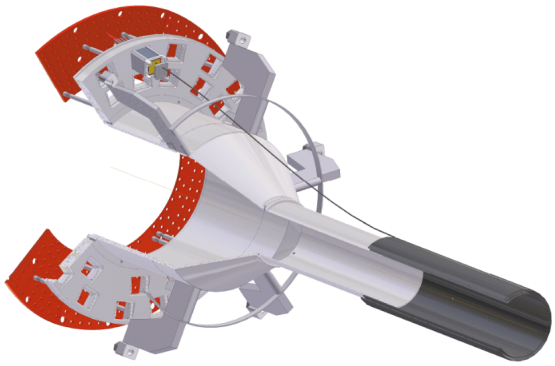


Figure 2.9: Inner detector (in black) with holding structures [Wal07].

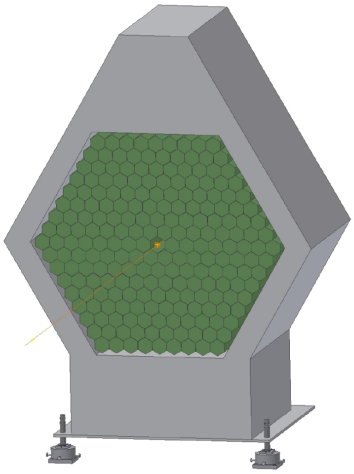


Figure 2.10: Graphical model of the MiniTAPS detector [Wal07].

2.5.2 Forward Veto Detector

With the inner detector in its designated position, the forwardmost 3 rings of the Crystal Barrel calorimeter (see section 2.8) lack angular coverage by the inner detector. Instead, thin plastic scintillator plates were installed in front of these crystals [Wen08]. They are read out by PMTs and can, just like the inner detector, provide position information for charged particles. This detector system is called a "veto" detector, because it discriminates charged particles from photons.

2.5.3 Cherenkov Veto Detector

One of the largest backgrounds in the measurement of photoproduction reactions stems from purely electromagnetic reactions. Most notably, e^+/e^- pair production and Compton scattering lead to the emission of charged particles; predominantly at very small angles to the photon beam direction. So it is desirable to exclude reactions containing one or more electrons or positrons in the final state.
The Cherenkov detector of the CBELSA/TAPS experiment serves exactly this purpose. It covers all angles below 12.8° and is placed in between the Crystal Barrel and MiniTAPS. It consists of a large aluminum box which is filled with CO_2 gas at atmospheric pressure. Under these conditions, electrons with an energy above 17.4 MeV produce Cherenkov light in the CO_2, which is detected by a PMT on top of the detector. The next lightest particle from the standard model, the muon, would need an energy above 3.5 GeV to be detected. Since such energies are not accessible in the experiment, the Cherenkov detector is therefore exclusively sensitive to electrons (and positrons). Its efficiency to detect these particles has been measured to be 99.97(5)% [Kai07]. This signal is then used as a veto signal to suppress electromagnetic background in the trigger system (see section 2.9 (Data Acquisition and Trigger)).

2.6 MiniTAPS

The MiniTAPS[14] calorimeter covers the polar angle from 1° to 12° and complements the coverage of the larger Crystal Barrel in the forward direction with good energy resolution and high granularity. It consists of 216 hexagonal BaF_2 crystals, with the innermost crystal left out to allow the beam to pass through (see Fig. 2.10). The crystals are 25 cm long, which corresponds to around 12 radiation lengths ($X_0 = 2.03\,\text{cm}$ [Gro+22]) and have a width (side-to-side) of 59 mm, which translates to an angular coverage of 1.6° per crystal at the measurement position of $d = 210\,\text{cm}$ from the target. In addition, the crystal ends have been ground to a cylinder with an additional length of 2.5 cm to which PMTs are attached [Nov91]. The signals are then digitized by Charge-to-Digital Converters (QDCs) (see section 2.8.1) and TDCs which provide both energy and timing information. The energy resolution has been empirically determined to be $\frac{\sigma_E}{E} = \frac{0.59\%}{E/\text{GeV}} + 1.9\%$ [Gab+94], while the angular resolution[15] is around 0.2° in the polar angle θ [Dah08].
Just like CsI(Tl) (see section 1.4), BaF_2 is another inorganic scintillator which has different components in the scintillation light: a very fast one with $\tau \approx 600\,\text{ps}$ and a slower one with $\tau = 630\,\text{ns}$ [Kno00], which could also be used for the purpose of particle identification [Mur+86]. This very fast decay time makes MiniTAPS predestined for use in the forward region where the short decay time ensures a lower probability of pile-up (overlapping signals) despite of the high event rate[16]. The according time resolution has been determined to be $(0.872 \pm 0.006)\text{ns}$ [Har08]
Similar to the Forward Veto detector, plastic scintillators are placed in front of the crystals and allow for the discrimination between charged and uncharged particles.

2.7 Gamma Intensity Monitor and Flux Monitor

In order to calculate cross sections, the total flux of photons impinging upon the target has to be known. For this purpose, the experiment is equipped with 2 specialized detectors at the end of the photon beam line: the Gamma Intensity Monitor (GIM) and the Flux Monitor (FluMo). The GIM consists of 16 PbF_2 crystals in which the photons start to develop an electromagnetic shower (see section 1.3.3). In contrast to other detectors in the experiment (MiniTAPS, Crystal Barrel), PbF_2 is not a scintillator[17]. Instead, Cherenkov light is produced by the electrons and positrons in the shower which is again read out by PMTs [McG08]. Unfortunately, the accuracy of the determined flux suffers for rates above a few MHz due to dead-time effects, which is why the FluMo was installed in addition [Die08]. The FluMo uses a 100 µm lead foil to convert impinging photons into e^+/e^- pairs which are then detected in plastic scintillators. Its advantage regarding high rates lies in the fact that for sufficiently thin foils, only a certain percentage of photons produce e^+/e^- pairs. After this percentage has been determined, the actual rate of photons can be calculated from the measured rate. To suppress any contribution by charged particles coming along the

[14] TAPS stands for "Two Arm Photon Spectrometer", a previous detector from which the MiniTAPS was adapted.

[15] The resolution is much better than the crystal granularity because electromagnetic showers extend over several crystals and the showers center of gravity can be more accurately determined than the size of a single crystal.

[16] By exploiting the different wavelengths of the scintillation light from both processes, it is also possible to remove the slow scintillation component entirely [Die+15]

[17] Or at least is a very dim scintillator with low light output [And+94].

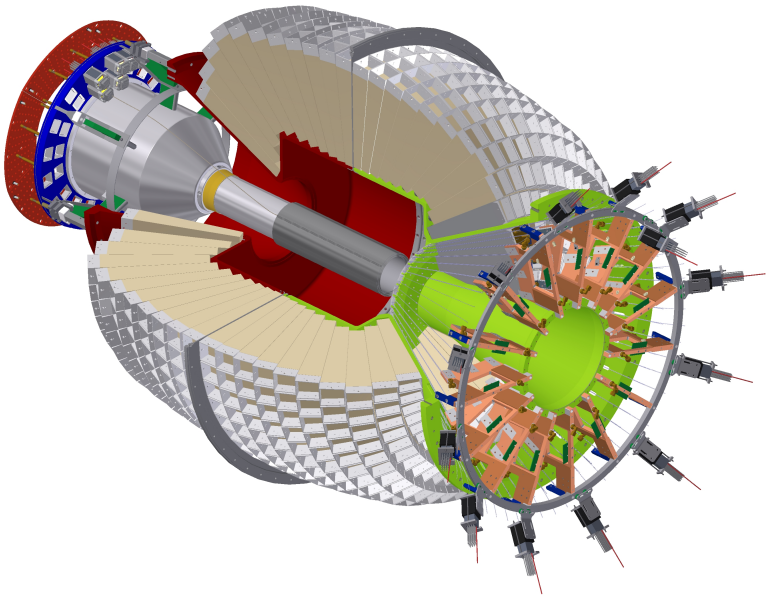


Figure 2.11: Setup of the Crystal Barrel calorimeter [Gru16]; the beam enters from the left. The green area on the right is the holding structure of the Forward-Plug, which still holds the PMT readout for the Forward Veto plates.

photon beam, another scintillator is installed in front of the FluMo.

2.8 The Crystal Barrel Calorimeter

The Crystal Barrel calorimeter is the main component of the detector system. It surrounds the target in a barrel-like arrangement (see Fig. 2.11) of 1320 Tl-doped CsI crystals with openings in the forward and backward direction, covering the whole 360° in the azimuthal angle Φ and the range of 11.2° to 156° in the polar angle θ.
It consists of two halves which can be separated. The half closer to the tagger, i.e. "up the photon stream", is canonically called upstream (US), the other being the downstream (DS). The crystals are arranged in 24 rings (13 in the downstream, 11 in the upstream). In the central part of the calorimeter, each ring consists of 60 crystals, each covering $6^\circ \times 6^\circ$ in azimuthal and polar angle. The exception are the four smallest rings (three in the downstream, one in the upstream) which consist of only 30 crystals, with each crystal covering 12° in θ. The crystals themselves are 30 cm ($\approx 16X_o$) long and have the shape of truncated pyramids with a trapezoidal base suited for the barrel-like arrangement. While all other crystals are facing the target in the center of the detector, the forwardmost 3 rings have been shifted 3 cm along the beam axis [Fun08] and are therefore slightly misaligned. The reason for this shift is that the scintillation light from the Forward Veto Plates (see section 2.5.2) has to be routed out of the barrel to 16-channel PMTs (see Fig. 2.11). In addition, the first 3 rings used to have a different readout than the rest of the calorimeter (so called "Forward Plug"), which explains the differing holding structure in this region. They were

read out by PMTs instead of the otherwise used PIN-diodes[18] in order to obtain a faster readout and to be able to include these signals into the trigger (see section 2.9). Nowadays, after the latest upgrade [Hon15], the readout is the same for each crystal. In general, large parts of the holding structures are internal (shown in red and green in Fig. 2.11), which means that particles have to traverse 1 to 2 cm[19] of aluminum before entering the crystals. Implications of this will be discussed in section 6.5.
The crystal length of $\approx 16 X_0$ ensures, that (almost) the entire energy of an electromagnetic shower is contained within the detector, which makes the Crystal Barrel detector – together with the high light yield of CsI(Tl) – ideally suited for the measurement of photons. Its energy resolution can be well approximated by

$$\left(\frac{\sigma_E}{E}\right)^2 = \left(\frac{2.39(5)\%}{\sqrt[4]{E/\mathrm{GeV}}}\right)^2 + \left(\frac{0.46(8)\%}{\sqrt{E/\mathrm{GeV}}}\right)^2 + \left(\frac{0.048(7)\%}{E/\mathrm{GeV}}\right)^2$$

where the first term is purely phenomenological and dominates the total error. The second term has its origin in statistical fluctuations in the shower formation and the third term is related to the (energy-independent) electronic noise in the readout [Hon+23].
Because a typical electromagnetic shower extends over several crystals[20], the center of gravity of a cluster can be determined with a sub-crystal accuracy. The corresponding angular resolution is better than 1.75° for photons over 200 MeV which are not hitting any detector boundaries [Got+21]. The resolution improves for higher photons energies, but is fundamentally limited by the finite target size (2 cm) along the z-axis. Because tracks are implicitly reconstructed towards the center of the target cell, reaction vertices further up- or downstream lead to inherent deviations of up to 2° for $\theta = 90^\circ$ between the reconstructed and the actual angle [Mül18].

2.8.1 Readout

The CB readout electronics underwent several changes during the last upgrade [Urb18] [Hon15] [Mül19]. Only the current state of the readout will be discussed.
Fig. 2.12 shows an overview of the entire signal chain: The scintillation light of the CsI(Tl) crystals is read out by 2 Avalanche Photo Diodes (APDs) each[21], which are glued directly onto the crystals surface [Urb18]. The signals are then integrated separately and amplified in charge-sensitive preamplifiers before the signals from both APDs are added and sent to the Buffer/Timing filter (BuffTi) [Hon15].
The BuffTi serves two main purposes: First of all, it has to actively split the signal into three paths. One path is dedicated to the precise determination of timing information and leads to the TDCs. Another path serves to provide energy information and leads to the newly established Sampling-ADC (SADC) readout system, which will be described in more detail in chapter 3. A third path connects to the former QDC readout system, which is also dedicated to the

[18] **P**ositive **I**ntrinsic **N**egative; refers to the doping of different regions in the diode
[19] The holding structure for the Forward Plug is thicker in front of the fourth ring and (geometrically) in the corner in front of the sixth ring [Grü22].
[20] As discussed in section 1.4, CsI(Tl) has a Molière radius of 3.53 cm, which means that an electromagnetic shower from a single photon usually extends over multiple crystals.
[21] 2 APDs are used for redundancy and to reduce the noise by a factor of $\frac{1}{\sqrt{2}}$ if both signals are added.

determination of energy. On the other hand, the BuffTi provides high-pass filtering for the signals in the timing branch, which is advantageous in shortening the pulses to generate fast signals. The signals in the two energy branches are left unfiltered at this stage, since both branches have their own dedicated hardware for signal shaping.

The QDC Readout In case of the still existing QDC readout, the shaping is achieved by an array of dedicated "CB-Shapers", which were developed in 1989 [Ess89]. The shaping is achieved by a combination of high- and low-pass filters, which manage to shorten the exponentially decaying preamplifier signal with a length of over 100 µs down to a roughly gaussian shape with approximately 4 µs Full Width at Half Maximum (FWHM). This mainly serves to avoid excessive pile-up in case of a high event rate. Without shaping, signals that were e.g. 50 µs apart would overlap and lead to wrong energy information.
The QDCs themselves are *LeCroy 1885F* 12 bit dual-range Fastbus-Analog-to-Digital Converters (ADCs). The working principle is that a capacitor gets charged for a given amount of time[22] and then the voltage of the capacitor is digitized as a measure of the deposited energy. It is clear that unless the baseline of the incoming signal sits at exactly 0 V, this measuring procedure will lead

[22] For all data taken in recent history, this gate length was set to 6 µs.

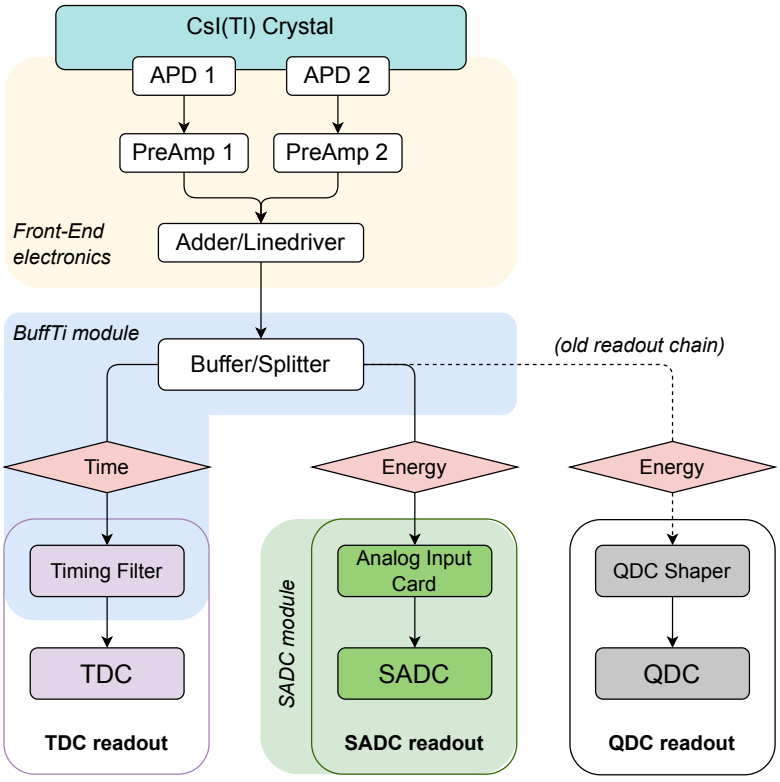


Figure 2.12: Schematic view of the readout chain of the CB calorimeter. The signal is generated in the CsI(Tl) crystals shown on top and then proceeds downwards. The background colors (ocher, blue and green) indicate spatial proximity, whereas the boxes at the bottom indicate functional groups.

to a non-zero voltage measurement at the capacitor, even if no actual signal was present. This so-called pedestal should be measured frequently to account for any potential drift and has to be subtracted from all of the output values.
One particular feature of this model of ADC is the use of a dual-range setup: The incoming signal is split by a voltage divider with a ratio of 8:1. Naively, the 12 bits ($2^{12} = 4096$) of resolution would, given a dynamic range of $\approx 2\,\text{GeV}$[23], correspond to only 0.5 MeV per ADC channel. But with the divider circuit, two separate capacitors can be supplied. Depending on the digitized value of the low-range capacitor, either its value or the value of the high-range capacitor will be used. Disregarding other effects, this results in an 8-fold improvement of the energy resolution in the lowest 8th of the dynamic range.

The TDC Readout The signals in the timing branch are read out by a Dual-Threshold Leading Edge (LE) discriminator whose output is sampled with very high precision in FPGA logic [Kla23]. The concept of a LE discriminator is simple: It detects the time at which a signal's amplitude first exceeds a given threshold. Assuming a constant signal shape and a finite rise time, this results in a so-called time-walk, i.e. the position of the extracted timestamp depends on the total amplitude of the incoming signal. To overcome this, the CB-TDCs feature two different thresholds which are canonically called "A" and "B". This dual-threshold information can be used to correct for this time walk [Sta19]. To account for different rates under different Θ-angles, both thresholds are increased for some rings in the forward (and backward) direction. Especially in the forward region, thresholds are deliberately higher due to the low-energetic electromagnetic background[24]. The values for the lower "A" threshold vary from 3.9 MeV to 11.3 MeV, while the "B" thresholds lie between 8.2 MeV and 23.3 MeV [Sta19].
The FPGA implementation which is used further possesses multihit capabilities, which means that it can record multiple timestamps per event. The input signal is sampled with a clock of 400 MHz, which results in a precision of 2.5 ns for the timestamps. By using a so-called carry-chain implementation, the time resolution can be substantially improved beyond this limit [Sta19] [Bie16].

2.9 Data Acquisition and Trigger

As presented, the CBELSA/TAPS experiment consists of many separate components. Each individual detector does not provide enough information to reconstruct physical events, so the data of all detectors has to be combined to get a complete picture of the reaction which occurred. There are in principle two ways in which a large experiment that involves many detector systems can be operated: In a triggerless experiment, data from all detector systems is continuously recorded and sent to a central unit that decides which events are worth writing to disk and which are not. In contrast, the CBELSA/TAPS experiment has a triggered readout. This means that data from all of the detectors is only actually recorded and saved to disk if certain criteria are met,

[23] The dynamic range of the QDCs varies over the detector. Due to the kinematics of a fixed target experiment, more particles of higher energy are expected in the forward region of the downstream. The dynamic range is therefore $\approx 2\,\text{GeV}$ only for the forwardmost rings and becomes gradually smaller for crystals in other parts of the detector; especially in the upstream half.
[24] This is necessary because the TDC timestamps are used for the generation of the trigger signal.

which saves network capacity. This requires the criteria, upon which the trigger decision is made, to be quickly available after the physical reaction took place. Because the energy determination takes time (the integration itself takes $\geq 6\,\mu s$ in case of the CB), the data from the aforementioned TDCs are used. At this point, the only type of information used is whether there was a hit or no hit at all[25].

Data Structure An *event* refers to the collection of all data recorded in the entire detector after a trigger signal was emitted. Data is usually recorded in chunks (typically on the order of 1 to 5 million events) called *runs*, where each run corresponds to one data file. The total amount data taken with the experiment is subdivided into so-called *beamtimes*, i.e. sets of several hundreds of runs, which provide the necessary statistics for the analysis of rare reaction channels. Each beamtime has its own set of parameters (target material, energy and polarization of the beam, trigger conditions, etc.) and usually takes multiple weeks to record.

2.9.1 Trigger settings

The CBELSA/TAPS trigger system is based on two main criteria: the multiplicity of hits and the information from the Cherenkov detector. Due to the segmented nature of the detector, most particles (in particular those which cause an electromagnetic shower) will deposit their energy over several crystals. In order to determine the actual number of detected particles as well as their entire energy deposition, those crystals have to be grouped together into so-called clusters, which is done by the *Cluster Finder* [Hon15] [Kla23].

Multiplicity Usually, at least two clusters in either one of the two calorimeters are required (2 in CB, 2 in MiniTAPS, or 1 in each). This reflects the expected topology of final states (also see section 2.10). Provided that a neutral meson was produced which decayed into 2 photons, one would expect 3 particles (the two photons and the recoil proton) in the final state. Only two instead of three clusters are required, because especially low-energy protons might get "stuck" in other detector components or insensitive material before they reach one of the calorimeters.

Cherenkov It was already mentioned in sections 1.3.2 and 2.5.3 that the hadronic reactions of interest do not necessarily have the highest cross sections; especially in the forward direction the production of e.g. e^+e^- pairs is orders of magnitude larger. If the trigger logic did not take this into account, electromagnetic reactions would easily constitute a majority of the recorded events, which would decrease the achievable rate of actually relevant events and unnecessarily occupy disk space.
The gas Cherenkov detector is ideally suited to achieve a near perfect rejection of fast electrons or positrons in the final state, provided that they were emitted under small enough angles to reach the detector. Such particles would then continue to hit the MiniTAPS detector, which means that the Cherenkov veto information only has to be checked if the MiniTAPS detector was involved.

[25] The existence of two different thresholds would theoretically allow for a rudimentary discrimination based on energy. Alternatively, the *energy sum* might be included in the future (see below).

Energy Sum It is foreseen to include a third criterion into the trigger decision in the near future: an energy sum [Ciu18] [Mit19]. This would facilitate to select only those events in which the entire energy deposited in the calorimeters is in excess of a given threshold. For example, it does not make sense to store events in which the total energy deposition was significantly below the mass of the lightest meson, the π^0, which has a mass of 135 MeV. Since this information needs to be quickly available, it must use signals from the timing branch. It should be noted that also without the energy sum included, the hit detection already relies on the energy deposition per crystal surpassing the TDC thresholds (see section 2.8.1).

Other Trigger Settings While the trigger conditions outlined above are used to record actual photoproduction reactions, it is possible to trigger on other scenarios for a number of different purposes. One example would be to require just a hit in the tagger in two adjacent bars (`tagger_or`) for cross-section determinations and as a cross-check for the flux measurements. For the scope of this thesis, two other settings are of particular interest:
In preparation for beamtimes (e.g. for calibration purposes) or for general testing and debugging, it can be useful to trigger on cosmic particles in so-called *cosmic runs*. Although the topology of events is different[26], these runs provide a possibility to easily test new features or components with sizable, realistic energy depositions in the crystals. Since cosmic particles are predominantly muons[27], which are (close to) minimally ionizing at typical energies, they can be used for a rough first energy-calibration of the readout.
Another possibility is to simply record events at a constant, predetermined frequency (*clock runs*), typically while the accelerator is not running. Significant energy depositions in such runs would be purely accidental.

2.10 Reconstruction of Physical Events

2.10.1 Clustering

The clustering process used for the trigger decision is quite primitive in nature and only counts the number of physically separated clusters. Once the data has been recorded, the reconstruction of physical events can make use of more elaborate algorithms to determine the size, total energy and center of mass of each individual cluster much more accurately. In particular, it might occur that the showers from individual particles (Particle Energy Deposits (PEDs)) overlap and form a single cluster. By using the energy depositions in every involved crystal, it is possible to disentangle such clusters. For practical reasons, only energy entries above a certain threshold (default: 1 MeV) are considered. In addition, at least one crystal has to exceed an energy threshold of 20 MeV. Consecutively, the energies of all crystals contributing to the PED are summed to give an estimate of the total energy of the particle which caused the initial energy deposition. To account for the transverse shower profile, the center of mass of the PED is calculated as a weighted average of the geometric midpoints of all individual crystals [Sei09].

[26] Typically only 1 particle which traverses the detector from the outside instead of multiple particles originating from the target.

[27] Many different particles are created from the primary cosmic rays, but in the lower parts of the atmosphere nearly all remaining particles are muons.

Traditionally, the timing information from the TDCs was not used in this clustering process. In turn, this entails that the information of different particles which hit the detector at different points in time might be erroneously combined into one cluster. Recently, a *Time Clustering* algorithm was introduced [Sta24] which makes use of the time information as well; both from the TDCs and (for lower energies) from the SADCs (see section 3).
Subsequent to this clustering process, the information from the charge sensitive detectors is combined with the clustered data. In doing so, each cluster is assigned a charge information, which can either be *charged* or *uncharged*. The sign or magnitude of the particles charge cannot be determined with the current setup.

2.10.2 Reaction Selection

Since most of the relevant particles/resonances which are produced in photoproduction reactions are unstable, one usually detects not the particles themselves but their decay products. This is especially true for light neutral mesons like the π^0 and the η and η' mesons, which all decay predominantly to final states comprised of several photons. As simplest possible example, consider the following reaction:

$$\gamma + p \rightarrow R^* \rightarrow p\pi^0 \rightarrow p\gamma\gamma \tag{2.2}$$

In order to fully reconstruct such a reaction from the final state, 3 PEDs are required: 2 uncharged ones (from the two photons) and 1 charged PED (from the proton). However, the charge information might be faulty or can be deliberately ignored. In that case, all combinatorical possibilities are considered and the selection of the "correct" combination is done by the kinematic cuts alone (see below).
As already discussed, low-energy protons might get stuck or lose too much energy before they reach the calorimeter crystals such that their resulting energy deposition is insufficient to be recognized as a cluster (below 20 MeV). Even for protons with higher energy, the measured energy is unreliable, because the calorimeter is calibrated for photons (see section 1.4.4) and protons can punch through the crystals if their energy is sufficiently high (above $\approx$ 450 MeV [Sei21]). So the reconstructed *energy* of a (suspected) proton PED should not be used. However, the reconstructed *direction* remains usable. Even more so, the direction from the inner detector and/or the Forward Veto plates can still be used, even if the proton was not registered in the calorimeter itself[28].
In addition, the reaction is actually over-determined, since the initial state is known due to the tagger which provides the photon energy. Because the direction of the beam-photon (conventionally along the z-axis) as well as the 4-momentum of the target proton (approximately at rest) are given by the setup, this is all what is needed to completely determine the initial state. This assumes, that the correct beamphoton was identified. The typical tagger rate is around 10 MHz [Har17], so there remains a significant background from uncorrelated beam photons, even if the cut on reaction time (see below) is applied.
This over-determination entails that the final state baryon does not need to be detected in order to fully reconstruct a reaction. Instead, a "missing baryon" can be constructed from the other data,

[28] so-called "half-PEDs".

which also works if the final state contains a neutron instead of a proton.

2.10.2.1 Cuts

The 4-momentum of every PED is constructed by taking the total energy deposition as energy and assuming that the momentum vector is pointing from the target (the probable interaction point) to the weighted center of energy deposition of the PED. Under the implicit assumption that the detected particle is massless, i.e. $|\vec{p}| = E$, an invariant mass can then be reconstructed from any pair of the so determined 4-momenta.
In addition to the overall multiplicity and charge information (see above), a series of cuts is applied to further select reactions which match the respective final state under investigation:

- **Reaction time:** $\Delta t_{\text{reaction}} = t_{\text{fs}} - t_{\text{b}} \overset{!}{\approx} 0$
  An average time t_{fs} is constructed from the individual energy depositions in the final state. This time should correspond to one of the tagged beam photons t_{b}. Otherwise final and initial state would not correspond to each other. Due to the high multiplicity of beam photon candidates for a typical event, there will always remain some contribution from uncorrelated beam photons which randomly occurred at the right time. Although it is impossible to tell the "correct" and the "wrong" beam photon apart solely based on their timing information, their effect on the data can be mitigated in form of a **sideband subtraction:** In addition to the signal sample, another sample of certainly uncorrelated beam photon candidates[29] is selected and subjected to the same cuts as the signal sample. Whatever remains in the end is then subtracted from the signal sample with an according weight.
- **Invariant mass** (of the reconstructed meson): $m_{\gamma\gamma} \overset{!}{\approx} m_M$
  One[30] of the pairwise combinations of final state 4-momenta should match the mass of an actual meson (for example $m_{\pi^0} = 134.977\,\text{MeV}$ or $m_\eta = 547.862\,\text{MeV}$).
- **Missing mass:** $m_{\text{mb}} \overset{!}{\approx} m_{p,n} \approx 940\,\text{MeV}$
  The invariant mass of the calculated missing baryon should match the actual nucleon mass.
- **Coplanarity:** $\Phi_p - \Phi_M \overset{!}{\approx} 180^\circ$
  Because the reaction is constrained to a plane along the beam direction (z-axis), the azimuthal direction of the recoiling proton (p) and the reconstructed meson (M) should be exactly opposite of each other.
- **Polar angle difference**: $\Theta_p - \Theta_{mb} \overset{!}{\approx} 0^\circ$
  In case the proton was detected, its direction can be compared to the calculated direction of the missing baryon.

Specific cut limits around the expected values are deduced for all of these quantities to ensure, that most of the remaining data sample actually corresponds to the reaction under investigation.

[29] to the left and right of the peak around $t_{\text{reaction}} = 0$, hence the name "sideband".
[30] Or more, in case of more complicated final states such as $p\pi^0\pi^0 \to p4\gamma$

The cut widths (typically around 2–3 σ of the respective distribution[31]) have to be carefully determined in order to balance high purity and low background contamination of the sample on the one hand and a sufficiently large number of events (small statistical error) on the other hand.

[31] In case of a non-gaussian distribution, the determination of a "σ" value is not trivial.

CHAPTER 3

The CB-SADC Readout System

The hardware design of the CB-SADC board is based upon an early prototype, which was developed by P. Marciniewski for the PANDA[1] experiment [Mar12]. Since 2014, the lines of development for the CB-SADC and the PANDA-SADC have diverged, but continued in parallel. While the latter one is now suited to the particular needs of the PANDA experiment, the CB-SADC was adapted for CBELSA/TAPS in the scope of the PhD thesis of J. Müllers [Mül19]. The following pages will give a short summary of its main hardware and firmware features as well as some related developments.

3.1 The Use of Sampling-ADCs

As most macroscopic quantities in nature, signals coming from a detector are typically analog[2]. This means, the output is continuously varying over time. An ADC is any device which converts an analog input signal (typically a voltage) to a digital output, which is necessary in order to be processed using computers. In case of the CBELSA/TAPS experiment, three main types of ADCs (illustrated in Fig. 3.1) are used:

- **Time-to-Digital Converter** A TDC outputs a single value which is proportional to the time passed after a given reference point in time until the input signal fulfilled a given condition. In its simplest form, the input voltage has to exceed a given threshold in order to output the corresponding time (so-called leading-edge discriminator; see also section 2.8.1). In case of a *multihit* TDC, more timestamps are emitted if the condition arises again.

- **Integrating ADC** An integrating ADC (in case of CBELSA/TAPS: QDC) outputs a single value which is proportional to the integral of the input voltage within a predefined gate (see also section 2.8.1). In case of typical scintillator signals, the integral of the voltage pulse is proportional to the deposited energy (see section 1.4). If the baseline of the signal sits at a non-zero voltage, the so-called pedestal (integral without any pulse) has to be subtracted in order to get the integral of the pulse itself.

[1] antiProton ANnihilation at DArmstadt

[2] Assuming the signals are sufficiently large. Otherwise, quantum effects might become relevant, which are distinctly not continuous, but discrete!

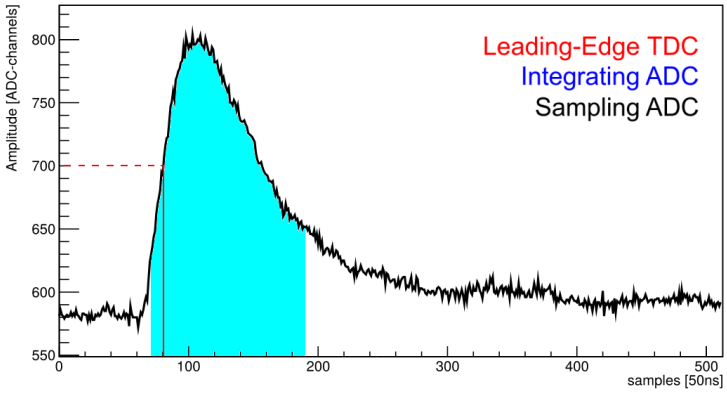


Figure 3.1: Working principle of a sampling ADC (black line). The incoming signal from a crystal of the CB calorimeter is sampled at N (512) points in time, at a constant rate (20 MHz). The operating principles of a leading-edge TDC and an integrating ADC are illustrated in red and blue, respectively.

- **Sampling ADC** An SADC samples the input signal with a given sampling frequency f_S and records the signals momentary amplitude A for N equally spaced points in time. In doing so, the continuous input voltage $V(t)$ is *approximated* by a series of N discrete values $A_0, A_1, A_2, ..., A_{N-1}$.

In order to have all of the necessary information included in the waveform, it is important to choose a suitable sampling rate and bandwidth. If the sampling rate is lower than double of the highest expected signal frequency, information will be lost[3]. On the other hand, sampling with a rate much beyond that also offers no additional benefits and only consumes more ressources.
In combination with a Field Programmable Gate Array (FPGA) for real-time processing of this data, this opens the door for a multitude of possibilities: Not only can the behavior of a QDC and TDC be recreated, but also entirely different, custom features can be extracted. For example, an individual baseline level can be determined for each event instead of a (run-based) pedestal value. In addition to that, the availability of the entire waveform facilitates the detection as well as the eventual correction of pile-up.
As a part of the recent upgrade, the former QDC readout was first complemented and later superseded by a faster, more versatile, modern SADC readout.

3.2 Hardware

Fig. 3.2 shows the CB-SADC board in its entirety. On the input side (right), it features $2 \times 32 = 64$ input channels. Every 8 adjacent channels are read out by one 8-channel LTM9009-14 ADC, which samples the incoming signal with a maximum sampling rate of 80 MS/s[4] at a resolution of 14 bit. The data from these ADCs is received by two Xilinx Kintex-7 FPGAs, which also control the behavior of these ADCs, provide network communication, perform feature extraction,

[3] This is known as the Nyquist-Shannon sampling theorem [Nyq28] [Sha48].
[4] Mega-Samples per second. Technically, the sampling rate might also be seen as a frequency of, in this case, 80 MHz.

and much more (see section 3.3 (Firmware)). Each FPGA is responsible for four of the eight ADCs, which correlates to the initial grouping of the input channels into blocks of 32. On the output side, two Gigabit Ethernet connectors enable the transfer of the extracted data to other parts of the Data Acquisition (DAQ) infrastructure. Among the adaptations for CBELSA/TAPS was the development of a dedicated power supply, a custom heat sink and the provision of further connectors for the integration into the existing trigger system via an extension card.
Similar to the QDC readout (see reasoning in section 2.8.1), it was chosen to configure the hardware for a dynamic range of 2.5 GeV in the downstream and half of that in the upstream[5]. As the calibration revealed later, the actual dynamic range came out to be closer to 2.2 GeV (DS) and 1.1 GeV (US) (see section 4.3.2).
Within the scope of several bachelor and master theses, additional hardware was developed. T. Poller developed an input card (r.h.s. of Fig. 3.2) for analog preprocessing of the signals during his bachelor thesis [Pol16] and beyond. Y.C. Wang designed a PCB and testing procedures to aid mass-testing of the assembled modules [Wan18]. Finally, J. Knaust designed a crate controller which allows to control up to 12 SADC modules via their front connector [Kna19]. Since the input card is most relevant for this thesis, it will be discussed in more detail in the following section.

[5] The ADC chips provide the opportunity (by soldering one resistors per eight channels) to perform an internal amplification of a factor of 2.

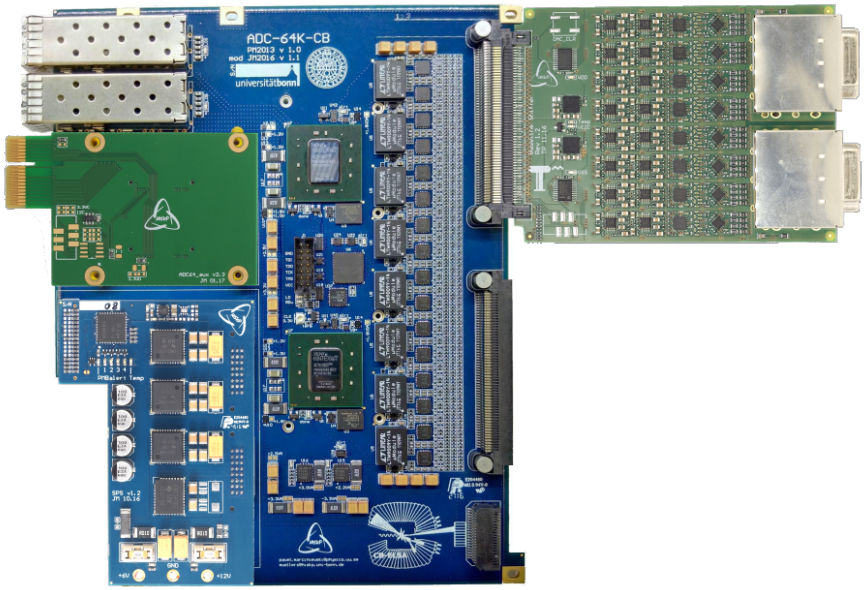


Figure 3.2: The CB-SADC module. The actual CB-SADC board [Mül19] is shown on the left and has a blue coating. One module requires two analog input cards [Pol16] with 32 channels each, which get plugged into the right hand side of the board. Only one card is shown in the upper slot. On the actual CB-SADC board, the two FPGAs are visible near the center. To their right, an array of 8 SADC chips can be seen. On the left hand side of the card (top to bottom), two Small Form-factor Pluggable (SFP) Ethernet ports, an extension card (green) and the (removable) power supply module (blue) are situated.

3.2.1 The Analog Input Card

Looking back at the readout chain discussed in section 2.8.1, both the TDCs as well as the QDCs have their own dedicated hardware for signal shaping. In case of the SADCs, this is achieved by the analog input card. But shaping is not its only purpose, it also provides functionalities to shift the signal baseline and to perform an adjustable Pole-Zero Cancellation (PZC). Both will be explained in the following.

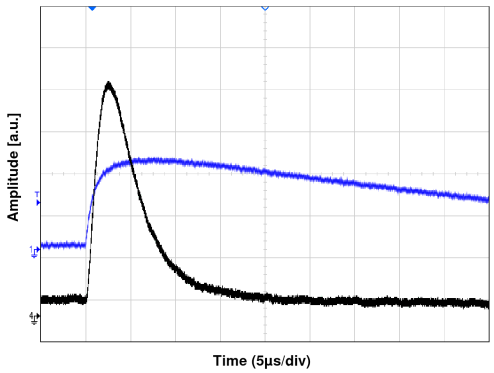


Figure 3.3: Effect of the shaping (black) on a recorded preamplifier signal (blue) [Mül19].

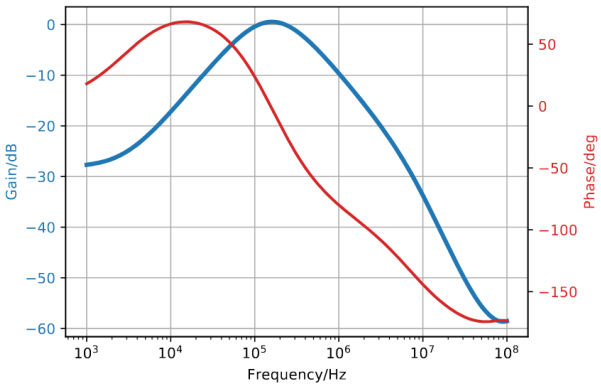


Figure 3.4: The frequency response of the shaper (blue) together with the phase shift (red) [Mül19].

3.2.1.1 Shaping

As with the QDC shapers, the goal of the SADC shaper is to shorten the very long preamplifier signals. The shaping is achieved by a combination of high- and low-pass filters. Fig. 3.3 shows the result. The long exponential tail with $\tau \approx 54\,\mu s$ is substantially shortened and the resulting pulse has a remaining FWHM of only $\approx 5\,\mu s$. While a simple high-pass filter might in principle be sufficient to shorten the pulses, the usage of an additional low-pass filter serves to increase the Signal-to-Noise Ratio (SNR). Since the preamplifier signal has no meaningful contributions above 1.5 MHz (see Fig. 3.5), frequency contributions above this value can be considered noise and are advantageous to remove. Therefore, the frequency response of the shaper has a broad maximum around 160 kHz and drops continuously from there on (Fig. 3.4). This leads to the loss of some of the higher frequencies that would have contributed to the signal, resulting in a slightly flatter rising edge of the shaped signal. But since analog filters cannot provide arbitrarily steep edges of the frequency response, these filter settings were found to be a good compromise between noise suppression and pulse shape retention.

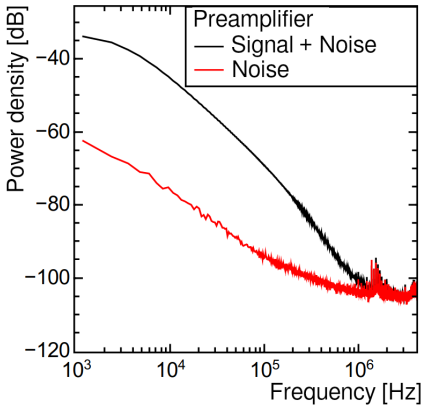


Figure 3.5: Power spectral density at the output of the preamplifier stage for cosmic particles ($E_{\text{deposited}} \approx 30$ to 50 MeV). Taken and translated from [Hon15].

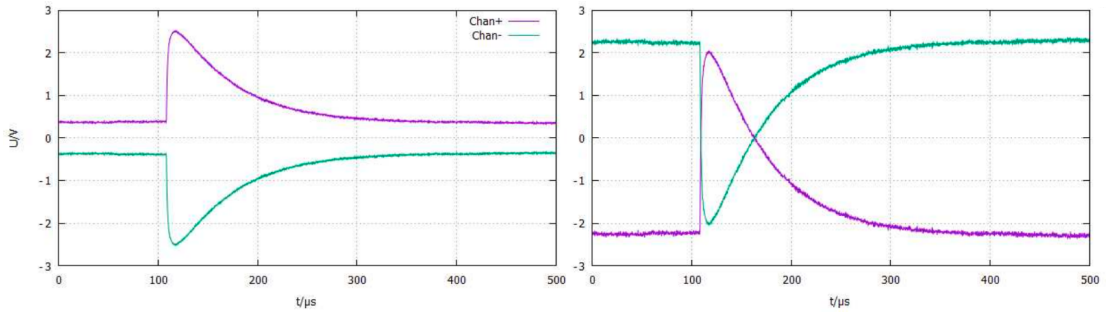


Figure 3.6: Differential transmission concept used to minimize power dissipation (left) and after the baselines were shifted to utilize the full dynamic range of the SADC chip again (right). [Pol16]

3.2.1.2 Baseline-Shifter

To avoid distortions of the electric signals, the transmission from the BuffTi to the SADCs is differential in nature. In contrast to ordinary, single-ended transmission, the signal is split between two adjacent conductors and inverted on one of the two. Subtracting the inverted signal from the other gives back the original signal. Any external distortion would act on both conductors equally and is thus automatically canceled by the subtraction. However, to minimize the (average) voltage difference between both conductors and thus the power dissipation at the terminating resistors, the implementation was modified in a way that the idle voltage of both conductors lies close to zero (see l.h.s. of Fig. 3.6). That way, the voltage difference is only large in the (relatively) rare case of an actual signal on the conductors.
The SADC chips are also set up to receive differential signals, but can handle both positive and negative voltage differences. If the transmission is implemented as described, only positive differences will occur, which would effectively mean that only have of the available dynamic range would be used. Therefore, it was the main objective to overcome this situation when the first revision of the analog input card was developed [Pol16].
Without going into too much detail, it was solved by adding a variable offset voltage to the input signal (see r.h.s of Fig. 3.6). In the end, this results in a shifted baseline at a negative voltage, which (after digitization) should ideally be close to ADC-channel zero. It is preferable to place the baseline not precisely at ADC channel zero, but leave enough of a safety margin such that potential underswings[6] or small voltage drifts do not lead to clipping[7]. To make the baseline shift adjustable, the voltage which is added/subtracted from the input signals is given by a Digital-to-Analog Converter (DAC) value that can be remotely set. A further bachelor thesis [Bie18] was, in part, dedicated to determine suitable, channel-specific DAC values under these constraints.

[6] A signal exhibits a positive pulse, but afterwards dips below the baseline value from which it started from.
[7] Clipping describes the situation when a signal exceeds the dynamic range of the ADC in either direction (too high or, in this case, too low).

3.2.1.3 Pole-Zero Cancellation

Without appropriate countermeasures, the shaped (positive) signal would exhibit a long, negative underswing (Fig. 3.7 provides an illustration). Although the theoretical description is much more complicated (see e.g. [Mül19]), this is fundamentally due to the presence of high pass filters in the shaping circuit. Electronically, this means that the output voltage has to pass through a capacitor (C_{diff}) at some point, which is effectively an insulator at low frequencies. This entails that there is no direct connection by which any Direct Current (DC) component could pass through the filter (so called "AC-coupling"). In other words, every positive pulse must be followed by an underswing of equal but opposite integral below the baseline. This is mitigated by adding some part of the input voltage back onto the output signal via a resistor (R_{pz} in Fig. 3.7). Some part of the DC voltage can flow through this resistor and prevent the underswing. In case of an exponentially decaying signal, this cancellation can be exact, provided that the correct resistance was chosen. In this case, the resistor was realized as a digital rheostat. Similar to the adjustment of the baseline, this also requires some optimization in order to find the ideal DAC values for this potentiometer. While the optimization was the other main goal of [Bie18], it was also improved and implemented in a more efficient manner during this thesis (see section 6.4).

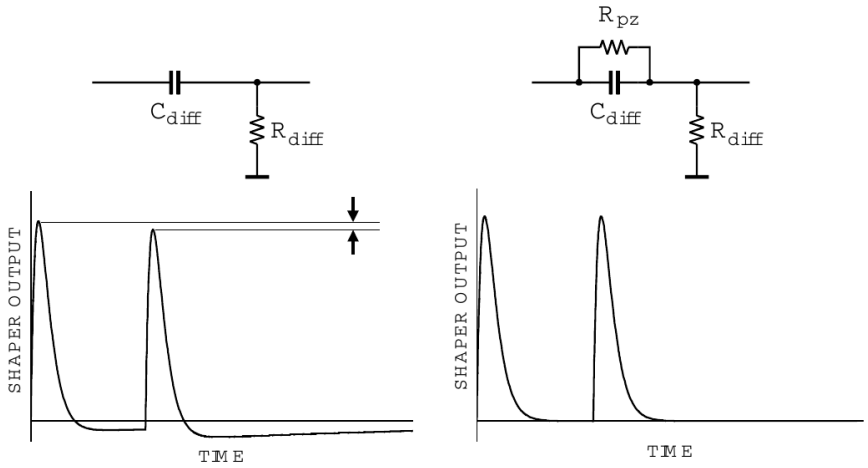


Figure 3.7: Functional sketch of a pole-zero cancellation circuit. Without PZC (l.h.s), the signal exhibts an underswing which reduces the measured amplitude of subsequent signals. By introducing a resistor R_{pz} (r.h.s), the underswing is cancelled. [Spi08]

3.3 Firmware

An FPGA is fundamentally just a collection of many different logic gates and some specialized building blocks. Without an appropriate configuration which specifies how these individual components should be connected and interact with each other, it is useless. In contrast to an Application-Specific Integrated Circuit (ASIC) in which the connections between the individual components are fixed in hardware, the behavior of an FPGA can be adapted to a specific task. Much like "ordinary" multi-purpose computing devices like a Central Processing Units (CPUs), for which software can be developed in a variety of different high-level programming languages, the configuration for an FPGA is called firmware. Such firmware is usually written in a language which is much more reflective of the actual hardware operations; these languages are generally referred to as Hardware Description Languages (HDLs). Among them, the Very High Speed Integrated Circuit HDL (VHDL) is one of the most widely used and is also the language in which the firmware for the CB-SADC is written. Most of it was developed by J. Müllers within the scope of his doctoral thesis [Mül19], but recently B. Otto [Ott22] has made substantial contributions as well.
A timeline of some key advancements made over the course of this thesis is compiled in Table 3.1. Most of these developments will be specifically addressed in sections 4 and 5.

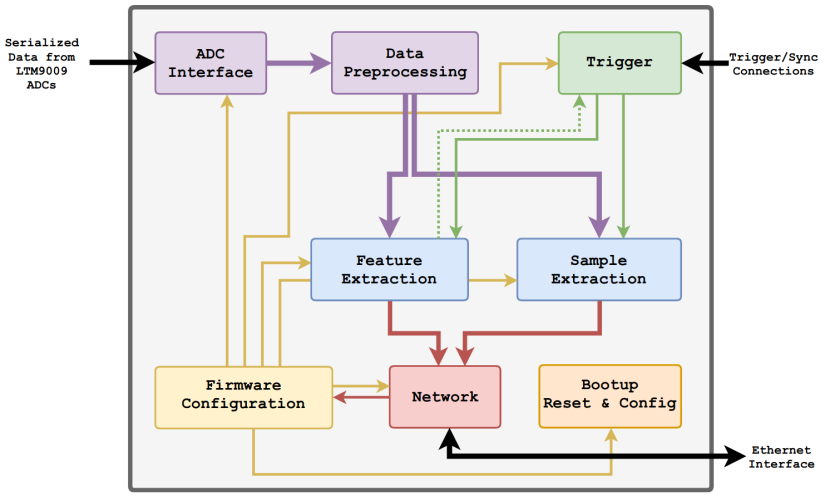


Figure 3.8: Modular structure of the CB-SADC firmware. Adapted from [Mül19].

Fig. 3.8 provides an overview of the general layout of the CB-SADC firmware in the form of different functional blocks (modules). Several aspects will be explained in more detail later, but the overall concept can be summarized as follows: The ADC chips provide a constant stream of sampled data. The data from one ADC chip consists of 8 channels each, which are sampled with 14 bit resolution at 80 MHz. It would be inefficient to have a dedicated electrical connection for each of these 14 bits, so instead the data is serialized, i.e. transmitted over a single connection, but at a substantially higher frequency of 640 MHz[8]. It is the purpose of the `ADC Interface` to receive and de-serialize this data. From there on, the data is fed to the `Data Preprocessing`

[8] Technically: A Double Data Rate (DDR) transmission at 320 MHz.

module. The data is decimated from the native 14 bit at 80 MHz down to a more sensible value of 16 bit at 20 MHz within this module. As a recent addition by B. Otto, the preprocessing module now also includes a Finite Impulse Response (FIR) filter [Ott22]. More on FIR filters and digital filtering in general can be found in section 6.2. From there on, the stream of data is passed onward to the `Feature Extraction` and `Sample Extraction` modules. The purpose of the latter is to structure the raw sampled data points into appropriately sized "waveform packets", which can then be sent to a receiving device via the `Network` interface.
In addition, different features are extracted from the waveforms, which serves to substantially reduce the amount of data, while still containing all relevant information (see section 3.3.2 (Mode of Operation)). The feature extraction will be discussed in more detail later, but to give a quick overview, there are 5 features which are extracted. In comparison to the existing QDC/TDC readout, three of these features (baseline, peak height, integral) are related to the measurement of energy, whereas the other two features (peak time and CFD time) fall into the category of time measurement. The algorithms behind these features are always running while sampled data is received, it is only upon arrival of an external `Trigger` signal, which is itself sampled with 160 MHz [Ott22], that the values of these features are determined for the event at hand and compiled into a "feature packet". A single feature packet contains data from all 32 channels connected to an FPGA. Additionally, the feature extraction module includes algorithms to detect pile-up (overlapping pulses). Whether or not pile-up was detected also influences how many waveforms are actually sent over the network (see section 3.3.5). On the receiving end of the connection sits a so-called Local EVent Builder (LEVB), which is a specialized program running on a normal computer/server. Its primary purpose is to receive the packets sent from all FPGAs and to format and/or compress the data in a way which enables them to be written to disk by the so-called saver. Since the data is transferred via User Datagram Protocol (UDP)[9], the LEVB also performs consistency checks on the data, and acknowledges the arrival of packets back to the SADCs.
Most of the parameters which determine the concrete behavior of each module can be configured by sending UDP packets of a specific format to the SADC. These parameters are parsed by the `Firmware Configuration` module and appropriately applied to the other modules.

3.3.1 Decimation and Downsampling

The expressions "decimation" and "downsampling" both describe similar concepts and are sometimes used interchangeably. The objective is in both cases to reduce the sampling rate of an already digitized input signal. As previously stated, it makes no sense to sample at frequencies much higher than double of the expected maximum frequency of the actual signal of interest. The most straightforward way to achieve this would be to simply use every N^{th} sample and just ignore the others. Within this thesis, this will be referred to as downsampling. Alternatively, every N samples may be replaced by their respective sum (here called decimation). While the former approach is a viable option, the latter has the added benefit of improving the SNR of the signal as

[9] In contrast to e.g. the Transmission Control Protocol (TCP) standard, UDP lacks any sort of "handshake" or acknowledging mechanism. This means that neither the arrival of packets per se, nor the preservation of their respective order is guaranteed by the standard. On the upside, it requires much less resources and overhead and is therefore easier to implement; especially on an FPGA. Depending on the implementation, the integrity of the data itself (header and payload) is verified by a checksum (so-called "frame check sequence (FCS)").

Month	Run no.	Modification
May 2018	201475	First production beamtime with SADCs
Oct 2018	203132	Zero-suppression in LEVB introduced (paket-wise)
Apr 2019	206059	Zero-suppression of individual channels in LEVB
Oct 2020	206215	CFD bugfix no. 1 (severe)
Apr 2021	206395	Full SADC coverage achieved
May 2021	206598	Pile-up unflagging in LEVB introduced
Jun 2021	206735	Improved baseline algorithm (box algorithm) implemented
Jul 2021	207838	CFD-Interpolation added + further (moderately severe) CFD bugfixes
Mar 2022	209511	FIR Filter introduced
	209515	Added more (192) samples in buffer before trigger
	209545	Improved version of peakfinder (accounting for “plateau”)
	209548	Zero-suppression now directly in FGPA
	209550	Multihit capabilities for CFD and Peakfinder introduced
	209568	Range of timestamps extended to negative values (before trigger)
	209626	Improved baseline algorithm (8-fold version v3)
Apr 2022	209631	Improved pile-up detection algorithm (LUT-based approach) in FPGA
	209633	Use multihit features for pile-up detection; check “relevance” of every peak
	209653	Added capabilites for flexible downsampling and number of samples
May 2022	209667	FPGA now sends already compressed waveforms
	209668	Introduced threshold for absolute peak height

Table 3.1: Timeline of selected key improvements to the SADC firmware and the LEVB. Changes affecting specific features were marked by color. Blue stands for changes related to the baseline-algorithm, red for the pile-up detection algorithm, green for the CFD-time and orange for the peakfinder algorithm.

well, which is why it is used in the firmware. To see this, suppose that every individual sample s^i consists of both an actual signal s^i_{sig} as well as some uncorrelated contribution of random electrical noise s^i_{noise}. By summing up N values, the relative noise contribution then grows with $\sqrt{N}$, since noise contributions are inherently uncorrelated. The actual signal amplitude on the other hand grows (if similar values are being added) by approximately N. This results in an increase of the SNR by $\sqrt{N}$. Details can be found in [Mül19].
Fig. 3.5 clearly shows that the signal itself contains significant frequency components only up to ≈ 1.5 MHz. Considering this, a sampling frequency of 4 to 5 MHz would seem appropriate, which would require 16 or 20-fold decimation.
On the other hand, the energy-sum signals (see section 2.9.1) are sampled with the SADCs as well. Since these signals come from the timing branch, they are considerably steeper and shorter than the normal signals in the energy branch. To retain a well enough sampling of these signals and to be conservative in general[10], a 4-fold decimation was chosen as the default. This has the added benefit of converting four of the initially 14 bit wide values to one 16 bit value, which coincides with the basic integer data type of `unsigned short` and as such eases efficient storage for later analysis.

[10] It is trivial to decimate the signal further at a later stage, but impossible to undo a decimation

3.3.2 Mode of Operation

In principle, all relevant information is contained in the sampled waveforms. But sending and saving the entire waveform (512 16-bit values) for every event from every channel (1320 crystals in CB) would be highly inefficient and produce enormous amounts of data with very little benefit, since only a few dozen crystals per event usually see any significant energy deposition at all (see section 4.3.3). Waveforms from the other channels contain no relevant information.
While the use of an SADC *allows* the user to extract the full waveform if needed, this is in most cases unnecessary. The relevant information content is usually limited to the amount of energy which was deposited and the time at which this deposition occurred. So it is much more economical to reduce the information within these waveforms down to a set of carefully chosen features, both in terms of disk usage and network traffic. In case of the CB-SADC readout, there are 5 different features (also 16-bit values), which naively[11] equates to a reduction of a factor of over 100.
However, in case of pile-up, i.e. overlapping signals from two different hits/particles, the information content of the full waveform is not easily reducible down to these 5 features. Depending on the way each individual feature extraction algorithm operates, the features for such a waveform might be either completely wrong, or at least not correspond to the correct[12] hit, but to an earlier or later one. In that case, saving just the extracted features would be insufficient to reconstruct the correct features; the full waveform is needed. To that end, a pile-up detection algorithm was implemented (see section 3.3.4). If pile-up is detected, the entire waveform of this crystal will be sent to the LEVB and stored on disk to facilitate offline[13] pile-up recovery.
In addition, it is important to save some waveforms regardless of pile-up in order to check whether the feature-extraction and pile-up detection algorithms are working as intended. For that reason, it was decided to save the entire waveforms of every channel for every N^{th} event (so-called "prescaler" events). For most of the data taken during this thesis, this was every 100th event, which attempts to find a compromise between having a representative data base and disk space usage. In recent months, with increasing trust in the performance of the employed algorithms, this value was occasionally reduced down to every 1000th event.

3.3.3 Feature Extraction

As mentioned already, the firmware and especially the feature extraction algorithms have undergone a number of changes and improvements during the work on this thesis. Several of these improvements were initiated and accompanied by findings from this thesis and developed in close collaboration with J. Müllers and B. Otto. Table 3.1 illustrates some of the progress.
It seems appropriate to start the discussion of the algorithms used with their respective state as of May 2018, and introduce improvements as they came. At this point in time, the algorithms used in the FPGA-internal feature extraction were mostly developed upon initial ideas by J. Müllers and tested and improved during the course of my master's thesis [Sch16]. The algorithms at

[11] Disregarding a possible compression of the waveforms, see section 3.3.5.1 (Compression) as well as the storage of features in a potentially compressed format.
[12] The one which is correlated to the trigger.
[13] "Offline" refers to an analysis which is not performed on the FPGA and is carried out after the data was already taken.

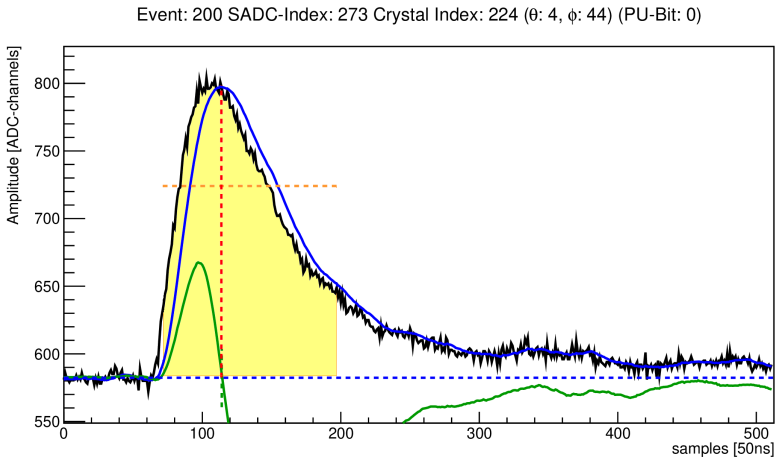


Figure 3.9: Example of features extracted by FPGA. The raw, digitized signal is shown in black; the reconstructed 16-fold moving average in blue (solid line) and the CFD-signal in green (solid line). The dashed blue line denotes the baseline, the red dashed line the peakheight and -position and the green dashed line the position at which the CFD zero-crossing was found. The yellow area illustrates the range in which the integration is performed, with the dashed yellow line representing the actual integral feature (integral divided by 128). The external trigger was detected at sample number 75.

this stage of development are described in detail in J. Müllers' thesis [Mül19]. This section is meant to provide a brief summary of the algorithms and the ideas behind them in order to define a starting point upon which the following improvements (discussed in section 4) are based. The relevant details of the specific implementations in the firmware will be elaborated when needed.

Nomenclature In order to ease understanding and avoid ambiguities, the following nomenclature is used: The entirety of datapoints sampled from a *signal* going into the SADC is called a *waveform*. Such a waveform can contain any number of *pulses*, corresponding to (in the simplest case) individual peaks in the signal amplitude from various energy deposits in the crystal. *Pulseshape* refers to the actual shape of one such pulse; usually normalized to unity. The collection of both the sampled waveform and the extracted features as well as data from other digitizers (TDC and/or QDC) for one specific crystal in an event shall, for the lack of a more descriptive word, be referred to as a *crystal hit*.

Indexing Conventions Each crystal in the CB calorimeter is assigned a unique index by which it can be identified. These indices are constructed from a crystal's position in the detector. Since each crystal covers 6° in Φ, the azimuthal position can be indexed using numbers from $k_\Phi = 1$ to 60. Similarly, the Θ position is indexed from $j_\Theta = 1$ (forwardmost ring) to 24 (last ring). The overall crystal index is then calculated as

$$i_{\text{crystal}} = 60 \times \left(j_\Theta - 1\right) + k_\Phi$$

In the first 3 rings as well as the last ring, crystals cover 12° in Φ, the indexing is however kept consistent by omitting every second index. As a result, the 1320 crystals cover an index range

from 1 to 1440.
The SADC channels on the other hand are naturally clustered into modules of 32 channels per FPGA. Each FPGA sends its own network packets and is assigned a unique, static IP address. It is therefore convenient to assign an "SADC index" as well, which is based on the FPGA's IP address and the internal channel number (0 to 31). 24 SADC modules are employed in total, with two independent FPGAs each. This results in a total SADC index range of 0 to 1535. A few SADC channels were left empty or carry the signal of the energy sum (see section 2.9.1). The mapping between crystal indices and SADC indices is not trivial, but can be deduced and is available at all times. Whenever an "index" is mentioned, it refers to the crystal index, unless explicitly stated otherwise.

3.3.3.1 Moving Average

After digitization in the sampling-ADCs, the raw signal still carries (electrical) noise. In order to reduce these noise contributions, the raw signal is filtered before any further feature extraction is performed. A Moving Average (MA) is a very simple case of a digital filter to suppress high-frequency noise. Under the assumption of white[14] noise, calculating a moving average of width N effectively reduces the overall noise level by a factor of $\frac{1}{\sqrt{N}}$. The moving average which is used for most applications in the SADC feature extraction is by default 16-fold ("MA_{16}") and reduces the average noise amplitude by a factor of 4, while leaving the signal shape itself approximately unchanged (see Fig. 3.9). A more elaborate alternative digital filter is a so-called finite impulse response (FIR) filter[15], which will be discussed in more detail in section 6.2. It was implemented in the firmware in March 2022 and has since superseded the moving average entirely.
With the exception of the integral calculation, all further steps in the feature extraction are based on the output of the moving average filter. Due to practical reasons, the moving average value at a given point x is calculated as the average of the N points before (including the actual point), resulting in an effective shift of $N/2$ samples with regard to the unfiltered pulse shape. At the same time, divisions in the FPGA can only be realized efficiently if the denominator is a power of 2 (hence the 16-fold moving average). In that case, a division can be done by a simple bit-shift on the binary representation of the dividend. Any remainder of the division will be dropped. An offset of half of the average length (8 in this case) is added before the division to get correct rounding to the nearest integer instead of the next lowest[16].

3.3.3.2 Baseline-Calculation

The baseline is calculated as the value of the moving average delayed by 32 samples (1.6 µs) and is read out when an external trigger is detected. This is done to ensure that the baseline is determined just before the onset of a pulse. The baseline value needs to be subtracted from the integral and the peak feature in order to compensate for the nonzero quiescent level of the

[14] "White" denotes, in analogy to visible light, a flat frequency spectrum in which all frequencies contribute equally.
[15] Technically, a moving average is also a FIR filter, albeit its most simple realization.
[16] Example: $173/16 \rightarrow$ 0b10101101 $\gg 4 =$ 0b1010 $= 10$ but $(173+8)/16 \rightarrow$ 0b10110101 $\gg 4 =$ 0b1011 $= 11$, which is closer to $173/16 = 10.8125$

setup. This implementation is very trivial, but already gives the great advantage of being able to perform an event-based baseline subtraction.

3.3.3.3 Integration

In analogy to the established QDC readout, the integral of the signal serves as the main energy feature. The integral is calculated by summing up 128 datapoints, which corresponds to an integration window of 6.4 µs, and then dividing the result by 128 again. The integration starts (depending on the configuration) a few samples before the external trigger is detected. In firmware, this is realized by calculating a 128-fold moving average ("MA_{128}") of the signal which is evaluated 128 samples after the start of the integration window[17] In order to get a meaningful energy information, the baseline needs to be subtracted from this value (see 4.3.1) and some sort of energy calibration (see 4.3.2) has to be applied. Earlier studies [Sch16] have shown that there exists an integration length of ≈ 100 samples or 5 µs which optimizes the achievable energy resolution. As the nearest power of 2, a value of 128 samples was chosen instead. In addition, this value more closely resembles the integration gate length of 6 µs of the QDC setup. The ideal position of the integration window can be determined by choosing the position in a way that maximizes the integral value for pulses that are correlated to the trigger, i.e. most pulses. Doing so ensures the optimal SNR, since the total noise contribution for a given integration length should be independent of the presence of any pulses.

3.3.3.4 Peakfinder

In contrast to the other algorithms discussed here, the peakfinder algorithm delivers not one, but two features at the same time: the peakheight and the position at which the peak was found. These features are inherently coupled to each other and always refer to the same peak. The peakfinder works by constantly evaluating the current value of the moving average and comparing it to the value one sample prior. If the pulse is rising, a counter is incremented; if the pulse starts to fall off, it is checked whether the counter reaches or exceeds an adjustable value (default: 20). In that case, the value of the previous sample (i.e. the last one before the pulse started falling off) is taken as the peakheight and the respective number of samples after the trigger is stored as the peakposition. Like the other feature extraction algorithms, it is always processing incoming data in real-time, but values are only actually read out once a trigger signal was detected. The algorithm will always output the first peak it finds, if it finds additional peaks within the sampling window, those are ignored. This serves to ensure that the extracted features have the highest probability of being actually correlated to the trigger. Multihit capabilities were added in 2022, see section 4.4.3.
Since the peakfinder operates on discrete values, it is worth noting that if the value of the current sample is equal to the previous value, the counter is neither reset nor incremented. If N rising samples are required to set off the algorithm, this automatically entails an effective threshold of N ADC channels for the minimal detectable peakheight above the baseline (provided that the observed rise actually started at baseline level). To overcome this, an addition threshold parameter was recently added, see section 4.4.2.

[17] If the integration starts 3 samples before the trigger, the MA_{128} has to be evaluated at sample number 125 after the trigger.

In the analysis, the peak features serve several purposes. On the one hand, the peak height can be considered to be an alternative measure of the deposited energy, albeit without being inherently fixed relative to the trigger position (like the integral is). Accordingly, the peak feature might refer to pulses which had nothing to do with the reaction that caused the trigger. In addition, the peaktime might serve as a backup timing information, should the CFD (see next section) not have found any valid zero-crossing. Apart from that, yet most importantly, the peak features play a crucial role in the pile-up detection process. Details will be elaborated in section 5.

3.3.3.5 CFD-Finder

Principle The straightforward way to get a timestamp from an arbitrary signal would be to build a simple discriminator in which a signal is emitted when the pulse exceeds a fixed threshold. One obvious drawback of this is that the timing signal will be dependent on the pulse height, even if all pulses have exactly the same shape in terms of rise and fall time. While it is possible to compensate for such a "time-walk" (see e.g. [Sta19]), this adds further complications and can be avoided entirely by using a Constant Fraction Discriminator (CFD). The idea is to detect at which point a signal reaches a certain constant fraction (hence the name) of its maximum value. This point in time should not depend on the pulse height at all.
In practice, this is done by splitting the pulse into two parts; one of which is attenuated by a certain fraction (e.g. $\frac{1}{2}$), while the other part gets inverted and delayed by a certain time (e.g. the time between the pulses half-maximum and its maximum). An illustration of this can be seen in Fig. 3.10: The original signal is shown in black; the attenuated pulse in green and the inverted and delayed signal in blue. If the attenuated and the inverted and delayed signal are added, the actual CFD signal (shown in red) is obtained, which crosses $y = 0$ at a time illustrated by the purple line. This point is called the zero-crossing[18] of the CFD signal and provides the desired height-independent time information. Its position in relation to the original pulse as well as its steepness is given by the attenuation coefficient and the delay. In this case, they were chosen such that it occurs when the original signal reaches its maximum, while the slope at this point is given by the original signals steepness at half height (which was estimated to be the steepest part of the signal).

Algorithm The CFD-Finder is without question the most elaborate feature extraction algorithm implemented in the current firmware version. It is implemented as a state machine, and depends on 3 adjustable parameters: A threshold value the cfd-signal has to exceed (h_{cfd}, default is 4 (DS) or 8 (US) ADC channels), a time it has to stay above said threshold (T_{TOT}[19], default is 12 samples or 0.6 µs) and another time parameter called the "defuse time" (T_{def}, default is 32 samples or 1.6 µs).

1. At first, the CFD-finder is in its idle state, which is called *CFD_WAIT*. For every sample at which the signal is above the threshold h_{cfd}, a counter (`timer_high`) is incremented. Once the signal drops below the threshold again, the counter is evaluated: If the counter exceeds T_{TOT}, it was sufficiently long above the threshold and the next state *CFD_WAS_ABOVE_THRESHOLD* is entered. Otherwise, the counter is reset and the state remains

[18] In case of a non-zero baseline, the point where the signal crosses this baseline would be chosen instead.
[19] TOT stands for Time-over-Threshold

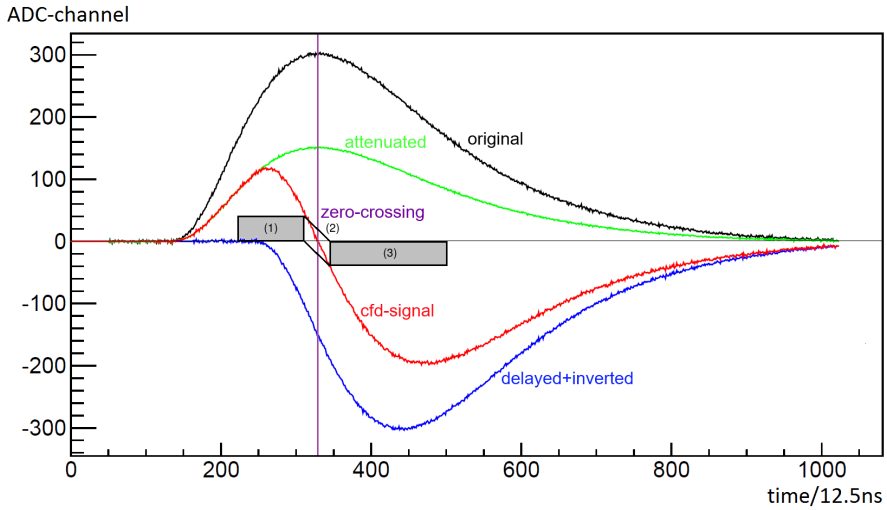


Figure 3.10: Working principle of a constant fraction discriminator. (1), (2) and (3) illustrate the various conditions and states checked in the algorithm

unchanged. Note that this is conceptually the same as checking whether a box of height h_{cfd} and width T_{TOT} fits under the signal (area (1) in Fig. 3.10).

2. In the *CFD_WAS_ABOVE_THRESHOLD* state, the CFD-signal is given a certain time T_{def} to drop below zero. This is again done by incrementing a counter; here called `timer_reset`. Once that happens, a timestamp according to the current sample number is created and the state changes to *CFD_ARMED*. Should the counter exceed the defuse time without the signal dropping below zero, the state is changed back to *CFD_WAIT*. In Fig. 3.10, this is represented by a parallelogram (2) of width T_{def} and height spanning from $+h_{\text{cfd}}$ to $-h_{\text{cfd}}$.

3. In the next state, *CFD_ARMED*, the signal is given whatever time of the defuse time T_{def} remains to drop below the negative threshold (note that the counter is intentionally not reset from the previous state!). If this condition is satisfied, the signal has to stay below that value for at least double the time-over-threshold, which is monitored by incrementing another timer called `timer_low`. Only if this is fulfilled as well, the previously determined timestamp will be marked as valid and is written to the feature packet. Should the signal jump above $-h_{\text{cfd}}$ again, `timer_low` is reset to zero and `timer_reset` is incremented further[20] until the signal drops below the threshold again or the reset timer exceeds T_{def}. In the latter case, the state is reverted back to *CFD_WAIT* and the search starts over again. As in the first state, this is conceptually again equivalent to fitting a box of height h_{cfd} and width $2 \times T_{\text{TOT}}$ in between the baseline and the signal (area (3) in Fig. 3.10).

Due to its relative complexity, there were some bugs in the first implementation. They are discussed in detail in section 4.4.1.1.

[20] This is done in order to retain smaller, potentially noisy signals which might cross $-h_{\text{cfd}}$ more than once directly after the zero-crossing. Without this, those timestamps would not be validated, even though an area of the necessary size might have followed.

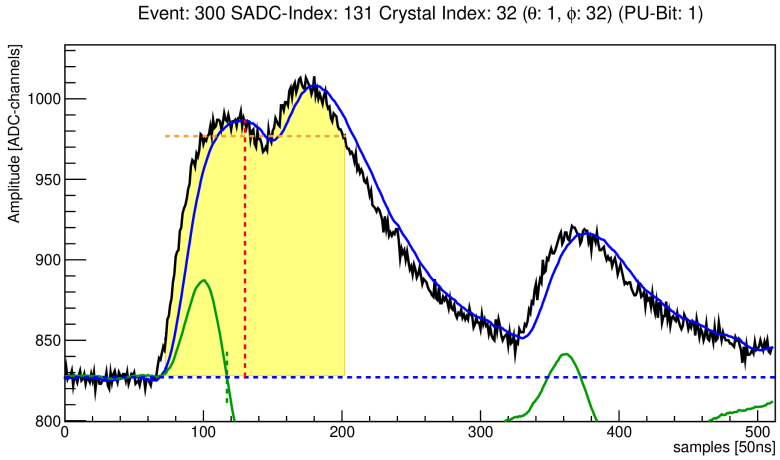


Figure 3.11: Example of a waveform containing pile-up. Features and color-coding are consistent with Fig. 3.9. The features of the first pulse are clearly influenced by the presence of the second pulse, whereas the third pulse further back does not have any significant impact on the extracted features.

3.3.4 Pile-Up Detection

Since the SADC feature extraction provides a number of different features, it is possible to detect cases of overlapping signals, so-called pile-up, by examination of the features alone. Fig. 3.11 shows an example of such a pile-up.
The idea behind the detection algorithm (v1.0) is, that under the assumption of a constant pulse shape, there should be a constant ratio $R_{\text{const}} \approx 0.64$[21] between the integral I and the peak height feature H (see Fig. 3.12). This ratio should not depend on the overall amplitude, but the influence of random deviations due to electrical noise will be larger at smaller amplitudes. The pile-up detection algorithm follows exactly this reasoning, by requiring the observed ratio $R_{\text{obs}} = \frac{I}{H}$ to be within carefully chosen limits, as shown in Fig. 3.12. The limits are parameterized as

$$-\theta_{\text{pu}} - \frac{s_{\text{lower}}}{H} \;<\; R_{\text{obs}} - R_{\text{const}} \;<\; \theta_{\text{pu}} + \frac{s_{\text{upper}}}{H} \tag{3.1}$$

where θ_{pu} (default value 0.02) denotes an absolute and $s_{\text{lower/upper}}$ (default value 5/10 in DS/US) a height-dependent relative threshold[22]. If the ratio is outside of that range, the waveform is flagged as pile-up. To avoid the excessive flagging of noise as pile-up, the extracted integral value further has to be above a configurable threshold θ_{int}, which was chosen to be 10/20 (US/DS) ADC channels ($\approx 0.5\,\text{MeV}$). This initial algorithm had some shortcomings, which will be extensively discussed in section 5.

[21] The numerical value of 0.64 is empirically determined.

[22] s stands for "slope", as these parameters determine the slopes of the limits; see Fig. 3.12. The meaning of θ_{pu} and s will be discussed in more detail in section 5.2.2.

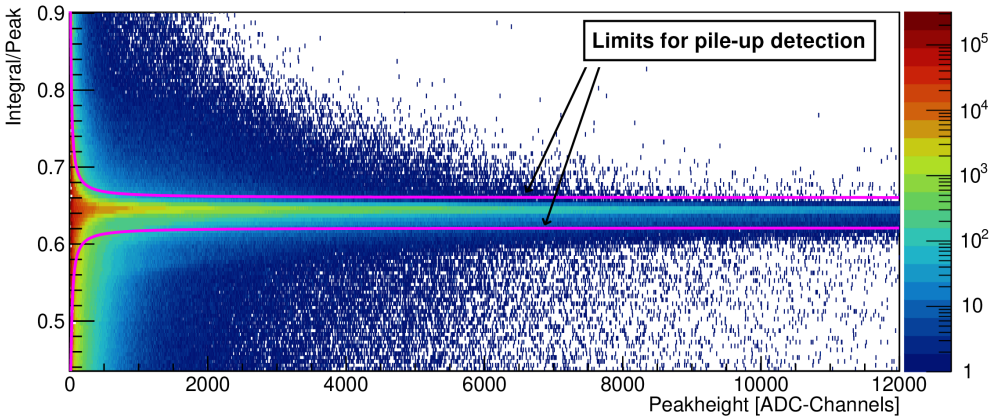


Figure 3.12: Plot of the ratio between integral and peak feature against peakheight.

3.3.5 Sample Sender

With the currently[23] used default settings, a sampled pulse sent from the SADC consists of 512 samples. With a sampling frequency of 20 MHz (after decimation), this gives a total pulse length of 25.6 µs. The sampling window is chosen such that the region of interest, i.e. the range over which the integration happens, is fully contained and approximately centered within this window. Several aspects of this module's behavior can be configured as needed:

1. In addition to the initial decimation, the *sampling rate* can be further reduced by one of two ways: By using only every N^{th} sample, the observable sampling window may be enlarged by an arbitrary factor (if the total number of samples stays the same). In the context of the CB-SADC firmware, this operation is referred to as *waveform stretcher*.
   A superior alternative was recently developed by B. Otto as well, which is to further decimate the signal by another factor of $2^n, n \in \mathbb{N}$. Summing the values up as before would result in the need for another datatype, since `unsigned short` is limited to 16 bit. Instead, and in order to leave the signal's amplitude effectively unchanged, the values are averaged, including correct rounding to the nearest integer. To ease the task of division (which is non-trivial on an FPGA), the available decimations are limited to powers of 2. This is operation will be referred to as *downsampling*.

2. The *number of sampled points* can (either in conjunction with an according waveform strectcher/downsampling or independently) be reduced from the default value of 512 by a factor of 2^m, where m is a natural number below 10[24].

3. The *position* of the sampling window (its beginning) can be modified as follows: The sampled data is buffered at all times with a First In First Out (FIFO) buffer depth of 320 samples, effectively delaying it by a fixed time of 16 µs. When the incoming trigger signal initiates the sample extraction procedure, the first sample that comes out of the buffer is

[23] Up to the year 2023. Meanwhile, the default is to send a FIR-filtered waveform with a 2-fold downsampling; i.e. 128 points covering a window of 25.6 µs.
[24] 2^m is already equal to 512 for $m = 9$

the earliest available sample point at 16 µs before the trigger. In that case, the sampling window would extend from [−16 µs..9.6 µs] around the trigger. By ignoring the first N samples coming out of the buffer before actually writing any values to the packet, it is possible to shift this window "to the right" by an arbitrary[25] amount of time.
It should be noted, that the buffer depth was initially limited to 128 samples (or 6.4 µs). This severely limited the obtainable amount of information for certain pile-up situations. In particular, situations in which the pulse of interest itself sits on top of another, previous pulse were harder to detect, because only 6.4 µs of the waveform were available in front of the trigger. At the same time, the entire second half (12.8 µs) the sampling window covered the region after the integration window, which is irrelevant for pile-up recovery. The buffer depth was only recently (see timeline) increased by another 192 samples to address this problem, which resulted in the current depth of 16 µs.

3.3.5.1 Compression

In addition, the latest version of the sample sender now also contains a compression algorithm. This algorithm was developed by J. Hartmann and was already used to compress the waveforms before they were written to disk. By moving the compression directly onto the FPGA, the amount of data which needs to be transferred over the network could be reduced significantly. Discussing the compression algorithm in detail would go beyond the scope of this work, but it is well described in B. Otto's thesis [Ott22]. In short: Instead of the values itself, the difference between subsequent values is stored and efficiently encoded instead. Typical compression factors for this application are in the range of 2–3[26] but vary somewhat with the exact type of waveform to be compressed. "Flat" waveforms (i.e. containing none or just a small pulse) tend to be more easily compressible than waveforms with large pulses or even pile-up situations because the difference between subsequent values tends to be small and can thus be more efficiently encoded.
In addition, the compression is even more efficient if the FIR-filtered waveform (see section 6.2) is used, because the waveform is significantly less noisy and consecutive samples are thus more likely to be closer together [Ott22].

[25] For practical purposes, an upper limit of 3 ms was introduced.

[26] Currently implemented is an improved version of this algorithm which achieves an even better compression of typically 4–5.

CHAPTER 4

Characterization of SADC Data and Improvements of Feature Extraction

This chapter serves to characterize some key aspects of the extracted features and discuss their respective implications. At the same time, it will provide an overview over the development of the firmware algorithms from the initial implementation in [Mül19] to the most recent improvements [Ott22].The continued development and improvement of the feature extraction happened in a fruitful and productive exchange between the author and J. Müllers and B. Otto. While the feature extraction algorithms ultimately had to be implemented in firmware, all of the upcoming improvements were either proposed or implemented and tested in offline analysis already within the scope of this thesis. The first implementations of both the baseline extraction algorithm and the CFD finder in particular had some serious issues, which were identified and will be discussed in detail, together with their respective mitigation in offline analysis and their eventual improvement in firmware.
Because the changes to the several algorithms were made over time and not all at once, the development of each algorithm of the feature extraction is presented on its own, while connections and dependencies between the different aspects are discussed as needed. For orientation, Table 3.1 provides the chronology to some of the key developments.

4.1 Overview of Recorded Data

The first production beamtime in which a significant amount of data was recorded with SADCs took place in May 2018. At this point in time, only 6 SADC prototype modules were ready to use, which equates to 384 usable channels.
In order to make the most of this limited number of channels, it was decided to connect them in such a way that, among others, every crystal from the first 4 rings was read out by SADCs. At this stage, the SADC readout was still only used in parallel to the established QDC readout. Having SADC coverage in addition was particularly relevant in the forward region where, due to being a fixed target experiment, the highest rates are to be expected. The specific grouping of channels into cables connecting the BuffTis to the SADCs led to the peculiar coverage shown in Fig. 4.1.
Two other blocks of data taking followed in the same year in October and November, under largely the same general conditions as in May (see Table 3.1) albeit with some different parameter

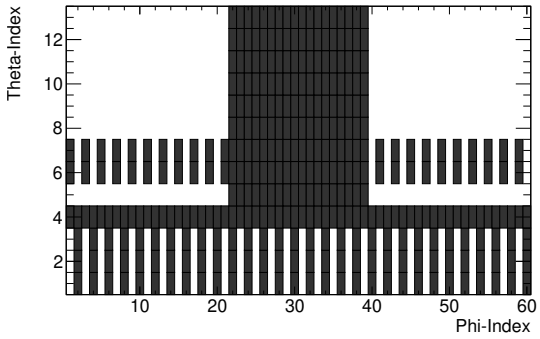


Figure 4.1: Crystals covered by the SADC readout in the Crystal Barrel in the May 2018 beamtime. Only the downstream half is shown.

settings. The next beamtime in February 2019 had to be finished without the SADCs, due to a persistent problem with the firmware.
The first beamtime with full SADC coverage for all channels took place in June 2021. It was quickly followed by two more beamtimes in November and December of the same year. Before and during this phase, B. Otto also started to work on his thesis project [Ott22]. As a result, several improvements which were previously suggested already or emerged from the analysis of the newly obtained data could be implemented and tested during this period of time and (especially) in the first half of 2022. Further production beamtimes took place in 2024, but were not taken into account for this thesis.

4.2 Baseline

Due to a number of reasons, the actual baseline, i.e. the amplitude level one would expect in absence of any signals, might vary from one event to another. Reasons include random noise, the presence of low-frequency oscillations (especially at multiples of the power grid frequency of 50 Hz; see Fig. A.1), and slow voltage drifts, for example due to temperature changes. By using an event-based baseline subtraction, these effects are of no concern, provided that they act over a timescale which is large in comparison to the integration window of 6.4 µs. In contrast, the baseline (pedestal) of the previous QDC readout used to be determined only once for every run (≈ every 20-30 minutes) and can as such not account for any of the mentioned effects.

4.2.1 Problems with the Initial Approach

While the original baseline finder algorithm (see section 3.3.3.2) was working exactly as intended, it gave an incorrect baseline value if another pulse occurred shortly before the actual pulse; i.e. the one correlated to the event which caused the trigger signal to be emitted. Since it simply took the value of the moving average 32 samples before the trigger, whatever that value might have been, it could happen that this value is much higher than the actual baseline would have been (see Fig. 4.2 for an illustrative example). Ideally, such a case would result in the crystal hit being flagged as pile-up and the waveform being written to file. While the shortcomings of the initial pile-up detection algorithm will be discussed at length in section 5, it is obvious already from the brief introduction in the previous chapter, that the detection algorithm relies on certain conditions to be met in order to function properly:

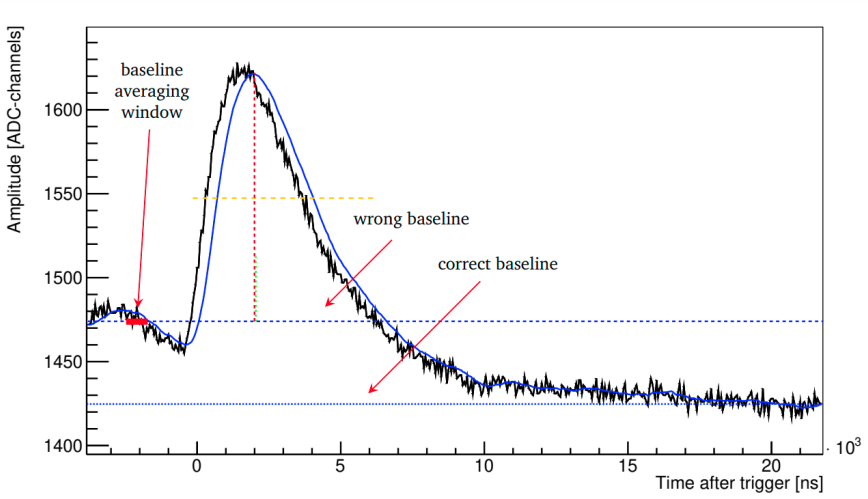


Figure 4.2: A pulse correlated to the trigger which sits on top of a smaller, preceding pulse. The baseline determination is affected by the other pulse and leads to a value which is too high (own work, annotations from [Ott22]).

(a) It relies on the presence of a peak-feature to compare the integral to. If no peak was found (for whatever reason), the algorithm can not work.

(b) It requires the integral feature to surpass a certain threshold above the baseline in order to suppress noise-hits. If the baseline is too high, this can result in an integral value which is (incorrectly) below the threshold or, in extreme cases, even below the baseline.

Due to these reasons, one cannot reliably count on the pile-up flagging of such cases.
Fig. 4.3 depicts a typical distribution of resulting baseline values during data taking (shown in gray). Assuming no systematic effects, one would expect a symmetric distribution of fluctuations around the MPV. As evidenced by the plot, this is not the case. While most of the histogram follows a roughly gaussian[1] spread, a significant asymmetry can be observed, tending towards higher values. It was the topic of a recent bachelor thesis [Zyw19] to deal with this situation and to develop an improved baseline algorithm which is not susceptible to pile-up effects. Since large amounts of data were already taken with the initial algorithm at this point (everything before June 2021, see Table 3.1), a method to retroactively correct this data had to be developed as well.

4.2.2 Retroactive Correction Algorithm

A very simple method would be to resort back to the same method that was used for the QDCs, which is to determine a fixed baseline value (e.g. by determining the MPV for each channel) and use this for every event in a run. This would however completely eliminate any advantage from having an event-based baseline subtraction in the first place (see Fig. 4.6 for a comparison). Instead, and in order to preserve most of the baseline values which did not show significant

[1] The randomness of noise and the fact that values from all channels entered the histogram lead to a gaussian distribution via the central limit theorem.

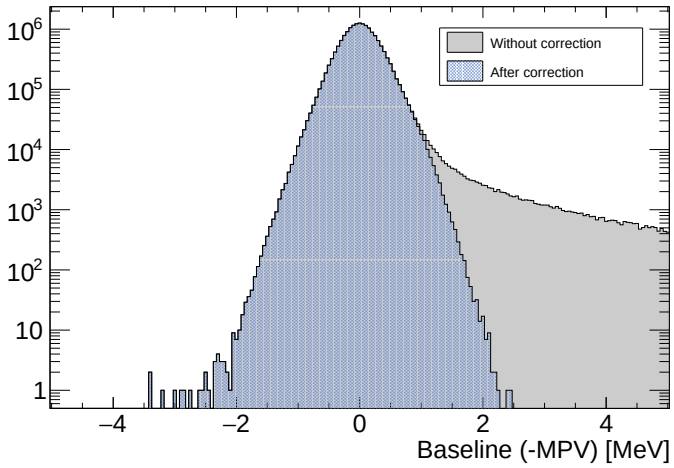


Figure 4.3: Calibrated baseline values of all channels with and without retroactive correction. The MPV of each channel was subtracted. The threshold value of 20 ADC channels corresponds to $\approx 1\,\text{MeV}$. Fitting a gaussian function to the gray histogram (limited to $x < 1\,\text{MeV}$) gives a width of $\sigma = 0.30\,\text{MeV}$ (not shown).

deviations in the first place, a retroactive correction algorithm was developed:
Every new baseline value is compared to the median value of a number (default: 15) of baselines that were found immediately prior to the current event. If the value in question is below the median or exceeds it by no more than an adjustable threshold, the value is assumed to be okay and nothing happens. If the value is above that threshold, it is assumed to be invalid and gets replaced with the respective median value[2].
The result of this correction algorithm with a default threshold value of 20 ADC channels ($\approx 1\,\text{MeV}$[3]) is shown in Fig. 4.3. Evidently, the correction has no effect on the vast majority of values, while the "tail" towards higher values is nicely removed.

Comment The median of the last 15 values is taken to allow for some low frequency oscillations or drifts of the baseline. The originally proposed algorithm used a weighted mean calculation instead of the median. The advantage of using the median over the weighted mean is that values deemed "incorrect" can enter the median-calculation for the next event without any issues. In case of using the mean, extra care has to be taken in order to ensure that the mean does not drastically increase if a particularly large baseline value was found. This was done in [Zyw19] by weighting deviating baselines before entering the mean calculation, but brings with it the necessity to specify and optimize this weighting factor, which is not trivial to do. On the other hand, deviating values can also not be completely ignored, as the mean value could theoretically get stuck at a lower value than adequate if the actual baseline should spontaneously increase

[2] There exists a – probably very small – subset of events which are unrecoverably lost: In case an affected waveform was part of a feature packet which was zero-suppressed due to the integral value(s) being to small in relation to the (deviating) baseline(s), no features at all are present to be recovered. This should be rare, because any other energy deposition above the zero-suppression threshold within this block would have prevented the (entire) packet from being suppressed (see section 4.3.3).

[3] Corresponds to a $\approx 3\sigma$ interval of the gaussian fit from Fig. 4.3

over just a handful of events. In that case, all subsequent baselines would be deemed "too high", leading to a situation from which the algorithm would not recover on its own. Simply using the median instead has proven to be more effective while at the same time getting rid of the need to determine said weighting factors altogether.

4.2.3 Improved Version

The algorithm which was proposed in [Zyw19] was (after some minor adjustments) recently incorporated into the Firmware as well. The idea of this new approach is, simply put, to avoid the determination of baselines in the presence of pulses in the first place. This is ensured by requiring the signal (in this case, the MA_{128} was used instead of the MA_{16}) to fit entirely within a "box" of specified dimensions (see Fig. 4.4). Given a sufficiently small box height h and an equally suitably long box width w, this ensures the needed flatness and practically precludes the presence of any pulses within this box. The baseline value itself is then calculated as the mean of the first and last sample within this box. It is stored and used for all subsequent events until a new baseline is found, in which case the stored value will be updated.
Ideally, multiple baseline boxes should be found in between events. However, due to constraints put on the complexity of the algorithm by the available resources of the FPGAs, boxes are stacked "end on end" along the incoming signal. This means that there might be areas in which the signal would have fit inside a box which were missed because the "test boxes" did not coincide with the flat area. Fig. A.2 (appendix) shows such a situation. B. Otto devised a way to mitigate such situations by effectively implementing four baseline finding algorithms running in parallel with test boxes shifted by $w/4$ each. Details can be found in [Ott22].

It cannot be ruled out that some diminishingly small part of the falling or rising edge of another signal might fall into the box. Based on an exponential approximation $\left(f(t) = A \cdot e^{-\frac{t}{\tau}}\right)$ of the falling edge of the pulses, the largest possible deviation from the "true" baseline may be estimated for a given set of parameters (neglecting the influence of noise):

$$f(t-w) - f(t) \leq h \quad \text{(Criterion to find baseline box)} \tag{4.1}$$

$$\Rightarrow A\left[e^{-\frac{t-w}{\tau}} - e^{-\frac{t}{\tau}}\right] = A \cdot e^{-\frac{t}{\tau}}\left[e^{\frac{w}{\tau}} - 1\right] \leq h \tag{4.2}$$

$$\Leftrightarrow t \geq \ln\left(\frac{A}{h}\left[e^{\frac{w}{\tau}} - 1\right]\right) \cdot \tau \tag{4.3}$$

Defining the smallest possible value of t as t_0 (denotes the earliest possible position at which a box may be found after a pulse), the actual baseline value is computed as the mean of the first and last sample within the box. It is then given by the following expression:

$$b(t_0) = \frac{f(t_0 - w) + f(t_0)}{2} = \frac{A}{2} \cdot e^{-\frac{t_0}{\tau}}\left[e^{\frac{w}{\tau}} + 1\right] = \frac{A}{2} \cdot \frac{\left[e^{\frac{w}{\tau}} + 1\right]}{\frac{A}{h}\left[e^{\frac{w}{\tau}} - 1\right]} = \frac{h}{2} \cdot \coth\left(\frac{w}{2\tau}\right) \tag{4.4}$$

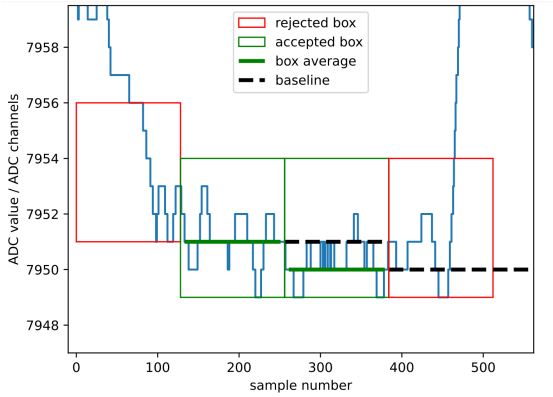


Figure 4.4: Working principle of the new baseline algorithm [Ott22]. The signal did not fit into the red boxes, but did fit into the boxes shown in green. The appropriate baseline value is being used subsequently (dashed black lines) until it gets overwritten.

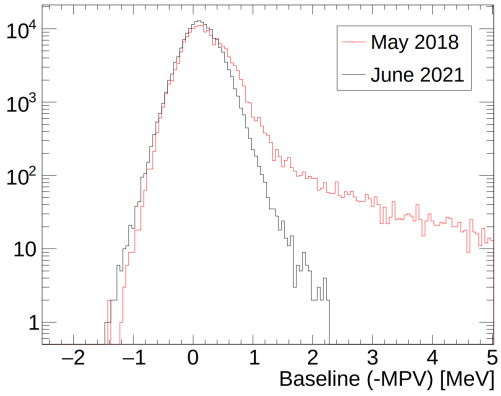


Figure 4.5: Results of the new (black) versus the old (black) baseline determination algorithm (own work; used by [Ott22]).
Baselines of all channels (minus their respective MPV) were plotted into one histogram.

Notably, this result does not depend on the size of the preceding pulse at all[4].
Assuming $\tau = 8\,\mu s$ as the longest decay component (see section 6.3), a box width of 6.4 µs (128 samples) and a box height of 5 ADC channels[5], the earliest baseline box which could be found on such a falling edge would lead to a baseline value of ≈ 7 ADC channels ($\approx 0.34\,MeV$) above the true baseline. While these were the initially proposed parameter values, it turned out that a length of 256 samples (12.8 µs) is suitable as well (see discussion below) and became the new default. With these settings, the theoretical deviation is only 0.20 MeV. Note that these values constitute the theoretical limit, and are not necessarily representative of the actual performance (which is mostly limited by noise).
The effect on the overall energy determination will be discussed in section 4.3.1.1. Nonetheless, Fig. 4.5 already shows a clear improvement with regards to the distribution of baseline values relative to the respective MPV. Large deviations over 2.5 MeV above the MPV no longer occur at all. The distribution still shows a slight asymmetry, this is most likely due to the discussed influence of pulse edges within the boxes. As pulses are always positive, this can only lead to positive deviations, which explains the asymmetry. This could be reduced by decreasing the box height or increasing the box width, at the risk of producing a situation in which the noise level alone makes the finding of new baselines difficult, resulting in the baseline values being updated too rarely. A situation like that has to be avoided at all costs, as it would again lose any advantage an event-based baseline determination has. It is closely monitored during data taking that such a situation does not arise. Since it is difficult to get real-time debug output from the algorithms running on the FPGA, this is achieved by comparing baseline values from subsequent feature packets for each channel. A value that differs from the previous one necessarily indicates that at least one new baseline box was found since the last event[6]. Each change in value is counted as

[4] This holds if the pulse is large enough to be well described by a single exponential before a box could be found. The independence of the amplitude is a result of the properties of the exponential function.
[5] Downstream value. 10 ADC channels in the upstream half.
[6] The opposite is not true: Two subsequent baseline values might be identical simply by chance.

an update, which can then be plotted against the channel index. Too tightly chosen parameter settings or an unexpected increase of the noise level would result in a significant drop in update frequency. Fig. A.3 (in the appendix) shows an extreme example, which was recorded with a deliberately low box height.

4.3 Integral

4.3.1 Energy Determination

The process to calculate the integral as the main energy feature is very simple and straightforward. As demonstrated in earlier work [Sch16], the achievable energy resolution is at least comparable (in the QDC low-range), or substantially better (in the high-range) than what is possible with the old QDC setup (see Fig. A.4). This has been confirmed to be the case under actual experimental conditions as well. Neither the peak of energy depositions from MIPs, nor the peak of reconstructed neutral pions [Sal18] showed any significant difference between the SADC and the QDC readout. It was further shown, that an ideal integration length exists which minimizes noise contributions and maximizes the achievable resolution power. The integration length used by the SADC readout is 128 samples or 6.4 µs at the moment. While the integration itself poses no challenges, it should be pointed out that the energy determination depends just as much on the baseline as it does on the integral.
Especially the advantage of having an event-based baseline subtraction is substantial, even with the old baseline algorithm (after correction). The effect it has on the resolution of the integral as the main energy feature can be seen in Fig. 4.6. It shows the uncalibrated integral values, minus their respective baseline. In order to demonstrate the effect a fixed (i.e. run-based) baseline would have, another histogram was constructed, using the constant MPV of each channel as a substitute for the determined baseline values. Since the majority of channels in an ordinary event are free of any actual signal[7] and only the baseline is measured, a distribution centered around zero is to be expected. In addition, both distributions exhibit a long tail towards higher values, which is of course unsurprising since those are the actual signals one is usually interested in. The FWHM[8] of the black distribution is 12 ADC channels. By using an event-based baseline subtraction (and the aforementioned baseline correction algorithm), the width is reduced to 7 ADC channels. For reference, fitting the distribution with a gaussian while taking only values with $x < \sigma$[9] into account yields $\sigma \approx 5.5$ ADC channels for the black and $\sigma \approx 3.0$ ADC-channels for the blue curve. For a normal distribution, FWHM and σ are related by $\text{FWHM} = 2\sqrt{2\ln 2}\sigma \approx 2.355\sigma$, which fits well with the observed values.
The σ values would, in analogy to the established QDC readout and using an average gain factor of 0.051 MeV per ADC channel, correspond to a "pedestal width" of ≈ 0.28 MeV for the fixed baseline and ≈ 0.15 MeV for the event-based baseline determination. [Mül19] reported a width of 147.6 keV for a data sample collected with clock-trigger, where the presence of actual signals (and therefore also deviating baselines) is inherently greatly reduced. Since those results are in

[7] Also see discussion in section 4.3.3 (Zero-Suppression).
[8] The FWHM is used, since it is largely independent of the right-hand side tail.
[9] The upper limit σ was taken from the fit result itself and was determined in an iterative process until convergence was reached.

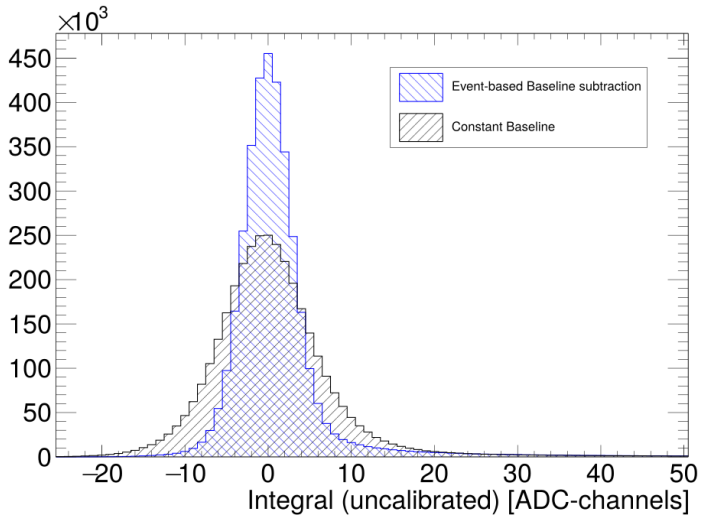


Figure 4.6: Effect of the normal event-based baseline subtraction (blue) as opposed to a subtraction of the (constant) most probable baseline-value for each channel (shown in black). Plotted are integral-features after baseline correction; summed over all channels. The blue curve represents the ordinary baseline subtraction, whereas the black curve was calculated artificially for comparison purposes. Data is from May 2018 and thus only from downstream channels. 20 ADC channels correspond to ≈ 1 MeV.

good agreement, it can be concluded that the baseline correction algorithm seems to have no detrimental effect on the energy resolution, while resorting to a constant baseline value per run definitely would.

4.3.1.1 Effect of the Baseline Box Algorithm

It is worth noting, that the new algorithm also has an effect on the energy resolution which has nothing to do with the avoidance of deviations due to pulse edges. By subtracting a baseline value that is "only" the value of a 16-fold moving average from the (carefully optimized) MA_{128} used for the integral, the theoretically achievable energy resolution gets worsened. This can be understood as follows:
Under the (oversimplified) assumption that the only distortion of an otherwise ideal signal consists of white noise with a mean amplitude of ν, the value of an N-fold moving average is expected to have a mean noise amplitude of

$$\sigma_{\mathrm{MA}} = \frac{\sqrt{N}\nu}{N} = \frac{\nu}{\sqrt{N}} \tag{4.5}$$

at any given point. From this consideration follows an uncertainty of $\frac{\nu}{\sqrt{128}} \approx 0.088\nu$ for the integral value itself, just from integrating the noise. To get a usable energy value though, the baseline value has to be subtracted, which, if calculated from the MA_{16}, has a much larger uncertainty of $\frac{\nu}{\sqrt{16}} = 0.25\nu$. So a large improvement of the new algorithm already lies in the fact that it uses the MA_{128} as well, i.e. the same moving average that is used for the integral feature. Fig. 4.7 showcases this improvement. Data was taken from clock runs (see section 2.9), which

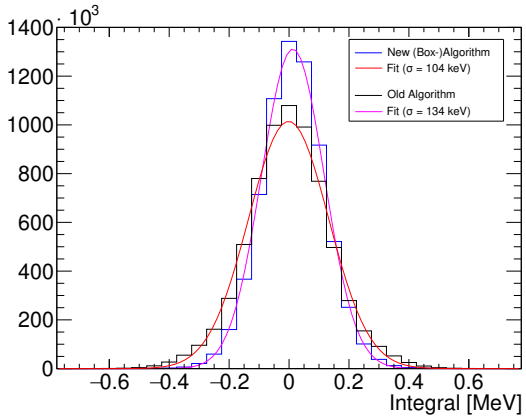


Figure 4.7: Comparison of integral values from clock runs, minus their respective baseline values. All downstream channels (equipped with the final SADC readout) contributed to the histogram. Baselines were determined with the old (black) and new (blue) baseline algorithm. A gaussian is fitted to estimate the widths of the distributions.

means that almost no pulses[10] are to be expected within the sampling windows. Under these conditions, one would expect the integral features to have exactly the same value as the baselines, which would result in a sharp peak, centered at 0 MeV. One might expect the old algorithm to perform quite well in this case, since the point in time at which the baseline was determined lies (by design) very close (1.6 µs) to the beginning of the integration window. The last found box from the new algorithm on the other hand could, theoretically[11], be quite far away. The figure shows, that even under these "favorable" conditions, the new algorithm has a better energy resolution, which is evidenced by a narrower distribution: 104 keV with the new algorithm versus 134 keV with the old algorithm. The latter value is again in fairly good agreement with the value of 147.6 keV given by [Mül19] [12].
As discussed, most of this improvement is most likely to be attributed to the usage of a wider moving average. In hindsight, it would have been possible to use the MA_{128} as input for the old algorithm as well, but the problem of deviating baselines would have occurred anyways.

4.3.2 Energy Calibration

In order to extract physically meaningful energy information from the data, the features have to be calibrated. Due to a number of reasons, the appropriate calibration factors to calculate an energy (typically in MeV) from the raw integral features given in ADC channels might vary from channel to channel. These reasons include:

1. Different hardware: Although identical in principle, the exact specifications for any given hardware component might vary from one unit to another, depending on the tolerances specified by the manufacturer.

[10] Only some cosmic particles which might have been recorded by chance.

[11] If no pulses are present, the chosen box parameters should ensure that baselines are updated with almost every new box.

[12] A potential explanation for the slight discrepancy (since both results come from the analysis of clock runs) could lie in the fact that J. Müllers' determination stems from a time when only part of the downstream was read out by SADCs (see Fig. 4.1) and the energy calibration was still preliminary. In addition, more channels with potentially different characteristics contributed to the overall histograms for this work.

2. Intentional gain differences: As stated before, the gain for channels in the upstream half of the detector was deliberately set to double of the downstream gain, since the expected average energy of particles in the upstream is lower due to kinematic reasons.

3. Position of crystal in the detector: Electromagnetic showers typically extend over several crystals (see section 1.3.3), which have to be clustered together to obtain the entire energy deposition of a particle. Especially at the edges of the detector and in the vicinity of holding structures, not all energy might be deposited in sensitive material (the crystals). To compensate for such losses, an energy- and Θ-dependent Energy Correction Function (ECF) [Mül18] was introduced, which is applied before the calibration takes place. However, the ECF might not be perfect over the whole barrel and residual deviations will be absorbed into the calibration factors.

The method employed for the calibration of CB data involves reconstructing full reactions in a larger data set and then iteratively adjusting the calibration factors for each individual channel in a way that the mean of the peak of reconstructed π^0-mesons coincides exactly with the experimentally very well established mass of 134.977 MeV [PDG22]. For the current calorimeter configuration, this method was first implemented by J. Müller [Mül18] and later substantially improved by B. Salisbury [Sal16]. It was used successfully for the established QDC readout in the past and is now employed for the SADC data as well. Fig. 4.8 shows a histogram of all deduced calibration factors. The two visible peaks belong to the up- and downstream channels, respectively. As anticipated from section 3.2, the calibration values lie close to 0.05 MeV per ADC channel in the downstream and around half of that in the upstream[13]. However, this calibration method requires data from all channels; a condition which could only be met since spring 2021, when full SADC coverage was achieved. Since a substantial amount of data was taken beforehand, this required a different ad-hoc solution (see section 4.3.2.1 for the calibration in order to make the SADC data usable for further analysis[14]

Dynamic Range Naively, a gain value of ≈ 0.05 MeV per ADC-channel would, given that there are 2^{16} ADC-channels in total, correspond to a dynamic range of over 3.2 GeV. While this is true in principle, the integral value is typically a factor of ≈ 0.64 smaller than the peak value of a given pulse (see section 3.3.4). Therefore, the largest pulse which fits entirely within 2^{16} ADC-channels has an integral value of only $\approx 0.64 \cdot 2^{16}$. Above that, pulses will be clipped off at the top. Multiplying this value with the respective gain values yields an effective average dynamic range of 2.14 GeV in the downstream and 1.07 GeV in the upstream.
In principle, it is possible to go beyond this limit and use clipped pulses as well, since their energies can still be unambiguously mapped to uncalibrated integral values. However, linearity will be lost beyond this point and applying the existing calibration will lead to a steadily increasing error for larger energies.

[13] A gaussian fit yields mean values of 0.0256 (US) and 0.0511 (DS) MeV per ADC channel.

[14] On the one hand, a calibration is needed to compare results from the standalone SADC readout with existing data. On the other hand, the data from the established QDC readout could be seamlessly supplemented with SADC data in order to correct pile-up in the forward region of the calorimeter (since both readouts ran in parallel).

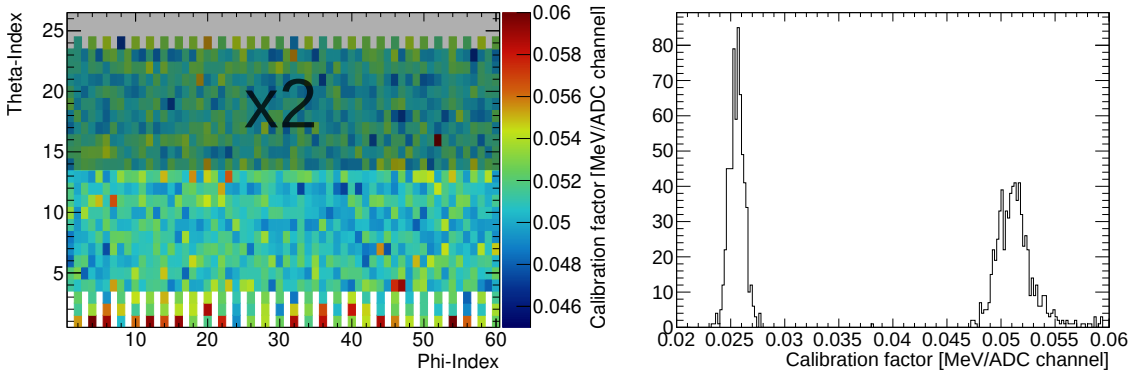


Figure 4.8: Distribution of calibration factors for the SADC integral feature, as obtained after calibration on the π^0 mass (data from December 2021). For the purpose of having a nuanced color scale, upstream values were multiplied by 2 in the left plot.

4.3.2.1 Relative Calibration

The established QDC setup was read out in its entirety until full SADC coverage was achieved and beyond[15]. The latter decision was made in order to facilitate a detailed comparison between the SADC and QDC data and to ensure the comparability and validity of the data produced by the SADC readout beyond any reasonable doubt. Therefore, a full calibration of all crystals was always possible for the QDC data. Instead of trying to achieve an independent SADC calibration with an incomplete coverage, it was much simpler to use the data from both setups together. Since both were read out in parallel, there exists a well-calibrated QDC energy for every SADC integral feature. Assuming linearity between both energy measures (see next section), it is easy to extract a specific factor for every channel by which the integral values have to be multiplied to match the calibrated QDC energies. This is simply done by fitting a linear function (y-axis intercept fixed to zero) to the distribution of calibrated QDC energies plotted against the uncalibrated SADC integrals for every channel. The overall impact of this calibration can be seen very clearly in Fig. 4.9: after calibration, energies from the SADC and QDC readout are in near perfect agreement. Considering the method (integration of the same signals over a very similar gate length), this was of course to be expected.
While being very simple in the implementation and hence less error-prone, this method of calibration offers the added benefit that it makes the energies extracted by SADC and QDC as compatible as possible; to the point that they can be used interchangeably without any noticeable impact on further analysis results. This becomes especially important if one wishes to use the pile-up detection and recovering capabilities of the SADCs in the forward region, but still use the QDC values for every other channel to achieve full angular coverage.

QDC High-/Lowrange Factors As mentioned before, the QDC readout was operated in a "dual-range" setup (see section 2.8.1). Due to hardware tolerances, the actual (internal) amplification factor will not always be exactly 8, and has to be deduced from on the recorded

[15] Meanwhile (2024), the SADC readout has superseded the QDC readout entirely, which is not actively used anymore but still remains as a backup option.

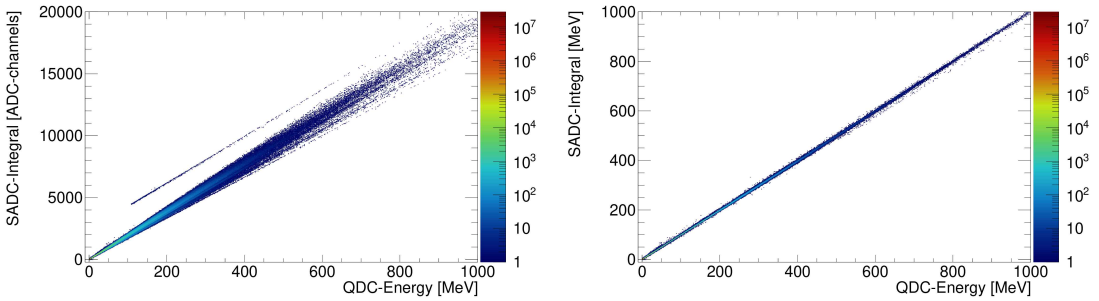


Figure 4.9: Left: Uncalibrated (left) and calibrated (right) SADC integrals of all channels plotted against the respective, calibrated QDC energies. Data was taken from the May 2018 beamtime. Only data points not flagged as "pile-up" and not in need of baseline correction were taken into account. The apparent "outlier" above the main diagonal could be attributed to a wrong QDC high-range calibration and was fixed in the right-hand side plot.

data. Likewise, the pedestal value of the high range is also unknown beforehand, but has to be determined separately. To avoid these complications, only low-range values are used for the calibration process, which means that there exist no "direct" calibration factors for high-range values. Instead, high-range values (x_{HR}) are converted into an appropriate low-range equivalent value ($\tilde{x}_{\mathrm{LR}}$) first:

$$\tilde{x}_{\mathrm{LR}} = x_{\mathrm{HR}} \cdot \texttt{rangefactor} + \texttt{rangeoffset}$$

There are different methods to determine these parameters. Without straying further from the main topic of this thesis, this matching was not always without problems in the past (see also the apparent outlier in Fig. 4.9). It was also the topic of a recent bachelor's thesis to tackle those problems and develop a new, more reliable method [Dav19].
The SADC setup has the advantage that is does not need to be operated in a similar dual-range setup in order to achieve a comparable (or even better) resolution to the QDC [Sch16]. This also implies, that as a "byproduct" of the actual calibration process, the SADC integral feature can be used to cross-check whether the high-/lowrange matching was successful. As can be seen in Fig. A.5, this was not the case for all channels. Fortunately, it was easy to calculate corrected range factors by analyzing the width and position of the apparent gap for a total of 3 QDC-channels[16].

4.3.3 Zero-Suppression

Once a trigger signal was emitted, data from all detectors is recorded and stored. In case of the SADC readout, the reception of a trigger signal results in a feature packet with data from all connected 32 channels being sent by every FPGA. The fact that only the extracted features are sent instead of the full waveforms already constitutes a great reduction, but this is still highly inefficient (both in terms of network traffic and disk usage) since only part of the 1320 channels actually detected a significant energy deposition. Fig. 4.10 shows that is rare to have events with more than 200 hits ($\approx 15\%$ of all channels) under ordinary trigger conditions. It would be much more efficient to only send and store information from channels that detected a hit. In this case,

[16] Index 40 [most noticeable], 70 and 164.

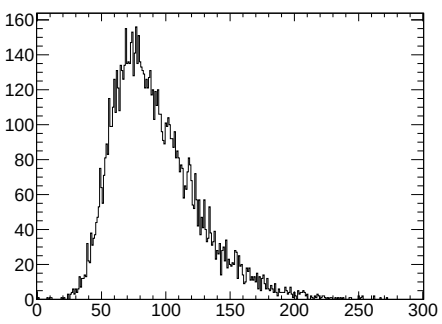


Figure 4.10: Number of hits ($E > 0.5\,\text{MeV}$) per event. Data was recorded during the May 2022 test beamtime under the usual trigger conditions.

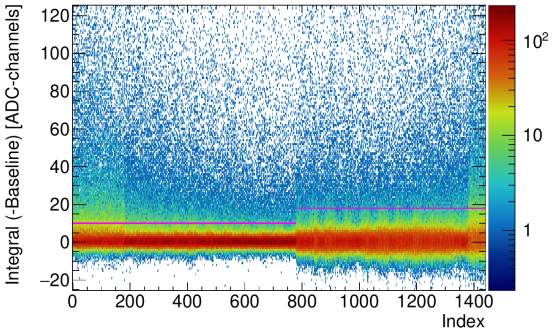


Figure 4.11: Distribution of uncalibrated integral values (after baseline subtraction). Magenta lines depict the implemented zero-suppression thresholds.

an energy deposition above 0.5 MeV was considered to be a "hit". This specific energy was chosen based on the distributions shown in Fig. 4.6 and 4.7. Since it corresponds to 3 to 4 σ of these distributions, accidental noise hits should be sufficiently rare. At the same time, a threshold of 1 MeV is typically used in the reconstruction of physical events (see section 2.10.1). Therefore, no relevant information is lost if energy depositions under 0.5 MeV are disregarded.
In a first step, the actual zero-suppression was realized by not saving entire packets only if all 32 channels were below threshold. This was realized by an external check of the packet's features in the LEVB. In addition, it was further required that neither a peak, nor a CFD zero-crossing was found. While not very selective, this approach was practically the easiest to realize. In a second step, the zero-suppression was applied to individual channels, yet still at the level of the LEVB. This means that all of the superfluous information was still sent as feature packets over the network. This only changed recently, when the zero-suppression was implemented directly in the firmware [Ott22] (see Table 3.1).
Since the appropriate calibration factors are available neither in the LEVB nor within the firmware, the zero-suppression must operate on the uncalibrated values. For the downstream, 0.5 MeV corresponds to approximately 10 ADC channels. For the upstream, double of that would have been appropriate, yet due to the lower overall hit rate it was decided to go with a value of 18 ADC channels instead. Fig. 4.11 shows the effect of the zero suppression with respect to the distribution of integral values. Note that the values one would like to suppress are expected to be symmetrically distributed around $y = 0$. The magenta lines denoting the thresholds correlate nicely with the visible edges of the distributions below the y-axis.
In addition to the integral being below threshold, it was again required that no peak and/or CFD zero-crossing was found in order for a crystal hit to be suppressed. These requirements ensure that no data that could have potentially been of any interest is ever lost.

4.4 Timing Capabilities

Given the time-resolution of the existing TDC readout (see section 2.8.1), its operation in the designated timing branch and native multihit capabilities, the extraction of timing information using the SADCs might seem superficial. However, due to the improved Signal-to-Noise Ratio in the energy-branch, the SADC readout is able to extract timestamps down to the order of ≈ 1 MeV (see section 4.4.4); an energy range which is not reached with the current TDC readout (see section 4.4.6). Since 2021, the SADC timestamps are utilized for the purpose of *time-clustering*; i.e. clustering which not only groups adjacent energy depositions together, but also takes into account their respective times [Sta24].

4.4.1 Constant Fraction Discriminator

The respective entries in Table 3.1 already alluded to the fact that the original implementation of the CFD-finder suffered from a few bugs. Especially the first one was quite severe and was discovered within this thesis as part of the offline analysis of the features. It affected data up until October 2020 (see Table 3.1). The following section will provide a short discussion of the causes. During his thesis, B. Otto discovered another problem, which was fortunately not quite as severe and harder to detect [Ott22].

4.4.1.1 Problems in the First Implementation

As detailed in section 3.3.3.5, the CFD-finder operates as a state-machine. Once a CFD zero-crossing[17] is found, it is stored internally and gets written to the feature packet if certain additional conditions are fulfilled. Unfortunately, there was a bug in the firmware that prevented any previously found zero-crossing (e.g. from events before) to be updated, unless a relatively large pulse was present. The issue is most clearly visible, if the difference of consecutive cfdtimes from one event to the next is considered; see Fig. 4.12.
Besides keeping the old timestamp in memory in the first place, the bug consisted of incorrectly going from the *CFD_WAS_ABOVE_THRESHOLD* state directly to the *CFD_ARMED* state, without allowing for the intended transition time. So unless the CFD-signal dropped from +`CFD_THRESH` below zero within one sample, the value where the zero-crossing occurred never got updated. In effect, this resulted in a cdftimes which were at energies at which no valid cfd zero-crossings should have been found (i.e. below 1 MeV) and at times which did not seem to correlate to the sampled waveform. Thankfully, due to the steepness of the CFD signal near the zero-crossing and the relatively low threshold value (4 ADC channels), this seems not to be an issue for pulses above ≈ 5 MeV.
At the same time, another (functionally unrelated) bug prevented `timer_low` from being reset, thus resulting in an incorrect validation of much more zero-crossings than intended and accordingly writing the corresponding, potentially "old", zero-crossing time to the feature packet. This also results in additional timestamps beyond the end of the range of what should have been possible: Due to the way the algorithm is designed, the last possible cfdtimes should be $2 \times$ `CFD_TOT` $= 24$ samples before the end of the region in which the algorithm is active[18], since any initially found

[17] Also referred to as "cfdtime" in this thesis.
[18] By default, the algorithm is for 200 samples (10 µs), starting with the arrival of the trigger signal.

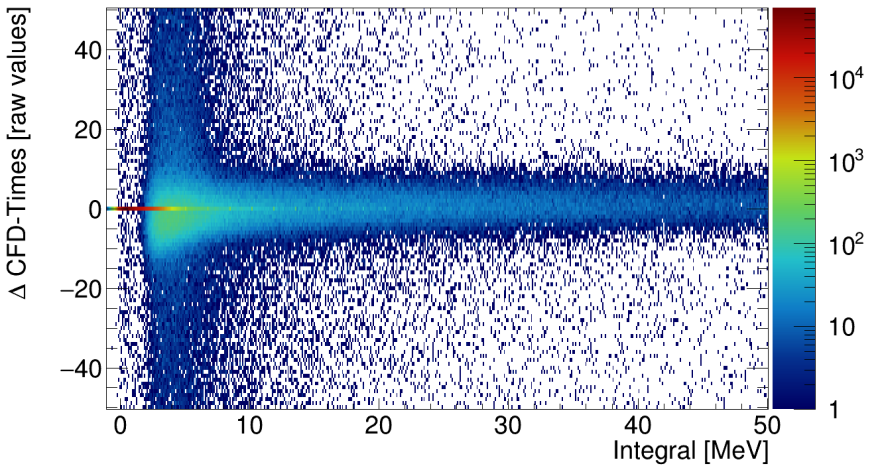


Figure 4.12: Difference of cfdtime versus previous value (from the same channel) against integral.

zero-crossing must be validated by the algorithm.
Luckily, the detrimental effect of these unfortunate bugs seems to be limited: The biggest issue arising is that in some cases where no timestamp should have been generated, a timestamp from the previous event was incorrectly written to the feature packet. As stated above, as long as a sizable pulse was present, the CFD timestamps appear to be correct, which is also why this bug went unnoticed for a relatively long time.

CFD Veto In order to mitigate the effect of the bugs in the CFD-finder, an algorithm to remove invalid cfdtimes was developed for the offline analysis of affected data. The algorithm operates as follows:

1. Cfdtimes in the "forbidden region" (last 24 samples of the active region) are removed altogether.

2. Every timestamp that was incorrectly set must be exactly equal to the last timestamp previously found in this channel (since the timestamp was never updated). Note that while this is a necessary condition to identify incorrect timestamps, it is not sufficient. The inverse of this statement is obviously not true: Even in case of a correctly functioning CFD finder, a lot of events where the timestamp matches the previously found value simply by chance are to be expected. (In Fig. 4.12 this would correspond to the events lying "behind" the prominent horizontal streak, which was caused by the bug.)

3. To further identify timestamps which were likely produced incorrectly, one can take advantage of the fact that the SADC readout has two (in principle redundant) time-features: The cfdtime and the peaktime (see Fig. 4.13). If they agree within ±10 samples (= 500 ns), the cfdtime might still be valid, even if it fulfilled criteria 1. and 2. from above. Otherwise, the cfd feature is marked as invalid and is not used for any further analysis.

The impact of this correction can best be visualized by comparing both available time features as shown in Fig. 4.13. Looking just at the upper diagram, it seems as if both time features are in

disagreement to one another quite often. Ideally, one would expect just a diagonal line at y=x, with a pronounced maximum somewhere around (35,35). This "prompt peak" corresponds to the expected position of ordinary pulses which are correlated to the trigger. The fact that two very distinct bands (one horizontal, one vertical) can be seen in addition indicates that there is more to discuss. Apparently, both bands lie at the same respective x- or y-value of $\approx 30-40$ samples after the trigger. Because the cfdtime feature is "stuck" at its previous value in a significant number of cases, the expected distribution is overlaid with a distribution of essentially uncorrelated features. Since both features independently are still much more likely to be found at the prompt peak position, this gives rise to the observed band-like structures. Additionally, a region of 24 samples (which is precisely $2 \times$ `CFD_TOT`, see reasoning above) at the upper end of both axes is visible in which seemingly no sensible cfdtimes were found at all[19].
While the performed vetoing operation is obviously quite crude (the 10 sample (±500 ns) cut on the time difference for example can be easily spotted in the lower picture), the corrected plot still gives a lot more insight as to how the different time features compare to each other:
First of all, one notices the absence of entries in a wide band directly above the main diagonal, indicating that is is almost impossible to find a CFD zero-crossing 0-50 samples after a peak was found. This has an obvious explanation in the way the CFD-finder operates. By design, the CFD combines the region of steepest increase with the rather flat region around the maximum to achieve a steep zero-crossing of the CFD signal. So even if the region around the maximum is distorted (e.g. by pile-up), the CFD zero-crossing has very little possibility to happen shortly after a peak, since the CFD signal would be substantially below zero in almost all cases. The peakfinder on the other hand is very sensitive even to slight distortions near the peak. Especially in relation to the cfdtime, the peaktime can be delayed due to various reasons such as pile-up, noise, integer roundoff effects (see section 4.4.2.1) etc., which is why some entries lie to the right of the main diagonal. The same holds true for the small subset of crystal hits in which the cfdtime was found at the prompt peak position, whereas the peaktime was found at a later point in time (this refers to the horizontal band extending to the right-hand side of the maximum). It has the exact same explanation as the other entries to the right of the diagonal, but cfdtimes at the prompt peak obviously occur more often.
Additionally, it must be noted that the distribution of cfdtimes at the prompt peak position seems to be much narrower than for the peaktimes. Along with what was discussed before, this clearly indicates that in order to get accurate timing information, a CFD algorithm is indeed much better suited than a peakfinder[20].

4.4.1.2 Improvements

Further improvements improvements to the entire CFD algorithm in the firmware were made in [Ott22]. For completeness and without repeating too much of the explanations and discussions provided there, some key improvements shall be summarized:

- A few other bugs were found, one of which could have led to erroneous timestamps by accumulating enough time over threshold from incoherent regions in which the signal

[19] Both a peaktime and a cfdtime had to be present in order for an entry to show up in the histogram.
[20] Getting very precise timing information was never the main objective of the peakfinder anyways.

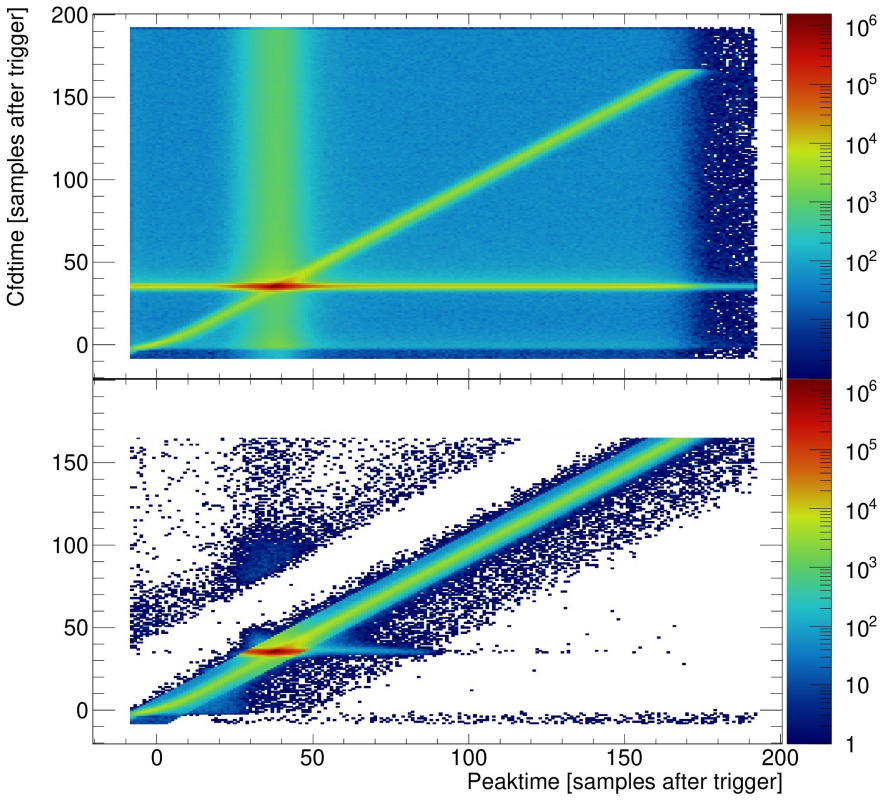


Figure 4.13: Cfdtime versus peaktime before (upper plot) and after (lower plot) the vetoing algorithm. Data from May 2018 beamtime. Values are uncalibrated.

was above the positive threshold. The problem was again, that the respective timer (`timer_high`) was not properly reset under specific circumstances.

- An interpolation algorithm was implemented, which allows for the determination of timestamps with sub-sample accuracy. Time resolution will be separately discussed in section 4.4.6.
- Multihit features were implemented (see section 4.4.3).
- Not related to the CFD algorithm directly, the improved baseline algorithm also has an effect on the CFD performance, since the "zero"-crossing is actually the point at which the CFD signal crosses the baseline.
- The usage of a FIR filter instead of the MA_{16} helped to improve the CFD's performance further, albeit to an almost negligible degree (see Fig. A.7 in the appendix).

4.4.2 Peakfinder

The peakfinder provides both an alternative energy measure in form of the peak height as well as an independent timestamp, the peaktime. As for the CFD, the extracted timing information should ideally be completely independent of the height of the incoming signal. After all, it is one of the advantages of the CFD to provide timestamps free of this so-called time-walk.

4.4.2.1 Time-walk of Peakfinder

Fig. 4.14 shows, that this is only approximately fulfilled in case of the peakfinding algorithm. The observed walk has to be attributed entirely to the algorithm, since the maximum position of the filtered waveforms as well as the CFD finder show no significant walk at all; see Fig. A.6 as well as [Ott22] (for the improved version). Two main effects are apparent:

1. Especially for intermediate to smaller pulses ($\widehat{=}$ small integrals), the peak is found at a slightly later position than for bigger pulses. This is immediately obvious from the way the peakfinder operates (see section 3.3.3.4): In order to find a peak, the value of the moving average has to *drop* below its previous value to mark the timestamp after rising for the required number of samples. Since the internal MA_{16} is subject to integer round-off, a plateau of identical values might exist around the actual maximum. This plateau has the tendency to be wider the smaller the initial pulse was, because the absolute rate of change (Δ ADC-channels / sample) decreases for smaller pulses. By design, the first drop will be found at the end of this plateau, which explains the delayed timestamps.

2. In case of very small pulses ($< 2\,\mathrm{MeV}$) however, the timestamps seemingly occur earlier again. This can be attributed to small fluctuations due to noise (which has a larger effect on the moving average for smaller pulses) which in turn means that the first drop in the moving average can occur near the beginning of the aforementioned “plateau”, causing a systematic bias towards earlier values.

4.4.2.2 Improved Version

The issues with the time-walk could be resolved recently [Ott22] (see table 3.1), which was accomplished by two changes.
On the one hand, the MA_{16} was entirely superseded by a FIR filter (see section 6.2). The effect of this change is twofold: As an intermediate result, Fig. A.7 (appendix) shows a direct comparison of peaktimes before and after the FIR-filter was introduced. Consistent with the explanation provided above, this indeed led to a mitigation of the early peaktimes for very small pulses (effect no. 2), because noise is better suppressed by the FIR filter. In addition, much more peaks are now found for smaller pulses as well.
At the same time, the problem of later timestamps due to the plateau (effect no. 1) was exacerbated[21], especially for the smallest pulses. A simple solution for this problem was implemented as well: Since the pulses are not too asymmetric in the vicinity of the maximum, it is to be expected that the actual peak should lie somewhere close to the middle of this plateau.

[21] Because the filtered values are less noisy, but still integers, the plateau gets (on average) longer.

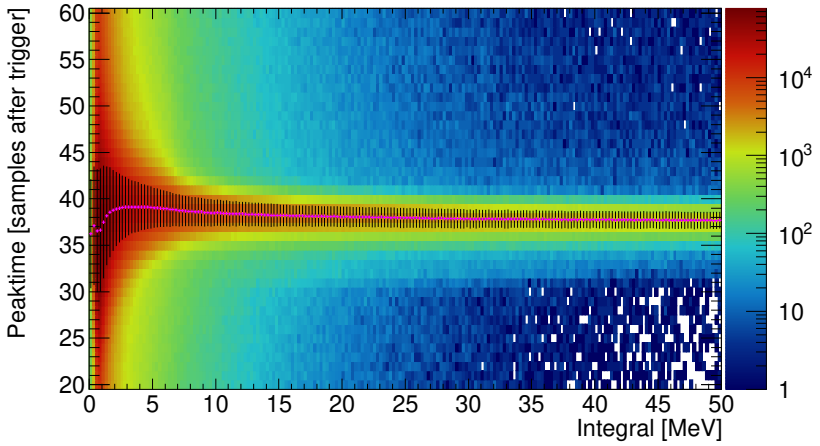


Figure 4.14: Time-walk of peaktime depending on the observed integral. Pink markers indicate the mean value of a gaussian which was fitted in a 3σ-environment around the MPV for each vertical slice. Black bars indicate the σ-width of the fitted function. Data from May 2018; i.e. with MA_{16}-filter and the initial peakfinder algorithm.

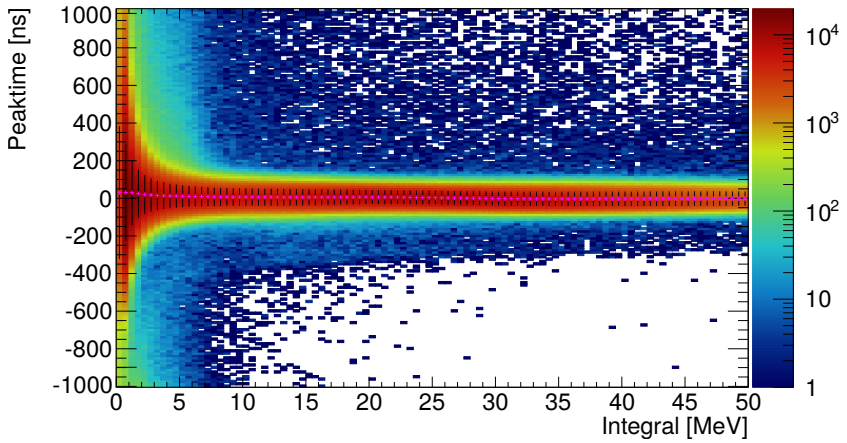


Figure 4.15: Time-walk of calibrated peaktimes depending on the observed integral. The improved version of the peakfinder (see section 4.4.2.2) was used. Pink markers indicate the mean value of a gaussian distribution which was fitted to each vertical slice. Black bars indicate the σ-width of the fitted functions.

Therefore, the latest implementation of the algorithm now calculates the width of this plateau and subtracts half of this width from the determined timestamp.
Fig. 4.15 shows the improvements from both of these adaptions: the observed time-walk could be (almost) completely eliminated.

Threshold Furthermore, another long-standing issue was fixed which had nothing to do with time-walk. Remember that the peakfinder used to depend on only one parameter, the number of consecutively rising samples. Since full SADC coverage was achieved, this posed a problem

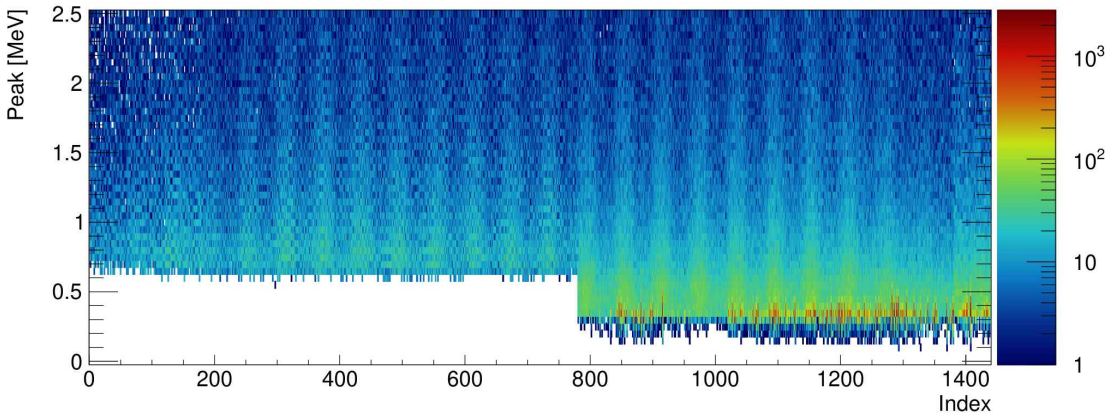


Figure 4.16: Peak values (determined without an explicit height threshold) against crystal index. Only a number of 20 rising samples was required in both up- and downstream. The data were obtained from a cosmic run, which explains the seemingly oscillating structure along x.

for the consistency between up- and downstream channels. Since the pulseshape is in principle identical (up to a scaling factor of 2), it makes sense to require the same number of rising samples in the upstream as in the downstream[22]. On the other hand, this effectively imposes an energy threshold on the upstream values which is only half as large as in the downstream. This resulted in more low energetic (possibly unwarranted) peaks found in the upstream, as can be seen in Fig. 4.16. The peak features were converted to an energy equivalent by using the calibration factors for the integral and the expected ratio (see section 3.3.4) between peak and integral feature. With this conversion, 20 ADC channels in the downstream correspond to an energy threshold of approximately 0.65 MeV, which seems to be nicely obeyed. However, the according threshold ($\approx$ 0.325 MeV) in the upstream seems to be too low. Some, presumably noisier, channels show a significant number of found "peaks" below 0.5 MeV. By the introduction of another, explicit threshold parameter which the peak feature has to clear before being accepted, this problem was easily solved. For the downstream, the value of this threshold was set to "20" as well, in order to remain consistent with data recorded so far. In the upstream, the value was accordingly set to "40", which now equates to an energy of 0.65 MeV as well (see Fig. A.8).

4.4.3 Multihit Features

It can easily happen that more than one pulse is present in the sampling window. In that case, it would be desirable to have a multihit peakfinder (same for the CFD finder) which gives an individual timestamp for every pulse. The TDCs in the existing setup were already operated in a multihit mode, but the SADCs were not.
As will be discussed in the following chapter 5, the lack of multihit capabilities was particularly relevant in certain pile-up situations, which ultimately contributed to the need for an improved implementation: If the respective algorithm is only capable of recording a single timestamp, it

[22] The average (y-)difference between two adjacent samples is twice as large, but the *number* of rising samples remains the same.

should be prioritized to get the timestamp from the pulse correlated to the trigger[23]. Because the first detected timestamp is the one which gets recorded, the region in which timestamps could be found had to be limited to [0..200] samples after the trigger. This practically[24] ensures, that if a pulse at the "right" time is presence, its features will be registered. Expanding this region further towards earlier times would have run an increased risk of detecting timestamps from previous pulses, which would then have prevented the timestamp from the "relevant" pulse from being recorded.
In principle, it would have been straightforward to implement multihit capabilities in the SADC feature extraction from the start, since the algorithms are inherently capable of finding arbitrarily many peaks or cfdtimes. The reason why this was not trivial to implement is that in order to actually send multihit features over the network to the saver, the structure of the feature-packets would have had to be changed. Previously[25], this structure had always remained static and with a fixed size; reserving space for exactly one feature of each kind per channel. This changed in March 2022 (see Table 3.1); the firmware is now capable of handling up to two timestamps of each kind [Ott22].
As a result, this now allows for the detection of timestamps before the trigger, extending the overall sensitive range to [-10..+10 µs] relative to the trigger. However, to avoid similar problems to before, the first timestamp may only be found in the 10 µs (200 samples) before the trigger and (in case there are multiple such timestamps) the *last* such value is recorded. On the other hand, the second timestamp may only be found within 10 µs after the trigger, where the first value is taken as before.

4.4.4 Efficiency of Timestamps

Besides the accuracy of a certain method to determine timestamps, its *sensitivity* and *specificity* are equally important. A very precise algorithm would be of no practical use, if it only produced timestamps in e.g. half of the cases or gave a timestamp for every signal, even if no actual pulse was present. An ideal method should be able to determine a timestamp for each and every pulse, provided that its height is above a certain threshold which is determined by the noise level.
Fig. 4.17 shows the relative percentage of signals of a certain energy for which a valid timestamp was found. Ideally, this percentage should follow a step function, which jumps from 0 to 1 at the threshold value. The latest versions of CFD and peakfinder are capable of finding more than one timestamp per crystal hit (see section 4.4.3 (Multihit Features)), which is why the sensitivity occasionally exceeds 1 [26]. For the peakfinder, this threshold is given by the required number of rising samples (or, in the newer implementation, an actual threshold). The default value for this is 20, which (see section 4.4.2.2) corresponds to an energy of about 0.65 MeV. This is indeed nicely fulfilled in the sense of a sharp increase at this value, and above 1.0 MeV the probability to

[23] Provided of course that there was such a pulse in the respective channel.

[24] Notice, that due to the differences between energy- and timing-branch (upon which the trigger signal is based) both cfd- and peaktime of a pulse correlated to the trigger are detected ≈ 30–40 samples (1.5 to 2 µs) after the trigger arrives (see e.g. Fig. 3.9).

[25] i.e. before the zero-suppression was implemented directly on the FPGAs, which also required the structure to be modified.

[26] The cosmic runs used to create this diagram were quite short, which is why some bins show values above 100%, even though the pile-up probability is generally quite low for cosmic particles.

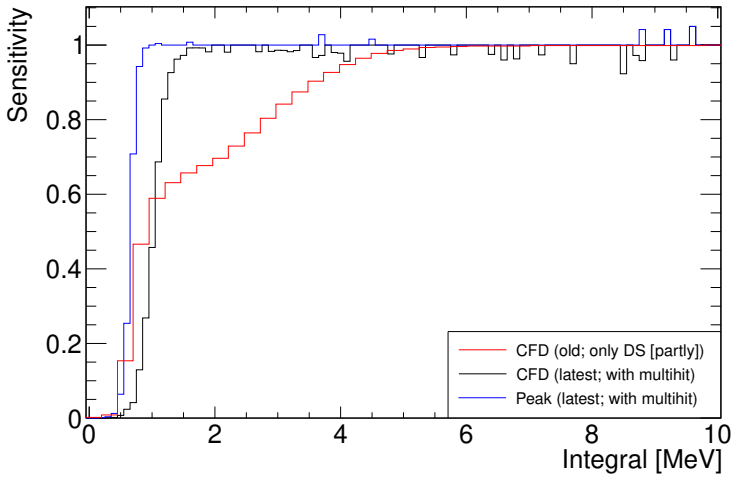


Figure 4.17: Sensitivities of different timestamp generation methods. The "old" CFD version refers to before the bugs in the algorithm were fixed; likely invalid CFD timestamps were excluded by the vetoing algorithm. Data is from cosmic runs with inherently very low pile-up probability (new algorithms) or explicitly without crystal hits marked as pileup from the May 2018 beamtime (old CFD algorithm).

have a crystal hit without a valid peaktime is indeed already close to zero.
Regarding the cfdtime, it is difficult to give a precise threshold value, due to the way the algorithm works. The threshold which the computed CFD signal has to clear in order to be accepted is typically chosen to be quite low[27]; most of its "restrictiveness" comes from the requirement to stay above the positive threshold and below the negative threshold for a longer time (see section 3.3.3.5). By estimations based on an extracted, mean pulseshape, the parameters were chosen such that the effective threshold lies around 1 MeV, which seems to fit nicely with the observed behavior for the latest implementation. In case of the old version, the aforementioned bugs in the firmware have prevented a lot of timestamps from being found, which results in a considerably lower sensitivity, especially in the lower energy region. The correction method detailed above only invalidates timestamps that are likely incorrect, but it cannot restore timestamps that were not found in the first place. With the bug, CFD timestamps are reliably found only above 5 MeV.

4.4.5 Time Calibration

To reconstruct physical events, timing information is almost as important as the energy information. Without proper time-information, it is impossible to tell whether a detected energy deposit was actually correlated to the trigger or not (especially important for time-clustering [Sta24]). In order to properly utilize the SADC timing information, the raw values have to be calibrated; preferably in a way which is consistent to the established timestamps from the TDC readout. The calibration is done by the following formula:

$$v_{\text{calib}} = v_{\text{raw}} \times \text{gain} + \text{offset} \tag{4.6}$$

[27] Default: 4 (DS) or 8 (US) ADC channels.

In contrast to the calibration of energy features, the gain (calibration factor) for the time features is already known. Since the original sampling rate is 80 MHz (the default sampling rate of 20 MHz is obtained after decimation; see section 3.3.1), the natural base unit in which timing information is encoded is 12.5 ns. However, since the extraction of timing information is based on the sampled waveforms, the actual features were usually discretized in steps of 50 ns[28]. This changed recently when interpolation capabilities were added [Ott22]. In order to represent timestamps based on the 20 MHz sample data with 32 subsamples (2^5), the base unit had to be reduced by a factor of 8 to 50 ns/32 = 1.5625 ns.
What is non-trivial is the determination of individual *offsets* of the timestamps. The value by which the SADC timestamps are represented is always to be understood in relation to the trigger. Due to the substantial rise time of pulses in the energy branch (see e.g. Fig. 3.9), typical timestamps of pulses correlated to the trigger occur several microseconds later. The values of the TDC readout on the other hand are calibrated to a value of $t_{\text{calib}} = 0\,\text{ns}$[29] for such pulses.
Accordingly, the task of calibration effectively comes down to determining the prompt peak position and subtracting this as offset from the values. This is easily achieved by analyzing recorded data and finding the "center" of the resulting distribution of timestamps for each channel. Since the straightforward MPV of any given distribution is (depending on the available statistics) subject to random fluctuations and the arithmetic mean is sensitive to outliers and depends on the chosen range of values, another method is used which combines the advantages of both approaches: the truncated mean of a range of values; chosen to be centered around the MPV. A third alternative would have been to fit each distribution with a gaussian and using its mean value μ as the center. This might have worked as well, but is inherently error prone should the fit not converge due to inappropriate start values or a lack of statistics.
While this calibration itself is not really noteworthy, one peculiar discovery was made in the process of this calibration: After a power cycle, some channels showed relative shift of ±12.5 ns in their relative prompt peak positions. While some progress was made in this regard (see [Ott22], p. 29-31) the ultimate reason for this behavior has not yet been understood. This however entails that for a precise timing calibration beyond ±12.5 ns, the calibration process has to be repeated after every power cycle; - a situation which is far from ideal[30].

4.4.6 Time Resolution

The most important measure upon which the performance of the CFD algorithm might be tested is the time resolution. Of particular interest is the question whether the CFD timestamps can provide a better or comparable resolution than the TDC readout.
In order to directly measure the resolution power of the algorithm itself, a reference signal of variable size which occurs at a precisely known point in time would be needed. Since such a signal is not readily available in the experiment, the resolution might be easily judged by the width of the prompt peak, i.e. the distribution of timestamps from pulses that are correlated to the trigger. It should be noted that the trigger signal itself is based upon information from the TDC branch of the readout. So the actual resolution of extracted timestamps (which are always

[28] This implies, that only every 4th numerical value actually occurred.
[29] The time at which photons from a reaction reach the detector is calibrated to 0 ns.
[30] As of finalizing this thesis, the issue seems to have been fixed [Ott24].

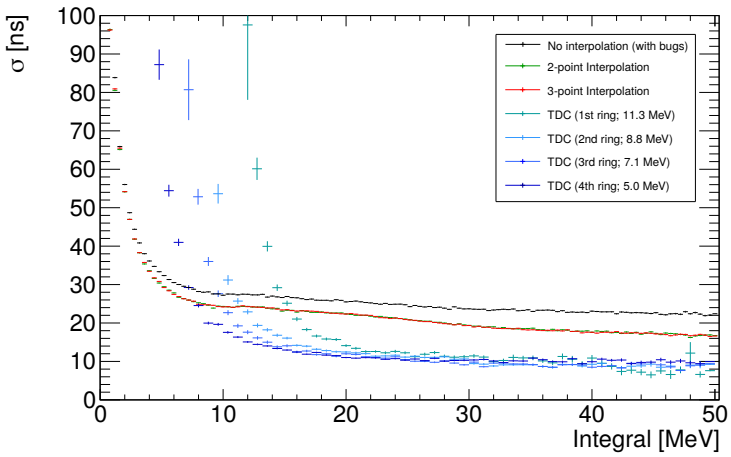


Figure 4.18: Time resolutions (sigma of gaussian fitted to prompt peak) of different iterations of the CFD algorithm. The resolution (same method) of the TDC readout is also shown; broken down into different rings of the detector with different discriminator thresholds. Threshold values taken from [Sta19].

given relative to the trigger) gets convoluted with the resolution of the trigger signal itself if the prompt peak width is considered. However, for a relative comparison of both methods, this does not matter.
Fig. 4.18 shows this comparison. Two conclusions are apparent:

1. The time resolution of the CFD algorithm improved with each iteration of the algorithm (as expected), although the improvement of a 3-point over a 2-point interpolation seems to be negligible.

2. The time resolution of the CFD algorithm is better than the resolution of the TDC readout for small energies.

The energy below which the CFD algorithm offers superior time resolution depends on the region of the detector and seems to lie between $\approx 8\,\text{MeV}$ and $15\,\text{MeV}$. The TDC thresholds (in this case, only the "A" threshold was used) were set to different values depending on the Θ-index in the detector (see section 2.8.1). Evidently, no timestamps are generated below these thresholds and the time resolution for pulses with heights close to threshold is severely degraded.
Furthermore, the resolution with which the trigger signal is *sampled*[31] (160 MHz / 6.25 ns; see section 3.3) seems to be sufficient to not have a significant impact on the achievable time resolution (e.g. $\sigma_{\text{CFD}} \approx 17\,\text{ns}$ at $E = 50\,\text{MeV}$).

[31] Not to be confused with the *resolution* of the trigger signal itself.

CHAPTER 5

Pile-Up Detection

5.1 Evaluation of Initial Approach (v1.0)

The approach of using the ratio between integral and peak feature presented in section 3.3.4 is obviously quite primitive. A few of its potential shortcomings were already discussed in [Sch16] when it was first proposed. At this point in time, it was pointed out that using just this one criterion would lead to some amount of false negatives in the data, i.e cases of actual pile-up that were not correctly identified as such. In particular, two possibilities for false negatives are conceivable:

1. If no peak was found[1], the waveform will never be flagged as pile-up since there is no ratio to evaluate.

2. If the relative height and position of two pulses follow a specific relation, it is possible to match the required ratio by accident (see Fig. 5.1 for an example).

A solution for the latter case was proposed in [Sch16] already, which will be presented shortly. However, the entire discussion as well as this "solution" have to be approached with care. Due to the lack of any real pile-up data at this point (i.e. before any part of the CB calorimeter was equipped with SADC readout), all conclusions were drawn from a simple simulation which investigated possible topologies of pile-up and the algorithm's response to such cases. The simulation was based on a few assumptions which later turned out to be partly unjustified or overly simplistic. On the one hand, it was assumed that every recorded event would consist of one primary[2] pulse which is correlated to the trigger. As we will see, this is not necessarily the case. On the other hand, only pile-ups in which the secondary pulse occurs *after* the primary pulse were considered. Furthermore, the peak height which would have been extracted from the combined waveform was not determined realistically. Instead, the plain (global) maximum of the simulated waveform was used, which differs from the actual implementation if two peaks are

[1] The specific circumstances under which this might occur will be discussed shortly.

[2] **Nomenclature:** The pulse correlated to the trigger will be referred to as the "primary" or "main" pulse (whether it was the first/most significant pulse within the sampling window or not) and the other pulse(s) will be called "secondary" or "pile-up" pulses. Note that this implies no temporal order; the secondary pulse could also occur before the primary pulse.

discernible in the final waveform and the first one is smaller than the second one[3].
Under these assumptions, a topology as shown in Fig. 5.1 resulted in an extracted peak/maximum position[4] of the total signal which was shifted from the normal peak position of the primary pulse (without the secondary pulse, the peak would have been found around 1.5 to 2 µs). So to solve the problem of false negatives, it was proposed to require all signals whose ratio was deemed unsuspicious to have a peak position within some tightly chosen limits in addition [Sch16].
In preparation for the first production beam time in May 2018, it was decided to drop this second criterion. On the one hand, it was suspected that the resulting algorithm could be too restrictive overall and produce a lot of false positives. On the other hand, the second criterion would have required choosing suitable limits for the position of the peak as well (in addition to the limits for the ratio), which was at this point in time not studied well enough to justify a particular choice. Lastly, an incomplete pile-up detection is still far better than no pile-up detection at all. The established QDC readout was running in parallel anyways, with full coverage of the entire Crystal Barrel.
During and after this first beamtime, it became obvious that this method of pile-up detection still produced a large amount of false positives. Fig. 5.2 shows the topologies of events which were flagged as pile-up by just using the aforementioned ratio (enumeration corresponds to the figure):

1. Cases of actual pile-up. These can be split further into three categories:
   (a) The secondary pulse sits on top of the primary pulse and contributes to the integral. Depending on height and position of the secondary pulse, the detected peakheight might be unaffected or enhanced as well.
   (b) The primary pulse itself sits on top of the secondary pulse, whose falling edge is still present within the integration window. With the old baseline algorithm, the baseline which is extracted will be too high; see section 4.2.1).
   (c) More than one pulse is present, but neither is correlated to the trigger.
2. The pulse is displaced, i.e. is not at the position expected for signals correlated to the trigger. Since the peakfinder can find peaks anywhere within its window of sensitivity, the extracted peakheight will be the actual height of the peak, whereas the integration window has a fixed relation to the trigger. The integral value will therefore be smaller than anticipated, which results in a ratio that is too small.
3. A particle deposited so much energy in a crystal that it exceeds the dynamic range of the readout. In that case, the pulse is clipped off at the top, the extracted peakheight is therefore too small for the integral and hence the ratio is too big[5].
4. The sampled waveform deviates so much from the ordinary pulse shape, that the integral is considerably smaller than expected for a given peakheight.

[3] In that case, the actual peakfinder would return the features of the first pulse, whereas the height of the global maximum will be dominated by the second, larger pulse.
[4] In this case it does not matter, as the final waveform exhibits only a single discernible peak.
[5] The maximum available energy of a single beam photon is 3.2 GeV, while the dynamic range is 2.2 GeV (DS) and 1.1 GeV (US). Taking into account the reaction kinematics and the fact that typical energy depositions are spread out above several crystals, such cases are very rare and of no practical concern

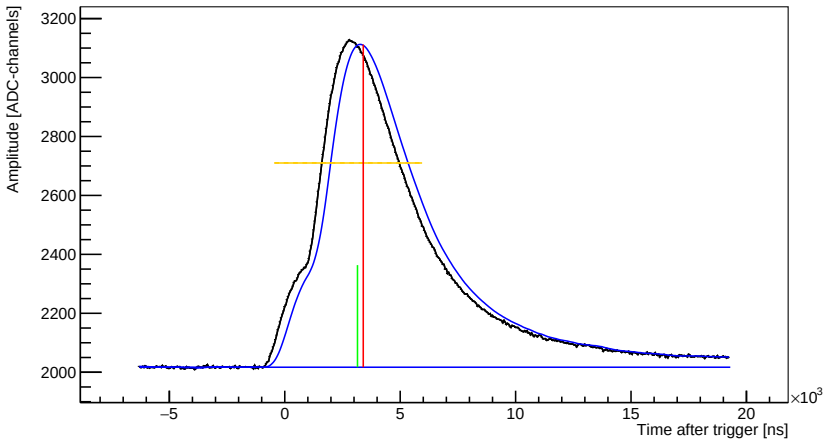


Figure 5.1: Example of a pile-up which was not detected based on the ratio-criterion (recorded in June 2022). The extracted (raw) features were 2017 for the baseline, 3112 for the peakheight and 2710 for the integral (uncalibrated values given in ADC channels), which results in a ratio of $\frac{2710-2017}{3112-2017} = 0.633$. The integral corresponds to a calibrated energy of 35.45 MeV

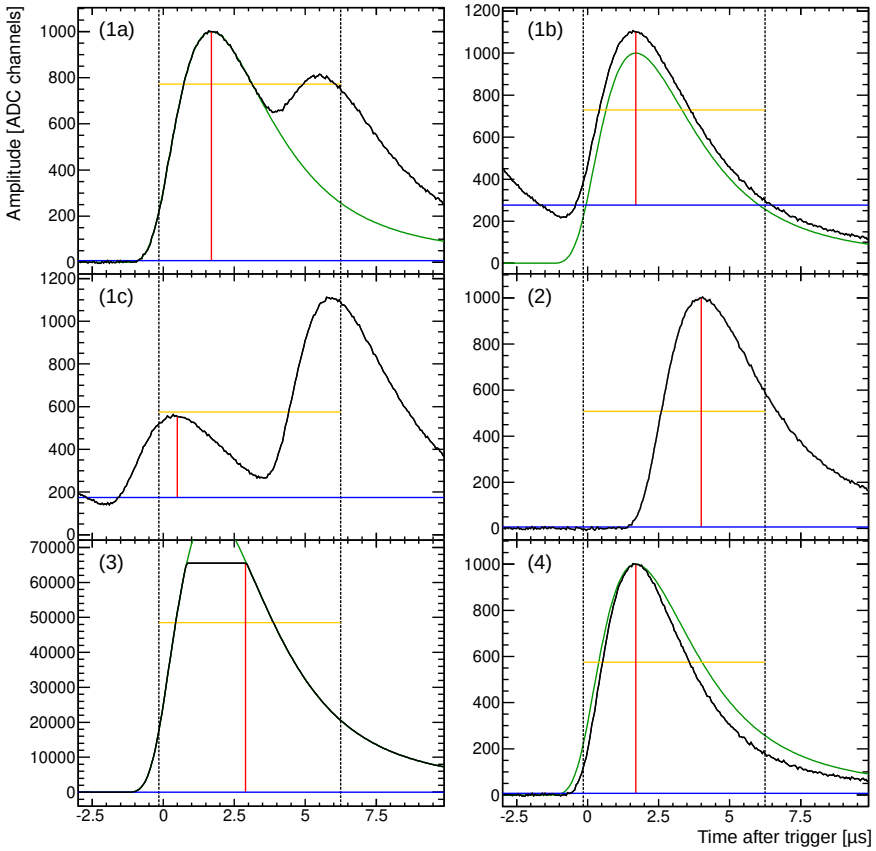


Figure 5.2: Mock-up examples of the different pulse topologies marked as pile-up. These were artificially created solely for illustrative purposes. The vertical black lines denote the integration limits, the red line the peak feature, the integral value is depicted in orange and the extracted baseline (old algorithm) in blue. The green curve shows how an undisturbed signal correlated to the trigger would have looked like (where applicable).

The most relevant cases for this treatise will be cases 1a) and 1b). It is only in those cases that a secondary pulse within the integration window actually leads to a deterioration of the data of interest, i.e. the features of the primary pulse. In order to get such cases, a secondary pulse has to coincide with a primary pulse within a time frame of a few µs. Considering that over 90% of crystals do not see a significant energy deposition in a typical event (see Fig. 4.10), it makes intuitive sense that secondary pulses will occur more often on their own than together with a primary pulse. This implies, that case 2) should be (and indeed is!) the dominant contribution to the overall number of alleged "pile-ups" found with the initial approach. Obviously, such cases are not in need of any recovery whatsoever and are therefore the dominant source of false positives[6]. In addition, a substantial number of waveforms fall into category 4) and were erroneously flagged as pile-up as well. Such cases will be discussed in depth in section 6.5.

5.2 Enhanced Algorithm (v1.1)

As discussed in the previous section, most of the signals flagged as pile-up with this algorithm do not actually contain more than one pulse, but only a single pulse that is simply uncorrelated to the triggering event. Since the entire waveform is saved for offline recovery in cases of alleged pile-up, this situation is highly undesirable as a lot of disk space and network traffic will be occupied by these false positives. Unfortunately, this was the case for the first two beamtimes[7] recorded with (partial) SADC coverage.
So a solution to identify such cases both in offline analysis and (for following beamtimes) during data taking had to be developed. To keep the computational overhead small, the solution still had to be based upon the analysis of feature data alone.
The idea is simple: Assuming a constant pulseshape and the presence of just one pulse in the sampled waveform, the integral should depend only on the peakheight H and – since the integration window is fixed in time – the peakposition P of this pulse. So an *expected integral* I_{exp} should be determinable as a function of H and P. The so calculated expected integral might then be compared to the *observed integral* I_{obs} to determine whether pile-up was present or not.
The actual function to estimate the expected integral from the peakposition can be calculated in the following way: The integral feature of a specific pulse is simply the value of a 128-fold moving average, evaluated at a fixed point T_0 in time[8]. But given a sampled waveform, it is also possible to evaluate this MA_{128} at any other point T in time. If $T < T_0$, this corresponds to the situation in which an earlier part of the waveform would have been inside the integration window and vice versa. Shifting the integration window like this (by $T - T_0$) is equivalent to keeping the integration window fixed, but shifting the signal in the opposite direction (by $-(T - T_0) = T_0 - T$)[9].
Applying this idea, Fig. 5.3 shows an overlay of the MA_{128} signals for all sampled waveforms,

[6] While they are in no need of a pile-up correction, they should not be used for reconstruction of physical events, as they are not correlated to the trigger (see time-clustering [Sta24].
[7] May 2018 and October 2018.
[8] T_0 is the point 128 samples after the begin of the integration window, which is fixed relative to the trigger.
[9] To predict the integral for a pulse peaking at a time which is "too late", the integral of a correctly timed pulse has to be evaluated deliberately "too early".

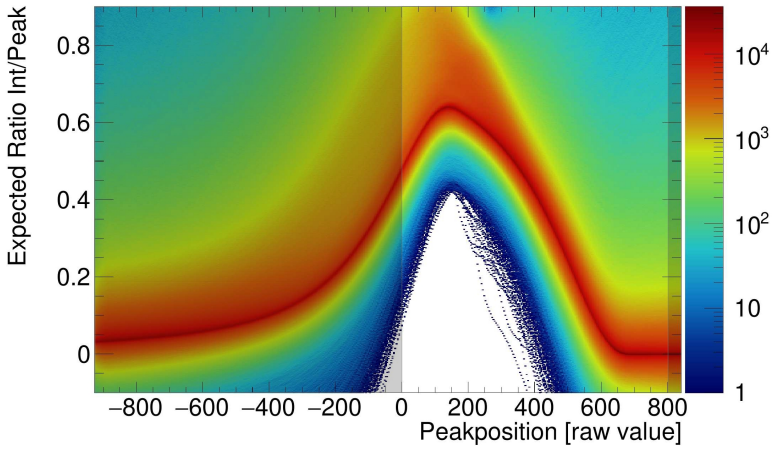


Figure 5.3: Time-inverted and shifted calculations of the 128-fold MA of the sampled waveforms, normalized by the respective peak feature. To ease the lookup operation, the peakposition is given as a raw value (as it would appear in the feature packet), i.e. in units of 12.5 ns after the trigger. The greyed out regions denote areas where no peak can be found. (At the time of development, multihit timestamps were not yet implemented and therefore also no timestamps before the trigger.)

but inverted in time[10], normalized by their peakheight and shifted according to their respective peakposition. Similar to the extraction of the CB reference pulseshape (will be described in section 6.3), the most probable value for each x-bin can be determined from this plot. These values can then be used to form a Lookup Table (LUT), from which the expected ratio R_{exp} can be obtained for any peakposition of a given signal. Multiplying this ratio by the peakheight H then gives the expected integral I_{exp}.
Without pileup, the expected and the observed integral should be identical:

$$I_{\text{exp}} = H \cdot R_{\text{exp}} \stackrel{!}{=} I_{\text{obs}} \tag{5.1}$$

Obviously, this equality will only hold within some margins of tolerance. It seems natural to assume a constant allowable deviation (θ_{abs}) and in addition some fixed percentage (θ_{rel}) of the total observed integral value by which it might deviate in addition:

$$-\theta_{\text{abs}} - I_{\text{obs}} \cdot \theta_{\text{rel}} \;<\; I_{\text{obs}} - I_{\text{exp}} \;<\; \theta_{\text{abs}} + I_{\text{obs}} \cdot \theta_{\text{rel}} \tag{5.2}$$

5.2.1 Comparison to Old Approach

It is worth noting, that the parameterization of acceptance thresholds presented in section 3.3.4 can be brought to a similar form as equation 5.2. Multiplying equation 3.1 by H yields:

$$-s_{\text{lower}} - H \cdot \theta_{\text{pu}} \;<\; I_{\text{obs}} - R_{\text{const}} \cdot H \;<\; +s_{\text{upper}} + H \cdot \theta_{\text{pu}} \tag{5.3}$$

[10] To account for the "inversion" between peaktime and position of the integration window

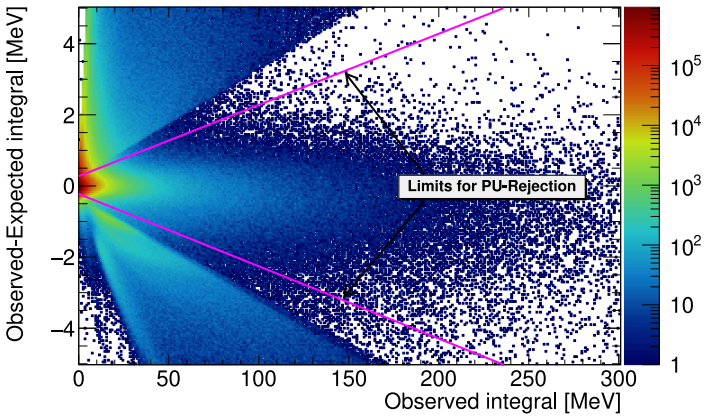


Figure 5.4: Difference between observed and expected integral (based on a signals peak features) for all signals initially marked as pile-up. Magenta lines denote the limits applied for the pile-up rejection.

Identifying the slope parameters with θ_{abs}, θ_{pu} with θ_{rel} and R_{const} with R_{exp} reveals the similarity. The only real difference is that in the previous approach, the independent variable was the peakheight H, whereas it is I_{obs} in the new approach.
From this angle, the main issue with the previous approach becomes apparent: Checking whether the observed ratio R_{obs} lies at a constant value $R_{const} \approx 0.64$ essentially reduces R_{exp} to a constant; independent of the peakposition P. Looking at Fig. 5.3, this is obviously only true for pulses correlated to the trigger, since those peak around a raw value of 140 (1.75 µs after trigger), i.e. precisely where the depicted function reaches its maximum of $y \approx 0.64$. So the other dimension of this problem, i.e. the peakposition, was implicitly omitted in the previous approach[11].

5.2.2 Removal of False Positives

For the already recorded data, it is simple enough to check every alleged pile-up whether its observed integral can be fully explained by its extracted peakheight and the predicted ratio for its peakposition. Fig. 5.4 illustrates this process. The apparent edges in the distribution come from the thresholds in the initial pile-up detection[12]. The absolute threshold θ_{abs} was set to a value of 5/10 (DS/US) ADC channels (≈ 0.25 MeV), which corresponds to the default value of $s_{lower/upper}$. The previous threshold θ_{pu} of 0.02 has to be divided by $R_{const} = 0.64$ to obtain a corresponding value of θ_{rel} of $\approx 3\%$. To avoid edge-cases and only remove pile-ups which were undoubtedly false positives, the actual value of θ_{rel} was chosen smaller, at 2% allowed deviation.

5.2.2.1 Pile-Up Rejection in LEVB

Ideally, the enhanced algorithm (the "LUT-approach") would have been implemented in the CBSADC-firmware right away. However, the actual implementation proved challenging at first.

[11] Under the initial assumptions in [Sch16] (see previous section), this would have been less of a problem as there would have always been a primary pulse correlated to the trigger.
[12] See reasoning above why they are functionally also linear functions in this representation.

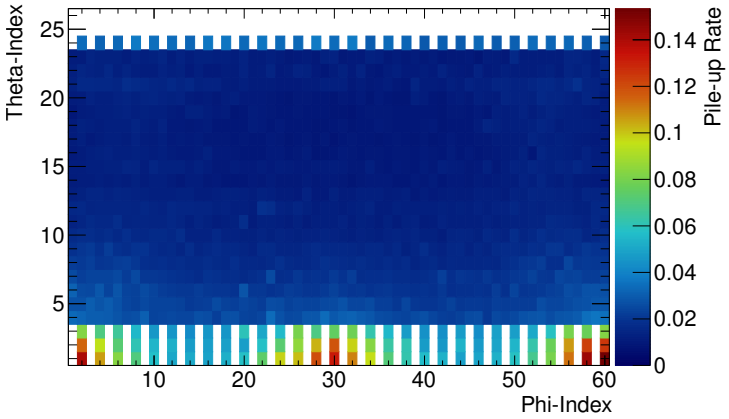


Figure 5.5: Rate of events flagged as pile-up after removal of false positives in the LEVB. For the rate calculation, all hits with an energy $E > 0.5\,\text{MeV}$ were considered. Data taken from June 2021 beamtime.

Since the "ratio-approach" was expected to be strictly more restrictive than the other algorithm, it was decided to employ a two-staged approach instead: The firmware was left unchanged, but the check from section 5.2.2 was moved to the LEVB, which sits in between the SADC modules and the data saver units (see section 3.3 and also Table 3.1). If a false positive is detected in the LEVB, the pile-up flag is removed and the according waveform is not written to disk. This requires the appropriate LUT to be available while the data is taken, which in turn has to be updated after every change to the feature extraction or the pulses shape and/or position.

5.2.3 Performance

With this configuration (which was in place since May 2021, see Table 3.1), i.e. a primary check of the simple ratio in firmware and a LUT-based, secondary check in the LEVB, it was expected that most false positives were filtered out and that only a very small number of pile-ups went undetected. At least this was the result of the preliminary analysis of the May 2018 beamtime data, which followed precisely the same approach[13].
Significant amounts of data (in particular the beamtimes in June 2021 and November/December 2021) were taken with this two-staged pile-up detection setup. While $\approx$ 14% of all hits (energy above 0.5 MeV) were initially marked as pile-up, this was reduced down to an overall rate of 3.4% by the pile-up rejection process. As Fig. 5.5 shows, the pile-up rate is distributed very unevenly over the whole detector. While most crystals have a detected pile-up rate below 2%, the rate goes as high as 15% in some of the forwardmost crystals[14]. Actually reconstructing all individual pulses from the corresponding waveforms (see section 7) confirmed, that those events are indeed predominantly ($>$ 80%) cases of actual pile-up[15].

[13] The only difference was that the secondary check was now done directly in the LEVB (while the data was being taken), whereas it was previously performed during offline analysis, i.e. after the data was already taken.

[14] The distribution along the azimuthal direction Φ is a result of the transversal magnetic field of the target, which deflects particles from electromagnetic background reactions (which presumably cause a lot of the pile-up) into opposing direction, depending on their charge. The distribution is not entirely symmetrical, because in addition to e^+e^- production, Compton scattering plays a significant role as well.

[15] i.e. waveforms which consisted of more than one pulse. Of the remaining "false positives", a substantial fraction ($\approx$ 40%) were caused by narrower pulses.

5.2.3.1 False Negatives

Due to the seemingly promising performance, it came as a surprise when a parallel analysis of N. Stausberg [Sta21] [Sta24] revealed that this approach seemed to miss a significant number of pile-ups. In his analysis of the June 2021 beamtime, he used multihit TDC data for an independent, albeit much less sensitive way of pile-up detection. It was found that while this method missed far more pile-ups (only $\approx 5\%$ of cases detected by either method were detected using the TDCs[16]), there is a small but significant number of cases ($\approx 2.5\%$) which were apparently missed by the SADC pile-up algorithm. Closer investigation revealed that these seemed to be predominantly cases in which the primary pulse sat on top of a preceding pulse (c.f. (1b) in Fig. 5.2). Looking ahead to the next chapter, Fig. 5.6 depicts the reconstructed "topology" of those cases in a way which is equivalent to the upcoming simulations[17].

[16] This is largely a result of the high TDC thresholds. For low energy pile-ups, there is simply no TDC data available from which a pile-up could be detected.

[17] The diagram was produced as a cross-check of the results of the simulations. To ease understanding, it is advisable to read the following section first.

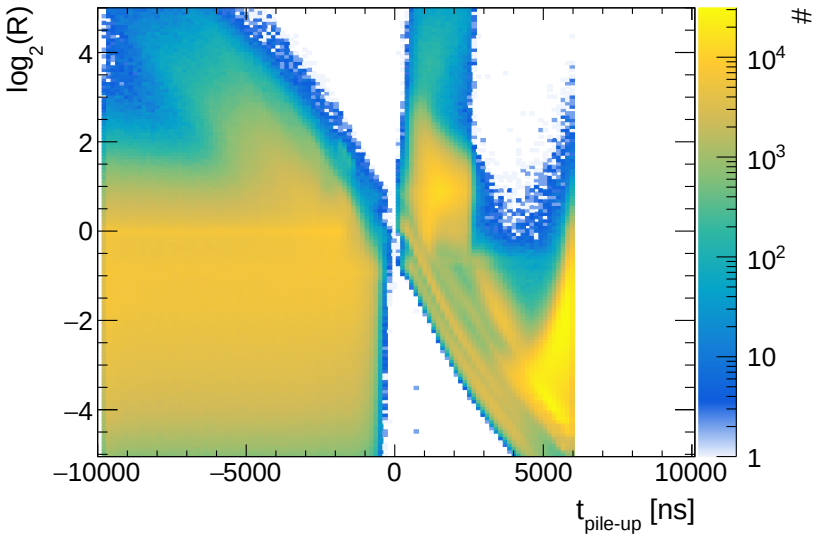


Figure 5.6: Reconstructed topology of false negatives in which a "pile-up" (2 peaks) was detected based on TDC information, but not by the SADCs [Sta24]. The diagonal bands in between 0 ns and 5 000 ns are suspected to be artifacts of the detection method (see [Sta24]). The x- and y-axes of the histogram are (despite slightly different nomenclature) functionally identical to Fig. 5.9 (see there). However, it should be noted that in contrast to the simulation (where every possible topology was tested), the histogram is not populated uniformly: Smaller pile-up pulses are inherently more probable than large ones, which results in the upper part of the histogram being only sparsely populated. Data is taken from the June 2021 beamtime.

5.3 Detailed Investigation

Faced with this unexpected discovery, a more thorough investigation of possible pile-up topologies and the sensitivity of the respective methods to such cases was performed. Similar to the initial studies in [Sch16], but this time more in-depth and under more realistic assumptions.
With a good understanding of the extracted features (see previous chapter) and knowing what pulseshapes to expect (see section 6.3), it is possible to artificially construct arbitrary pile-up topologies and study the expected response of the pile-up detection algorithm under investigation. This time, a realistic reproduction of the actual peak finding algorithm was used.

5.3.1 Simulation

To keep the simulation manageable and the results understandable, only pile-ups consisting of two pulses shall be studied[18]. The pulses were generically labeled "A" for the primary/main pulse as before and "B" for the secondary/pile-up pulse. In principle, the parameter space for all possible combinations of two pulses is 4-dimensional. Both pulses have a height $H_{A,B}$ and a position/time $T_{A,B}$ which is independent of the other parameters. To exhaustively study all possible pile-up scenarios, the height of both pulses would have to be varied from 0 to 2^{16} ADC channels and their position (in relation to the trigger) from $-\infty$ to $\approx 7\,\mu s$[19]. The complexity can be reduced by two observations:

- Only cases in which there is a primary pulse at $T_A = 0\,\mu s$, i.e. correlated to the trigger, are relevant for the investigation of false negatives. If no pulse was correlated to the trigger, the waveform is of no particular interest for the following analysis anyways and will be sorted out in the time-clustering [Sta24].
- The overall scale of the simulated pile-up waveform is irrelevant[20]. Only the size of the secondary pulse in relation to the primary pulse is relevant. For practical purposes, the studied range was limited to $-5 < \log_2\left(\frac{H_B}{H_A}\right) < +10$.

With those, the problem can be reduced to two dimensions. Fig. 5.7 shows the resulting ratio for all possible pile-up combinations in the studied range of parameters. Fig. 5.8 provides illustrative examples for different topologies marked by magenta points in the diagram. Several observations can be made:

1. There exists a large empty area on the top left, i.e. for large pile-up pulses which occurred before the main pulse. This is a result of the ratio not being evaluated in the first place, which can have two possible reasons:
   a) No peak could be found in the waveform. This happens if a peak either occurred too early to be detected, i.e. before the trigger ($T_B < -1.8\,\mu s$ in this diagram) and thus outside of the region in which the peak finder is sensitive, or the required number of rising samples (20) was not met.

[18] Two pulses are by far the most probable case; see section 7 and especially figures 7.6 and 7.8. Furthermore, the probability to produce false negatives is inherently much lower with three or more pulses than in the two pulse case.
[19] Pulses with a (calibrated) peaktime $T > 7\,\mu s$ do not contribute to the integral anyways.
[20] The overall energy (or height) of the resulting waveform is relevant for energy-dependent thresholds, but not for the "topology" itself.

b) The integral value was smaller than the extracted baseline, which is a result of the old baseline algorithm being used (see upcoming section 5.3.2).

2. The region in which the Integral/Peak-ratio lies close to the expected value (green area) can be subdivided into three distinct topologies:

   a) Areas below and to the right of the lowest, dotted black line. In these areas, the influence of the secondary peak is miniscule (<3%) either way. On the right edge of the diagram, the secondary pulse is too late to contribute anything to the integral.

   b) If the secondary pulse lies very close to the primary pulse, i.e. between $\pm 1.5\,\mu s$ for small pulses and $\pm 0.5\,\mu s$ for the biggest pulses, the ratio lies very close to the expected value. This is immediately obvious, since the resulting waveform is practically indistinguishable from a single pulse.

   c) Around an amplitude-ratio of $\log_2\left(\frac{H_B}{H_A}\right) \approx 1$ and T_B around 0 to 3 µs, a green "branch" is formed. Within this area, the expected ratio is reproduced by accident[21].

3. The rest of the r.h.s. of the diagram is either blue (ratio too low) or red (ratio too large), indicating a well-working pile-up detection in this area. A striking edge, running almost vertical along $T_B = 3\,\mu s$, can be observed. This is a result of the peakfinding process: To the right of this edge, both pulses are separated far enough such that the second pulse does not influence the peak of the primary pulse. Instead, only the integral is enhanced, which results in ratios which are firmly too large. To the left of this edge, the opposite is true: both pulses overlap and produce only one detectable peak. The ratio then depends on the relative sizes of both pulses. If $\log_2\left(\frac{H_B}{H_A}\right) < 1$, then the detected peak is still dominated by pulse A and the integral is enhanced more strongly due to pulse B than the peakheight. The situation for slightly larger pulses B was discussed in point 2c) already. Finally, for even larger pile-up pulses, both the detected peak as well as the extracted integral are dominated by pulse B. Since this situation is similar to a single, displaced pulse, the ratio will be too low[22].

4. The same is true for the l.h.s. of the top half of the diagram, where the resulting waveform is also dominated by pulse B. The rest of the blue area on the l.h.s. is largely a result of the baseline being too high: In a simplified picture, the preceding pulse can be considered to just enhance the baseline, without contributing anything to the peak or the integral[23]. If the undisturbed main pulse has a peak height H_A, its integral would be $0.64 \times H_A$. So since the integral is numerically smaller than the peakheight, subtracting a baseline $B > 0$ from both features will always impact the numerator of the ratio more strongly, resulting in a ratio which is too small.

[21] Fundamentally the same problem/topology which was pointed out in [Sch16]; albeit this time with a more realistic simulation.

[22] This is the same effect as for the previously discussed false positives.

[23] This is obviously an oversimplification (see section 5.3.2), but other effects are not relevant for the argument.

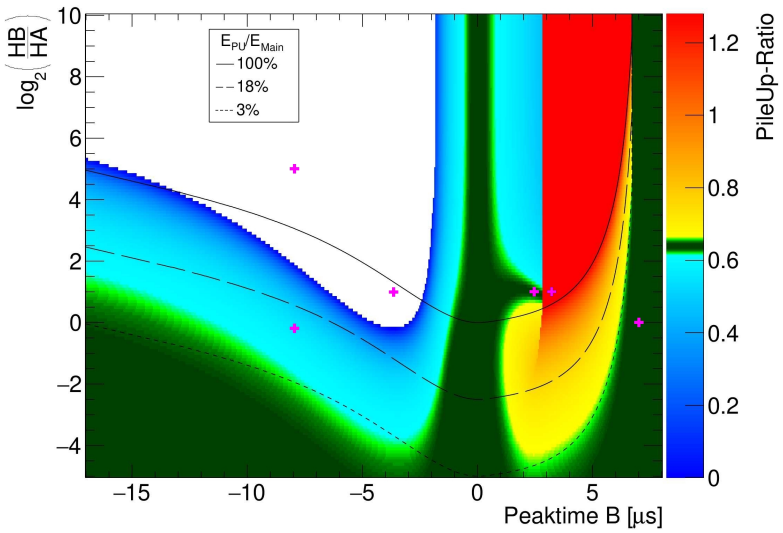


Figure 5.7: Ratio of integral to peak feature for simulated pile-up cases. The "old" baseline algorithm (see section 3.3.3.2) was used. The peak position ("Peaktime") of pulse A was fixed to 0 µs, i.e. perfectly correlated to the trigger. The resulting ratio is plotted color-coded for each simulated (x,y) bin. Dark green corresponds to the expected value of ≈ 0.64, whereas lighter shades of green were chosen for values still within ±3% of this value (corresponds to the energy-independent threshold value). The black lines indicate at which position the added energy by the pile-up pulse B equals a certain percentage of the energy of the main pulse A. The points marked in magenta correlate to example waveforms provided in Fig. 5.8.

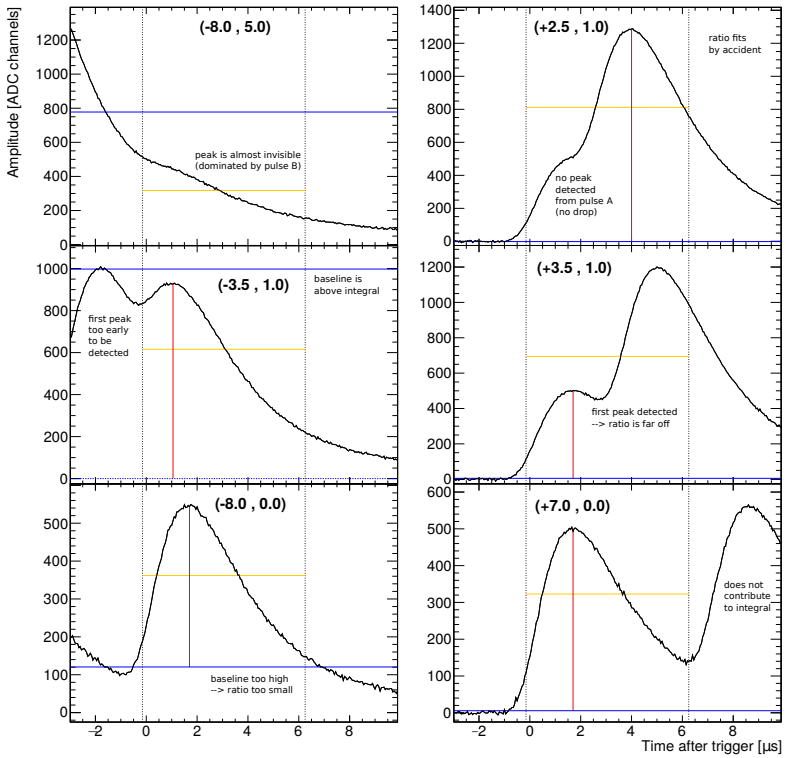


Figure 5.8: Example waveforms for certain topologies (marked as magenta points in Fig. 5.7). These were artificially created solely for illustrative purposes. The vertical black lines denote the integration limits, the red line the peak feature, the integral value is depicted in orange and the extracted baseline (old algorithm) in blue.

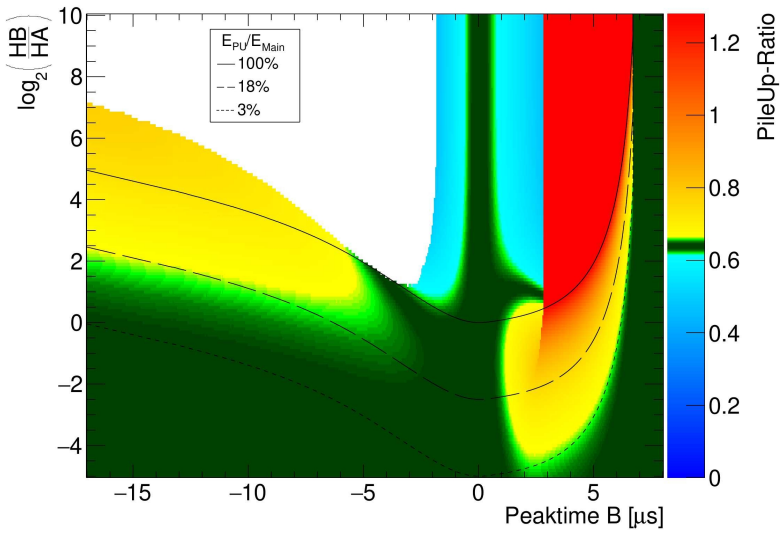


Figure 5.9: Ratio of integral to peak feature for simulated pile-up cases (same as Fig. 5.7) if the new baseline (box-)algorithm is used. Instead of implementing the actual algorithm, the simulated baseline was used; i.e. the algorithm was assumed to work perfectly.

5.3.2 Interplay with Baseline Determination

As just discussed, the baseline determination has a significant effect on the pile-up detection process, since both features (integral and peak height) require the baseline to be subtracted in order to be usable. The effect of an incorrectly high baseline can be severe, as the integral threshold required for the pile-up detection to be triggered (default value: 0.5 MeV) might not be surpassed. By applying the retroactive correction algorithm presented in section 4.2.1 and re-evaluating the ratio after the correction, this effect might be mitigated to some degree. But since the waveforms of hits not flagged as "pile-up" will most likely[24] not be available anymore, a pile-up correction can not be done.

It is for this reason, that the introduction of the new baseline-algorithm was expected to greatly reduce the area in which no pile-ups could be detected. Fig. 5.9 shows the result if the new algorithm is used. Unsurprisingly, the right-hand side is entirely unaffected, since in such cases the old baseline algorithm was working perfectly as well. Furthermore, the size of the "hole" on the top left is indeed significantly reduced. The hole was not closed entirely, because there are still topologies for which no peak will be found [25] due to the large size of the preceding pile-up pulse.

Despite these advantages, a major drawback is apparent as well: The area in which the ratio fits by accident was inadvertently greatly expanded. In addition to the entire lower left area of the diagram[26], a peculiar region which resembles the already discussed branch on the right-hand side is formed in the upper left quadrant as well. The reason can be illustrated by the middle

[24] Unless the event at hand was, by chance, a prescaler event (every 100th), for which the waveforms are written to disk in any case.

[25] (At least not within the default window in which the peakfinder algorithm operates.)

[26] "Only" the area above the dotted 3%-line is of actual concern.

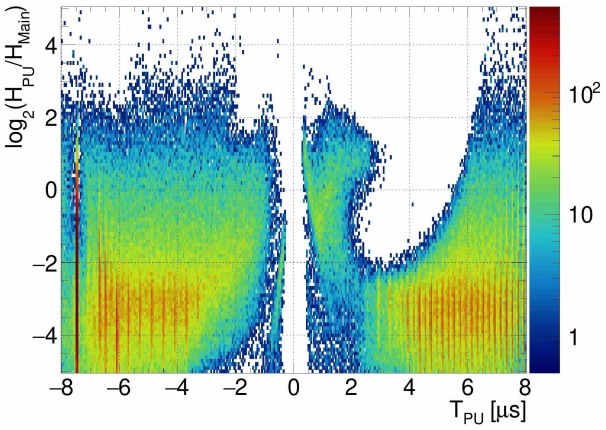


Figure 5.10: Reconstructed topology of prescaler-waveforms (June 2021) which contained exactly 2 pulses (as determined by the fit) and were not flagged as pile-up by the "old" detection algorithm. The main peak had to be around 0 µs. To avoid strong smearing due to the energy-dependent thresholds, a minimum (total) energy of 2 MeV was required. Similar to Fig. 5.6, the population of the histogram is again not uniform. The vertical bands (especially the one at −7.5 µs) are artifacts of the fitting procedure without further relevance for this discussion.

left example (-3.5 , 1.0) in Fig. 5.8, which would lie in this left branch: Assuming a correctly determined baseline (indicated by the dashed line at $y = 0$), the ratio between integral an peak again fits by accident.
As indicated by the black lines in the diagram, the effect of false negatives from these areas can be very severe: E_{PU} can reach up to 100% of E_{Main}, i.e. half of the total extracted energy would be due to pile-up. This perfectly correlates with the findings in [Sta21]; the problematic topologies from that approach were depicted in Fig. 5.6.
A very similar histogram can be extracted from recorded prescaler waveforms (see Fig. 5.10). The process (fit) by which the "true" number of pulses and their respective heights and positions were extracted will be detailed in section 7. It is, however, not necessary to understand the diagram. Although the histograms are a little less clear and show some artifacts which will not be discussed here, the most densely populated areas show a striking similarity between both histograms and to the the corresponding green areas in the simulation.
The question arises, how severe those missed pile-ups actually are. Fig. 5.11 provides the answer: While entries in which either E_{PU} or E_{Main} were below 0.5 MeV and those below the "3% line" may be safely disregarded as irrelevant, some sizable pile-ups were missed by the algorithm as well. As anticipated from Fig. 5.9, the most severe cases (especially those above the 100% line) come predominantly from waveforms for which no peak was found. If E_{Main} is projected out (see bottom left of Fig. 5.11), it is revealed that pile-ups with energies below $(2^{2.5} \approx 6)$ MeV are actually more likely to be overlooked than detected.

Comment and Explanation In hindsight, the effect of the new baseline algorithm onto the pile-up detection should have been expected. Given some thought, the existence of a second branch to the left of $T_B = 0$ should not come as a surprise, since the problem exhibits a certain symmetry. It was only due to the previous, somewhat primitive, baseline algorithm that this problem was hidden.
As the large green area beneath shows, the ratio criterion seems to be poorly suited in general for the detection of pile-up pulses occurring before the main pulse. Naively, there is no reason why this should be the case, as in order to leave the ratio unchanged, both the integral and the peakheight would have to be enhanced by the same factor.

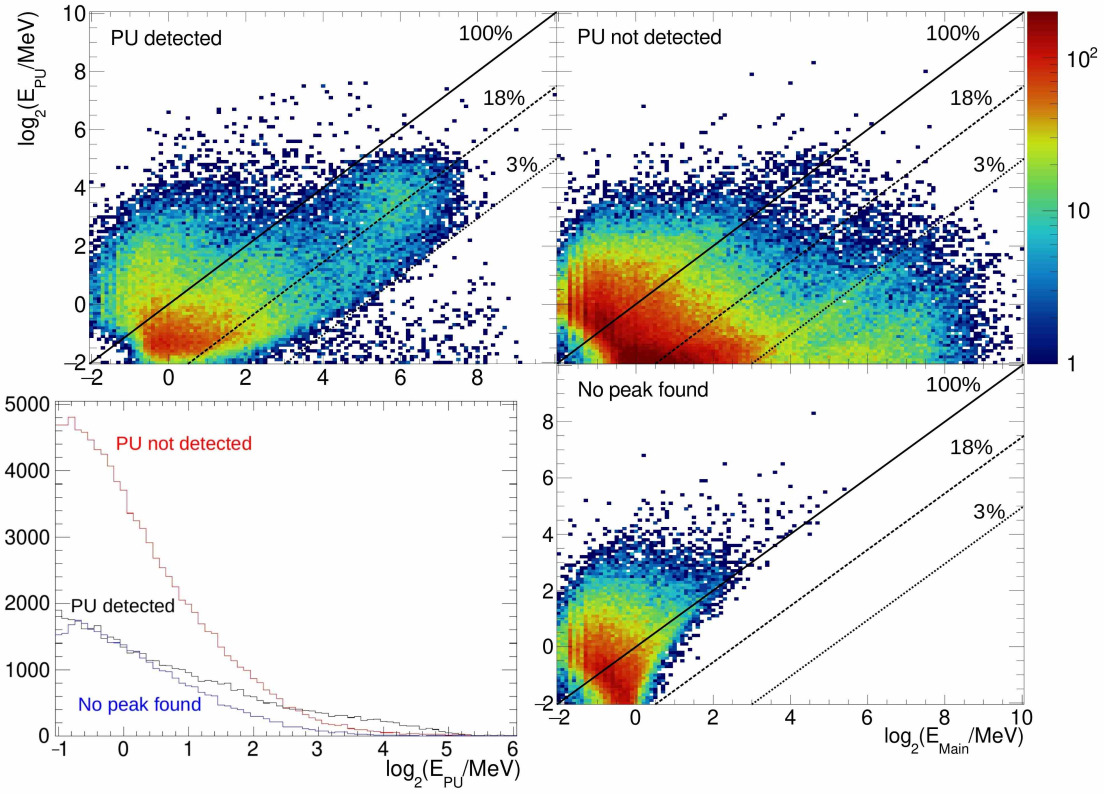


Figure 5.11: Energies of main and pile-up pulses for waveforms containing precisely 2 peaks, as determined by fit (prescaler waveforms from June 2021). The position of the pile-up pulse was projected out; see Fig. 5.10 which is another result from the same analysis. The diagonal lines denote the same relative error ($E_{\mathrm{PU}}/E_{\mathrm{Main}}$) as in figures 5.7 and 5.9 but are now straight lines in this representation. In contrast to those diagrams, not a respective pulse's heights, but its expected contribution to the overall integral is shown on the axes. Entries in the lower right histogram are a subset of the histogram above. The lower left diagram is a simple y-projection of the other three histograms and does not account for "irrelevant" pile-up situations (see text). A probable explanation for the peculiar accumulation of entries around $(6, 4)$ visible in the top left diagram will be provided soon in section 5.6.

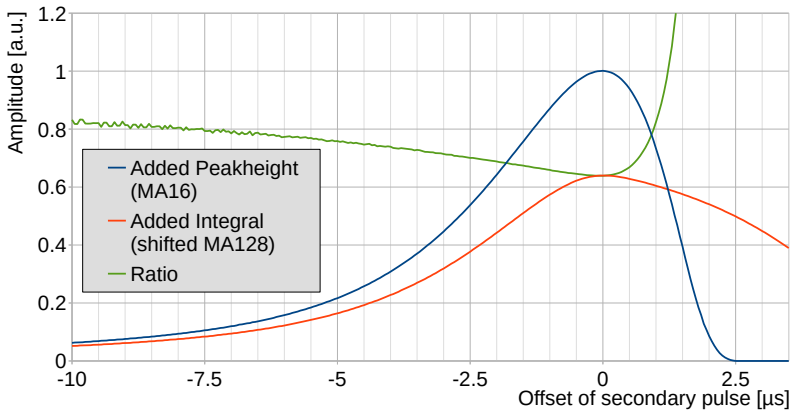


Figure 5.12: Theoretically expected peakheight (normalized to unity) and integral which would be *added* by a pile-up pulse; depending on its offset with regards to the main pulse. In addition, the green curve shows the respective ratio between the two.

This behavior can be understood as follows: Assuming a pile-up situation as before (pulse A correlated to the trigger; $T_A = 0$) and assuming further that the pile-up pulse B is only a "small" distortion to the main pulse, the apparent peak height of the main pulse H_A will be enhanced by the respective amplitude (MA_{16}) of pulse B *at the peakposition of pulse A*[27]:

$$H_{\mathrm{add}} = \mathrm{MA}^{B}_{16}\left(T_A\right)$$

At the same time, the integral I will be enhanced by the respective value of pulse B's 128-fold MA, evaluated at the same time (T_A + 128 samples):

$$I_{\mathrm{add}} = \mathrm{MA}^{B}_{128}\left(T_A + 128\right)$$

Taken together, the final ratio can be expressed as follows:

$$R = \frac{I_{\mathrm{obs}}}{H_{\mathrm{obs}}} = \frac{I_A + I_{\mathrm{add}}}{H_A + H_{\mathrm{add}}} = \frac{I_A + \mathrm{MA}^{B}_{128}\left(T_A + 128\right)}{H_A + \mathrm{MA}^{B}_{16}\left(T_A\right)} \tag{5.4}$$

In order for this ratio to equal 0.64, the *added ratio* between added integral I_{add} and added peakheight H_{add} would have to equal 0.64 as well. Fig. 5.12 shows that this ratio for the most relevant case of $T_A = 0$: It is indeed only 0.64 if both pulses overlap exactly. But while the added integral quickly outweighs the added peakheight for positive offsets, the increase is not that substantial for negative offsets, i.e. pile-up pulses occuring before the main pulse (ratio only around 0.7 to 0.8). Especially for comparatively small pile-up pulses, the total amount which is added to both features is not that large to begin with, which explains why the resulting pile-up ratio in the end is barely larger than 0.64. This problem is exacerbated by the fact that the detected peak will not be found precisely at the position of the main pulse T_A, but (due to the falling edge beneath it) at a slightly earlier position. This earlier position entails that the added peakheight is even larger, which shifts the added ratio even closer to 0.64 ($H_a dd$ is in the denominator).
Unfortunately, these effects were initially completely obscured in the data by the issues with the baseline. Only by fixing the shortcomings of the baseline algorithm, the problematic nature of the ratio-approach was uncovered.

[27] If pulse B is indeed only a small distortion, the peakposition of the overall signal will be close to the "undisturbed" case ; i.e. T_A.

5.4 Improved Algorithm (v2.0)

As discussed in the previous section, the initially used pile-up detection algorithm (the "ratio-approach") had some serious shortcomings. To reiterate the main points:

1. The sensitivity for pile-up pulses occurring before the main pulse is poor.

2. There are areas in which the observed ratio fits purely by accident.

3. Based on the ratio alone, it is impossible to tell whether a waveform contained actual pile-up, it simply consisted of a displaced pulse, or whether the waveform was narrower than usual (c.f. section 6.5).

4. A large fraction of pile-ups is missed entirely, because no peak could be found.

The third point was addressed and solved by the two-staged process (secondary check in the LEVB) outlined before. The last point was mitigated to some degree by the new baseline box-algorithm, but remained problematic. Depending on the energy, up to $\approx 1/3$ of all pile-ups (see Fig. 5.11) were missed that way.
While the conception of a new pile-up detection algorithm was highly desirable to address the other three points, it fell together with the introduction of dualhit-features (see section 4.4.3). This required a modified approach either way, since it would have been unclear which peak should be used for the computation of the ratio. Due to the ability to detect peaks before the trigger now, the topology of "no peak found" was effectively eliminated[28].

Naively thinking, one might assume that the problem of pile-up detection would be solved by simply counting the number of detected peaks. There are two problems with this idea: First, there are scenarios in which two or more pulses overlap such that only one peak is detectable. Second, the sole presence of two peaks may not imply that both pulses actually contributed anything significant to the measured integral (which is still fixed to the trigger).
Instead, the new algorithm makes use of the already introduced "LUT-approach" (see section 5.2, but this time it was directly implemented on the FPGAs): For every detected peak (default acceptance window: $[-10..+10]$ µs around the trigger), an *expected integral* is calculated based on the respective peakheight[29] and -position. Only peaks with an expected integral in excess of a fixed threshold plus some energy-dependent part are considered to be "significant". The default threshold is:

$$\theta_{\text{sig}} = 0.5\,\text{MeV} + 1\% \times E_{\text{tot}} \tag{5.5}$$

Based on the number of significant peaks, three cases have to be considered:

- No peak is significant. In that case, no pile-up flag is assigned.

[28] At least under the assumptions used for the simulation. The size of the remaining hole/white area is entirely determined by the width of the acceptance window for timestamps (c.f. section 4.4.3).

[29] Important note: The peakheights extracted from the sampled waveform are (if pile-up is present) in general larger than the actual height of the contributing pulse. For this reason, the expected integral might *overstimate* the actual impact on the integral, but it will never underestimate it.

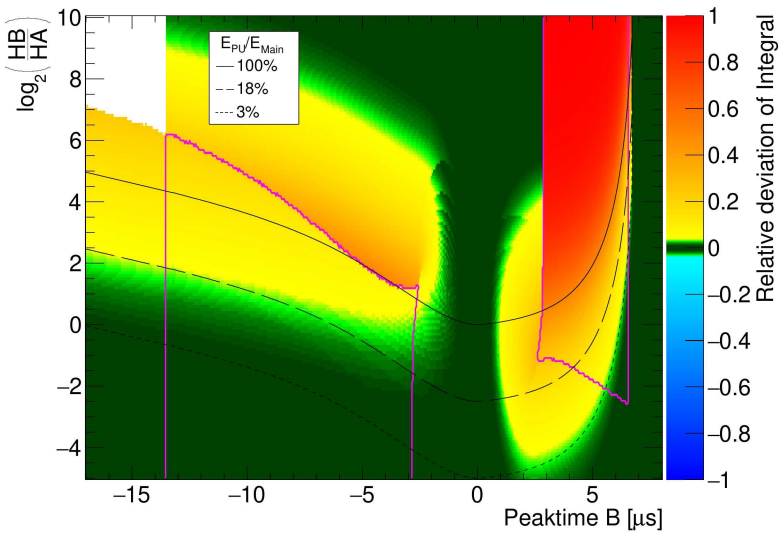


Figure 5.13: Simulation for the new LUT-based pile-up detection approach (see text). In this case, the color scale displays the relative deviation of the integral from the expected value: $(I_{\text{obs}} - I_{\text{exp}})/I_{\text{obs}}$. In case of two peaks, the integral deviation was (for this representation) only evaluated for the peak related to pulse A. Depending on the topology at hand, this peak might have been shifted from its original position. The areas in which two detectable peaks could (given sufficiently large pulses in absolute terms) possibly occur are outlined in magenta. Due to the lack of an overall energy scale, the number of detected peaks which would be considered significant cannot be provided. The apparent edges in the green area are a result of the extracted peaktime changing from one discrete value to the next.

- Two peaks are significant. This waveform does, by definition[30], contain significant pile-up.
- One peak is significant. In this case, the expected integral is compared to the observed integral. If the difference between observed and expected integral is above a certain threshold (same as above: $0.5\,\text{MeV} + 1\% \times E_{\text{tot}}$), the waveform is classified as pile-up.

Fig. 5.13 shows the result of the same simulation as before for this new approach. It solves all of the aforementioned issues of the old approach:

1. Pile-up pulses occurring before the trigger are not an issue anymore. If the previous peak/pulse was classified as significant, the pile-up will be detected. If not, the pile-up will not have been relevant anyways.
2. There are no more peculiar areas in which the predicted integral would fit the observed integral by accident.
3. In contrast to the previous approach, only deviations in one direction (i.e. only detected integrals which are *larger* than expected) need to be considered[31]. As a result, narrower pulseshapes (see section 6.5) will no longer be erroneously flagged as pile-up. Furthermore,

[30] Saying both peaks are significant is equivalent ot saying that both peaks are expected to have contributed significantly to the measured integral.

[31] Notice how there are no blue areas in the diagram.

displaced pulses are inherently accounted for in the expected integral, which means they will no longer cause false positives and an "unflagging" of pile-ups in the LEVB (v1.1) will no longer be necessary[32].

Only a small area in which no peak can be found is left. The size of this area is entirely determined by the range in which the peakfinder is allowed to operate. If necessary, this range could be easily expanded further. The large green area on the top of the diagram is of no concern. On the one hand, it is very unlikely to have such huge pile-up pulses occurring before the trigger during actual data taking. On the other hand, such topologies look, for all intents and purposes, just like a single pulse. The miniscule contribution from pulse A would be challenging to detect with any conceivable method. Although large parts of this area appeared in blue with the previous approach, it should be noted that this was simply due to the the apparent displacement of the resulting pulse. In the unflagging process in the LEVB, such cases would have been removed as allegedly false positives anyways.

Comment While the LUT-approach has several advantages on its own, it should be noted that it is not inherently more sensitive to pile-up pulses occurring before the trigger (see large green area in bottom left of Fig. 5.13). Because the shift in pulse A's peak position is relatively small, calculating the expected integral for a peak enhanced by pile-up at $T \approx 0\,\mu s$ essentially yields a value which is just 0.64 times the peakheight. So a similar argument as for the ratio applies, which explains its similarly poor sensitivity. The new approach draws its sensitivity in this region entirely from the detection of two separate peaks.

5.4.1 Performance

After its conception, the new algorithm outlined above was implemented in the SADC firmware as well (see [Ott22]). Some initial results using cosmic particles (see Fig. A.9 in the appendix) confirmed that the implemented algorithm was indeed working exactly as intended [Ott22]. A

[32] The fact that displaced pulses are uncorrelated to the trigger can be accounted for elsewhere, for example in the time-clustering.

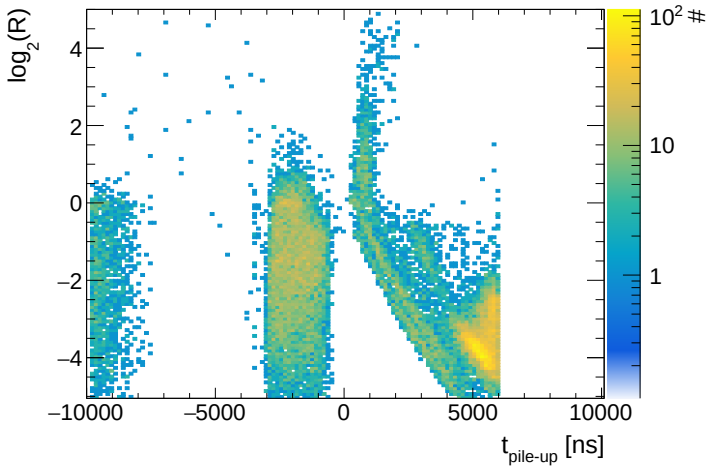


Figure 5.14: Same diagram as Fig. 5.6, but this time with the newest pile-up detection algorithm in place. The (sparse) amount of data was gathered from the May 2022 test beamtime [Sta24].

subsequent analysis by N. Stausberg (see Fig. 5.14) has, despite limited statistics, provided clear indications that the new algorithm had the desired effect onto the reduction of false negatives. His results are largely consistent with the simulation.
To add to that, the performance of the algorithm shall be studied in more detail in the following section. Due to a lack of data containing significant pile-up recorded with the new algorithm[33], it is studied based upon previously recorded prescaler waveforms (recorded for every 100th event, whether pile-up was detected or not). By employing the same pile-up recovery process which will be presented in section 7, it is possible to reliably determine the "true" pile-up related information – in particular the number of pulses involved and their respective height and position – for every such waveform.
Although the overall scale of a particular waveform was implicitly omitted in the previous simulations (by only considering the ratio of peak heights), it is in practice not at all irrelevant. While useful and instructive in the study of theoretical weaknesses of a particular algorithm, it would be misleading to just display the ratio of pulse heights for actual data, because such a representation would be inevitably smeared by threshold effects[34]. Furthermore, it would be dominated by low energy waveforms, since those are by far the most common. Therefore, this discussion needs to consider three dimensions: The height of the main pulse, the height of the pile-up pulse, and its position. The main pulse is still required to be correlated to the trigger (as in the simulations): $T_A \in [-500\,\text{ns}, +500\,\text{ns}]$. In addition, both pulses were required to be significant[35] for the observed integral: $E_{\text{exp}} > 0.5\,\text{MeV} + 1\% \times E_{\text{obs}}$ (see equation 5.5).

As three-dimensional data is challenging to present in two-dimensional diagrams, Fig. 5.15 depicts a projection into two dimensions[36] of all such pile-up situations which were reconstructed from the data. It can be clearly seen that pile-up pulses are almost[37] uniformly distributed in time and occur predominantly at low energies. In place of a pulse's height, its "energy", i.e. its expected impact onto the measured integral may be used instead, which is fundamentally the more relevant quantity. Note that this "energy" is inherently position dependent and must not be confused with the actual, deposited energy of whatever particle caused the pile-up pulse (which

[33] As of writing this thesis in 2023.
[34] Fixed thresholds like the number of rising samples for the peakfinder or the absolute threshold in the pile-up detection itself are much more difficult to overcome for smaller pulses.
[35] In accordance to the previously defined thresholds used in the pile-up detection process itself
[36] The height of the main pulse was projected out, complimentary information in following figures.
[37] The accumulation of entries to the top right of the middle gap will be discussed shortly.

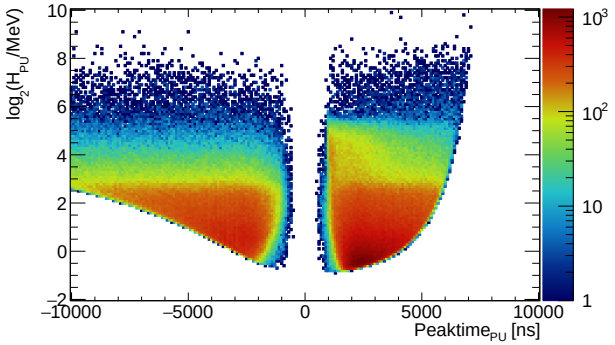


Figure 5.15: Topology of all reconstructed pile-up scenarios with precisely two (significant) peaks. The lower boundary is given by the minimum threshold of 0.5 MeV required for pile-up pulses. No entries are found in a gap in the middle due to the limited resolution power of the fit. (Data from prescaler waveforms; June 2021).

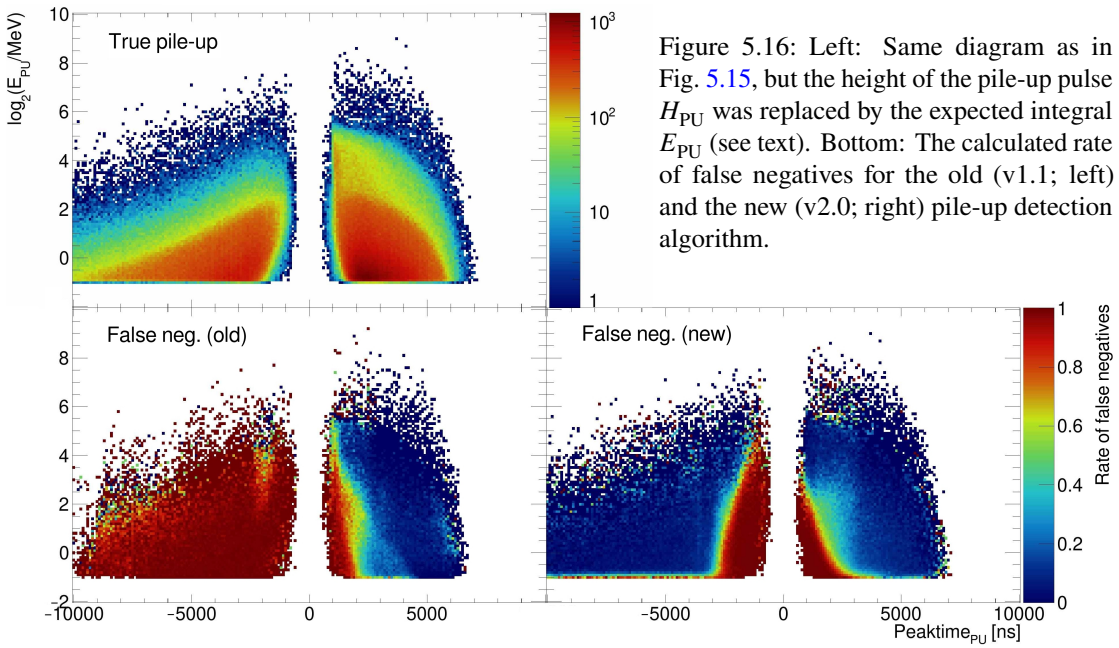


Figure 5.16: Left: Same diagram as in Fig. 5.15, but the height of the pile-up pulse H_{PU} was replaced by the expected integral E_{PU} (see text). Bottom: The calculated rate of false negatives for the old (v1.1; left) and the new (v2.0; right) pile-up detection algorithm.

would be equivalent to its height H_{PU}). The upper diagram of Fig. 5.16 shows precisely this kind of (equivalent[38]) representation.
The lower parts of Fig.s 5.16 and 5.17 finally show the calculated rate of false negatives for all occurring topologies for both the old (v1.1) and the new (v2.0) algorithm (both with the baseline box algorithm). The different energy-dependent thresholds (3% for the old and 1% for the new algorithm) make for a slightly unfair comparison which is most easily spotted at the lower right edge of the distributions in Fig. 5.17. Nonetheless, the significant algorithmic improvement of the new algorithm is still clearly visible. Especially striking is the comparison in Fig. 5.16: While the rate of false negatives for $T_{\mathrm{PU}} < 0\,\mathrm{ns}$ is close to 100% with the old algorithm, it is in most areas close to zero with the new approach.
Overall, it is expected that around 65% of all pile-ups would be detected by the new algorithm, in contrast to only 42% with the old algorithm. The only areas in which the new algorithm misses a significant number of pile-ups are:

- Pulses which are close together ($-3\,\mu\mathrm{s} < T_{\mathrm{PU}} < 3\,\mu\mathrm{s}$; see Fig. 5.16 bottom right)
- Either the main or the pile-up pulse is small in comparison to the other ($E_{\mathrm{Main}} \gg E_{\mathrm{PU}}$ or $E_{\mathrm{PU}} \gg E_{\mathrm{Main}}$; see Fig. 5.17 b.r.)
- Both main and pile-up pulse are small overall ($E_{\mathrm{Main}}, E_{\mathrm{PU}} \leq 1\,\mathrm{MeV}$; see Fig. 5.17 b.r.)

Furthermore, the areas in which the new algorithm is insensitive are entirely in accordance to the expectations from the simulation shown in Fig. 5.13. Ultimately, it seems unavoidable that *any* conceivable detection algorithm will have a limited resolution power under these "unfavorable" conditions.

[38] $E_{\mathrm{PU}} = H_{\mathrm{PU}} \times R_{\mathrm{exp}}^{\mathrm{LUT}}(T_{\mathrm{PU}})$; i.e. the energy is position-dependent and a linear function of H_{PU}.

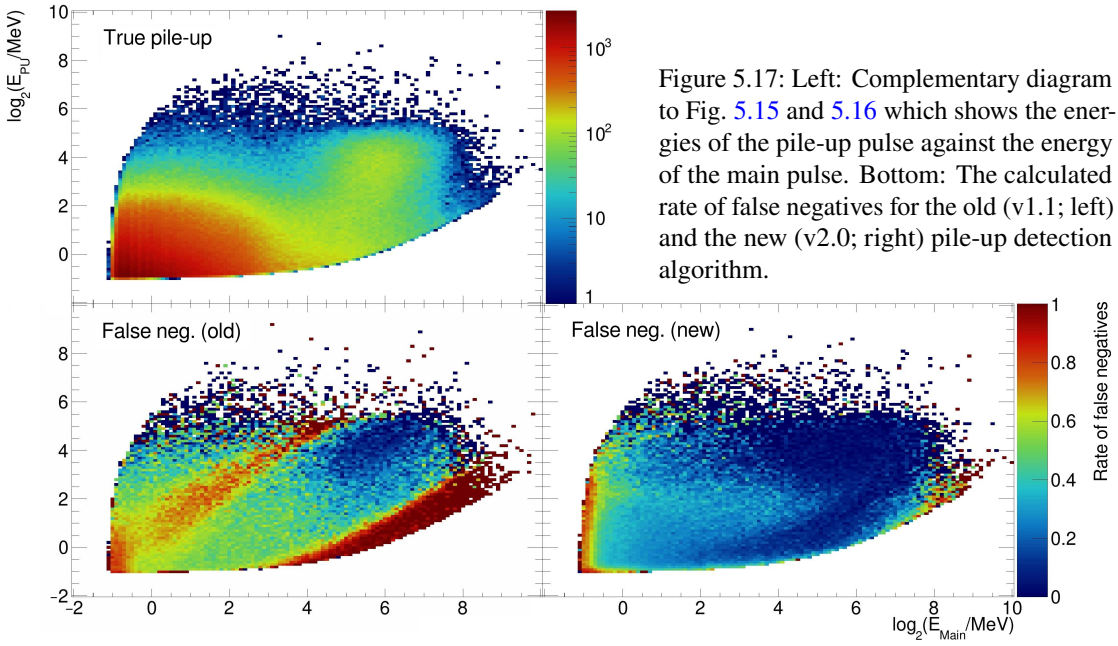


Figure 5.17: Left: Complementary diagram to Fig. 5.15 and 5.16 which shows the energies of the pile-up pulse against the energy of the main pulse. Bottom: The calculated rate of false negatives for the old (v1.1; left) and the new (v2.0; right) pile-up detection algorithm.

5.5 Projected Pile-Up Rate

As the amount of false negatives is expected to be greatly reduced by the new algorithm, the rate of detected pile-ups will increase. Instead of $\approx 3.5\%$ of all hits (see Fig. 5.5) with the old algorithm, the new algorithm is projected to produce an average pile-up rate of $\approx 5\%$. This will inevitably imply that a larger amount of waveforms will have to be written to disk. In contrast to the already recorded data, where each waveform consisted of 512 samples taken at 20 MHz, the recent changes of the firmware (see Table 3.1) enable a much more sensible sampling rate of 5 MHz, such that only 128 samples need to be stored for each waveform (also see section 3.3.1). This already reduces the amount of data by a factor of 4. On the other hand, the thresholds of the new algorithm are parameterized in a much more intuitive way, such that the value of the applied thresholds is directly reflective of what potential error is deemed acceptable. It goes without saying that tighter thresholds will automatically lead to a higher pile-up rate and vice versa.

5.6 Particular Sources of Pile-Up

Upon closer investigation, the distributions of true pile-ups exhibit a strange accumulation of entries in certain areas (see Fig. 5.15 and top left diagrams of Fig. 5.16 and 5.17). Its defining characteristics in three dimensions can be summarized as follows:

- $H_{\mathrm{PU}} \approx 10$ to $50\,\mathrm{MeV}$ $\quad\left(\text{visible in Fig. 5.15: } 3 \lesssim \log_2\left(H_{\mathrm{PU}}/\mathrm{MeV}\right) \lesssim 5.5\right)$
- $E_{\mathrm{Main}} \approx 20$ to $200\,\mathrm{MeV}$ $\quad\left(\text{visible in Fig. 5.17: } 4.5 \lesssim \log_2\left(E_{\mathrm{Main}}/\mathrm{MeV}\right) \lesssim 7.5\right)$
- T_{PU} is strictly positive and the density of entries decreases with T_{PU} (see Fig. 5.15).

Most pile-up is assumed to be caused by different particles coincidentally striking the same crystal. The pile-up particle usually comes from either electromagnetic background reactions or other, possibly hadronic, reactions. Another source of pile-up would be cosmic particles ($\mu^{\pm}$), which might strike the detector at any time. In the latter case, typically only one or two crystals are affected by the pile-up, as they do not produce an electromagnetic shower. Few crystals are affected as well if the energy of the secondary particle is small ($\lesssim 10\,\mathrm{MeV}$). The Molière radius of 3.8 cm in CsI(Tl) (see section 1.4) dictates, that only one or two crystals will likely see a significant energy deposition[39]. This seems to be the case for most pileups, see e.g. Fig. 5.15.
In relatively rare cases, both primary and secondary particle might also possess a substantial ($\gtrsim 20\,\mathrm{MeV}$) kinetic energy; such that they both leave a significant amount of energy in several, neighboring crystals. Fig. A.10 in the appendix shows a particularly striking example of such a case.
However, neither of those sources of pile-up seems to be a likely cause for the observed behavior; in particular because there is no reason to assume any correlation between the pile-up pulse and the main pulse or the trigger. The following discussion will provide a possible explanation:

In addition to the production of neutral mesons which most analyses at the CBELSA/TAPS experiment are focused on (π^0, η, ω, η', etc.), charged mesons ($\pi^{\pm}$, $K^{\pm}$, $\rho^{\pm}$, etc.) are produced as well. Depending on the exact reaction, these cross sections are comparable, if not larger[40].
A simple and, due to the high production cross section, very relevant example would be the decay of the $\Delta(1232)$ resonance:

$$\gamma p \to \Delta(1232) \to p\pi^0$$

Instead of $p\pi^0$, the final state could just as well[41] be $n\pi^+$. Unlike their neutral counterparts however, charged pions do not decay electromagnetically, but must decay via the weak interaction. Therefore, they have a significant lifetime of $\tau = 2.6 \times 10^{-8}$ s [PDG22], which is long enough to reach the detector ($c\tau = 7.8$ m). Under the assumption of only electromagnetic interactions, charged pions with kinetic energies up to 200 MeV should not be able to penetrate the entire 30 cm of the CsI(Tl) crystals[42]. Given their origin at the target, they will most likely deposit their

[39] If $\approx 90\%$ of the total energy are contained within R_M, at most 1 MeV will be spread beyond this radius. In order to produce discernible pile-up, those depositions have to line up with the cluster of the primary particle as well.
[40] e.g. $\sigma(\gamma p \to \pi^+\pi^-)$ ($\sigma_{\mathrm{max}} \approx 75\,\mu\mathrm{b}$) [ABB68] is much larger than $\sigma(\gamma p \to \pi^0 + \pi^0)$ ($\sigma_{\mathrm{max}} \approx 10\,\mu\mathrm{b}$) [Sok+15].
[41] The strong interaction is, in principle, independent of a particle's charge state. The exact branching ratio will be determined by the different isospin couplings and (small) differences in phase space.
[42] Muons (which are slightly lighter than pions) of this energy have an mean range of 29 cm in CsI(Tl) [Gro+22].

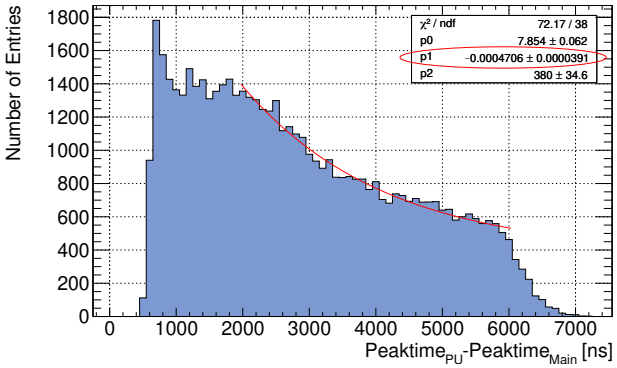


Figure 5.18: Time difference between the extracted peaktime of the pile-up and the main pulse for all of the entries in the top right of Fig. 5.15 $\left(T_{\mathrm{PU}} > 0\,\mu\mathrm{s} \wedge \log_2\left(H_{\mathrm{PU}}/\mathrm{MeV}\right) > 3.25\right)$. An exponential function was fitted to the data in between 2 000 ns and 6 000 ns.

entire energy in just a few crystals (potentially only one) and then come to rest within the detector. These pions subsequently decay into a muon and an (anti-) neutrino (BR $\approx$ 100%). Due to the small phase space, those muons will have very little kinetic energy ($\approx$ 5 MeV) and will not get far in the material; likely staying within the same crystal. Due to its miniscule cross-section, the neutrino will escape undetected in almost all cases.
The muons then decay as well; almost exclusively into 2 neutrinos and an electron/positron:

$$\mu^+ \rightarrow e^+ \bar{\nu_\mu} \nu_e \qquad \mu^- \rightarrow e^- \nu_\mu \bar{\nu_e}$$

As all three particles are very light in comparison to the muon, the resulting kinetic energy of the $e^\pm$ will be distributed between 0 MeV and $\frac{1}{2} m_\mu \approx 50$ MeV. Even at the maximum possible energy of 50 MeV, the electrons/positrons have a mean range of only 3.55 cm which means they will likely deposit their entire energy (or at least a substantial part of it) within the same crystal in which the muon decayed.

So far, the results of this discussion (E_{Main} sizable, but below 200 MeV and H_{PU}[43] below 50 MeV) align perfectly well with the observations. The argument becomes even more compelling when the time-distribution of the entries in question is examined. Fig. 5.18 shows the time between the main pulse and the pile-up pulse for all of the entries under investigation. The exponential fit yields a time constant τ[44] of $(2.12 \pm 0.18)\,\mu$s which is in near perfect agreement to the established decay time of the muon: $\tau_\mu = 2.20\,\mu$s [PDG22].
The limits for the fit were determined heuristically as the region in which the distribution looked "exponential". For smaller time differences, the impact of the limited resolution in the pile-up recovery process might have been too big. For larger time differences, the available data simply ends, because such pile-ups would no longer be relevant (see fig. 5.15).
In summary, the provided explanation seems to be highly plausible. Additionally, some low energy cosmic muons could get stuck in the material as well and contribute. An in-depth investigation of this topic would have gone beyond the scope of this thesis, but could be interesting for further studies.

[43] The position-independent "energy" of the pile-up pulse
[44] Inverse of fit parameter "p1".

CHAPTER 6

Digital Signal Processing and Pulseshape-Analysis

While the previous chapter was dedicated to the *detection* of pile-up in great detail, the final goal is to not only detect, but to also *recover* the corrected features from waveforms marked as pile-up. To that end, it is necessary to study, analyze and filter the arising pulseshapes first, which will be the topic of this chapter. The recovery process itself is described in the following chapter 7.

6.1 Fourier Transform

Fourier transformation is a very useful mathematical concept which has numerous applications in physics and other sciences. The core idea is to decompose any integrable function $g : \mathbb{R} \to \mathbb{C}$ in the time domain into an (infinite) series of complex exponentials. The function which represents the coefficients for every individual exponential in the frequency domain is known as the Fourier transform[1] $\mathfrak{F}[g(t)]$, which is often denoted as $\hat{g}(f)$.
There exist different conventions for the definition of the actual transformation operation. Depending on the convention, the variable in the transformed representation might be interpreted as an actual frequency, an angular frequency $\omega = 2\pi f$ or it might simply be seen as another variable ξ entirely. A common[2] definition of the Fourier transform is [Bra00]:

$$\mathfrak{F}[g(t)] = \hat{g}(f) = \int_{-\infty}^{+\infty} g(t) \cdot e^{-i2\pi ft} \, \mathrm{d}t \tag{6.1}$$

where f denotes an actual frequency. What makes the Fourier transform so useful, is that every signal $g(t)$ in the time domain has a corresponding Fourier transform $\hat{g}(f)$ in the frequency domain and vice versa. If a particular problem is difficult to solve in the time domain, it might be easier or more elegant to solve in the frequency domain. Apart from that, the Fourier transformation itself has some convenient properties [Bra00]:

[1] Both the method itself as well as the result are often called "Fourier transform". To avoid confusion, the method will be referred to as "transformation" and the result as "transform".

[2] Different definitions may omit the 2π in the exponent, but then include a factor of $1/(2\pi)$ in the backtransformation (or $1/\sqrt{2\pi}$ in both) to ensure that $\mathfrak{F}^{-1}[\mathfrak{F}[\mathfrak{g}]] = g$.

1. **Invertibility:** The Fourier transform is invertible by performing the same integral as in equation 6.1, but with a positive sign in the exponent and integration over frequency instead of time:
$$\mathfrak{F}^{-1}\left[\hat{g}(t)\right] = \int_{-\infty}^{+\infty} \hat{g}(f) \cdot e^{+i2\pi ft}\,\mathrm{d}f = g(t) \tag{6.2}$$

2. **Linearity:** For $a, b \in \mathbb{R}$ and integrable functions $g, h : \mathbb{R} \to \mathbb{C}$, the following relation holds:
$$\mathfrak{F}\left[a \cdot g(t) + b \cdot h(t)\right] = a \cdot \hat{g}(f) + b \cdot \hat{h}(f) \tag{6.3}$$

3. **Convolution theorem:** The Fourier transform of the convolution of two functions g(t) and h(t) is given by the product of their individual Fourier transforms. Therefore, the convolution of two functions might be conveniently calculated as:
$$g(t) * h(t) = \mathfrak{F}^{-1}\left[\hat{g}(f) \cdot \hat{h}(f)\right] \tag{6.4}$$

4. **Time shifting:** Shifting a real signal $g(t)$ in time results in an additional complex phase factor in the frequency domain:
$$\mathfrak{F}\left[g(t - t_0)\right] = e^{i2\pi ft_0} \cdot \hat{g}(f) \tag{6.5}$$

5. **Relationship to δ-distribution:** The Fourier transform of the δ-distribution yields unity:
$$\mathfrak{F}\left[\delta(t)\right] = 1 \tag{6.6}$$
The transform of a shifted delta distribution trivially follows from equation 6.5:
$$\mathfrak{F}\left[\delta(t - t_0)\right] = e^{-i2\pi ft_0} \tag{6.7}$$
This result makes intuitive sense. Just like the transform of a perfect sine wave would be zero everywhere except for the frequency of this sine wave, the transform of a single peak at position t_0 and vanishing width is equal to a single sine wave in the frequency domain.

6.1.1 Discrete Fourier Transform

For practical applications, a signal (e.g. from the SADC) to be transformed is usually not given as a continuous, smooth function $s(t)$ but in a discrete form $[s_0, s_1, s_2, ..., s_N]$, e.g. from sampling an electronic signal with a fixed sampling frequency F. Obviously, for any real-world application, only a finite number N of sampling points will be recorded. In that case, the integral collapses into a sum and yields a finite number of coefficients for the frequency spectrum. The formula to calculate these coefficients is know as a Discrete Fourier Transform (DFT) and closely resembles the continuous form [Bra00]:
$$\hat{s}[k] = \sum_{n=0}^{N-1} s[n] \cdot e^{-i2\pi \frac{k}{N} n} \tag{6.8}$$

This DFT shares all of the properties of the continuous version. In case of real-valued signals, it can be proven that the second half of the N values one gets from performing this transform is redundant by complex conjugation:

$$\hat{s}[N-k] = \hat{s}[k]^*$$

In the context of digital signal processing, one is mostly interested to see which actual, physical frequencies are present at which magnitude and phase in a given signal and how certain components (e.g. filters) affect the frequency distribution of the signal. Comparing formula 6.8 to the continuous version, it is easy to see how one might associate a coefficient at index k with a frequency $f = \frac{kF}{N}$.
To summarize, performing a DFT on a real-valued input signal of length N results in $N/2$ non-redundant complex coefficients[3] which correspond to the same number of discrete frequencies evenly spaced in between 0 and $F/2$. As mentioned previously in section 3.1, $F/2$ is known as the Nyquist-frequency and expresses the fundamental upper limit of reconstructable frequencies.
Instead of representing complex numbers $z = x + iy$ by their real part x and imaginary part y, a completely equivalent (and sometimes more instructive) way is to write complex numbers in terms of their *magnitude* r and *phase* ϕ. The following relations hold:

$$r = \sqrt{x^2 + y^2} \tag{6.9}$$

$$\phi = \text{atan2}(x, y) \tag{6.10}$$

$$x = r\cos\phi \tag{6.11}$$

$$y = r\sin\phi \tag{6.12}$$

This gives rise to a more intuitive understanding for the actual physical meaning of complex-valued coefficients: The magnitude r describes the amplitude of the respective frequency in the input signal, whereas the phase ϕ encodes the relative time shift of contributing frequencies with respect to each other.
In order to actually compute the DFT values, it is often unpractical or tediously slow the calculate all coefficients from the above definition. Instead, there exist various advanced algorithms, commonly known as Fast Fourier Transform (FFT). One of the fastest and most popular packages to do so is the FFTW[4] library [FJ05]. Within the ROOT framework [BR97], which was used for most of the analysis in this thesis, the FFTW algorithm is accessible via the `TVirtualFFT` interface.

6.1.1.1 Windowing

All FFT algorithms implicitly assume the input signal to be periodic. If the first and last sampling point are not equal to one another, a discontinuity is created. This is of particular concern if periodic signals are recorded over a necessarily limited time which is not an integer multiple of the period length T. This discontinuity then results in artifacts in the frequency spectrum, an effect which is also known as *spectral leakage* [OSB99].
Another way to understand the problem is to consider the time over which the signal is sampled

[3] It is instructive to see how no information is lost when N real data points are transformed to $\frac{N}{2}$ complex numbers, each possessing a real and an imaginary part.
[4] "Fastest Fourier Transform in the West"

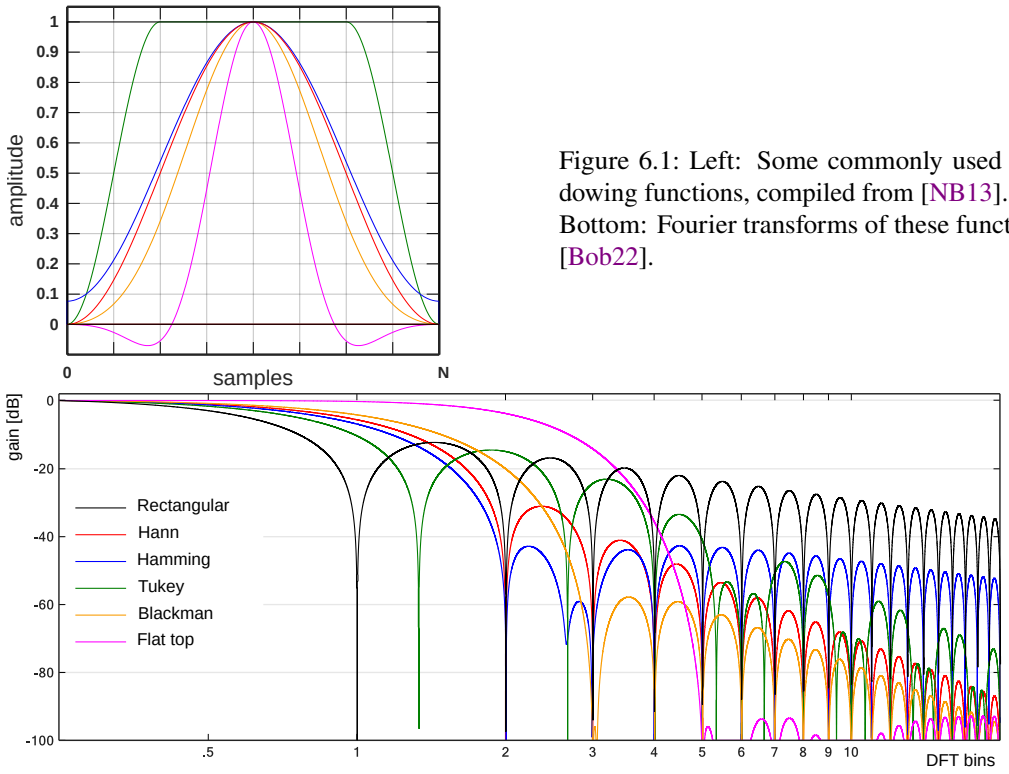


Figure 6.1: Left: Some commonly used windowing functions, compiled from [NB13]. Bottom: Fourier transforms of these functions [Bob22].

to be a rectangular *window function*[5] with which the hypothetical full signal is multiplied prior to the transformation. Via the convolution theorem (equation 6.4), it follows that the result of the transformation is a convolution of the original frequency spectrum and the transform of the window function. The transform of the rectangular window is a sinc function ($\operatorname{sinc} x = \frac{\sin x}{x}$). Depending on the application, another window function might be applied instead. Most commonly used window functions vanish at the edges to prevent discontinuities and have a suitable, ideally very narrow[6], Fourier transform. Several such windows exist which each have their own advantages and disadvantages [Har78]. Fig. 6.1 shows a selection of popular choices. One particular disadvantage that all other windows share, is that they necessarily lead to a distortion of the signals. Therefore, the rectangular window still gives the best frequency resolution of all, despite the artifacts. This is reflected in the smallest width of the main lobe of the frequency spectrum, see black line in Fig. 6.1.

For the signals of interest in this thesis, spectral leakage is of only moderate concern. While the sampled waveforms typically start with some amount of baseline samples before the pulse and the amplitude has not entirely dropped back to baseline level after $\approx 25.6\,\mu s$ (default length of the sampling window), the discrepancy is only small in comparison to the overall pulse height (see e.g. Fig. 6.6). If precise knowledge of the frequencies is needed, such as in section 6.6 (Deconvolution), the pulseshape might be suitably extrapolated to avoid the problem of spectral leakage altogether.

[5] The rectangular window is 0 everywhere except for the sampling window, where its value is 1.

[6] The narrowest function possible would be the δ-distribution. The according window however would be 1 everywhere (see equation 6.6) and therefore not a useful "window".

6.2 FIR Filter

For a long time, the FPGA-internal feature extraction used a 16-fold moving average in order to suppress noise (see section 3.3.3.1). Recently, the switch to a non-trivial FIR filter was made in the firmware. While the advantages were already pointed out at an early stage of this thesis, the implementation proved challenging [Ott22]. The filter discussed here is (up to differences in the implementation) functionally identical to the one which was then finally implemented in firmware. This section is meant to provide the reasoning why a FIR filter is a better choice, how it was tested and used in offline analysis first, and how the parameters were chosen.

6.2.1 Concept

It is very straightforward to calculate an N-fold moving average from a series of points:

$$\mathrm{MA}_k = \sum_{i=k-N+1}^{k} \frac{1}{N} \cdot x_i$$

A FIR filter generalizes this concept by allowing for coefficients in the sum that might differ from the simple $\frac{1}{N}$. The main advantage of FIR filters is that by carefully choosing these coefficients in a certain way, on can design digital low-, high- or band-pass filters with in principle arbitrary steepness and bandwidth. There exist various different tools to conveniently calculate these coefficients; for this thesis the web-based tool by T. Roelandts [Roe20] was used. With this tool, a cutoff frequency f_{cutoff} and a transition bandwidth b_{trans} (also called "roll-off" or "steepness") can be specified for a low-pass filter. The former specifies the frequency at which the gain is exactly 0.5, while the latter describes the bandwidth over which the gain changes from 1 to (practically) 0.

6.2.2 Frequency Spectrum of CB-SADC Waveforms

In order to estimate what cutoff frequency and transition bandwidth might be suitable for the CB-SADC waveforms, one needs to study the signals themselves first, in particular their frequency spectra. The raw signals are shown in the top left diagram of Fig. 6.2. Since the set of waveforms was not constrained in any way, most waveforms do not actually contain any signal, which causes a broad band around $y = 0$. Its width is directly indicative of the amplitude of random noise oscillations around the baseline, because the pulses were not normalized for this plot. Above this band, a continuous distribution of actual signals can be observed. Most of these signals start at the trigger time, but some pulses which occurred at other times can be seen as well. The right hand side shows the corresponding overlay of Fourier transforms of the same pulses.
In analogy to the previous plot, a prominent band can be seen as well, which stems from those same pulses without actual signal content. Under the assumption of white, i.e. frequency-independent noise, one would have expected this band to be perfectly horizontal. However, due to the characteristics of the signal shapers discussed in section 3.2.1.1, this is not the case. While noise contributions above ≈ 1 MHz are suppressed, lower frequencies are in the pass band of the filters and therefore appear with a higher, unsuppressed amplitude. Above this band, the frequency spectrum of actual pulses can be seen. As expected, the pulses have frequency contributions

up to 1 MHz and nothing significantly different from noise above that[7]. The "ringing" structure around 1 to 2 MHz is due to spectral leakage (see section 6.1.1.1), since most pulses did not exactly reach the baseline level at the end of the sampling window. For the sake of simplicity and in order to keep the pulse shape distortion-free, no windowing was used in the FFT calculation.

6.2.3 The Moving Average Filter

The middle row of Fig. 6.2 shows precisely the same waveforms, but this time after the application of a 16-fold moving average. Starting with the left plot, one notices two things: Firstly, almost all pulses seem to pass through the point $(-1.6\,\mu\text{s}, 0)$. This is of course due to the baseline subtraction. In May 2018, the old algorithm was still used, which means the baseline was determined as the value of the MA_{16} 1.6 µs before the trigger. The fact that some pulses do not pass through this point is due to the application of the baseline correction algorithm[8]. A similar structure can also be seen in the other plots. In case of the raw waveforms however, it is somewhat obscured by noise. The second observation is that the width of the "noise band" is visibly reduced, which indicates a good noise-suppression by the filter.
The Fourier transform confirms this observation in that the mean magnitude of frequencies above 1 to 2 MHz is significantly lower than in the unfiltered case while frequency contributions below 1 MHz are left almost unchanged. However, it also reveals that the moving average filter introduces peculiar artifacts in the frequency spectrum. Fig. 6.3 provides an illustration: It is obvious and follows directly from its definition that a 16-fold moving average will suppress frequencies that are an integer multiple of $20\,\text{MHz}/16 = 1.25\,\text{MHz}$ completely. At the same time, it will be especially bad at suppressing half-integer multiples of 1.25 MHz.

6.2.4 Application of FIR Filter

Several aspects have to be taken into consideration when choosing parameters for the FIR filter. In general, there is a negative correlation between the desired transition bandwidth of the filter and the required number of coefficients[9]. Too many taps would exceed the resources of the FPGA, so 64 taps was considered to be sensible choice. The filter needs to be realizable not only in offline analysis, but also on the FPGA as well (which was the case [Ott22]). With this many taps, a transition bandwidth of 1.5 MHz can be achieved. From the discussion so far, it would seem that $f_{\text{cutoff}} = 1\,\text{MHz}$ is the logical choice. But due to the comparably large[10] transition bandwidth, this would result in an imperfect suppression of noise in the region between 1 MHz and 2 MHz. Instead, a cutoff frequency of 0.5 MHz was chosen. It leads to only minor distortions of the pulse shape but results in a near-perfect noise suppression for $f > 1.5\,\text{MHz}$. Fig. 6.3 provides the expected frequency response. In the following, "FIR" will always refer to this specific FIR filter. The overlay of resulting pulses in the time domain (bottom left in Fig. 6.2) looks very similar to the results of the moving average filter, but the picture in the frequency domain reveals the FIR

[7] This statement is somewhat redundant, since the pulses were already subjected to the filters of the shaping circuit. So even if the actual signals would have had significant contributions above 1 MHz, they would not show up in this plot.
[8] The correction threshold of 20 ADC channels is visible in the plot as well.
[9] Also commonly referred to as "taps".
[10] "Large" in comparison to digital filters with more taps. The filter is still much steeper than most analog filters.

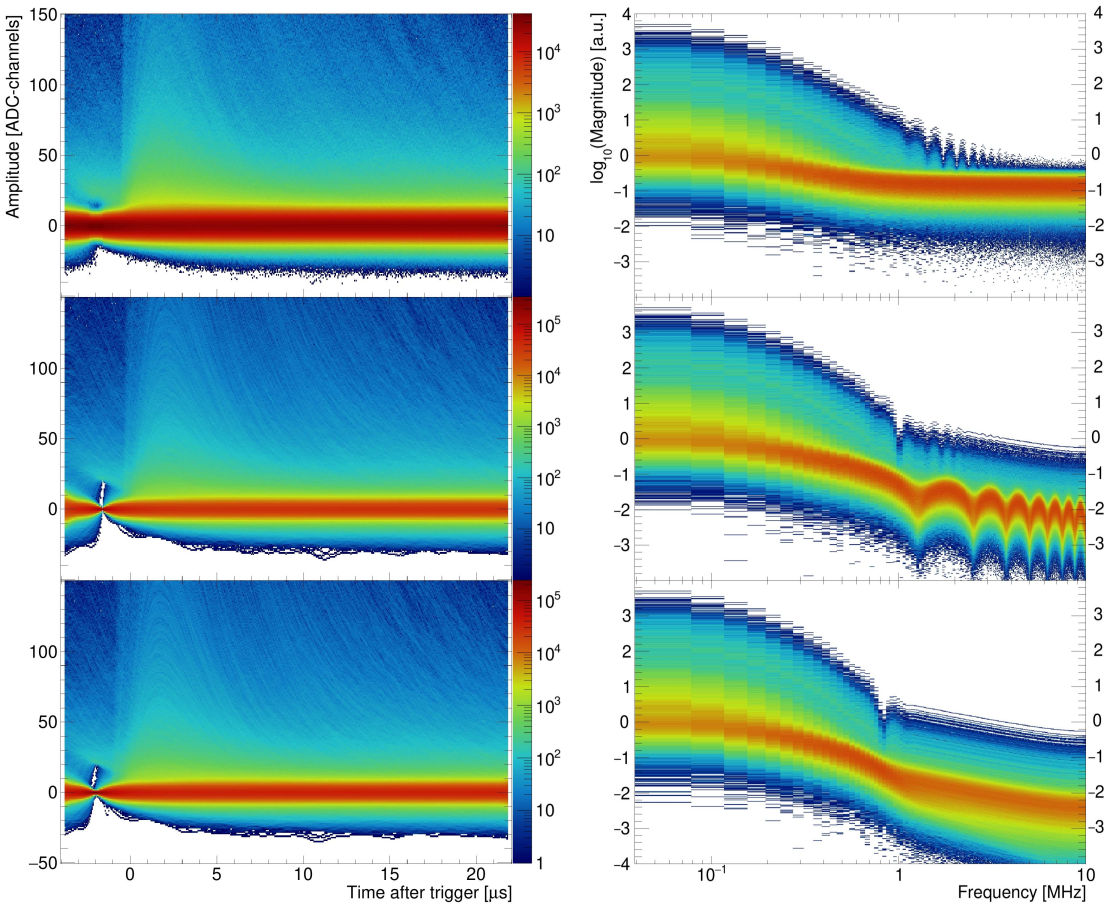


Figure 6.2: Overlay of pulses (left) and their respective Fourier transforms (right) that were recorded during the May2018 beamtime. The top row shows raw waveforms, the middle row the same waveforms after a MA_{16} filter and the bottom row after the FIR filter was applied. The baseline was subtracted, but pulses were *not* normalized to their respective height.

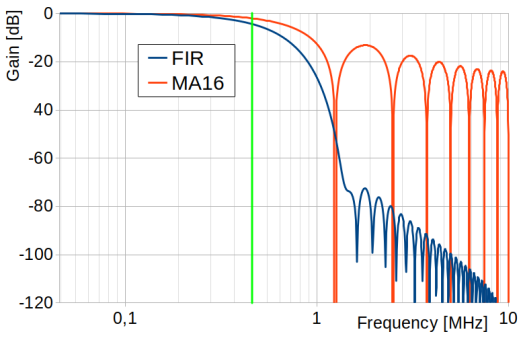


Figure 6.3: Calculated gain of a 16-fold moving average filter and of a FIR lowpass-filter with a cutoff frequency of 0.5 MHz (denoted by the green line) and a transition bandwidth of 1.5 MHz. Data taken from [Roe20]. As is common in the description of filters, gain values represent the *power* ratio. Accordingly, values have to be divided by 20 to get the decadic logarithm of the amplitude ratio.

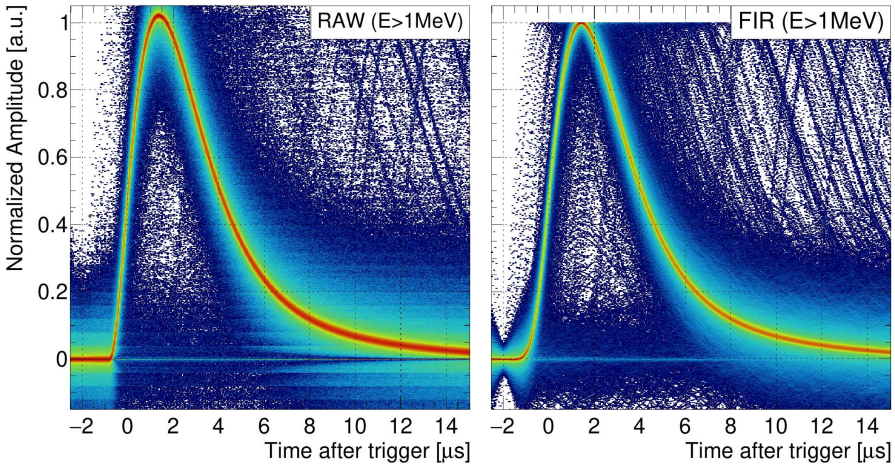


Figure 6.4: Overlay of raw and FIR-filtered pulse shapes exceeding a height of 25 ADC-channels ($\approx$ 1 MeV). Pulses in both plots were normalized to the maximum of the FIR-filtered waveform. Because the FIR-filter slightly "flattens" the maximum, the amplitude in the left plot overshoots a value of 1 in the maximum. Horizontal structures (spaced $\Delta y = 1/25 = 0.04$ apart) are due to the height threshold and the discrete nature of possible amplitude values.

filter's advantages: Not only is the suppression of high frequencies visibly stronger[11], but it also shows no artifacts as the moving average filter did. The only remaining artifact around 800 kHz seems to be a remnant of the implicit (rectangular) windowing. Fig. A.11 again underlines that both the MA_{16} and the FIR filter lead to a significant overall improvement regarding the mean noise amplitude, with the FIR filter being ever so slightly superior.
Due to these advantages, the entire offline-analysis is based on the FIR-filtered signal instead of the moving average signal. As mentioned, the filter does distort the pulse shape to some degree[12] (see comparison in Fig. 6.4). The distortion is most obvious at the onset of the pulse, which is not as sudden as in the unfiltered case, and near its maximum, which gets flattened a bit. While the former is of no concern, the latter effect has to be taken into account when features are extracted offline from a FIR filtered waveform (see section 7.3.3).

[11] The reason why the suppression of the "main band" is not as strong as might have been expected from Fig. 6.3 lies in the fact that also the filtered values were rounded to the next integer, which effectively deteriorates the achievable resolution for small (or non-existent) signals.

[12] Although the moving average leads to similar (slightly smaller) distortion as well.

6.3 Extraction and Characterization of Pulse Shapes

For much of the further analysis, a precise knowledge of the pulse shape is needed. Fig. 6.5 depicts an overlay of normalized, FIR-filtered waveforms (similar to what was shown in Fig. 6.4), but this time all waveforms have in addition been shifted in time such that their respective maxima coincide in the point (0,1). It is straightforward to take the most probable value (MPV) for every bin along the x-axis and save this dataset as reference for further usage. This is represented by the black points in the figure. Before sample number −59 (almost 3 µs before the maximum), the pulse shape is assumed to be *exactly* zero. Since the sampling window is not long enough to observe the signal actually dropping back to zero, the falling edge was approximated by a single exponential (enhanced view in Fig. A.14) over samples 300 to 400. The decay time of this exponential comes out to be approximately 8 µs. By combining the extracted mean waveform with this exponential extrapolation, a full description of the entire pulseshape can be achieved. However, the combination itself requires some caution. If one were to simply "attach" the extrapolation to the extracted values, discontinuities both in the pulseshape itself and its derivative could be created. In order to get a smooth transition from the extracted datapoints to the extrapolation ones, the `smoothstep`-function[13] S_1 is used:

$$S_1(x) = \begin{cases} 0 & t < 0 \\ 3x^2 - 2x^3 & 0 \leq x \leq 1 \\ 1 & t > 1 \end{cases} \tag{6.13}$$

[13] Family of sigmoid shaped interpolation functions commonly used in computer graphics.

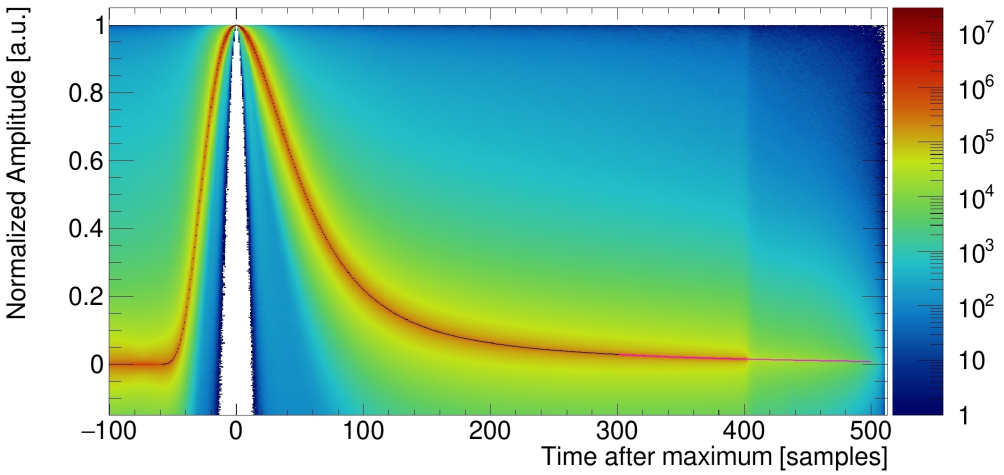


Figure 6.5: Overlay of FIR-filtered waveforms exceeding a height of 25 ADC-channels (≈ 1 MeV). Data taken from June 2021 beamtime. Pulses were normalized to their respective height and shifted to the same point in time. Waveforms that had their maximum directly at the start or the end of the sampling window were not taken into account. Black points represent the MPV of each x-axis bin. The magenta curve is an exponential fit which ranges from $x = 300$ to $x = 400$.

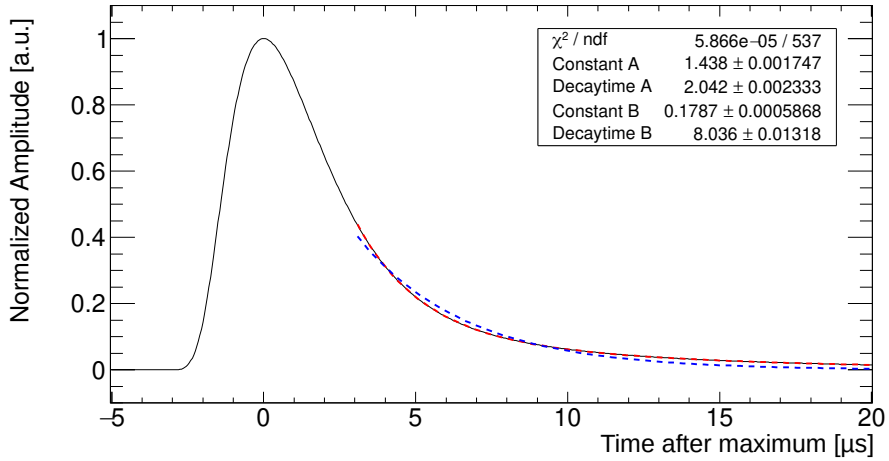


Figure 6.6: Reference pulse shape extracted from data, fitted with a single exponential (blue) and the sum of two exponentials (red). The single exponential fit yields a decay time of $\tau = (3.557 \pm 0.036)\,\mu s$

Let $\tilde{P}(t)$ denote the extracted datapoints, valid up to sample number 400 and let $E(t)$ be the extrapolation function valid from sample number 300 onwards. In this case, t is measured in samples after the maximum (with a sampling rate of 20 MHz). The complete reference Pulseshape $P_{\text{ref}}(t)$ is then implemented as the following expression:

$$P_{\text{ref}}(t) = \begin{cases} 0 & t < -59 \\ \tilde{P}(t) & -59 \leq t < 300 \\ \tilde{P}(t) \cdot \left(1 - S_1\left(\frac{t-300}{100}\right)\right) + E(t) \cdot S_1\left(\frac{t-300}{100}\right) & 300 \leq t < 400 \\ E(t) & 400 \leq t \end{cases} \tag{6.14}$$

In the end, one obtains a smooth, well-defined description of the most probable pulse shape to be detected in the experiment.

Comments To the best of the authors knowledge, there exists no simple analytic description of this shape. Instead of phenomenologically constructing a somewhat suitable function that involves a lot of parameters, it is much simpler and more precise to directly use the numerical description from above. Although it might seem that way at first glance, the falling edge does not follow a strictly exponential decay. While a single exponential is precise enough for the very last part of the decay to be used as an extrapolation, it is not a valid description of the whole descent. One has to use a combination of at least two exponentials ($\tau_1 \approx 2\,\mu s$; $\tau_2 \approx 8\,\mu s$) in order to get a satisfying description (see Fig. 6.6). For the very last part of the decay (see Fig. 6.7), an even longer decay time of 10.6 µs is needed for a suitably precise description[14] For the applications in which the reference pulseshape is used, the last few percent of the decay are not critical, which is why an extrapolation with $\tau = 8\,\mu s$ is still usable.

[14] While these two (three) distinct decaytimes are in the same ballpark as the literature values provided in section 1.4, they do not match the (already inconsistent) literature values. Considering the shaping and filtering which was applied to the signals, an exact agreement was not to be expected anyways.

6.4 PZC Optimization: A New Approach

It is crucial for the following pulse shape analysis as well as the pile-up detection and recovery process, that pulses from all channels have the same shape. Special care has to be taken when setting the DAC values for the Pole-Zero Cancellation resistors (see section 3.2.1.3), since this directly influences the pulse shape. M. Biel has developed a routine to determine suitable values for all channels during his bachelor thesis [Bie18]. His method has however proven to be rather tedious and requires a large data set to work. Since the DAC values had to be newly determined anyways for all the new channels after the full SADC coverage was achieved, it was decided to take this as an opportunity to review the algorithm and make it more robust. Apart from removing the underswing and staying clear of the overcompensated case[15], the focus was placed on ensuring the highest possible degree of similarity in between channels.

Method The simplest conceivable analytical function to describe a well-compensated pulse shape is a single exponential. For the PZC, only the decay characteristics of the falling edge far behind the maximum are relevant. To that end, waveforms with a *waveform stretcher* setting of "8"[16] were recorded. Fitting the last part ($> 20\,\mu s$ after the trigger, when the normal sampling window is already at its end) of the decay of pulses from the already optimized channels with an exponential function $f_{\text{ref}}(t) = \exp\left(a - \frac{t}{\tau}\right)$ yielded a time constant of $\tau \approx 10.6\,\mu s$. The objective of the optimization was then to tune the pulse shapes from every channel to look as similar to this exponential function as possible.
In order to do so, several datasets with different, global PZC settings ranging from 30 to 70 were recorded with cosmic particles[17] impinging on the detector. The available values for the DAC settings range from 0 (strongest compensation) to 127 (weakest compensation). Care was taken to ensure that the likely optimum (somewhere around 50) was well contained within this range (see Fig. 6.7). All pulses of each individual channel were then collected and superimposed into a histogram, from which a mean pulse shape $P(t)$ for each channel and each PZC setting x_{pzc} could be extracted. The goal of the optimization was then to find values which minimize the deviation from the reference exponential, i.e. the following metric M:

$$M(x_{\text{pzc}}) = \sum_{t=t_{\text{start}}}^{t_{\text{end}}} \left| P(t\,; x_{\text{pzc}}) - f_{\text{ref}}(t) \right|$$

It summarizes the absolute deviation of each point of the mean pulse shape from the reference exponential. $t_{\text{start}} = 20\,\mu s$ was chosen as start of the summation, since it is the same point from which the initial fit was performed. $t_{\text{end}} = 200\,\mu s$ constitutes a practical endpoint, as the enlarged sampling window extends barely beyond this point.
Fig. 6.8 shows the results of this process. One could go ahead and simply take the PCZ setting for each channel for which the lowest value of M was achieved. But due to a limited number of events, single values are always subject to some statistical fluctuation. At the same time, there are

[15] In the overcompensated case the pulse takes unnecessarily long to reach the baseline (see Fig. 6.7 for examples).
[16] Meaning only 8^{th} sample was used, effectively enlarging the sampling window by the same factor (see section 3.3.5)
[17] Mostly muons; following section 1.4.4, their recorded pulse shapes should be very similar, if not identical, to those of an electromagnetic shower. No evidence of a possible difference was found at any point in this work.

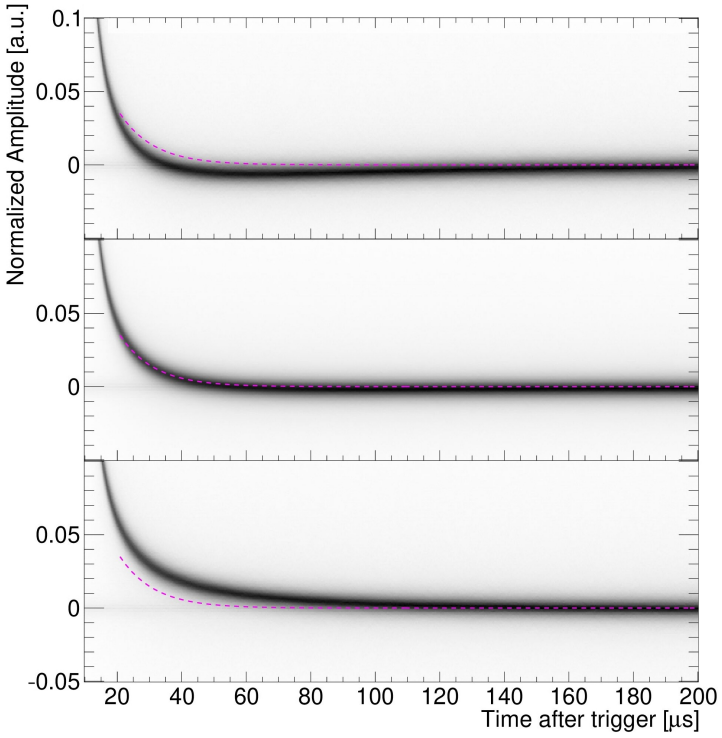


Figure 6.7: Overlay of multiple pulses (normalized to their respective height) recorded with a PZC setting of 70 (top), 50 (middle) and 30 (bottom). Data was recorded using cosmic particles and an 8-fold *waveform stretcher* setting to capture (almost) the entire pulse shape. Only the falling edge is shown. An exponential function with $\tau = 10.6\,\mu s$ was superimposed on all three plots.

some channels (see e.g. the lower right plot of Fig. 6.8) for which the ideal setting might not have been part of the sample of tested values[18]. To overcome both challenges at once, the minimum was calculated by fitting a second order polynomial to the data for each channel. A parabola was used, since the appropriate function to describe the distribution of values is unknown. However, any function exhibiting a local minimum can be usually well approximated in the vicinity of this minimum by a second degree polynomial[19].
After the appropriate DAC values for all channels were calculated this way, another run was recorded to evaluate the performance. The improvement over using a single global value is not huge, but can be clearly seen in Fig. 6.9, as the spread of pulse shapes is significantly reduced.

Limitations As discussed before (see section 6.3), the description of a pulse's falling edge by a single exponential is just an approximation. Fig. 6.7 shows, that while the reference function lies entirely above the signals at PZC setting 70, and entirely below for setting 30, the shape does not fit perfectly at the intermediate value of 50: There exists a small region where the function lies slightly below the data (around 20 to 30 µs after the trigger), followed by a large region where it lies slightly below. Optimizing on similarity to this function therefore still results in a miniscule underswing (around 10^{-3} of the signals amplitude at it's lowest point), as identifiable in Fig. 6.9. While not ideal, this is of no practical concern in any circumstance. It could be mitigated in the future by carefully constructing another (presumably more complicated) reference function which describes the pulse's decay better than a single exponential.

[18] In order to save time and effort, regions of unlikely PZC values were not sampled as densely.
[19] Equivalent to a Taylor expansion up to second order of the actual (unknown) function.

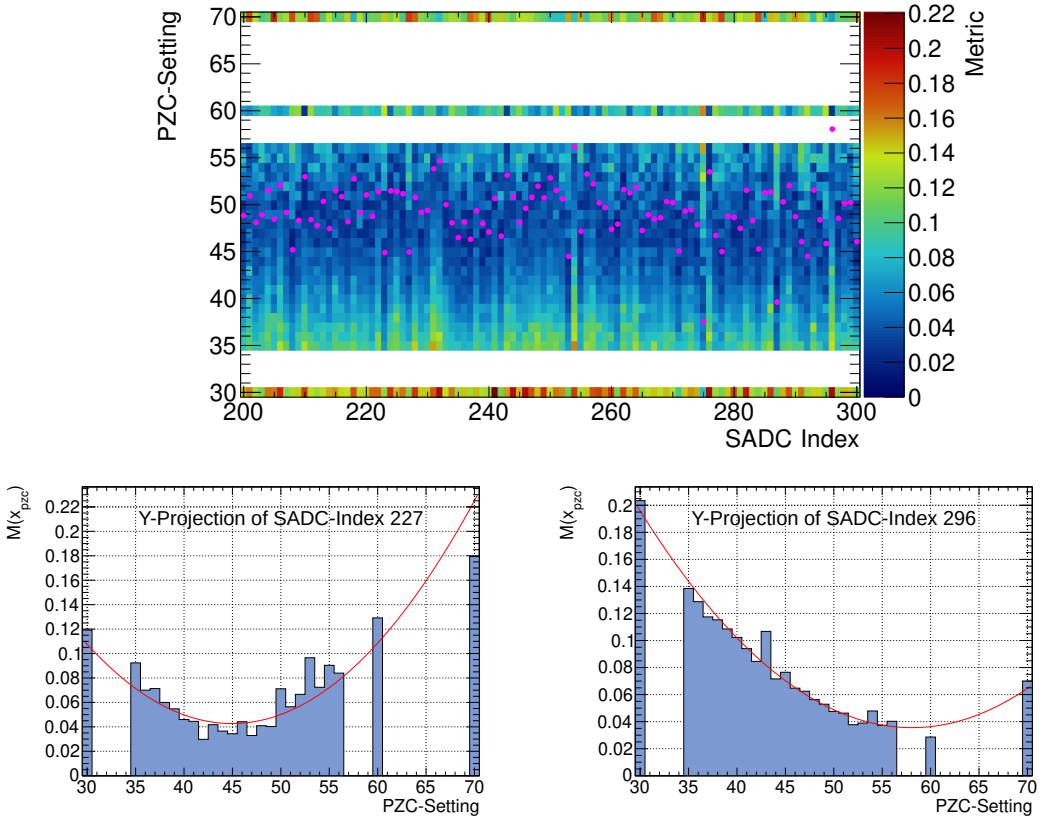


Figure 6.8: Result of the calculations for the ideal PZC setting. The upper plot shows the respective value of the figure of merit M, color coded along the z-axis. It is plotted against the respective SADC channel (only a subset of channels is shown for the sake of readability) on the x-axis and the respective PZC setting on the y-axis. Quadratic fits were performed for each bin along the x-axis; two examples are shown below. The respective minima of the fitted parabolas are overlaid in magenta onto the upper plot.

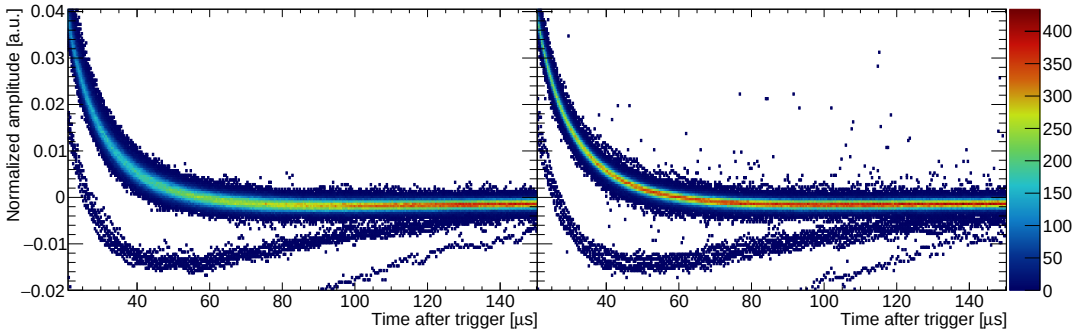


Figure 6.9: Overlay of the extracted mean pulse shapes from each of the 1230 channels. Left plot shows a global PZC setting of 50, whereas the right plot shows the results of the channel-wise optimization. A handful of channels had faulty hardware and were stuck at one PZC setting of (presumably) 127, leading to large underswings[a].

[a] The repair proved difficult, and was only accomplished during a break in the June 2021 beamtime. The PZC-values were provisionally set to "50" before the proper determination could be done.

6.5 Differing Pulseshapes

Even without explicitly studying the shape of the pulses themselves in greater detail, a closer look at the ratio of integral to peak (which was used for the first pile-up detection algorithm) already reveals that the assumption of a constant pulse shape might not always be exactly fulfilled. The circled area in Fig. 6.10 indicates that there seems to be at least one, possibly more, subsets of entries which exhibit a smaller ratio than the majority of pulses. Examining the pulse shapes of these events closer (see Fig. A.12) showed that those are, in fact, narrower than ordinary pulses. This narrowness implies that for an identical peak height, the associated integral is smaller for such pulses, which immediately explains the reduced value of the ratio.
Looking back at the theory of scintillation light output in CsI(Tl) (see section 1.4.3), this is likely caused by different particles traversing the crystals. Provided those particles possess a higher ionization density than photons and electrons or positrons (which cause most pulses), their scintillation light would possess more of the fast component and therefore result in a narrower pulseshape. Such particles would most likely be protons or other heavy charged particles.

6.5.1 Previous Work

It was the topic of a recent bachelors thesis [Grä19] to investigate further, ideally verify this hypothesis and to examine whether the observed difference could be usable for a potential PID. The results, however, were not entirely conclusive. One confounding factor might have been that the "narrowness" of individual pulses was evaluated based on simple cuts on the sampled pulses themselves, which might not have provided the best discrimination possible. An imperfect pole-zero cancellation (see section 6.4) at that point in time might have also led to some variations in the pulseshape and therefore a reduced discrimination power based on the sampled waveforms. Still, some interesting results were produced:
Photons[20] indeed seemed to produce a "normal" pulseshape. Protons on the other hand could be identified to be responsible for *some*, but not all, of the narrow pulses. The correlation was however much less clear than initially anticipated and the differences seemed insufficient for any kind of potential particle identification. To complicate matters further, the scintillation

[20] Photons were identified as decay product of a π^0 meson in a fully reconstructed $\gamma p \rightarrow p\pi^0$ reaction.

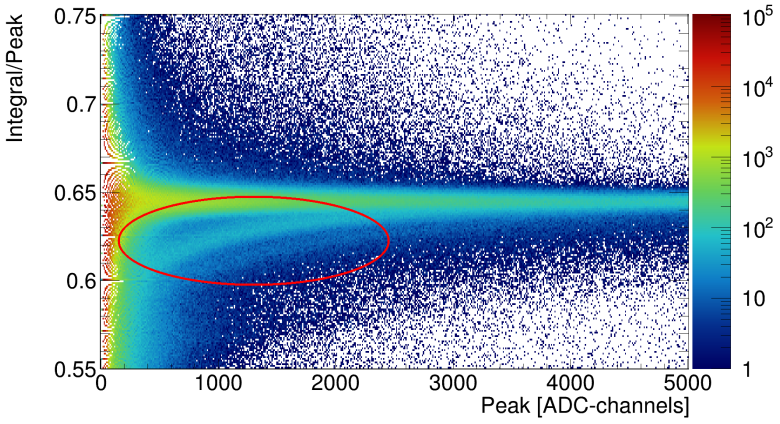


Figure 6.10: Ratio used for pileup detection. Area of narrower pulses is highlighted by red circle. Since discrete values are being divided, not every y-value is realized on the l.h.s. of the diagram. Data from June 2021 beamtime.

characteristics of protons strongly depend upon whether they were stopped in the crystal or not (see section 1.4.3 again). It was shown, that stopped protons lead to narrow pulses, while unstopped protons lead to signals which were mostly indistinguishable from photon pulses. These findings are in agreement with those in [Ben14].

6.5.2 New Findings

In the meantime, while working on the new pile-up detection algorithm, a more elegant way to investigate differences in pulseshape was found: By employing the difference between expected and observed integral, several relatively well separated regions of presumably narrower regions can be identified; see Fig. 6.11. Not only 2, but 4 to 6 (depending on what is considered to be suitably well separated) regions are now identifiable. Interestingly, the entries in these regions seem to follow an energy distribution which has its maximum somewhere around 20 to 30 MeV.

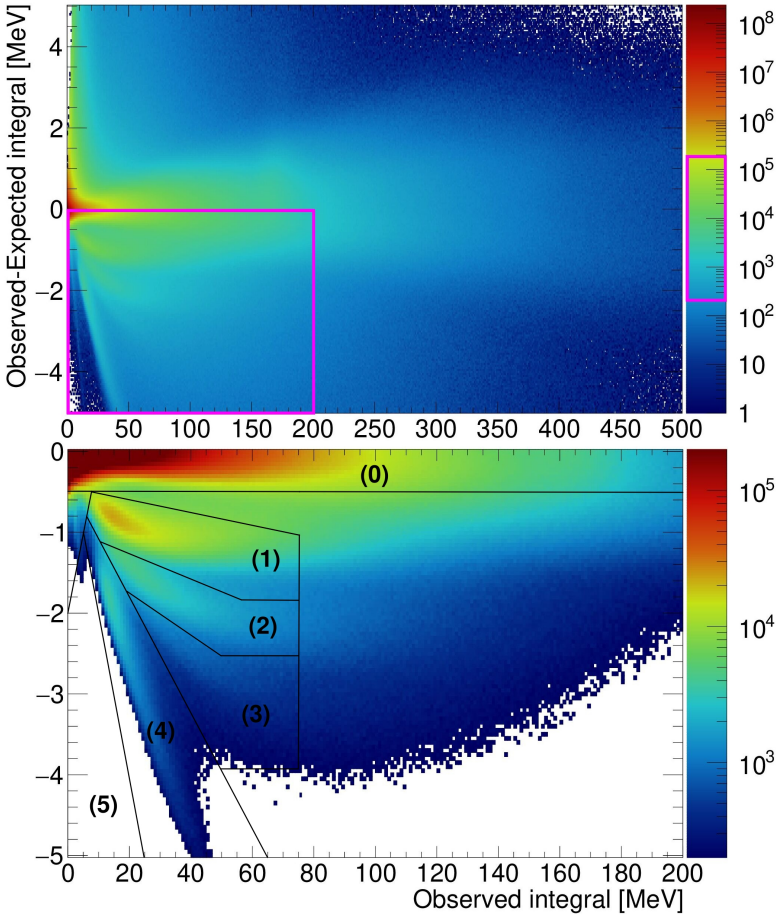


Figure 6.11: Difference between observed and expected integral against observed integral. The "expected" integral was calculated from a pulses peak height and position by using the respective LUT (see section 5.2). The lower plot shows an enhanced view of the lower left region of the upper plot. The z-axis (color scale) was zoomed in as well in order to highlight the different regions. Within the lower plot, several distinct regions were identified and differentiated by simple, polygonal cuts. Data from June 2021.

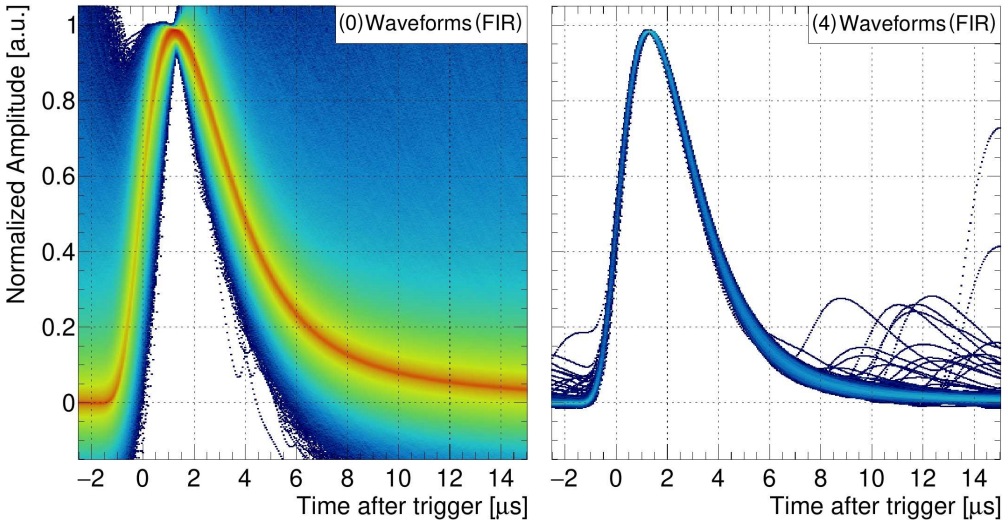


Figure 6.12: Comparison of (FIR-filtered) waveforms from regions (0) and (4); see Fig. 6.11. To aid the comparison, pulses were shifted in time such that their respective peak positions (determined by the feature extraction) coincide in one point. In reality, the pulses from region (4) reach their peak slightly earlier; see also Fig. 6.17.

Fig. A.13 shows a projection from Fig. 6.11 at $x = 30\,\text{MeV}$. 4 peaks are clearly identifiable, the existence of a 5th peak (labeled 3a) seems very likely. An insufficient fit result without the inclusion of a 6th peak (3b) seems to indicate the existence of yet another region[21], which would lie between regions marked (3) and (4).
For energies above $\approx 40\,\text{MeV}$, the populations of regions (2)-(4) seems to slowly decline until they disappear into the background. Region (1) seems to fade as well, but also shows an upwards trend. Above 80 to 100 MeV, the observed pattern seems to disappear altogether.
In light of these new findings and since large parts of the analysis in this thesis hinge on a precise description of the pulseshapes, it seemed imperative to dive deeper in order to gain a better understanding of the origin of such deviating pulseshapes.

6.5.3 Detailed Investigation

Even though this new method of discrimination relies on the extracted features only[22], the different regions in Fig. 6.11 indeed all correlate (from top to bottom) with increasingly narrower pulses. As an example, Fig. 6.12 shows a comparison of pulses from the regions labeled (0) and (4)[23]. The pulseshapes of the other regions fall in between those two extremes, except for region (5) in which pulses were even narrower (this region was not investigated in detail due to its low population).
This existence of more than two regions seems to suggest the presence of more than the two species (γ, p) investigated in [Grä19]. The following particles, ordered by their ionization density

[21] Indiscernible to the human eye (at least with the given resolution power).
[22] As opposed to the studies in [Grä19], which were based on the sampled pulseshapes.
[23] The difference in shape is best visible by comparing with the underlying grid lines; especially on the falling edge.

at typical energies, might reasonably be expected to reach the detector in a hadron spectroscopy experiment such as CBELSA/TAPS[24]:

0. Photons, electrons/positrons, muons, charged light mesons ($\pi^{\pm}$, $K^{\pm}$), protons (not stopped)
1. Protons (stopped in the detector)
2. Deuterons
3. Nuclei with mass number $A = 3$: tritons (a) and ^{3}He-nuclei (b)
4. α particles
5. Recoil nuclei or nuclear fragments (from disintegration of heavier nuclei)

In addition, neutral particles (neutrons and neutrinos) might be expected as well, although they would not leave a signal in the detector[25]. Following the reasoning from section 1.4.4, it is expected that muons as well as light mesons (singly charged and of similar mass) should be roughly as weakly ionizing as an electromagnetic shower from photons or electrons/positrons. The same holds true for unstopped protons, which is why this very heterogeneous collection of particles was subsumed as one item. Based on the discussion so far, the following hypothesis is put forward:

The different regions in Fig. 6.11 are caused by different particle species. The numbering of regions matches the above-mentioned grouping of particles.

For reasons which will be elaborated in the following, it was not possible to conclusively prove or disprove this hypothesis within the scope of this thesis. However, several convincing findings will be presented over the following pages.

6.5.3.1 Reconstructed Particles

Of the particles listed above, only photons and protons are routinely reconstructed in a typical analysis (see section 2.10). As a first test of the hypothesis, reactions with a final state of 1 proton and 2 photons (from a π^0 decay) were investigated. An explicit cut requiring the proton to be charged and the photons to be uncharged was not applied. The specific reconstruction process is outlined section 7.4.1.1 which used the same cuts and data from the same beamtime (June 2021). Fig. 6.14 shows the data from all hits which belonged to a cluster from either the photons or the baryon. Due to the missing charge cut, some amount of neutrons might be included as well, although their probability to leave a detectable (secondary) signal through interaction with protons in the crystals is quite low.

[24] Particles might either originate from the reaction in the target itself (in June 2021, a polarized butanol target was used), or from secondary reactions.

[25] While neutrons themselves do not leave a signal, they might interact with e.g. protons in the material which can then be detected instead.

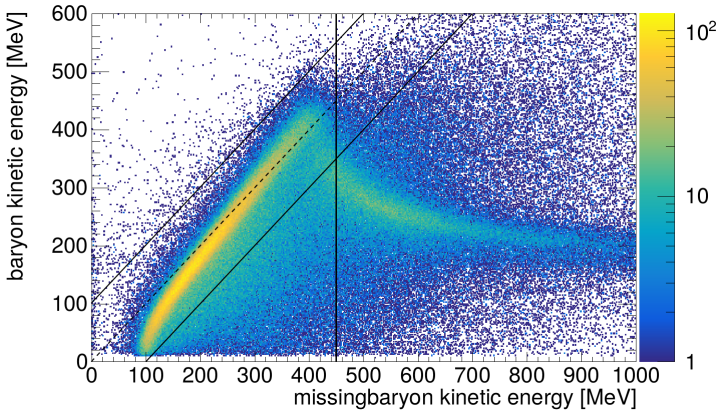


Figure 6.13: Energy deposited by particles kinematically identified as protons (or neutrons) against the kinetic energy of the missing particle (reconstructed from inital state and measured photons). The solid lines denote the respective cut limits (see text) and the dashed line goes along y=x for illustrative purposes. Data from June 2021.

The proton clusters were further subdivided into those which deposited their entire kinetic energy within one single crystal (and nothing in the neighboring crystals) and those which had a kinetic energy in excess of 450 MeV; i.e. too much energy to be fully stopped within a crystal. The selection was based upon the kinematically missing energy as an estimate of the actual baryon energy (see section 2.10). Their relation is shown in Fig. 6.13 (the criterion to have just a single crystal hit is not included).
For the stopped protons, up to 100 MeV less or more were allowed to to account for energy losses and a potential over-estimation of the deposited baryon energy (see section 1.4.4[26]), respectively. For the unstopped protons, the actually measured energy was left unconstrained and no distinction was made whether the energy was distributed over multiple crystals or not.
The following conclusions can be drawn:

- Photons behave as expected and fall almost entirely into region (0).
- Fully stopped protons indeed exhibit narrower pulseshapes and fall into region (1) at lower energies (up to $\approx$ 100 MeV), but show an upward trend towards region (0) for higher energies. This behavior is consistent with the expectations, since for higher energies, a smaller fraction of the entire energy will be deposited near the end of a protons track where the ionization density is largest.
- Unstopped protons show two distinct branches:
  - The upper branch shows an upward trend as well, but seems to exhibit similar or even slightly broader pulseshapes than in the photon clusters. This is to be expected, since high-energy ($E_{\text{kin}} > 450$ MeV) protons already approach the region of minimum ionization; see Fig. 1.13.
  - The lower branch seems to behave in a way similar to stopped protons (seemingly contradicting the hypothesis), but does not extend much further than $\approx$ 100 MeV.
- In addition, some faint indication of a third branch (in region (2)) might be perceived in the baryon clusters. But given the limited size of the sample at hand, this observation seems speculative.

[26] Also notice, how the distribution of entries in Fig.6.13 exceeds the dashed line (y=x) at the highest energies; underlining the fact that the CB calorimeter is calibrated for photons, not protons.

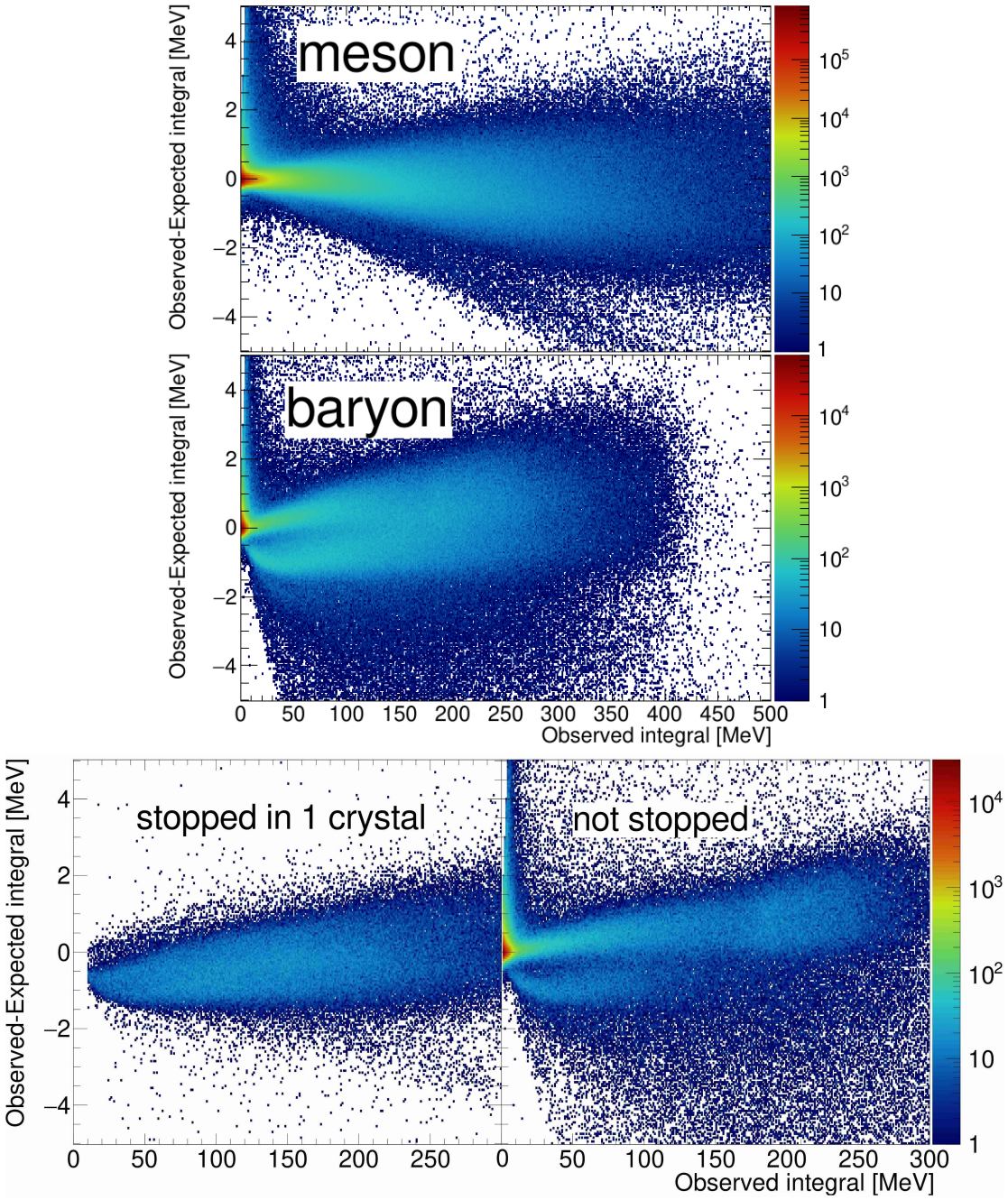


Figure 6.14: Difference of observed and expected integral for all hits involved in clusters which were reconstructed as photons from a π^0 decay (top) or belong to the respective baryon cluster (middle). The bottom plots both show a subset of baryon clusters: Those which involved only a single crystal in which the baryon's entire kinetic energy was deposited are shown on the left. The lower right plot shows clusters of baryons which had a kinetic energy in excess of 450 MeV; see explanations in the text. At least one hit of the cluster had to exceed 10 MeV in order to start the clustering process. This threshold is clearly visible as a cutoff in the lower left plot.

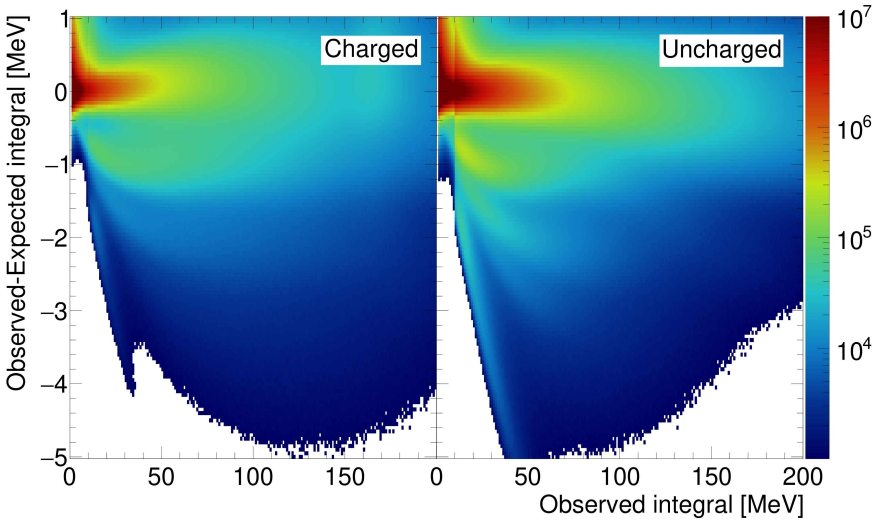


Figure 6.15: Difference of observed and expected integral for all hits involved in a cluster marked as charged or uncharged. The apparent edge in both diagrams is due to the clustering threshold of 10 MeV. At least one hit of the cluster had to exceed this threshold in order to start the clustering process.

6.5.3.2 Charge Information of Clusters

While the results for reconstructed photons and fully stopped protons were consistent with the expectations, the reason for the lower branch(es) in case of unstopped protons remains to be explained. At the same time, the origin of particles populating regions (2)–(5) in Fig. 6.11 has not yet been understood. To investigate further, the information of the charge sensitive detectors might be used directly. Their data is however limited to "uncharged" or "charged"; i.e. there is not way to know the sign or multiplicity of a particle's charge.
Fig. 6.15 again shows the same diagrams as seen before, but this time only for hits which belonged to a charged or an uncharged cluster. No further selection (e.g. on $p\pi^0$ final states) was performed. At first glance, there seems to be no particularly striking difference between charged an uncharged clusters. This would appear to contradict the hypothesis already, since only photons should have produced uncharged clusters and hence regions (1) to (5) should not be populated beyond background level. Upon closer inspection, some subtle differences become apparent:

(a) The main branch of the distribution exhibits a slight upward trend in the charged clusters. No such trend is visible in the uncharged case.

(b) The distribution of assumed protons (region (1)) extends slightly further right in the charged clusters.

(c) All regions except for region (0) seem more densely populated in the uncharged clusters.

Observation (a) already fits well to the just observed behavior for unstopped protons. The subtle upward trend might be explained by a potentially even slightly lower ionization density for unstopped protons in comparison to photons / electromagnetic showers (from which the underlying LUT was deduced). But the other findings are in need of a convincing explanation.

6.5.3.3 Geometrical Considerations

The geometrical detector setup leads to further complications: The inner detector is placed right around the target and the forward veto plates are placed on the target-facing side of the holding structure, whereas the crystals are installed behind these holding structures. The holding structure consists of 1 to 2 cm of aluminum i.e. insensitive material (see section 2.8). Such a material budget should naively provide an effective shielding for heavy charged particles at the observed energies. As an example, an α particle would need an energy of ≈ 200 MeV to even penetrate 1 cm of aluminum [Ber+05]. However, the most probable energy of supposed α particles in Fig. 6.11 seems to lie around 20 MeV. To produce such a distribution from α particles generated in or somewhere around the target, an energy distribution with a (local) maximum around 220 MeV would be needed, which seems highly artificial. So on top of the previous confusion, the question arises how such heavy particles (if produced in the target) could even reach the crystals; especially given their respective energy distribution.

6.5.3.4 Nuclear Interactions

After careful investigation and extensive literature research, the most likely reason for these observations seems to lie in the possibility of nuclear interactions; mostly in the CsI(Tl) crystals themselves and, to a lesser extent, also in the Al holding structure. The most abundant particle species which could cause these reactions are (in a simplified model) photons, neutrons[27] and protons. Only the latter would leave a signal in the charge sensitive detectors, while the other two would not.
Fig. 6.16 shows the respective reaction cross sections in ^{133}Cs[28]. It is immediately obvious that there exists a substantial cross section on the order of 100 to 1 000 mb for the production of one or more neutrons from all three species at energies around 10 to 30 MeV. This already strongly correlates with the observed maxima of the respective energy distributions[29]. In addition, sizable cross sections on the order of 1 to 10 mb exist for the production of protons, deuterons, tritons, α particles and even ^{3}He.
The respective conversion probabilities are directly proportional to the cross sections:

$$p_{\text{conversion}} = 1 - e^{d\cdot\rho\cdot\sigma} \approx \underbrace{d\cdot\rho}_{=:\kappa}\cdot\sigma \quad (\text{for } d\cdot\rho\cdot\sigma \ll 1) \tag{6.15}$$

For a $d = 30$ cm long crystal of CsI[30], the conversion factor κ comes out as $\approx 0.62\,\text{b}^{-1}$. For example, a neutron entering the crystal with an energy of 20 MeV would, given a cross section of 12 mb, have a probability of $\approx 0.75\%$ to lead to the emission of an α particle within the crystal.

[27] In principle, neutrons should be expected in roughly similar numbers as protons in the experiment. On the one hand, several neutrons are found in the target material of butanol (C_4H_9OH implies 42 protons and 32 neutrons per molecule) as well, from which the photoproduction reaction could have taken place. On the other hand, resonances like the $\Delta(1232)$ could also decay into final states involving a neutron instead of a proton.

[28] The cross sections obtainable for ^{127}I in [Kon+19] seem to be almost identical, presumably due to the similar nuclear charge and mass number.

[29] The exact energy distribution of the produced particles (ejectiles) could not be obtained. Also see following section 6.5.3.6 (Discussion).

[30] Average molar mass $M \approx 130\,\text{g mol}^{-1}$, density $\rho = 4.51\,\text{g cm}^{-3}$

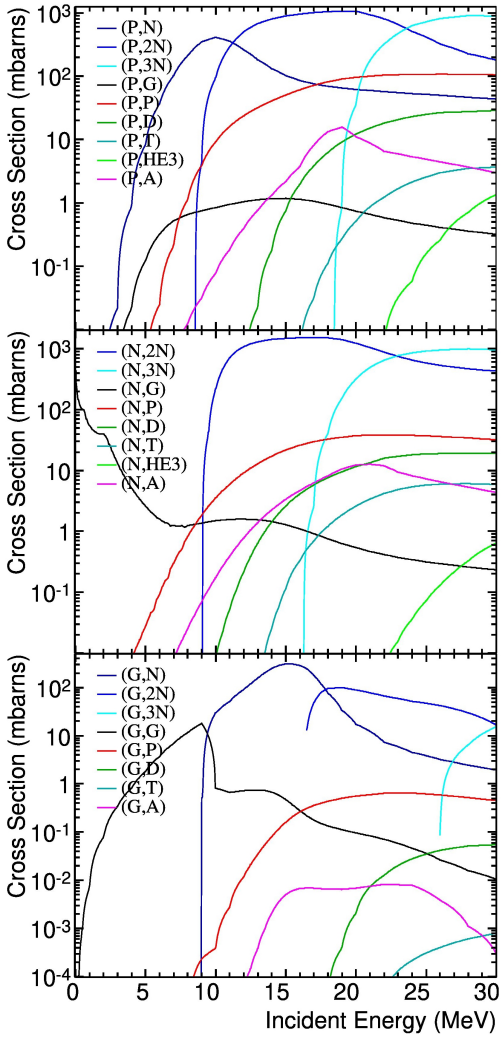


Figure 6.16: Various cross sections of possible nuclear reactions in ^{133}Cs. Notation (X,Y) means X is the projectile which induces a reaction and Y are the respective ejectiles. (N=neutron, P=proton, D=deuteron, T=triton, HE3=^{3}He, A=α-particle, G=photon (γ)). The color of the respective cross sections is related to the ejectiles and is kept consistent over all three plots.
The plots were obtained from www-nds.iaea.org/exfor/endf.htm and use data from the TENDL-2021 library (latest update not yet published; previous publication: [Kon+19]). Unfortunately for this investigation, energies above 30 MeV were not part of this library. Likewise, the resulting energy distributions of the ejectiles could not be obtained.

6.5.3.5 Interpretation

As seen in Fig. 6.16, reactions leading to the emission of one or more neutrons have the largest cross sections. Neutrons in turn have significant probabilities in the order of 0.1 to 1% to produce heavy charged particles in the crystals. The proton-induced direct production of other heavy charged particles is (despite similar cross sections) probably less relevant, since the cross sections are strongly energy dependent. Protons continuously lose energy through ionization in the scintillator crystals, so their effective range in which they would have a suitable energy for the relevant reactions will be substantially lower than the full length of 30 cm. Neutrons on the other hand suffer no such energy losses and can (unless they undergo a nuclear reaction) be assumed to traverse the medium with a constant energy.

Firstly, the most plausible explanation for the well-populated regions (1) to (4) in the uncharged clusters seem to be reactions from neutrons. In addition to primary neutrons, which probably account for most of these events, some secondary neutrons from (N,N) or (G,N) reactions might also contribute. However, their contribution will be suppressed, since 2 cross sections ((X,N) and (N,Y)) enter. As seen in Fig. 6.16, the direct production of heavy charged particles by photons is possible as well, but orders of magnitude more unlikely. This also aligns well with the previous observation, that *more* heavy charged particles seem to be present in *uncharged* clusters than in charged clusters.
The additional branch(es) seen in the diagram of unstopped baryons seem to be consistent with secondary protons, which were (unlike their "parent" particle that induced the reaction) stopped inside the crystal. Both protons and neutrons might come into question as primary particles in this case. Although the cross sections at energies above 450 MeV could not be obtained for this discussion, especially inelastic proton-proton scattering (i.e. scattering of a projectile proton off a proton (initially) bound in a nucleus) seems to be a likely candidate for the production of lower-energy protons which can be stopped in the crystals.
Finally, the presence of deuterons, tritons, ^{3}He and α particles in charged clusters cannot be attributed to reactions induced by primary neutrons, since neutrons are obviously uncharged. Instead, they must have been produced by primary protons, either directly or via secondary neutrons as well. Should any significant contribution due to primary particles of the respective particle species exist at all, it is most certainly overshadowed by secondary (or tertiary) particles, which also lead to the particular energy distribution observed.

6.5.3.6 Discussion

The line of reasoning provided above is obviously massively simplified. For example, the energy distribution of the produced particles was not discussed. Despite an apparent lack of data and/or theoretical predictions[31], a first-order approximation would be to assume that the (total) energy of the ejectile(s) is approximately equal to the primary particle's energy, since the nuclear recoil energy should be small in comparison. In addition, the direction of emission might be entirely different from the initial particle's direction, a fact which was also not discussed[32].
The exact role of photons also seems somewhat unclear. On the one hand, no secondary particles were observed in the explicit photon clusters (see Fig. 6.14). This could be explained by the fact that secondary reactions might lead to a somewhat different total energy deposition or reconstructed angle, which would in turn result in an invariant mass which does no longer match the π^0 mass anymore. At the same time, the seemingly large maximum neutron production cross section of ≈ 300 mb has to be compared to the pair production cross section of ≈ 7.6 b [Ber+10]; i.e. only every 25$^{\text{th}}$ photon of a suitable energy would result in the emission of a neutron. This would be somewhat compensated by the multiplication of photons within an electromagnetic shower, but the exact effect is difficult to estimate.
Another confounding factor is the aluminum holding structure. Of course, nuclear reactions are possible here as well but were, for the sake of simplicity, omitted in the discussion. Besides the

[31] Due to research in radiation therapy, some data for the energy distribution of ejectiles in water or other surrogate materials for human tissue exists. But to the best of the author's knowledge, no such data exists for a material like CsI.

[32] As long as the ejectiles stay within the detector, this should not matter anyways.

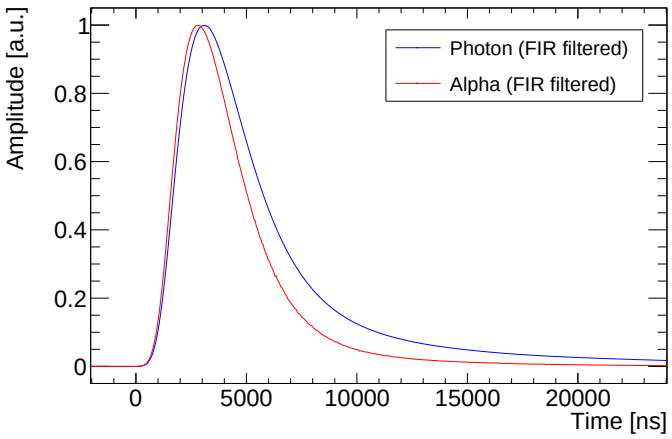


Figure 6.17: Extracted pulseshapes of supposed photons (blue) and alpha particles (red). The pulses were subjected to a FIR-filter (see section 6.2) prior to the extraction, which was performed analogous to section 6.3. Data was taken from the June 2021 beamtime.

production of further neutrons, the impact is questionable anyways, since e.g. α particles with an energy of 20 MeV would most likely not get out of the aluminum in the first place[33]. In any case, this does not take away from the reasoning, as including the holding structure would only lead to more secondary particles in the crystals. Exploring all possible reactions and discussing their several interconnections would go beyond the scope of this thesis.
The main conclusion to be drawn from this discussion is that the presence of heavy charged particles which lead to narrower pulseshapes is highly plausible. Besides primary protons, the likely origin for most of these particles lies in nuclear reactions within the detector itself.

6.5.4 Parametrization

Regardless of how these pulseshapes come about and whether they might eventually be used for the purpose of particle identification, they have to be accounted for in the description of pulse shapes. In order to do so, a mean pulseshape was extracted both from region (0) and from region (4). The former one is dubbed the "photon pulseshape" P_γ (which is by far the most common one) and the latter one the "α pulseshape" P_α (see Fig. 6.17 for a direct comparison). Since those are the limiting cases in both directions, the general pulseshape for any particle can be assumed to be reasonably well described as a linear combination of the two:

$$P(t) = (1 - S) \cdot P_\gamma(t) + S \cdot P_\alpha(t) \tag{6.16}$$

The dimensionless parameter S ranges from 0 (pure γ pulseshape) to 1 (pure α pulseshape), although slightly wider ($S < 0$) or narrower ($S > 1$) pulseshapes might be permitted as well. It should be further noted that P_α reaches its maximum somewhat earlier than P_γ. This is again due to more of the fast scintillation component which results in a steeper rising edge. In effect, a variation of S not only modifies the pulseshape per se, but also slightly shifts the maximum of the pulse.

[33] Their effective range in Al is just 0.2 mm.

6.6 Deconvolution

Deconvolution is the reverse operation to convolution. While the convolution of two functions

$$f(t) = (g * h)(t) = \int_{-\infty}^{\infty} g(\tau) \cdot h(t - \tau)\, \mathrm{d}\tau \tag{6.17}$$

is easily[34] computable, the same is not true for the reverse operation: Given the result $f(t)$ and knowing $h(t)$, it is in general non-trivial to deduce $g(t)$.
In the context of digital signal processing, the need for deconvolution often arises. For example, suppose a filter with a known transfer function $h(t)$ is acting on a raw signal $g(t)$ to produce a measurable output $f(t)$. Deconvolution attempts to answer the question how the output signal $g(t)$ would have looked without the filter.

6.6.1 Fundamental Considerations

From a purely mathematical standpoint, assuming precise knowledge of the output signal $f(t)$ and the transfer function $h(t)$, deconvolution is relatively easy to achieve. Instead of trying to compute the deconvolution in the time-domain, it is advantageous to exploit one of the properties of the Fourier transformation (see section 6.1): Convolution in the time-domain is equivalent to a multiplication in the frequency-domain and vice versa. Consequently, the deconvolution operation is equivalent to a simple division in the frequency-domain. The following diagram attempts to visualize the process:

	Time-Domain	Frequency-Domain
Convolution	$f = g * h$	$\hat{f} = \hat{g} \cdot \hat{h}$
	$\downarrow$ hard	$\downarrow :\hat{h}$ (easy)
Deconvolution	$g = f(*)^{-1} h$	$\hat{g} = \frac{\hat{f}}{\hat{h}}$

$\mathfrak{F}$ (easy) $\rightarrow$ (Convolution, Time-Domain $\rightarrow$ Frequency-Domain)
$\mathfrak{F}^{-1}$ (easy) $\leftarrow$ (Deconvolution, Time-Domain $\leftarrow$ Frequency-Domain) (6.18)

where $(*)^{-1}$ denotes a ficticious "deconvolution operator".

6.6.2 Use Case at CBELSA/TAPS

Fundamentally, there are only two questions which the feature extraction aims to answer: How much energy was deposited, and when did this energy deposition happen? The answer to those questions comes down to the determination of the height (or integral) and position of a pulse[35]. This task becomes non-trivial if pulses overlap and produce pile-up (see section 5), which would be less likely if the pulses were significantly shorter. The (relatively "long") pulseshape which is observed is the result of a series of amplifiers, integrators, filters, etc. which all acted upon the light output signal from the CsI(Tl) crystals.

[34] Not necessarily analytically, but the numerical calculation is straightforward.
[35] Disregarding differences in pulseshape for a moment.

Mathematically, any recorded waveform $y(t)$ can be modeled as a convolution of a series of scaled and shifted original pulses $g(t)$[36] with a common transfer function h that encapsulates all effects acting upon the signal:

$$y(t) = B + \left(\sum_i H_i \cdot g\left(t - T_i\right) \right) * h(t) \tag{6.19}$$

where B is the baseline and H_i and T_i denote height and position of pulse i.
From a signal processing point of view, it would be advantageous to "undo" everything which leads to an elongation of the pulses. In this context, the transfer function h can be considered a "nuisance" which would be advantageous to remove. The narrower the individual pulses are in the result, the easier it would be to determine the height and position of each pulse and the lower the pile-up probability would be. For reasons which will become apparent shortly, the achievable width is fundamentally limited, but different approaches to shorten the pulses substantially will be presented.

6.6.3 Moving Window Deconvolution

Principle There are special cases in which a deconvolution can be calculated analytically, without the need to perform back-and-forth Fourier transformations. One such algorithm is commonly known as a Moving Window Deconvolution (MWD). It was introduced in 1994 [GG93] and has since been used regularly in digital signal processing. Due to its simplicity, it is the method of choice for implementations on an FPGA. A summary and exemplary application of the method can be found in [Sto12].
In case of a signal which is composed of a series of exponentially decaying pulses, each with a common decay time τ, the MWD is capable of removing the exponential decay exactly. It is based on the fact that the integral of an exponential function $\exp\left\{-\frac{t}{\tau}\right\}$ is equal to the original function itself; up to factor of $-\frac{1}{\tau}$. The algorithm can be broken down into 2 steps: The actual deconvolution process and a numerical differentiation to avoid excessive pile-up. The principle is depicted in Fig. 6.18.
The first step revolves around getting rid of the exponential decay. It depends critically on precise knowledge of the decay constant τ. Mathematically, the calculation is straightforward. Starting from a single exponential with an instantaneous rise at t_0:

$$A(t) = \begin{cases} 0 & \text{for } t < t_0 \\ N \cdot e^{-\frac{t-t_0}{\tau}} & \text{otherwise} \end{cases} \tag{6.20}$$

The deconvolution function is defined as follows:

$$D(t) = A(t) + \frac{1}{\tau} \int_{-\infty}^{t} A(\tilde{t})\, \mathrm{d}\tilde{t} \tag{6.21}$$

[36] The precise nature of this "original" signal is deliberately left unclear at this point. For example, the exponential decay of the scintillation signals themselves may be considered to be part of g or of h, depending on the context (i.e. whether the preamplifier is considered to be fundamentally part of the signal generation or of the following electronics). See section 6.6.4 for further discussions.

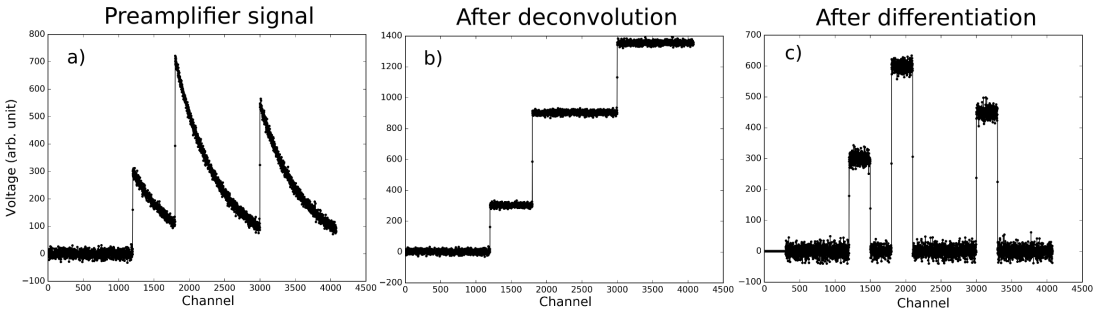


Figure 6.18: The principle of MWD illustrated [Rot16]. After the actual deconvolution, a staircase-like signal is created, which gets then transformed to rectangular pulses of width w by the differentiation.

Exploiting that the integral from $-\infty$ to t_0 vanishes anyways, the integral can be easily calculated:

$$\frac{1}{\tau}\int_{t_0}^{t} A(\tilde{t})\,\mathrm{d}\tilde{t} = \Big[-A(\tilde{t})\Big]_{t_0}^{t} = N - A(t) \tag{6.22}$$

It immediately follows that

$$D(t) = \begin{cases} 0 & \text{for } t < t_0 \\ N & \text{otherwise} \end{cases} \tag{6.23}$$

Since integration is a linear operation, this will apply to sums of single exponentials as well. In case of a sampled, non-continuous signal, the integral in the definition (6.21) has to be replaced by a summation[37]. The value of the deconvolution function at a point t_k will be given by

$$D(t_k) = A(t_k) + \frac{1}{\tau}\sum_{i=-\infty}^{k-1} A(t_i) \tag{6.24}$$

At this point, the exponential decay is completely removed, but the resulting signal will tend towards infinity the more signals arrive (see stage (b) in Fig. 6.18). To avoid this and to make the infinite sum computable, a numerical differentiation is added to get the actual MWD output:

$$\mathrm{MWD}(t_k) = D(t_k) - D(t_{k-w}) = A(t_k) - A(t_{k-w}) + \frac{1}{\tau}\sum_{i=k-w}^{k-1} A(t_i) \tag{6.25}$$

Where the width w of the *moving window* was introduced. Ideally, w would be just one sample, thus achieving the optimal deconvolution which results in a series of sharp peaks and nothing else. In any real world application however, the original signal will take some finite time ω before the situation of an actual, pure exponential decay is reached. In order to avoid artifacts, w must not be chosen smaller than ω.

[37] To ensure the correct "step size", i.e. $D(t_0) \overset{!}{=} N$, the summation must run to $k-1$ instead of k. In case of the continuous integration, this distinction does not have to be made, because in the limit $\epsilon \to 0$ ($\epsilon \in \mathbb{R}$) the integral $\int_a^{b-\epsilon} f(x)\,\mathrm{d}x$ is identical to $\int_a^b f(x)\,\mathrm{d}x$.

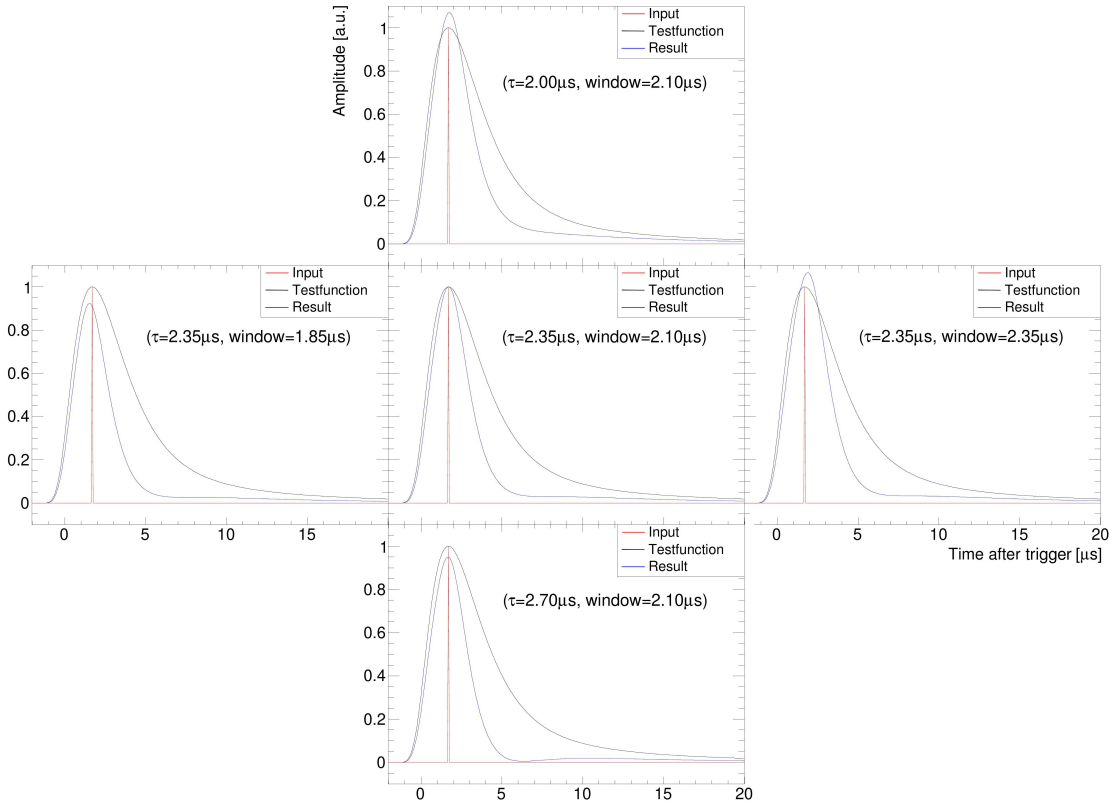


Figure 6.19: Test of different decay times τ (varied from top to bottom) and window lengths w (from left to right) for the MWD on the FIR filtered reference pulse shape. The values which were finally used are depicted in the central plot. All output signals (blue) were scaled down by a constant factor (≈ 1.5) to match the original height in the central plot.

Applicability to CB Waveforms As demonstrated in section 6.3, the average CB pulseshape cannot be described satisfactorily by a single exponential. Yet the MWD relies on exactly this property, so one might expect that the deconvolution could never work for those signals. But since the deviations from a single exponential are not too severe, it was still tried.
Testing different values for the decaytime and the window length (see Fig. 6.19) revealed, that while no choice of τ achieves a complete removal of the long falling edge, it can still be reduced to an appreciable degree. A choice of $\tau = 2.0\,\mu s$, corresponding to the faster decay time estimated in Fig. 6.6, produces an output which still shows signs of the longer decay component(s). This makes intuitive sense, as only the shorter decay component is removed by the MWD. Choosing $\tau = 2.7\,\mu s$ on the other hand leads to an output signal which quickly drops to the baseline level, but then rises back up to form a flat second peak which again stems from an incomplete removal of the long decay component. The empirically best value was found to be in-between at $\tau = 2.35\,\mu s$. It achieves the best possible removal of the falling edge without introducing a local minimum. On the downside, the signal still does not seem to drop back to the baseline level entirely.
The width w should be at least as big as the rise time of the pulse. Judging from Fig. 6.6, this

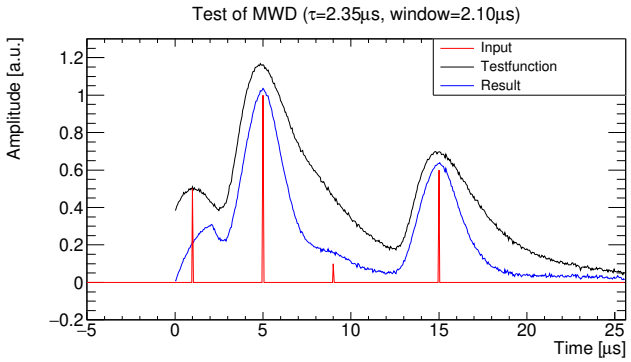


Figure 6.20: Performance of MWD with optimized parameter settings in a modeled pile-up situation. A small amount of random noise was added as well. The black curve shows the test signal and the blue curve the corresponding deconvolution result. The positions and heights of peaks used for constructing the input signal are shown in red.

should be around 2.0 to 2.5 µs[38]. A width of $w = 2.1$ µs was finally deemed appropriate, since it leaves the position of the maximum unchanged. Comparing the sub-plots from left to right, it is evident how a smaller width leads to an earlier and (relatively) smaller peak (indicating that not the entire signal amplitude was accumulated before the differentiation took place) whereas a bigger width does the opposite.
The question arises, how well this deconvolution would work in an actual pile-up situation. Fig. 6.20 shows an example: It is obvious, that the deconvolution constitutes a substantial improvement over the raw signal. The third (small) peak for example is, while still not completely separated, much more discernible in the deconvolution result than before. However, a crucial drawback is also apparent: The actual positions and heights of the individual peaks are not exactly recovered by the deconvolution! Apart from the first peak (which suffers from the fact that its rise was not completely sampled), the reason for this lies in the incomplete removal of the falling edge of the pulses. Accordingly, the peaks in the deconvolution result are still a bit higher than they ought to be (red peaks), depending on the height and distance to the previous pulse. As such, this method of deconvolution is of limited use to disentangle piled-up pulses.

6.6.4 Custom Deconvolution Method

Due to the suboptimal performance of the MWD, other options for a satisfactory deconvolution had to be investigated. One alternative was already presented in section 6.6.1, i.e. doing a Fourier transformation of the signal, dividing by the transform of the transfer function h and then doing the back-transformation.
There are two problems with this approach: On the one hand, the FFT itself is a rather complex operation which takes a certain amount of time to compute and is not trivial to efficiently implement on an FPGA[39]. On the other hand, and fundamentally more relevant, the transfer function h itself is unknown! While it might be possible, albeit tedious, to precisely measure the

[38] An exact value is hard to provide, since the onset of the pulse is smoothed out by the FIR filter.
[39] In case the deconvolution should be moved from offline analysis to the firmware in the future (presently not planned).

frequency response of every component in the signal chain, the following discussion will show why this is not even necessary.

6.6.4.1 Initial Idea

Lacking precise knowledge of the transfer function h, an alternative formulation of equation 6.19 might be employed:

$$y(t) = B + \left(\sum_i H_i \cdot \delta \left(t - T_i \right) \right) * P(t) \tag{6.26}$$

where $P(t)$ is the extracted pulseshape, see section 6.3. In this formulation, the original signal g was reduced to the simplest form possible: a series of scaled and shifted δ-distributions. Accordingly, the *entire* pulseshape $P(t)$ is treated as the transfer function h. This implies that even the exponential decay of the scintillation signal itself would be incorporated into h. But naively trying to remove P by taking the Fourier transform of the observed signal and dividing by $\hat{P}$ in the frequency domain would be ill-advised:
As seen in equation 6.6, the reconstruction of a δ distribution in the time domain requires frequencies up to infinity[40]. Given the frequency spectrum of the signals seen in Fig. 6.2, frequencies above 1 to 2 MHz are (apart from some small noise contributions) not part of the signal; neither of the sampled signal, nor of the reference pulseshape[41]. Doing a point-wise division $\hat{y}(f) \div \hat{P}(f)$ at frequencies above 1 MHz would mean dividing a very small value by another very small value. So even if some incredibly small contribution from the original signal still existed at such frequencies, it would need to be massively amplified. The same amplification would apply to any noise contributions at the same frequency, and would as such completely dominate the end result. It is simply *impossible* to successfully reconstruct frequencies absent from the sampled signal in any meaningful way.

6.6.4.2 Improvements

Even though a δ-distribution is unachievable, the approach itself has potential. There would not have been an issue if the "target" function g did not require such high frequency contributions. To be more precise: Only functions whose power spectral density shows a similar or steeper decline than $\hat{P}$ can be realistically expected as result of a deconvolution. But what could a suitable function be? Following the discussion so far, the purpose of a deconvolution has always been to restore a component of the signal which was convoluted with some other function. Modeling the original signal by a series of δ-peaks was already questionable in that regard, as it would have reduced the original signal to just an instantaneous deposition of the entire energy, even though this could conceivably not have been measured directly.
Taking this realization a step further, there is no reason to refrain from *choosing* any convenient function as initial[42] signal shape g, even though the actual signal most certainly never had that form at any point in the signal chain. The most natural choice for a simple, peaking function is a Gaussian (G). Its position, height, and width are already conveniently parameterized in its

[40] Or, practically speaking, at least up to the Nyquist frequency.
[41] This problem is exacerbated by the applied FIR filter, but even without it the argument would still apply.
[42] Whatever function is assumed as initial signal shape will also be the target of the deconvolution.

definition and it has the very useful property that its Fourier transform is again a Gaussian:

$$\mathfrak{F}\left[A \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}}\right] = \frac{A}{\sigma} \cdot e^{-\frac{f^2}{2\sigma^{-2}}} \quad \left(\cdot e^{-i2\pi f\mu}\right) \tag{6.27}$$

(The fact that the width of the transformed Gaussian is the inverse of the original width σ illustrates a more general principle: The narrower signals are in the time domain, the wider they are in the frequency domain and vice versa.)
In that case, the parametrization for a single pulse (with baseline already subtracted) reads as follows:

$$y(t) = P(t) = \left(\sum_i H_i \cdot e^{-\frac{(t-T_i)^2}{2\sigma^2}}\right) * h(t) \tag{6.28}$$

Provided that the chosen Gaussian matches the frequency spectrum of the sampled signal, the according transfer function h by which a Gaussian G can be converted into the sampled Pulseshape P can be calculated:

$$\begin{array}{ccc} P = G * h & \xrightarrow{\mathfrak{F}} & \hat{P} = \hat{G} \cdot \hat{h} \\ & & \downarrow :\hat{G} \\ h = \mathfrak{F}^{-1}\left[\frac{\hat{P}}{\hat{G}}\right] & \xleftarrow[\mathfrak{F}^{-1}]{} & \hat{h} = \frac{\hat{P}}{\hat{G}} \end{array} \tag{6.29}$$

Note that this h does no longer need to be related to the electronic components in the signal chain at all. Once h has been calculated (technically, the Fourier transform $\hat{h}$ would suffice), the deconvolution $D(t)$[43] of any sampled waveform $y(t)$ can be easily calculated by the process outlined in section 6.6.1: Calculate the Fourier transform $\hat{y}$, divide by $\hat{h}$ and do the back transformation.

The second challenge remaining to be addressed is the need to compute both a forward and a backward Fourier transform of every new signal at hand. Fortunately, there is an easy solution: The division by $\hat{h}$ in the frequency domain can be equivalently expressed as a multiplication by $\frac{1}{\hat{h}}$. While this might sound obvious, we remind ourselves that multiplication in the frequency domain is equivalent to a convolution in the time domain. So the actual deconvolution operation $\hat{y} \div \hat{h}$ might be carried out in the time domain as

$$\underbrace{g}_{=:D} = y * \underbrace{\mathfrak{F}^{-1}\left[\frac{1}{\hat{h}}\right]}_{=:H} \xleftarrow[\mathfrak{F}^{-1}]{} \hat{g} = \hat{y} \cdot \frac{1}{\hat{h}} = \frac{\hat{y}}{\hat{h}} \tag{6.30}$$

[43] i.e. a series of scaled and shifted gaussians.

where we defined the *deconvolving function* H as

$$H = \mathfrak{F}^{-1}\left[\frac{1}{\hat{h}}\right] = \mathfrak{F}^{-1}\left[\frac{\hat{G}}{\hat{P}}\right] \tag{6.31}$$

To summarize: The deconvolution of any sampled waveform down to a series of Gaussians can be achieved by a simple convolution of the signal with a specific function H, which has to be determined *once* by the process outlined above.

$$D(t) = y(t) * H(t) \tag{6.32}$$

Technical Remarks As evident by the middle plot in Fig. 6.21 (discussion in upcoming section), the original pulseshape[44] still exhibits some amount of noise above 1 MHz. Its average magnitude seems to lie in between 10^{-5} and 10^{-6}. On the one hand, this order of magnitude is around the expected suppression due to the FIR filter. Some part of it might also be caused by numerical imperfections and rounding in the extraction process; since the FIR filter was applied to all waveforms before they went into the histogram from which the pulseshape was extracted, such effects would not be mitigated by the filter.
Whatever the cause, this noise would have a deteriorating effect onto the calculation of H, since it would be disproportionately amplified due to the smallness of $\hat{G}$ at these frequencies. In order to mitigate this, a small number can be added to the denominator. This is a very simple case of what is more generally known in mathematics as regularization. More precisely, since $\hat{g}$ and $\hat{P}$ are complex functions, the denominator should be squared before the regularization parameter $\epsilon \in \mathbb{R}$ is added:

$$H(t) = \mathfrak{F}^{-1}\left[\frac{\hat{G}}{\hat{P}}\right] = \mathfrak{F}^{-1}\left[\frac{\hat{G}\cdot\hat{P}^*}{\hat{P}\cdot\hat{P}^*}\right] \approx \mathfrak{F}^{-1}\left[\frac{\hat{G}\cdot\hat{P}^*}{\hat{P}^2+\epsilon^2}\right] \tag{6.33}$$

This obviously implies some small distortion of the function; so one has to be careful to choose ϵ as big as needed, but not bigger than necessary. A value of $\epsilon = 10^{-5}$ has proven very effective in that regard. It is obviously larger than the observed noise amplitudes, while its impact on the end result will be miniscule: Since the dynamic range of the SADCs consists of $2^{16} \approx 0.65\times10^5$ ADC channels, a relative error of 10^{-5} would (in first order approximation) lead to almost unnoticeable deviations well below 1 ADC channel, even for the largest possible pulses.

6.6.4.3 Implementation

Fig. 6.21 illustrates the entire process. A Gaussian with a width on the order of 1 µs[45] was found to be a suitable target function. It is narrow enough to provide a substantial improvement over the original pulseshape, while at the same time being wide enough that its Fourier transform declines steeper than $\hat{P}$, as evidenced by the plot in the middle. While this plot shows only the magnitudes of the contributing frequencies, the entire calculation involves a real and an imaginary

[44] As it was extracted from the data, see section 6.3.

[45] 0.75 µs was used in this case (see upper plot) which seems to be close to the limit of what is possible. For the actual implementation, it was decided to use the more "conservative" value of 1 µs instead, which lines up with the significant frequency contributions being below 1 MHz as well ($(1\,\text{MHz})^{-1} = 1\,\mu\text{s}$).

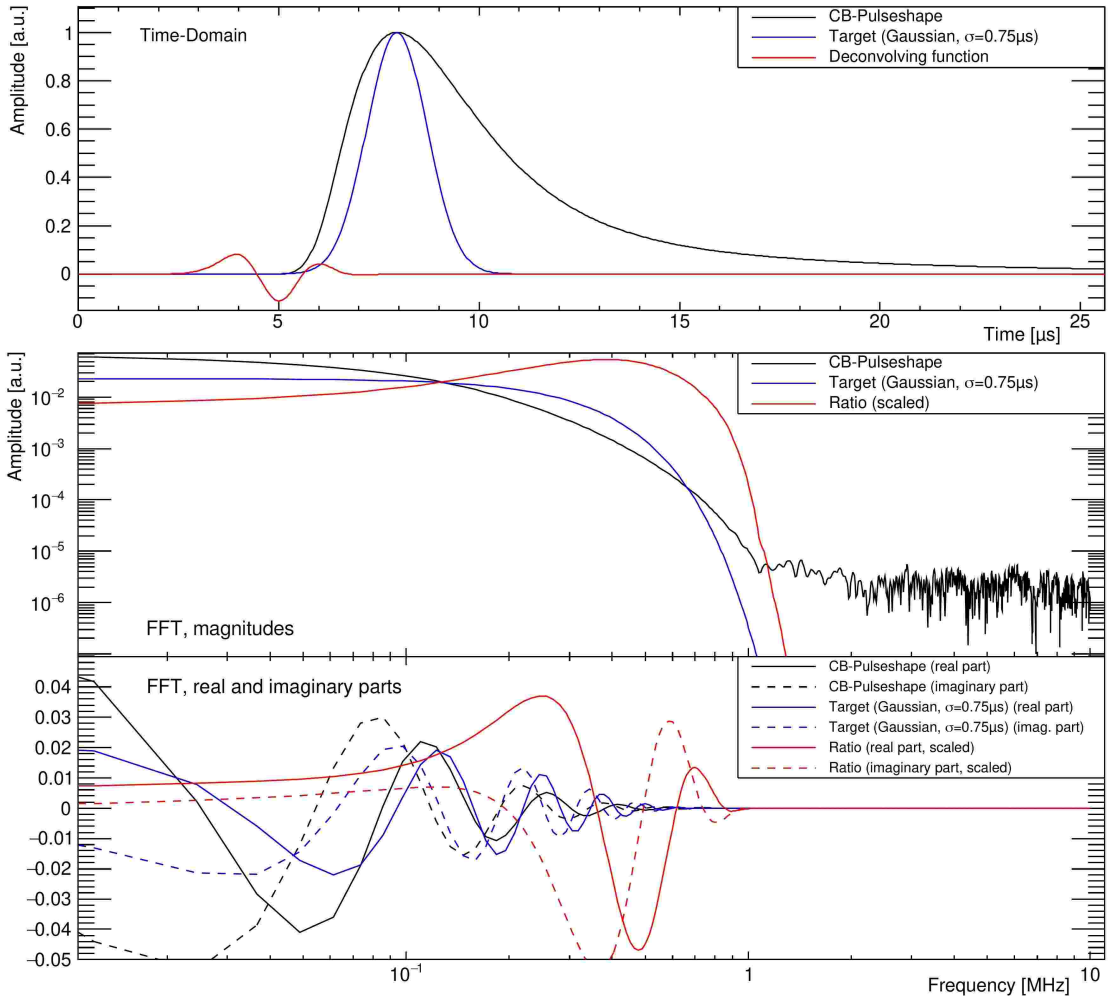


Figure 6.21: Calculation process of the deconvolving function in the time (top) and frequency domain. (middle and bottom). The actual deconvolving function (red) would be centered around $t = 0$ in the upper plot, but was shifted for clarity in this representation. With the function as depicted, the result of the deconvolution has to be shifted back in time by a fixed amount in order to line up with the input signal.

part. Those are depicted in the lower diagram. It can be seen that the ratio[46] between the target Gaussian $\hat{G}$ and the CB-pulseshape $\hat{P}$ rapidly approaches zero for frequencies above 1 MHz, as required. The actual backtransformation H of this deconvolving function can be examined in the upper plot again, shown in red. Convolving the original pulseshape with this function yields (by design) precisely the Gaussian that went into its calculation and which is shown in blue.

6.6.4.4 Application and Performance

Fig. 6.22 shows the same test scenario which was used to test the MWD before. Especially in comparison to those results, the performance is remarkable. Even the third peak (which sits on

[46] This ratio is precisely the Fourier transform of the needed transfer function, i.e. $\hat{h}^{-1}$.

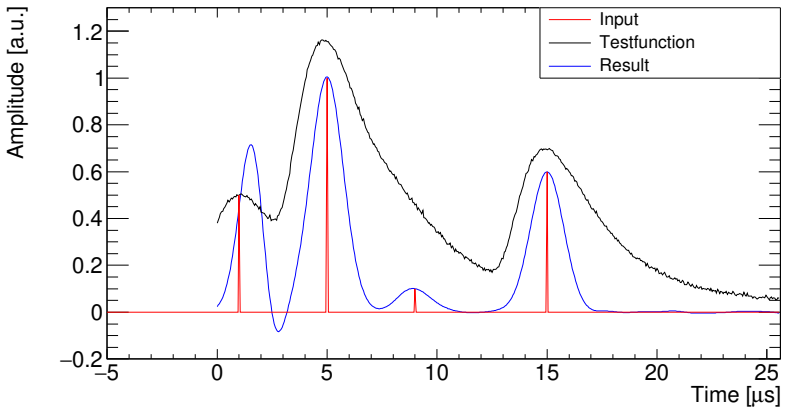


Figure 6.22: Performance of the custom deconvolution in a modeled pile-up situation (same input as Fig. 6.20).

top of the falling edge of a larger pulse and was barely noticeable in the original signal) is nicely recovered. In fact, the heights and positions of all peaks (apart from the first one) are precisely reproduced in the deconvolution. One small modification can be made to the pure algorithm in order to deal with partial pulses, such as the first one, that are cut off at the beginning of the sampling window. Because the algorithm expects the pulses to have a very specific shape, it does not handle those pulses well. A simplistic solution to deal with such cases is to extend these signals into the region before the sampling window, as shown in Fig. A.15. Depending on whether the first few samples are falling or rising, the signal is either extended linearly down to the baseline or extrapolated by attaching the left part (up to the maximum) of the reference pulse shape to the waveform. This method is admittedly rather crude and could easily be improved upon, but it does fulfill the requirement of mitigating the mentioned adverse effects to a certain degree, as evidenced by the plot[47].

Since the resulting pulse shape is an exact gaussian, it is possible to provide a theoretical resolution limit: Two pulses of the same size must be at least 2σ apart in time so that two distinct peaks can still be resolved[48]. Given frequency contributions up to ≈ 1 MHz, it is not possible to go much beyond a gaussian width of $\sigma \approx 1$ µs (see footnote in section 6.6.4.3).

Furthermore, it should be noted that a perfect gaussian is only produced if the shape of the pulse to be deconvolved matches the one used for the calculation. If the pulse is narrower (Fig. 6.23 shows the result for the *much* narrower α pulseshape), an underswing will follow the peak. At the same time, the height and position of the peak will not be reproduced exactly. For a wider pulse[49], the opposite will apply. This is a direct result of the deconvolution being designed for one specific pulse shape and cannot be avoided in this approach. Another drawback is the large number of datapoints required for H. To get rid of the long tail of the pulseshape P completely, about as many datapoints as there are in the signal to be deconvolved are required.

[47] This correction could have been applied before the MWD as well, although it would not have solved its other issues.

[48] This does not mean, that such pulses can not be separated by e.g. fitting.

[49] If any, only *slightly* wider pulses should be expected at all.

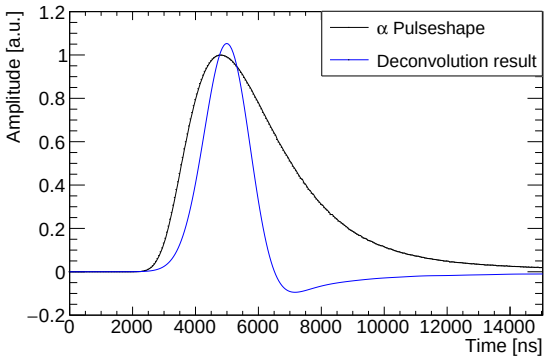


Figure 6.23: Performance of the deconvolution method for the extracted narrower "α pulseshape".

6.6.4.5 Disambiguation and Credits

If the transfer function h as well as the original function g is unknown, the problem is formally known as *blind deconvolution* [SCI75]. It generally relies on several assumptions and is a challenging problem; a typical use-case would be the de-blurring of an image. The process outline above is substantially different, since the target function g is deliberately chosen to be a gaussian. It was initially inspired by the Wikipedia article on the so called *Wiener deconvolution* [Wik22] which follows a somewhat similar approach, but assumes a-priori knowledge of the transfer function h. From there on, the above approach was developed by the author in a process of trial and error. A later literature research revealed that the exact process seems to be known already, as it is described in e.g. [Smi97] and [Lan21]. Unfortunately, both sources neither cite an original paper nor do they assign a name to the procedure. To the best of the authors knowledge, this specific approach seems to lack a descriptive name so far. After some thought, the name *Pulse Shaping Deconvolution (PSD)* seemed appropriate since it encapsulates the core idea that the observed signal can (within some limitations) be shaped to the users desire to give the best possible result for the specific problem at hand. The name PSD will be used henceforth within the scope of this thesis.

6.6.4.6 Outlook and Relation to FIR Filter

It should be noted, that a convolution is algorithmically the same operation as the application of a FIR filter (see section 6.2). The individual values of the deconvolving function H could conceptually be considered to be filter coefficients of a specialized digital filter which achieves the deconvolution!
Furthermore, since convolution is an associative operation, it should be possible to merge the application of the FIR filter and the deconvolution into one operation. By computing the convolution between the FIR filter coefficents and H first and then applying the result to the waveforms, both steps could be achieved at once. Due to the filtering effect of the deconvolution itself, an additional FIR filtering might not even be necessary at all. Either way, this would also lend itself well to an eventual porting of the deconvolution operation onto the FPGA. The exact implementation and the study of its implications could be an interesting research project for the future.

CHAPTER 7

Pile-Up Recovery

7.1 Overview

The challenges in detecting pile-up and its solution were discussed in detail in section 5. Regardless of the detection algorithm, every waveform which was flagged as pile-up will be saved for later, offline analysis. The primary goal of this analysis should be the reconstruction and disentanglement of all contributing pulses. The filtering and deconvolution operations described in section 6 can be seen as "preparatory actions" which serve to ease this task greatly. Ultimately however, the goal is the recovery of all features of the main pulse, i.e. the features that would have been extracted without the pile-up pulse(s). In accordance to section 5, the convention is to treat the pulse whose peaktime is closest to the trigger as the main pulse; regardless of its actual height. This is because in general, only the data of pulses which are related to the trigger are of interest for subsequent analyses.
So far, it was shown how the waveform can be transformed from the rather complicated and long-reaching CB pulseshape to a much shorter and more manageable Gaussian shape via deconvolution. By design, the heights and positions of these Gaussians should be identical to the respective properties of the original pulses. Deviations might occur if the pulses are narrower than anticipated by the deconvolution or if they were cut off in the beginning. An incorrect baseline[1] will also have a deteriorating effect onto the deconvolution. If the deconvolved pulses are far enough apart and do not overlap, the properties of all contributing pulses may be easily extracted. To that end, a new (offline) peakfinder algorithm was implemented.
After all peaks from the deconvolved waveform are identified, it has to be checked whether those peaks, together with the extracted baseline value, actually provide a complete and accurate description of the waveform at hand. This is done by constructing an estimated waveform from the extracted peaks and comparing it to the actual waveform. If they are in good agreement, no further pile-up recovery needs to be done. Otherwise, the waveform needs to be fitted or further peaks might need to be introduced.
Once a satisfactory description is reached, the features of the main pulse may be deduced and will be written to disk in exactly the same format in which the initial, pile-up affected features were stored.

[1] Severe deviations as with the initial algorithm are highly unlikely with the improved algorithm, but small deviations from the correct baseline might still occur.

7.2 A New Peakfinder Algorithm

It would seem natural to employ the same peakfinder algorithm as is used in the FPGAs (see section 3.3.3.4). However, this algorithm is far from optimal. One of the main considerations in its design was to develop an easy and reliable algorithm which works directly on the incoming data stream, i.e. without the ability to "look ahead"[2]. In offline analysis, no such constraints apply. Simply put, the offline algorithm searches for local maxima in the waveform. Two conditions are checked for each point (x, y):

1. Does the height y exceed a given threshold above the baseline? As the waveform is mostly free of noise after the digital filtering stage, a very sensitive threshold of 10/20 (DS/US) ADC channels ($\approx 0.35\,\mathrm{MeV}$) was chosen as default value.

2. Does the height y of this point exceed the height of every other point in a range with radius d around x? The parameter d[3] was consistently set to equal the width of the Gaussian used in the deconvolution of 20 samples or 1 µs. Choosing a smaller value would be nonsensical, as σ is the theoretical resolution limit for two identical Gaussian peaks.

There is no check on the depth of the valley in between two adjacent peaks. This is intentional and ensures that even peaks that are still overlapping after deconvolution can be identified; at least as long as they rise above the surrounding area at all. Overall, the thresholds were chosen to be quite low to ensure that every peak that might be present in the waveform gets found; obviously at the expense of possibly finding too many peaks in some cases.

7.3 Fitting Procedure

7.3.1 Preparation

The definition of the *estimated waveform* $f(t)$ – which will be used for fitting – follows directly from the pulse shape $P(t)$ defined in section 6:

$$f(t) = B + \sum_{i=0}^{N-1} H_i \cdot P\left(t - T_i; S_i\right) \tag{7.1}$$

Where B is the baseline, N the number of peaks/pulses and i enumerates the individual pulses. T_i denotes a pulse's position relative to the trigger, H_i its height and S_i its dimensionless shape parameter. While a good estimate of the baseline is readily available from the feature extraction[4] and the number of peaks as well as the different heights and positions are provided by the peakfinder algorithm, S is initially always assumed to be zero, i.e. a "photon pulseshape".
Crucially, only pulses which are expected to have any influence onto the measured integral (see section 5.4) need to be taken into account. Accordingly, all pulses which have a peak time later than the end of the integration window t_{end}, plus the rise time of a typical pulse (≈ 3 µs), do not

[2] This would have required some sort of additional buffering of data and additional comparison operations.
[3] Also referred to as "isolation", which is the corresponding term used in the topographic classification of mountain peaks.
[4] Potentially (if the data was taken with the old algorithm), after the correction took place.

enter $f(t)$. Similarly, parts of the sampled waveform which lie more than 2 µs after t_{end} may be ignored as well[5].
The agreement between estimated waveform $f(t)$ and actual waveform $y(t)$ (as usual: after FIR-filtering) is evaluated based on the *residual* $r(t)$:

$$r(t) = y(t) - f(t) \tag{7.2}$$

To easily quantify a good or bad description, the total, cumulative deviation R between $w(t)$ and $f(t)$ is calculated (which is called "residual" as well):

$$R = \sum_{t_k=t_{\text{start}}}^{t_{\text{end}}} \left| w(t_k) - f(t_k) \right| = \sum_{t_k=t_{\text{start}}}^{t_{\text{end}}} |r(t)| \tag{7.3}$$

where t_{end} again denotes the end of the integration window, i.e. the sample number at which the 128-fold MA is evaluated. t_{start} is taken to be the first available sample of $w(t)$.
To evaluate the size of a given residual in terms of physical energies, a "calibration" for the residuals may is needed. Because the residual R is effectively the integrated absolute deviation, the existing calibration factors for the integral (see section 4.3.2) are appropriate. Only the different lengths of integration window and the window over which the residual is calculated have to be taken into account: Since the integral feature is technically a moving average, the residual must be scaled accordingly:

$$R_E = R \cdot \frac{g_{\text{calib}}}{t_{\text{end}} - t_{\text{start}}} \tag{7.4}$$

Convention During this thesis, "time" has been expressed in a variety of ways. For one, the time associated with a specific *sample* of the waveform might be expressed in the number of samples or nano-/microseconds; either after the beginning of the sampling window or in relation to the trigger. Obviously, choosing the latter options is ultimately the more sensible choice, as it does not depend on the samplingrate or the position of the sampling window. Computationally however, the waveforms are fundamentally just a collection of samples (arrays). Knowing the sampling rate and the sample number at which the trigger was detected, both representations can be easily converted into each other.
Additionally, the time of a specific *pulse* might be characterized by different properties like the maximum, the onset of the pulse, the position of its detected peak, the CFD zero-crossing, and so on.
For this chapter, it was decided to assign a time of $T = 0$ ns to *pulses* which are perfectly correlated to the trigger; i.e. at the most commonly encountered (prompt-peak) position.

This convention also implies that the *calibrated* timestamps (peak- and cfdtime) which would have been extracted from such pulses are, by definition, also precisely 0 ns[6]. Any offset between

[5] In principle, everything directly after the end of the integration window could be ignored. Extending this range by 2 µs (the exact value is not hugely important) serves to better constrain the latest possibly contributing pulses.

[6] Provided that the pulses had a normal γ-pulseshape. In case of a narrower pulse at $T = 0$, the extracted timestamps will be found slightly earlier because the pulses rise time is shorter; see e.g. Fig. 6.17.

actual peak position and the trigger signal[7] is implicitly accounted for in the calibration already. The goal of the recovery process is to deduce the features of the single pulses how they would have been, had the pile-up not occurred. Under the above convention, a pulse's time T has to be converted back to the (uncalibrated) feature values. To that end, the calibration might be inverted and the uncalibrated feature values can be easily calculated. Narrower pulses are accounted for by a shape-dependent additional offset.

7.3.2 Fitting Procedure

Fitting is done by using the `Minuit2` minimizer implemented in the ROOT framework. In the default χ^2-minimization, the goal is to find values for the parameters such that the sum of quadratic residuals of every data point reaches a global minimum. Once this minimum is reached[8], the fit terminates. The entire recovery process (which consists of more than just the fitting) can be broken down into several steps or "iterations", which will be detailed below.
After each of these steps, the previously defined residual R_E is evaluated. If its value is below a carefully chosen, predefined threshold, the description is considered satisfactory and the process terminates. If not, the description has to be improved (see below). The threshold value is parameterized similar to previous thresholds as some absolute part and an energy-dependent part:

$$R_E < 0.15\,\text{MeV} + 1\% \cdot E_{\text{obs}} \tag{7.5}$$

On the Choice of Threshold It should be noted, that the choice of a suitable threshold is not trivial. Fig. 7.1 shows the resulting residuals after step 2 of the fitting procedure (will be explained soon). A clear band of small residuals, presumably indicating a "perfect" description of the waveform at hand, stretching from (0,0) to approximately (150,0.2) and beyond can be seen. At the same time, the apparent systematic resolution limit of the fit seems to lie directly below this band, which gives it a sharp limit in that direction. No such clear limit exists in the other direction.
The applied threshold (magenta line) lies quite far above this band, which is reassuring in the sense that the threshold seems high enough to accept all "complete" descriptions. On the other hand, it may be questioned if the threshold could have been chosen smaller. However, since the overall energy resolution of the calorimeter is well above 2% even at $E = 1\,\text{GeV}$ (see section 2.8), a relative deviation below 1% is already sufficient for all practical purposes. Choosing the threshold as described gives the fit enough leeway to account for e.g. noisier channels, slightly differing pulseshapes, or other confounding factors which do not impact the data that much, but which could easily put the resulting residual over a stricter threshold. A thorough investigation of the effect of different thresholds on the reconstructed data would be an interesting topic for future studies, but was not done during this thesis.

All examples and histograms in this section were taken from the June 2021 beamtime.

[7] Trigger is sufficiently precisely sampled with 160 MHz (see section 4.4.6). An offset is caused by the pulse's rise time, the delay of the 16-fold MA, and other factors.
[8] Technically: Once the point in parameter-space is close enough to the estimated minimum position.

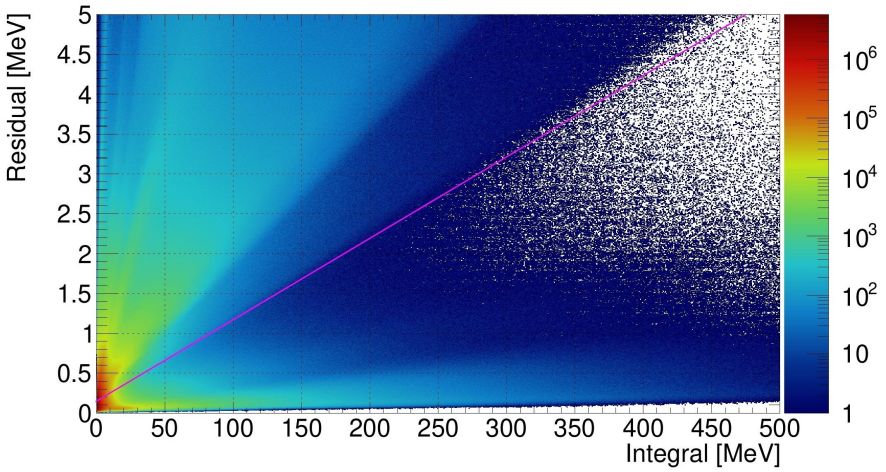


Figure 7.1: Residuals after the steps 0,1 and 2 of the fitting process (see text). The acceptance threshold is depicted in magenta. Steps were only executed until the residual was below threshold, which explains the (weakly pronounced) edge in the distribution which runs along the threshold.

Step 0 The very first iteration is essentially a "dummy" step. Before any fitting is done, it is checked whether the fit function $f(t)$ is a good enough description of the waveform already, just by setting the start values extracted from the deconvolved waveform. In the data from the June 2021 beamtime, this is the case for around 44% of all detected pile-up signals (see Fig. 7.6). This high percentage is indicative of a great performance of the deconvolution process: If the deconvolution was able to cleanly separate overlapping peaks and no other confounding factors like differing pulseshapes play a role, no fitting should be needed at all. Fig. 7.2 shows one example of such a case:
Three peaks were found in the deconvolved signal; after their respective values and the extracted baseline were inserted into the fitfunction (red), the description is already sufficiently good, as evidenced by the residual[9] (lower plot). The pulse *closest* to the trigger (dashed red line) is taken as the "main" pulse from which the corrected features are deduced. In this specific case, all three pulses (including the main pulse at $T = -1\,800\,\text{ns}$) seem to be uncorrelated to the trigger.
In terms of execution speed, it is very desirable if this first step is already sufficient, as this requires virtually no computation time in comparison to complex fitting operations. Especially at this stage, the exact choice of threshold parameters has a strong impact on how many waveforms can be dealt with in this way.

Step 1 Should setting the start values not suffice, the fitting procedure is executed for a first time, but the shape-parameters of all pulses are left fixed to 0 (photon pulseshape) for now. This accelerates the process and should still cover most cases, since narrower pulses are expected to be relatively rare. The other parameters (pulse heights, positions and the baseline) are left unconstrained; the only constraints applied are dictated by what is physically meaningful: negative baselines and negative pulse heights as well as pulses higher than 2^{16} samples were explicitly

[9] As the residual R is defined, $r(t)$ might occasionally exceed the dashed threshold lines, without the total residual R being above threshold.

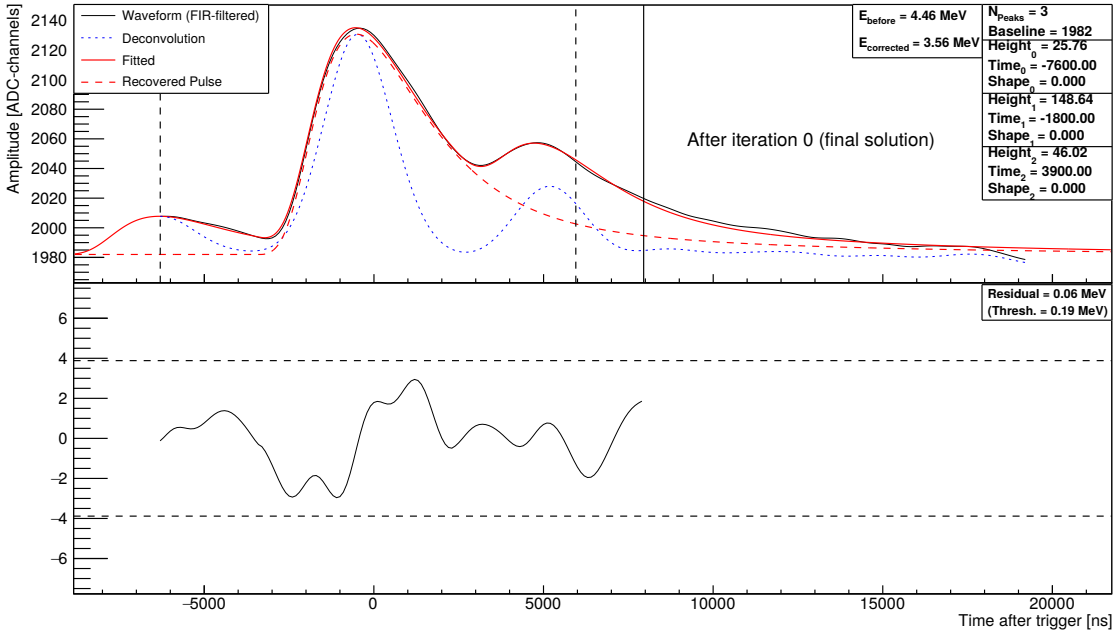


Figure 7.2: Graphical representation of a successful pile-up recovery after 0 iterations. The upper diagram shows the FIR-filtered waveform in black, the deconvolved waveform in blue and the fitfunction in red (solid line). The reconstructed "main pulse" (smallest $|T|$) is shown as dashed red curve. The dashed vertical black lines denote the relevant range $\left[t_{\text{start}}, t_{\text{end}}\right]$ from which the residual *value* R is calculated. The residual *function* $r(t)$ (see lower plot) is calculated in the entire fitrange (although no fitting was neccessary in this case) which is 2 µs longer and whose end is denoted by the solid black line. Samples after this line are completely irrelevant for the recovery process and are shown for illustrative purposes only. The horizontal dashed lines in the lower plot represent the residual threshold, divided by the length over which it is calculated.

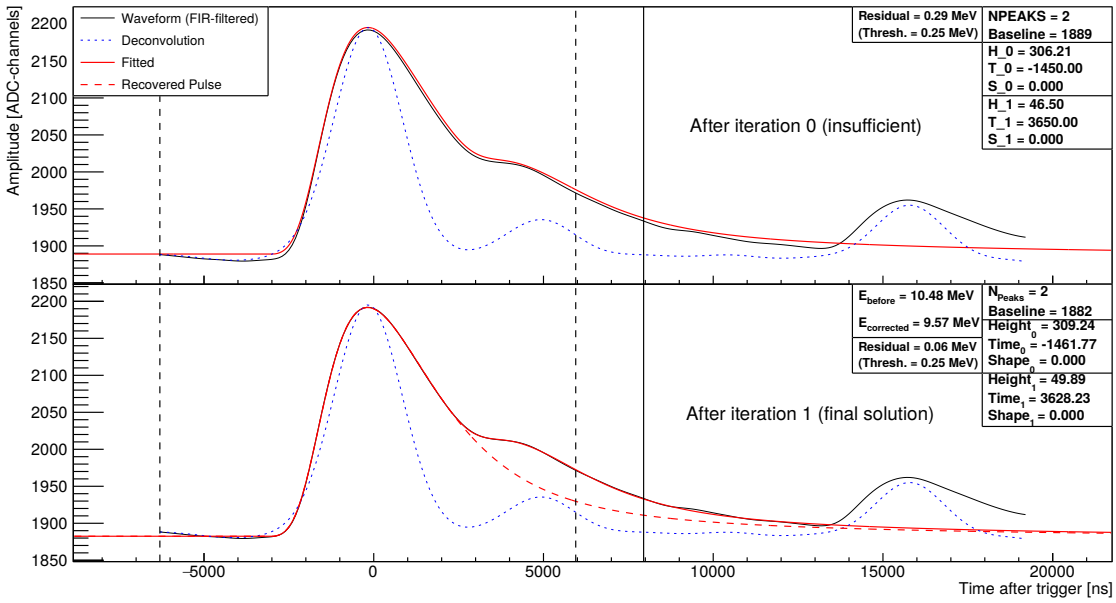


Figure 7.3: Graphical representation of a successful pile-up recovery after 1 iteration.

disallowed.
Fig. 7.3 provides an example of such a case: Visually, the description after iteration 0 is already quite good. But because the residual was still barely too high, fitting was required. In particular, the baseline value seems to have been a bit too high (1889 initially vs. 1882 ADC-channels after the fit). After the fit was performed, almost no discernible discrepancy between data and fit remains.
Cases like this that are acceptable after one iteration make up another 29% (June 2021) of pile-up-flagged waveforms, which again emphasizes the good performance of the deconvolution[10].

Technical Remark: Insignificant Pulses Every time a fit operation is performed in the recovery process, it is checked whether any of the pulses might have become insignificant to the fit or otherwise problematic. Such problems are rather technical in nature, but should be mentioned for the sake of completeness. The following conditions are checked:

(a) The pulse height assumes a value below a predefined threshold (default: 5/10 (DS/US) ADC channels or $\approx 0.15\,\mathrm{MeV}$). In that case, the respective pulse is removed from the fit.

(b) A pulse gets shifted to a position at which it no longer contributes significantly to the residual (i.e. too far right or left of the window). If the largest amplitude it contributes within the fit range is below 5/10 (DS/US) ADC channels (same threshold as above), it is removed.

(c) The fitted position of one pulse gets too close to another pulse, below a separation that should realistically be resolvable. This happens most often if the peakfinder has found more peaks than justified in the signal. In that case, the fitting algorithm tends to shift those incorrect peaks into the nearest actual peak. The default limit is chosen consistently to the required separation "d" of the peakfinding algorithm (which itself was chosen to equal the width of the Gaussian used in the deconvolution) of 1 µs. Should this condition be fulfilled, the smaller of both peaks is removed while the height of the other one is set to the sum of both heights. By construction, this results in a hard limit for the resolution of the recovery process: Pulses closer together than whatever limit was chosen will not be resolved[11].

After the respective pulses have been removed/merged, the fit is repeated and the outlined conditions are checked again. The process is repeated as needed until no more insignificant pulses are detected in the fit result.

Step 2 If fitting with the standard photon-pulseshape yielded no satisfactory result, the constraint on the shape parameter is lifted. The shape of the main pulse[12] is now allowed to vary in between 0 (photon pulseshape) and 2.5 (very narrow). The already very narrow alpha pulseshape would correspond to $S = 1$, but it is unproblematic to allow a (limited) extrapolation beyond this value for even narrower pulses. It was refrained from allowing for wider pulses because this would

[10] As in this example, the fit parameters are typically only slightly tweaked in step 1. With a slightly more lenient threshold, most of these could have been considered acceptable after step 0 as well.

[11] For the investigation which led to Fig. 5.18, this limit was artificially reduced to 500 ns, albeit seemingly to no great avail.

[12] If there are more than one pulses, this helps to reduce complexity.

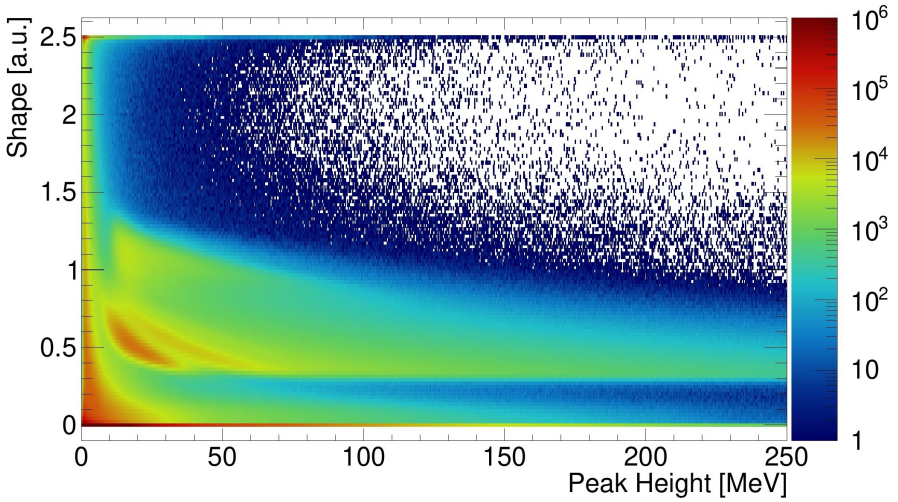


Figure 7.4: Histogram of shape parameters S (parametrization discussed in text) as determined in the recovery process from waveforms marked as pile-up. Each reconstructed pulse enters this histogram. June 2021 beamtime.

– while being physically unjustified anyways – involve the risk of accidentally describing two pulses which are very close together as one, wider pulse. Another 7% of all cases are successfully recovered in this way. However, over 90% of those cases were not true pile-ups to begin with, as they contained only one (narrow) pulse. This is illustrated by Fig. 7.6: the third row from the bottom (iterations = 2) has a pronounced maximum at $N_{\text{Peaks}} = 1$. Fig. A.17 (in the appendix) shows a typical example of a waveform that is narrower than normal and could therefore not be fitted in step 1.
As before, the given percentages apply to the June 2021 beamtime. With the new pile-up detection algorithm (v2.0), such cases will not be marked as pile-up in the first place. Accordingly, the percentage of cases requiring step 2 is expected to decline significantly.

Discussion The distribution of occurring shape parameters, depicted in Fig. 7.4, exhibits a striking similarity to the distributions examined in section 6.5. Again, several different regions are discernible: The vast majority of entries are exactly at $S = 0$. As $S = 0$ was fixed during the steps 0 and 1, there exists a heavy bias towards this value. Only if this did not result in a satisfactory description, other values were allowed. For this reason, relatively few entries fall into the direct vicinity of $S = 0$. Furthermore, 3-4 other regions exist, which all have a pronounced maximum around ≈ 20 MeV (also see Fig. 7.8) and tend towards smaller S-values for higher pulse energies. As $S = 1$ explicitly corresponds to the extracted mean pulseshape of α particles, it is only logical to expect a cluster of entries from the very same pulses around this value. In accordance to the discussion in the previous chapter, the most populated secondary branch around $S = 0.5$ would correspond to (stopped) protons, the one directly above it and at slightly higher energies to deuterons and the faint one even further above and to the right to tritons and/or ^{3}He-nuclei.

Step 3 In case the residual is still too high, even though all parameters are now varying freely, the fit function has to be changed fundamentally. Presumably, no satisfactory description could be reached because not all necessary pulses were included, so more pulses need to be added. This is done by analyzing the residual: Since it is defined as the difference between the datapoints $y(t)$ and the fitfunction $f(t)$, a pronounced peak in the residual must correspond to a region in the waveform where the fit is underestimating the data.
Fig. 7.5 shows an example of this approach in action: Despite the deconvolution, only the second, larger peak was found. After steps 0–2, the residual still exhibits a clear maximum at 900 ns after the trigger[13]; i.e. approximately at the position of the first (missed) pulse. The maximum of the residual at this position is 31.5 ADC channels. However, using this value as predicted pulse height is not ideal. Because the fit aims to minimize χ^2, it will try to compensate for missing peaks (underestimation of the data in this area) by shifting the baseline up to some degree. As a result, the actual height of the missing peak is empirically better estimated by the difference between the maximum and minimum of the residual, which gives a provisional height of 44.17 ADC channels in this case. The baseline is shifted *upwards* accordingly, by the minimum value of the residual. As such, the description is already looking promising before another fit operation has taken place (see third diagram). But even if this description would yield a sufficiently small residual, another fit operation similar to step 2 (i.e. the previous pulse is still allowed to vary in its shape) is carried out in any case. In doing so, a very good description is reached (bottom panel). Overall, a further 19% of all cases are successfully recovered after step 3.

Further Steps So far, over 99% of all pile-up-flagged waveforms could be recovered within steps 0 to 3. Should that not have sufficed, the routine of step 3 is repeated, i.e. the residual is analyzed for further peaks until a good description is reached at some point. It may occur that a pulse is inserted but yields no substantial improvement, presumably because it was at the wrong position. As insignificant pulses are continually checked for and removed after every iteration, a subsequent analysis of the residual might find the same maximum again. In that case, a region of radius 2 µs[14] around that point is excluded from the search and the maximum of the remaining range of the residual is used instead. If a subsequent step results in further repetitions of the same maximum position of the residual, the process is aborted, as no more progress is expected to be made in this way.
For practical reasons, the total number of iterations is limited; the default value for this is 7. Likewise, the possible number of pulses was also limited; in this case to a maximum of 5.

Challenging Cases In very rare cases, this approach might fail to find the best description of the waveform at hand. Should the final result contain the maximum number of peaks or yield a baseline which differs substantially (> 2 MeV) from the previously extracted value, this is taken as an indication that something might have gone wrong. In that case, every pulse but the largest one is removed and the entire process is rolled back to step 1. In this second round, the baseline is kept fixed at the predetermined value. The final residual from this second try is then compared

[13] A pulse which has its peak at 900 ns after the trigger is −400 ns too early w.r.t. the integration window / prompt peak. Therefore, $T_1 = -400$ ns is the proper time for this pulse.
[14] Heuristically, using double of the required separation "d" of the peakfinding algorithm has proven to be a suitable value.

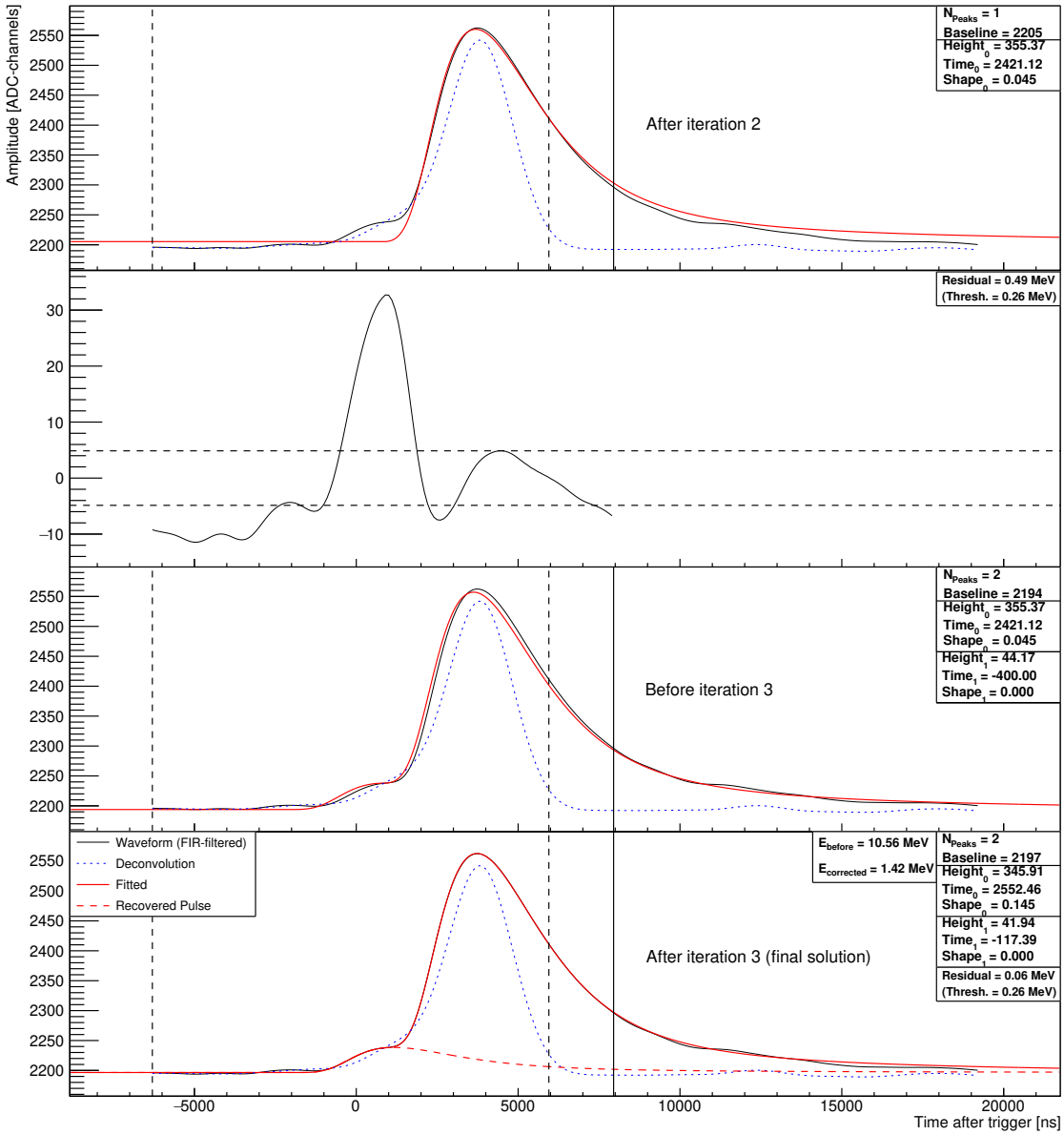


Figure 7.5: Waveform which was not satisfactorily described after iteration 2 (top) and required the addition of another pulse to be well described. To that end, the residual (2$^{\text{nd}}$ row) was analyzed and a second pulse inserted (3$^{\text{rd}}$ row; see explanations in text). After fitting again, the data is very well described (bottom). The final residual $r(t)$ was not shown in order to save space, but $R = 0.06\,\text{MeV}$ was well below the respective threshold of $0.26\,\text{MeV}$.

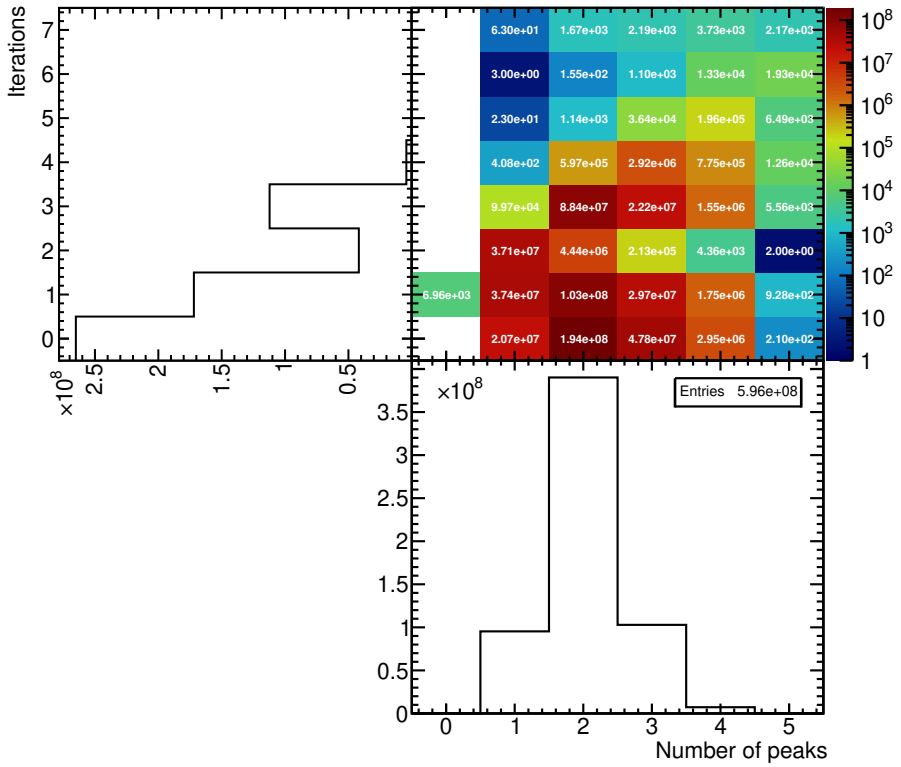


Figure 7.6: Number of iterations against number of reconstructed peaks (in the fit-range) of pile-up-flagged waveforms. Only the last step/iteration which yielded a satisfactory description enters the histograms. The top left and bottom right histograms are projections onto the respective axes. As expected, most detected pile-ups consist of two pulses, while 3, 4 or even 5 pulses are increasingly more unlikely (see also Fig. 7.8). Most cases are "solved" in iteration 0 and 1, and a few more in iteration 2 (after shapes $S > 0$ are allowed). Another large amount of cases required the addition of one more pulse than anticipated from the deconvolution (iteration 3), while only very few cases require 4 or more iterations. From iteration 3 upwards, a positive correlation between iteration and number of peaks is visible, which is to be expected because with every iteration ≥ 3, a new peak may be added.

to the one obtained in the first try and whatever yielded the smaller value is finally taken as the best (albeit potentially above threshold) solution.
Fig. A.18 (see appendix) gives an example of such a case: Although the residual after step 1 is already (barely) below the threshold, the fitted baseline of 2281 ADC channels lies substantially above the extracted value of 2229 ADC channels. For that reason, the result is considered suspicious[15]. The second round in which the initial baseline value was kept fixed finally results in a much better description. Conincidentally, this involves the maximum number of 5 pulses. 4 of those are visually easily recognizable as small bumps in the deconvolution result. The fit indicates the existence of a fifth pulse further in the past[16]. Two more pulses can even be seen later in the waveform; these are, however, irrelevant to the description in the range of interest.

[15] To the human eye, a brief look at $r(t)$ already confirms that the description is sub-optimal.

[16] The exact parameters of $T \approx -26\,\mu s$ and $H \approx 9000$ ADC channels are poorly constrained by the fit and should be approached with caution.

Failed Recovery In particularly challenging cases, the recovery process might terminate without a sufficiently small residual (about 1 in 10000 cases). If the final residual is at least close to the threshold (up to twice as much), the description is still deemed acceptable. Only if even this is not the case, the recovery is finally considered failed and no pile-up correction is done at all[17] (about 1 in 100000 cases).

7.3.3 Actual Feature Correction

Once a satisfactory description of the waveform is reached, the features of the main pulse, i.e. the one most closely related to the trigger, have to be determined. For data taken before 2022, a particular complication arises from the fact that the recovery process is based on the FIR-filtered waveforms, whereas the feature extraction was based on the raw waveform and the MA_{16}[18]. So it is not possible to calculate e.g. the integral of one particular pulse directly from the fitted function. The result would be close, but not exactly the same as if the raw pulse was integrated. To avoid these complications entirely, the expected integral is calculated from the pulse's height and position by again using the lookup table which was presented in section 5.2. To account for different pulse shapes, two such lookup tables were used, one for $S = 0$ and one for $S = 1$. Together with those lookup tables, the associated peak- and cfdtimes for these kind of pulses were stored as well[19]. The features for intermediate shapes and pulses with $S > 1$ are deduced by inter- and extrapolation, similar to the pulseshape itself (see eq. 6.16).
In order to get the raw, uncalibrated values as they would have appeared in the feature packets sent by the SADCs, the calibration of the respective features is inverted (as already mentioned above). As an example, the timestamps are initially given in units of 1.5625 ns[20] after the trigger. In the calibration, they are converted to ns and shifted such that a value of 0 ns corresponds to a pulse which is perfectly correlated to the trigger. Due to the way in which data objects holding the feature data are processed internally, the parameters used for this calibration (in particular the factor and the offset) are easily accessible and can be used to "un-calibrate" the values accordingly (see eq. 4.6):

$$v_{\text{raw}} = \left(v_{\text{calib}} - \text{offset}\right) / \text{gain} \tag{7.6}$$

After all "raw" features are determined, the data is written to disk in a format (ROOT-Tree) which is entirely compatible with ordinary, uncorrected data. This ensures, that it is possible to switch between either the uncorrected or the corrected data with minimal effort.

[17] In that case, the features are left uncorrected, i.e. as they were extracted during data taking.
[18] For data taken since, all features (including the integral / the MA_{128}) are based directly on the FIR-filtered waveform.
[19] By design, the (calibrated) timestamps of a normal ($S = 0$) pulse at $T = 0$ ns are precisely 0 ns as well. For $S = 1$, the peak occurs ≈ 300 ns earlier (see Fig. 6.17) and the cfd zero-crossing 100 ns earlier. The discrepancy between both timestamps is a result of the CFD parameters being tuned for the normal pulse shape; for narrower pulses, they lead to a zero-crossing which no longer coincides with the maximum.
[20] 12.5 ns before the cfd-interpolation was introduced.

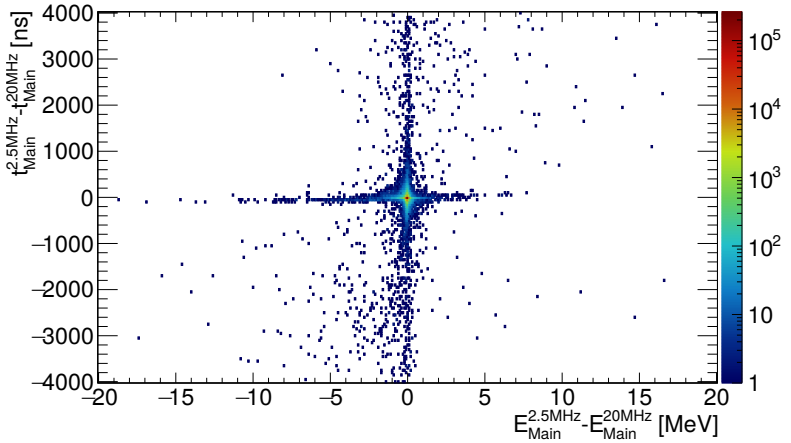


Figure 7.7: Effect on results (energy and time of the main pulse) of the pile-up recovery if only every 8th sample is used for the fitting process. The central bin (0.1 MeV wide and 50 ns tall) contains over 82% of all entries, while 99% fall within ±1.0 MeV and ±500 ns from (0,0).

7.3.4 Performance and Possible Improvements

As elaborated previously, the relevant frequency contributions for typical pulseshapes lie almost entirely below 1 MHz. This is all the more true after the FIR-filter is applied. For this reason, only every 8th sample of the FIR-filtered waveform was actually used in the fitting process, effectively reducing the sampling rate to 2.5 MHz[21]. Investigations have shown, that this has a practically negligible effect onto the results of the recovery (see Fig. 7.7). Despite some outliers, the results from the "accelerated" are effectively identical almost all cases. While it is theoretically certain that no significant information is lost, there are always edge-cases in which the fit might prefer a certain solution over another depending on the specific set of datapoints. In isolated cases, this can lead to single, small peaks being "overlooked" in either the full or (slightly more likely[22] the reduced sample, which causes the outliers. A commented example of such a case is shown in Fig. A.19 in the appendix.

With the hardware specifications of typical nodes of the CB computation cluster, the entire recovery process (including data preprocessing, the recovery itself, reading/writing to file and other overhead) runs at a speed of 100 to 300 events per second. However, most of the time (≈ 90%) is in fact expended for the recovery process itself, while the remaining 10% are to a large degree spent for the calculation of the deconvolution. By including an additional 4- or 8-fold decimation of the waveforms before they are processed any further, it should be possible to speed up the deconvolution significantly[23]. For the fitting however, this would change nothing, as the "downsampling" prior to the fitting process would have to be reduced accordingly, such that the number of samples used for fitting stays constant.

In the most recent beamtimes, data was recorded at a rate of 2 to 3 kHz. This entails that in order to achieve a pile-up correction directly while the data is taken, a total of ≈ 10 computation nodes would have to be dedicated to the recovery process. Whether such a "real-time" recovery would

[21] The number of samples was reduced *after* the FIR filter was applied, which helps to improve the SNR of the individual samples (because the ultimately ignored samples also contributed to the remaining ones).

[22] The left "arm" in Fig. 7.7 is more pronounced than the right one.

[23] To retain the precision of start values for the recovery process, this would require some sort of interpolation in the subsequent peakfinding process.

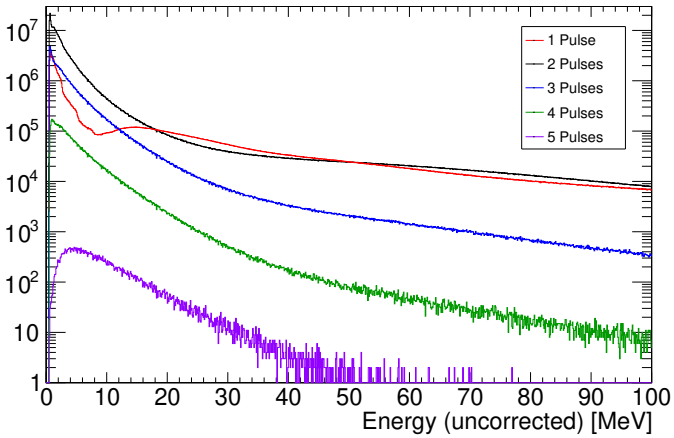


Figure 7.8: Number of reconstructed pulses against total, uncorrected energy (integral) in waveforms flagged as pile-up. Apart from the 1-pulse-waveforms, the *relative* probability of each topology is roughly independent of the energy. The structure in the first around $E = 20\,\mathrm{MeV}$ is due to narrower pulses being erroneously flagged as pile-up and is entirely consistent with the exhaustive discussion of the topic so far (see section 6.5 as well as "step 2").

be necessary or even desirable remains to be discussed[24].
Furthermore, the rate of detected pile-up is projected to roughly double due to the improvements in the pile-up detection algorithm if everything else is kept unchanged. As most of the detected pile-up is at very low energies, it should be discussed if wider thresholds for the detection of pile-up should be chosen.
At the same, it might be worthwhile to investigate whether the acceptance threshold for the residual could be increased as well, such that even more cases could be recovered without the need for fitting at all. Should the difference on data quality be insignificant, the process could be sped up by a factor of up to ten[25].

[24] In principle, there is no downside to the recovery being performed after the data has been taken.

[25] Even the most simple fitting operations in step 1 usually take around 100 µs, whereas step 0 takes typically less than 10 µs.

7.4 Results

The latest version of the recovery process runs very reliably and produces a good description of the waveform in almost all cases (see numbers quoted in the previous section). Further characterizations of the recovery's results (e.g. figures 7.4, 7.6 and 7.8) are entirely consistent with the expectations. The latter plot further illustrates that choosing a maximum value of 5 pulses seems to have been a suitable choice. Roughly extrapolating the numbers for 2, 3, 4 and 5 pulses (1 pulse is a special case which was already discussed in detail) would have resulted in only a handful of waveforms containing 6 pulses, while the data base is one tenth of an entire beamtime (June 2021). Either way, it might be questioned how well the recovery would work for waveforms with 6 or even more pulses anyways.

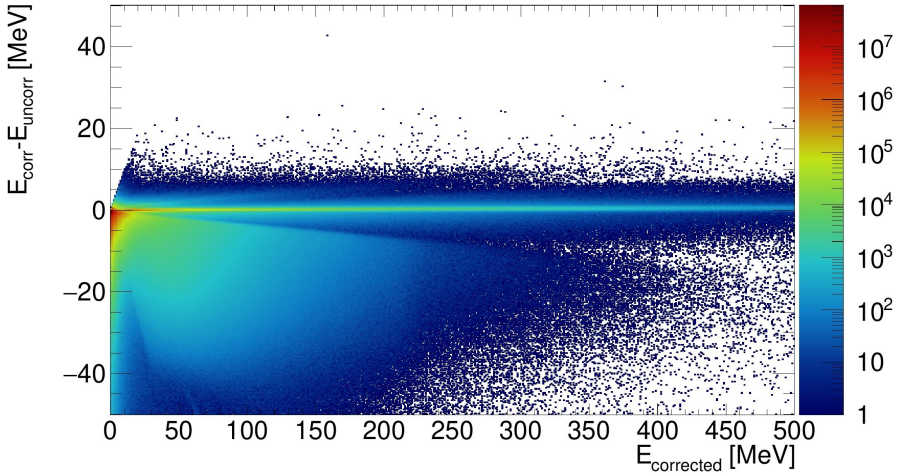


Figure 7.9: Amount by which the extracted energy from every treated waveform was corrected in the recovery process. Data from June 2021.

Stepping away from the recovery process itself, the most relevant question to be asked is how this correction ultimately impacts the data. Fig. 7.9 shows the amount by which the energy of single hits was corrected.
First of all, a pronounced band running along $y = 0$ is apparent. This is due to the previously discussed amount of false positives (mostly narrower pulses) in the data. Accordingly, this band exhibits a pronounced local maximum around 15 MeV, as expected (also see Fig. 7.8). It is logical that the recovery process did not change much about the extracted features (including the integral) in such cases.
Furthermore, it is apparent that the vast majority of entries lie below $y = 0$, i.e. cases in which the energy was corrected down to a smaller value. This is hardly surprising, since pile-up situations should only extra energy to the (correct) energy of the main pulse which is then removed in the recovery process. The upper edge of these entries running from approximately (0,0) towards (500,-15) is an artifact from the thresholds in the pile-up detection process[26]. The other apparent structure is the (relative) lack of entries in the lower left corner (below the edge running from ≈ (20,-20) to (40,-50)). Considering that the y-axis effectively shows $-E_{\mathrm{PU}}$ (in case of two-pulse

[26] $\theta_{\mathrm{pu}} = 0.02$ corresponds to ≈ 3% allowed deviation, see section 5.2.2.

pile-ups), this area correlates well with the "strip" of false negatives in the lower left diagram of Fig. 5.17, which runs (approximately) along $E_{\mathrm{PU}} = E_{\mathrm{Main}}$. Apart from that, the area lower area of the diagram seems rather featureless, as one would have expected.

Special cases Naively, no entries would be expected above the $y = 0$. All the more surprising are the few entries which lie precisely there, seemingly indicating that energy was *added* in the recovery process. There are three possible explanations for such entries: On the one hand, the baseline could have been too large before the recovery took place. Since the *extracted* baseline was subtracted for the uncorrected energy and the *fitted* baseline for the corrected energy, this would explain cases in which E_{corr} is somewhat larger than E_{uncorr}. On the other hand, the fitted baseline could have been too low and the extracted (approximately) correct. Ideally, neither should happen. The new algorithm does a very good job at minimizing the first problem, but some rare deviations are always to be expected. Likewise, the fit can be off in isolated instances: There seems to be no clear characteristic of waveforms in which the recovery leads to an increase in energy, other than that they seem to be convoluted edge-cases. Especially in cases with a lot of actual pulses, only small parts (if any) of the waveform are actually at baseline level, and the baseline is therefore not that well constrained. Fig. A.20 in the appendix shows another example example of such a case, where the fit got stuck in a local (but not the global) minimum.
Another possibility for entries above $y = 0$ are (very large) pulses which are (with regards to the residual) well enough described by $S = 0$ after iteration 0 or 1, but might still have some discrepancy to their true pulse shape. As their "true" pulseshape would most likely be narrower, the fit will thus overestimate their energies a bit. By design, the deviation has an upper limited in the residual threshold of $0.15\,\mathrm{MeV} + 1\% \cdot E_{\mathrm{uncorr}}$.
In any case, such exceptions are sufficiently rare and limited in their overall impact (note that the color scale is logarithmic!) that they are not expected to pose any significant problem in practice.

7.4.1 Impact onto the Reconstruction of Physical Events

Ultimately, the purpose of the pile-up correction is to reduce errors in the reconstruction of physical events in the experiment. As this thesis took part before the time clustering was introduced (see upcoming thesis of N. Stausberg [Sta24]), only the impact of the correction of the integral as the main energy feature was studied. In particular, changes to the measured energy will directly impact the reconstructed mass of neutral mesons. The π^0 at an invariant mass of $\approx 135\,\mathrm{MeV}$ is the most abundant meson detected in the experiment and is used as a benchmark[27]. As before, all data in this section is taken from the June 2021 beamtime.

7.4.1.1 Event Selection

As introduced in section 2.10, several so-called "cuts" are applied in order to obtain the cleanest possible sample of π^0 particles. More precisely, reactions involving a recoil proton and two photons shall be pre-selected by the following cuts:

[27] One example would be the energy calibration (see section 4.3.2) which works by optimizing calibration factors in order to get the MPV of the distribution of reconstructed pions to coincide with the physical value.

- Only events which consisted of precisely 3 clusters were considered. The charge information of the individual clusters was disregarded.
- The time between the "reaction time" and a tagged beam photon was required to be within $-15\,\text{ns} < \Delta t < 10\,\text{ns}$. Without a fitting beam photon in addition to the initial target proton (which is assumed to be at rest), the initial state would not be known and a meaningful selection of events (see following cuts) could not be done. For the sake of simplicity, a sideband-subtraction was not performed[28].

In addition, kinematic cuts on the coplanarity, the polar angle difference, and the missing mass were applied (see section 2.10). A subtraction of As the width of the corresponding distributions tends to vary for different kinematic configurations, specific cut limits were determined for individual bins. To parameterize the available phase space, the beam energy E_{beam} and the cosine of the polar angle of the reconstructed meson candidates $\cos\left(\theta_{\text{meson}}\right)$ in the Center-of-Mass System (CMS) were chosen. The entire phase space was subdivided into a total of 400 bins (20 in each dimension), in a way which results in (approximately) equal numbers of entries in each bin. The cut limits were then determined in an iterative process for each bin, such that they correspond to $\mu \pm 2\sigma$, based on a Gaussian fit of the distribution at hand. As an exception, the cuts on the polar angle difference are 3σ wide[29]. Fig. 7.10 shows a visual representation of these cuts, while Fig. A.22 provides the respective cut limits. Taken together, these cuts lead to the selection of a very clean, almost background-free collection of reactions. The resulting masses of the decaying mesons are shown in Fig. 7.11, which will be discussed shortly.
All cuts were optimized based on uncorrected data. To keep the comparison between uncorrected and corrected data as clear as possible, the deduced cut-limits were then applied to the corrected data as well. In first order approximation, the distributions of coplanarity and back-to-back should be unaffected anyways, since only the energies are acted upon by the correction. In second order, an increase in energy of one or more particular crystals will of course affect the reconstructed angle of a specific cluster at hand; which in turn could lead to an event passing or failing the cut. As it depends on both the energies and the reconstructed angles, the missing mass will be affected as well.
All these effects were deliberately disregarded in this analysis. A preliminary inspection revealed no significant differences in the relevant spectra; the largest discrepancy was observed in case of the missing mass, see Fig. A.21 in the appendix. However, determining other cut limits for the corrected data set would have involved the risk of interfering with the effects of the pile-up correction itself. Should there have been any relevant systematic shift, the used cut limits should be to the disadvantage of the corrected data, as they were not optimized for this particular data set.

7.4.1.2 Effect of Pile-Up Correction

As stated previously, a crystal hit by a pile-up pulse will suffer from an erroneous increase of the measured energy. Accordingly, the immediate effect of a pile-up would be an increase in the mass of the reconstructed meson. Ideally, the pile-up correction should be able to reverse this.

[28] The purpose of this section is to study the *relative* impact of the pile-up correction, for which no correlation to random time coincidences is to be expected.

[29] This was done to account for a not perfectly Gaussian distribution in some bins, which is mostly due to discretization artifacts (the resolution of the direction of the detected proton is typically limited by the finite crystal size).

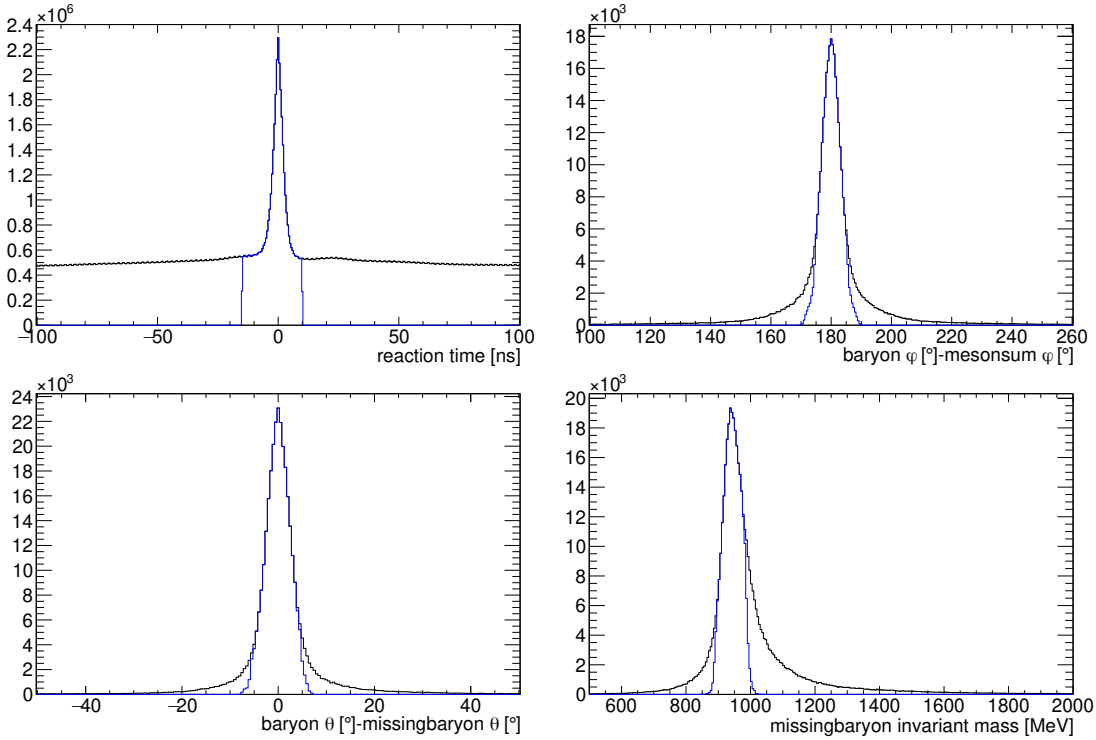


Figure 7.10: Visual representation of cuts which were applied in the event selection process. Except for the reaction time (upper left), all panels show the distribution after all other cuts were already applied in black and the result after the current cut in blue. Since specific cut limits were applied for all 400 $\left(E_{\text{beam}}, \cos\left(\theta^{\text{CMS}}_{\text{meson}}\right)\right)$ bins (see Fig. A.22), the cuts do not appear as straight drop-offs in these diagrams. In case of the reaction time, the cut limits were chosen globally (i.e. identical for all 400 bins) and no other cuts were applied beforehand for the plot.

Fig. 7.11 depicts the resulting spectrum of reconstructed meson masses, both for the uncorrected data (black) and for the corrected data (red). A very clear peak is formed around 135 MeV, the physical mass of the π^0. The second largest belongs to the η meson, with a mass of 548 MeV. Further peaks at higher energies belong to the ω[30] and the η' meson (m = 958 MeV), but are irrelevant for this analysis.
As both spectra seem to be very similar at first sight, it makes sense to subtract the uncorrected from the corrected spectrum and study the residual. This is shown in Fig. 7.12 for the regions around the π^0 and the η. While more entries fall below the nominal mass in the corrected spectrum than in the uncorrected one, the opposite is true for masses above this value. Since energy is taken away by the correction, this was to be expected. Both the π^0 peak and the η peak are slightly asymmetric (see enhanced views in Fig. 7.11) with a tail extending towards smaller masses. This tail is largely a result of particles losing energy in insensitive material. On the flip

[30] Due to its C-Parity, the ω meson cannot decay into two photons. A small peak at approximately the same mass (783 MeV) appears regardless, because two of the three decay photons from $\omega \to \pi^0\gamma$ might overlap and form a single cluster. Another possibility could be a topology in which one of the three photons has a very low energy and does not get detected, leaving the overall invariant mass almost unchanged.

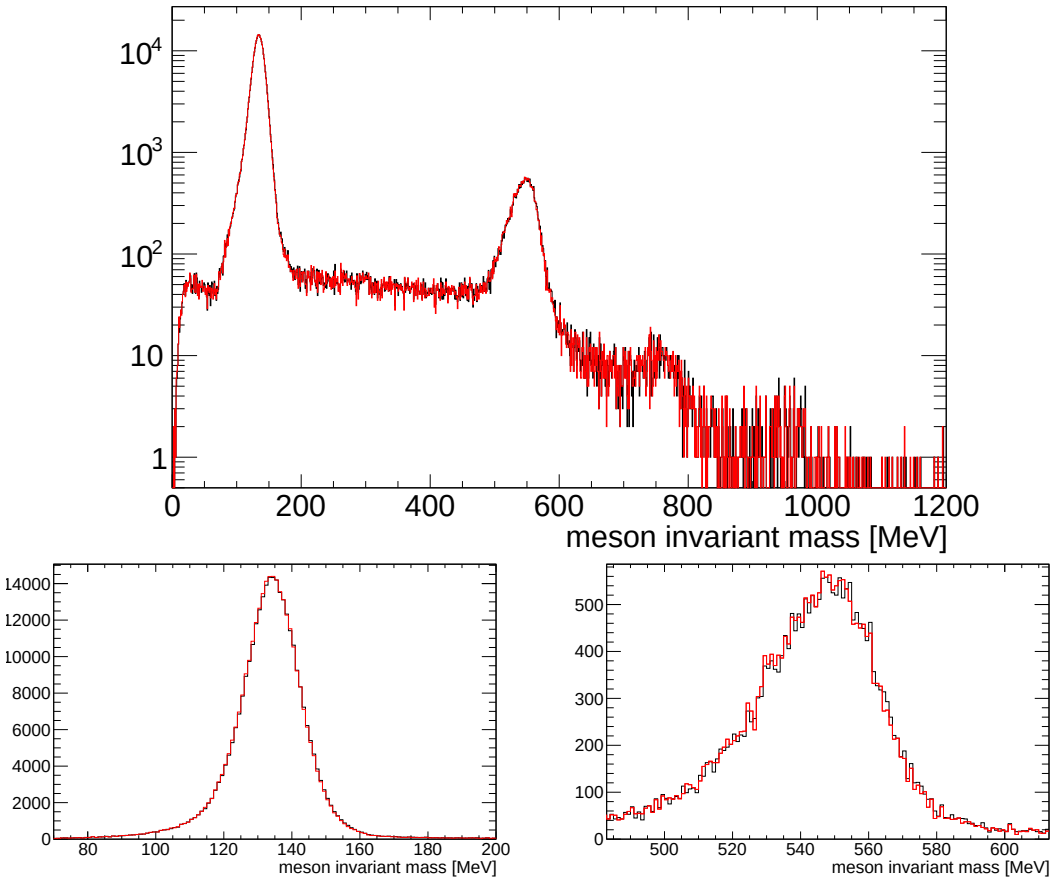


Figure 7.11: Distribution of reconstructed invariant meson masses after all other cuts were applied. Black shows the result for uncorrected data, red for pile-up corrected data. Shown below are enhanced views of the respective regions around m_{π^0} and m_η.

side, there is no way (besides pile-up) by which the detected energy could ever be too high, which explains the asymmetry. While the pile-up correction cannot change anything about lost energy, it does correct deviations towards higher masses.

Naively, the number of reconstructed mesons should increase by the pile-up correction, because the added energy from a pile-up could have been enough to shift the reconstructed invariant mass to a value which is entirely inconsistent with a π^0 or η meson. On the other hand, the opposite effect could occur as well: A photon from a π^0 decay could have lost enough energy in insensitive material such that the reconstruction would have produced a mass significantly below m_{π^0}, but due to the added energy from the pile-up, the reconstructed mass could have come close to the nominal value. So it is a priori not clear which effect would dominate.
Furthermore, it is fundamentally impossible to decide whether any given event actually involved the decay of a π^0 or any other meson, or if it the invariant mass of two uncorrelated photons just happened to fall into the observed π^0-peak. This implies that the "number of reconstructed mesons" is actually not easy to extract. In the analysis, this problem was circumvented: Fig. 7.13 shows the relative increase of entries within a given radius from the nominal meson mass. It is

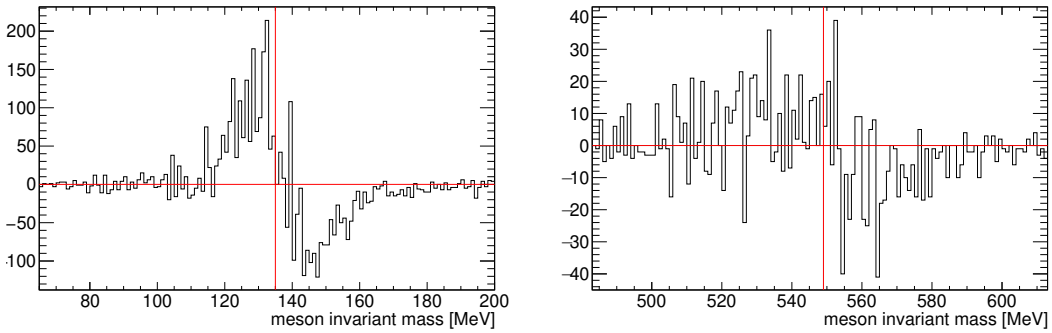


Figure 7.12: Difference between the corrected spectrum of reconstructed meson masses and the uncorrected one (red/black in Fig. 7.11) in the region of the π^0-peak (left) and around the η-peak (right). Red lines denote equality between both distributions (horizontal line) and the nominal meson mass (vertical line). Data from June 2021.

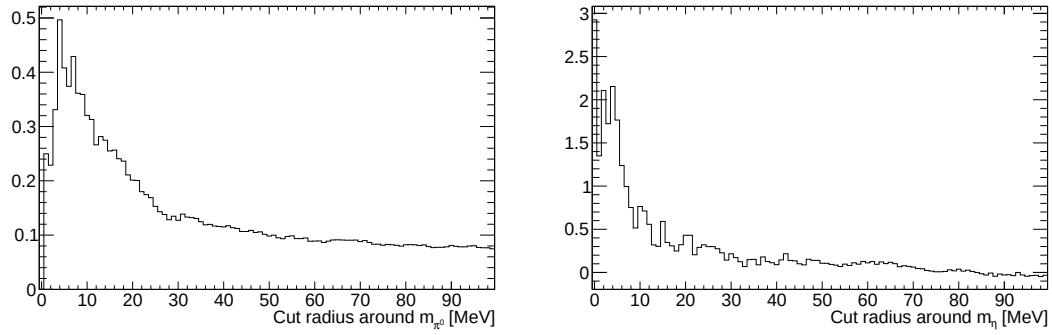


Figure 7.13: Relative increase of entries (in %) in a region of variable radius around the nominal meson mass. For these plots, the integrated number of entries in the residual (see Fig. 7.12) was divided by the corresponding number of entries in the uncorrected spectrum from Fig. 7.11.

obvious, that the overall number of entries around the respective nominal masses does, in fact, increase! By how much exactly does seem to depend strongly on the chosen radius. To provide some context to these values, the FWHM of the π^0-peak is $\approx 18.5\,\text{MeV}$ while the width of the η-peak lies around 38 MeV. In terms of a Gaussian width σ (which is a rather poor description of these peaks), these values would correspond to $\approx 8\,\text{MeV}$ for the pions and $\approx 16\,\text{MeV}$ in case of the η mesons. Applying the commonly used 3σ criterion for a hypothetical cut on the meson mass would lead to an increase of 0.17% more entries in the π^0 peak and around the same for the η peak.

Besides the number of reconstructed particles, the impact onto the mean and width of the distributions may be studied as well. As the effects might vary for different regions of the phase space, the phase space was split into 49 (7x7) different regions of roughly equal population. Due to the limited statistics in the test-sample, this was only done for the π^0 peak. In each region, the distribution of reconstructed meson masses is reasonably well described by a combination of a Gaussian and a linear background. Both the mean and the width of the Gaussians was extracted and compared between the corrected and the uncorrected data set. The differences in parameters are shown in Fig. 7.14. As expected, the mean masses in the corrected data set are found at

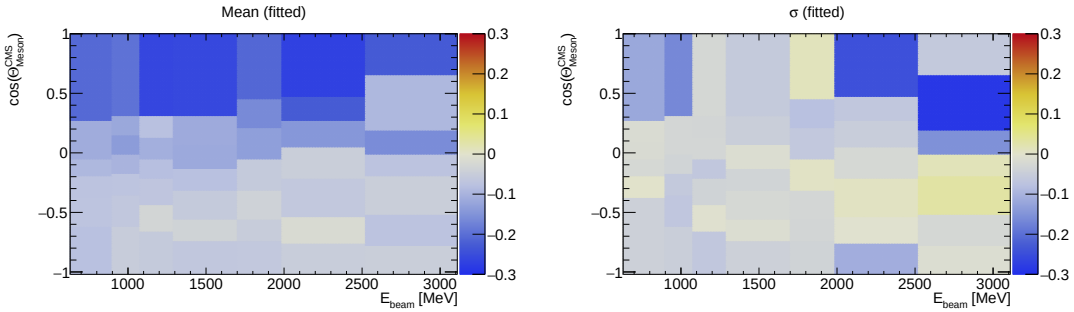


Figure 7.14: Difference in extracted mean (left) and width (right) between the pile-up corrected and the uncorrected data set. Values were extracted from a Gaussian which was fitted to the data, together with a linear background. The color-coded values are given in MeV.

slightly lower or equal values. Interestingly, the correction seems strongest in the very forward area close to $\cos\left(\theta^{\mathrm{CMS}}_{\mathrm{meson}}\right) = 1$. Considering that this is also the region where most pile-up occurs, this makes intuitive sense. Furthermore, the correction seems to result in slightly smaller widths overall, but no clear structure is observed.

7.4.1.3 Discussion

At an overall relative increase of 0.2% in the number of reconstructed mesons and only very minuscule changes to the shape of the distribution on general, the impact of the entire pile-up correction process onto the reconstruction of physically relevant reactions seems to be almost negligible. However, it must be noted that this is reassuring for past analyses which were done at the experiment. Had there been a large impact due to undetected pile-up, those analyses would have to be critically re-examined.
On the other hand, only approximately half of the actual pile-up was detected so far (see section 5), so the impact of the correction is expected to roughly double in size in the future. Furthermore, having a well-tested pile-up detection and correction routine available might enable a future increase of the event rate alongside which the pile-up rate is expected to grow simultaneously.
With the recently developed time-clustering [Sta24], the data quality might be improved further. In addition, the time-clustering itself is expected to profit from the pile-up correction as well, since the correction encompasses not only the energies, but also the respective timing features by which a specific energy deposit can be assigned to a corresponding cluster. First results[31] seem to indicate that the time-clustering has a comparable impact onto the reconstruction as the pile-up correction.

[31] N. Stausberg 2024, private communication.

CHAPTER 8

Summary and Outlook

8.1 Summary

By now, all 1320 CsI(Tl) crystals of the Crystal Barrel calorimeter are read out by the new SADCs. Due to the continued improvement of the firmware, the feature extraction has reached a status which can be considered final. It has been thoroughly tested and evaluated over the course of this thesis. Several of the presented improvements were inspired and accompanied by results from this work.

A digital FIR filter has been implemented, upon which the rest of the feature extraction is based. It achieves a significantly better suppression of frequencies above $\approx 1\,\mathrm{MHz}$ compared to the previous moving average. Among other things, this leads to an improved performance of the peakfinding algorithm and facilitates a more efficient compression. The baseline algorithm has been completely reworked and has been running very reliably ever since it was introduced. Unlike the previous algorithm, it produces no deviations towards higher values and at the same time offers an improved resolution of up to 104 keV (under ideal conditions) over the previous algorithm. The integral as the main energy feature has been calibrated in relation to the former QDC readout and, later, as a standalone readout. The integral feature offers, together with the event-based baseline subtraction, an improved energy resolution over the former QDC readout.

The CFD algorithm has been substantially improved as well, up to a point that it now offers a superior time resolution over the dedicated TDC readout for energies below 8 to 15 MeV, depending on the region of the detector. The resolution σ_{CFD} reaches typical values of $\approx 17\,\mathrm{ns}$ at $E = 50\,\mathrm{MeV}$ and gets even better for higher energies. At $E = 5\,\mathrm{MeV}$, below which no TDC times are measured, the resolution is still around 29 ns.

The peakfinding algorithm measures both the height and the position of pulses in the sampled waveforms. By employing the FIR filter and slightly tweaking the algorithm, the position of the peak is now determined much more accurately, almost completely eliminating the previously observed time walk. While the peak features may be used as secondary measures for energy and time, respectively, they are particularly relevant for their application in the pile-up detection process, which was a main focal point of this thesis.

The initially employed pile-up detection algorithm was discovered to have systematic weaknesses for certain topologies of pile-up, which were verified and investigated in an extensive analysis. In order to improve the algorithm for further data taking, a new approach was needed, which was developed based on simulations. By using a LUT to calculate the expected integral from a given pair of peak features as well as employing the recently introduced multi-hit capabilities of the peakfinder, the rate of false negatives in the pile-up detection process was greatly reduced. Based on the simulation, the new algorithm seems to exhibit no systematic weakness for any relevant topology. The algorithm was also tested using actual data and was found to be compatible with the simulations. As part of the detailed investigation, it was discovered that muons decaying at rest in the calorimeter constitute a significant source of pile-ups which exhibit a distinctive topology.

The following pile-up recovery process required a careful analysis of the occurring pulse shapes beforehand. As a major byproduct of these investigations, it was observed that the pulse shape of the occurring signals is not constant. Consistent with the expectations for a CsI(Tl) detector, it was shown that the pulse shape instead depends on the energy and type of particle, leading to different ionization densities. However, it seems unlikely that these findings could be used for a successful particle identification in the future, as there seems to be a strong influence of particles produced in secondary reactions in the detector itself or in insensitive materials in front of it.

To ease the task of disentangling overlapping pulses in pile-up situations, a custom "pulse shaping deconvolution" was developed, since a standard moving window deconvolution was shown to give unsatisfactory results. The PSD was employed with great success in the pile-up recovery process and provides a fast and precise algorithmic solution to almost half (44%) of all recorded pile-ups. Almost another third (29%) of all pile-ups could be resolved by only slightly tweaking the description obtained by the deconvolution, leaving only a quarter of all pile-ups for more elaborated routines in the recovery/fitting process. In the end, over 99.99% of all detected pile-ups were successfully disentangled, leaving only rare edge-cases in which the recovery did not work.

Finally, the impact of the recovered data onto the reconstruction of physical reactions was investigated. It was shown that the pile-up recovery has an overall positive impact on the distribution of reconstructed mesons (π^0 and η) in a simple analysis of the $p\gamma\gamma$ final state. The effect is, however, quite small, as the number of successfully reconstructed mesons increased by only $\approx 0.2\%$. Given the shortcomings of the initial pile-up detection algorithm, the effect is expected to roughly double in size in the most recent beamtimes (which were not yet analyzed) and beyond.

8.2 Outlook

Large amounts of data were already recorded with the new readout system and further data is projected to be taken (or was already taken) in the year 2024 and thereafter. By employing the pile-up detection and recovery capabilities, the event rate might be substantially increased without a loss of data quality. First results for the extraction of (double-)polarization observables with the new SADC setup are expected to be published soon.

8.2.1 Potential for further investigations

Due to the large breadth of topics studied during this thesis, there remains some potential for further investigations in the future. Some of these projects might also be suitable topics for future Bachelor's or Master's theses.

For one, it could be worthwhile to verify the proposed explanation of the origin of the different pulse shapes experimentally once and for all. While the proposed explanation seems to be at least very plausible, the large number of potentially confounding factors (secondary reactions, 1 cm of Al in front of the detector, energy depositions often split over several crystals, etc.) makes an unambiguous interpretation of the data difficult. It would be advantageous to directly study the pulse shapes resulting from different particles in a controlled environment. This could be done by shooting particle beams of suitable energy directly into CsI(Tl) crystal, which would need to be attached to the exact same readout system as in the actual experiment. Conveniently, the institute at which this thesis was written features its own cyclotron[1]. It is capable of delivering protons, deuterons and α-particles with energies of up to 14 MeV, which would be almost ideally suited for such investigations.

As alluded to in the last chapter, another fruitful project might be to investigate the interplay between the newly developed time-clustering (see section 2.10.1) [Sta24] and the pile-up correction process. As of writing this thesis, it is unclear to what degree the pile-up correction will be beneficial to the time-clustering process. An overall positive effect is to be anticipated, but if this is actually the case remains to be verified. Likewise, the size of this effect would have to be investigated as well.

Furthermore, the revised pile-up detection algorithm was recently applied in large scale during the March and April beamtimes of 2024. While preliminary results show no indications of the algorithm not working exactly as intended, a final analysis of the so-recorded data remains to be done.

8.2.2 Related developments at CBELSA/TAPS

After the Crystal Barrel is now fully equipped with an SADC readout, a similar project is underway for the MiniTAPS detector [Ott]. It is currently planned to use similar (if not the same) hardware for most components. A particularly challenging factor is the short signal length in BaF_2 ($\tau \approx 600$ ps for the fast component), which makes it impossible to do a meaningful

[1] https://www.zyklotron.hiskp.uni-bonn.de/

sampling of the signal even if no decimation at all is employed and the native 80 MHz sampling frequency is used. For that reason, the signal has to be artificially elongated using specialized analog filters. It remains to be seen to what degree the potential for particle identification based on the scintillation light in BaF_2 (see [Sta23] for application at CBELSA/TAPS) will be retained. In the end, a similar feature extraction process to CB (with similar possibilities to detect and correct pile-up) could be employed; making the MiniTAPS readout faster, more versatile and more comparable in nature to the CB readout.

While the initial design of the CB-SADC has been adapted from the antiProton ANnihilation at DArmstadt ($\bar{\text{P}}$ANDA) experiment, the knowledge transfer has recently gone both ways. During her master's thesis, C. Frenkel has applied offline feature extraction algorithms to both simulated and actual waveforms of the first fully assembled detector modules of the forward endcap. In addition, she also investigated potential applications of the pulse shaping deconvolution developed in this thesis. In both fields (feature extraction and deconvolution) she was able to draw inspiration from this work and provide valuable feedback in a very fruitful exchange.

Coincidentally, since the completion of the $\bar{\text{P}}$ANDA experiment is expected to be delayed by several years, the opportunity to employ the forward endcap of the $\bar{\text{P}}$ANDA experiment at the facility in Bonn has arisen. According to the most recent proposal, the forward endcap will replace the MiniTAPS detector in the future as part of a major upgrade and reorientation of the CBELSA/TAPS experiment[2]. As part of this upgrade, it is also foreseen to equip the experiment with a forward spectrometer, including a large dipole magnet, tracking detectors[3], a Time-of-Flight (TOF) wall in front of the endcap and a silicone pixel detector around the target (inside the crystal barrel). In case this new experiment is approved and realized[4] it would be very advantageous that the readout and feature-extraction algorithms in both main calorimeters are already very similar. Among others, the knowledge obtained in the pile-up detection and recovery process for CB could be very beneficial for similar projects regarding the forward endcap.

[2] The name of the future experiment is still to be discussed.

[3] Both Gas Electron Multipliers (GEMs) and straw tube detectors.

[4] Which is not yet finally decided, as the application is still ongoing.

List of Acronyms

AC Alternating Current
ADC Analog-to-Digital Converter
APD Avalanche Photo Diode
ASIC Application-Specific Integrated Circuit

BR Branching Ratio
BuffTi Buffer/Timing filter

CB Crystal Barrel
CFD Constant Fraction Discriminator
CMS Center-of-Mass System
CPU Central Processing Unit

DAC Digital-to-Analog Converter
DAQ Data Acquisition
DC Direct Current
DDR Double Data Rate
DFT Discrete Fourier Transform
DNP Dynamic Nuclear Polarization
DS Downstream

ECF Energy Correction Function
ELSA Electron Stretcher Accelerator

FFT Fast Fourier Transform
FFTW Fastest Fourier Transform in the West
FIFO First In First Out
FIR Finite Impulse Response
FluMo Flux Monitor
FPGA Field Programmable Gate Array
FWHM Full Width at Half Maximum

GDR Giant Dipole Resonance
GEM Gas Electron Multiplier
GIM Gamma Intensity Monitor
GPU Graphics Processing Unit

HDL Hardware Description Language

IP Internet Protocol

LE Leading Edge
LEVB Local EVent Builder
LINAC LINear ACcelerator
LQCD Lattice QCD
LUT Lookup Table

MA Moving Average
MIP Minimum Ionizing Particle
MPV Most Probable Value

P̄ANDA antiProton ANnihilation at DArmstadt
PDG Particle Data Group
PED Particle Energy Deposit
PID Particle Identification
PIN Positive Intrinsic Negative
PMT Photo Multiplier Tube
PSD Pulse Shaping Deconvolution
PWA Partial Wave Analysis
PZC Pole-Zero Cancellation

QCD Quantum Chromo Dynamics
QDC Charge-to-Digital Converter
QED Quantum Electro Dynamics

SADC Sampling-ADC
SFP Small Form-factor Pluggable
SNR Signal-to-Noise Ratio
STE Self Trapped Exciton
STH Self Trapped Hole

TCP Transmission Control Protocol
TDC Time-to-Digital Converter
TOF Time-of-Flight

UDP User Datagram Protocol
US Upstream

VHDL Very High Speed Integrated Circuit HDL

WIMP Weakly Interacting Massive Particle

Bibliography

[ABB68] Aachen-Berlin-Bonn-Hamburg-Heidelberg-München Collaboration, *Photoproduction of Meson and Baryon Resonances at Energies up to 5.8 GeV*, Phys. Rev. **175** (5 1968) 1669, URL: https://link.aps.org/doi/10.1103/PhysRev.175.1669 (cit. on p. 100).

[Ahr+06] J. Ahrens et al., *Measurement of the Gerasimov-Drell-Hearn Integrand for ^{2}H from 200 to 800 MeV*, Phys. Rev. Lett. **97** (20 2006) 202303, URL: https://link.aps.org/doi/10.1103/PhysRevLett.97.202303 (cit. on p. 25).

[Ake+92] E. Aker et al., *The crystal barrel spectrometer at LEAR*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **321** (1992) 69, ISSN: 0168-9002, URL: https://www.sciencedirect.com/science/article/pii/016890029290379I (cit. on pp. 12, 15).

[AKN+15] A. V. Anisovich, E. Klempt, V. A. Nikonov, et al., *Bonn-Gatchina Partial Wave Analysis*, 2015, URL: https://pwa.hiskp.uni-bonn.de (cit. on p. 6).

[Ams+02] C. Amsler et al., *Temperature dependence of pure CsI: scintillation light yield and decay time*, Nucl. Instrum. Methods Phys. Res. A **480** (2002) 494, ISSN: 0168-9002 (cit. on p. 12).

[And+94] D. Anderson, J. Kierstead, P. Lecoq, S. Stoll, and C. Woody, *A search for scintillation in doped and orthorhombic lead fluoride*, Nucl. Instrum. Methods Phys. Res. A **342** (1994) 473, ISSN: 0168-9002 (cit. on p. 27).

[ATL12] ATLAS Collaboration, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, Physics Letters B **716** (2012) 1, ISSN: 0370-2693 (cit. on p. 1).

[Ben14] M. Bendel, *Entwicklung einer neuartigen Nachweismethode hochenergetischer Teilchen im CALIFA-Kalorimeter*, Dissertation: Technische Universität München, 2014 (cit. on pp. 13, 14, 17, 117).

[Ber+05] M. Berger, J. Coursey, M. Zucker, and J. Chang,
ESTAR, PSTAR, and ASTAR: Computer Programs for Calculating Stopping-Power and Range Tables for Electrons, Protons, and Helium Ions (version 1.2.3),
National Institute of Standards and Technology, Gaithersburg, MD. [Online; Accessed October 13, 2022], 2005, URL: http://physics.nist.gov/Star
(cit. on pp. 18, 123).

[Ber+10] M. Berger et al., *XCOM: Photon Cross Section Database (version 1.5)*,
[Online; 27.7.2022], 2010, URL: http://physics.nist.gov/xcom (cit. on pp. 10, 125).

[Bet30] H. Bethe,
Zur Theorie des Durchgangs schneller Korpuskularstrahlen durch Materie,
Annalen der Physik **397** (1930) 325 (cit. on p. 8).

[BGH+13] M. Bendel, R. Gernhäuser, W. Henning, et al.,
RPID — A new digital particle identification algorithm for CsI(Tl) scintillators,
The European Physical Journal A **49** (2013) 69 (cit. on p. 16).

[BH34] H. Bethe and W. Heitler,
On the stopping of fast particles and on the creation of positive electrons,
Proc. R. Soc. Lond. A **146** (1934) 83 (cit. on p. 9).

[Bie16] J. Bieling, 2016, URL: https://github.com/jobisoft/jTDC/wiki (cit. on pp. 23, 31).

[Bie18] M. Biel, *Entwicklung einer Analysemethode zur Bestimmung und Optimierung der Baseline-Einstellung und Pole-Zero-Kompensation für die Analogstufe der Sampling-ADCs des Crystal Barrel Experimentes*,
Bachelor thesis: Rheinische Friedrich-Wilhelms-Universität Bonn, 2018
(cit. on pp. 41, 42, 113).

[Bob22] Bob K., *Comparison of spectral leakage of several window functions*,
Creative Commons CC0 1.0 license, 2022,
URL: https://commons.wikimedia.org/w/index.php?curid=117573737
(cit. on p. 106).

[BR97] R. Brun and F. Rademakers, *ROOT — An object oriented data analysis framework*,
Nucl. Instrum. Methods Phys. Res. A **389** (1997) 81, New Computing Techniques in Physics Research V, ISSN: 0168-9002 (cit. on p. 105).

[Bra+99] C. Bradtke et al., *A new frozen-spin target for 4π particle detection*,
Nucl. Instrum. Methods Phys. Res. A **436** (1999) 430, ISSN: 0168-9002,
URL: https://www.sciencedirect.com/science/article/pii/S0168900299003836
(cit. on pp. 24, 25).

[Bra00] R. Bracewell, *The Fourier Transform and Its Applications*, 3rd edition,
Boston: McGraw-Hill, 2000 (cit. on pp. 103, 104).

[Ciu18] S. Ciupka, *Test and calibration of a module for the new Energy-sum trigger for the Crystal Barrel Calorimeter at ELSA*,
Master thesis: Rheinische Friedrich-Wilhelms-Universität Bonn, 2018
(cit. on p. 33).

[CM97] D. G. Crabb and W. Meyer,
Solid polarized targets for nuclear and particle physics experiments,
Annual Review of Nuclear and Particle Science **47** (1997) 67,
URL: https://doi.org/10.1146/annurev.nucl.47.1.67 (cit. on p. 24).

[CMS12] CMS Collaboration, *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*, Physics Letters B **716** (2012) 30, ISSN: 0370-2693 (cit. on p. 1).

[Dah08] T. Dahlke, *Bestimmung einer winkelabhängigen Energiekorrekturfunktion für das TAPS-Kalorimeter des Crystal-Barrel/TAPS-Experimentes an ELSA*,
Diploma thesis: Rheinische Friedrich-Wilhelms-Universität Bonn, 2008
(cit. on p. 27).

[Dav19] N. Davis,
Das neue Crystal Barrel Lichtpulser-System: Untersuchung und Optimierung verschiedener Methoden zur Kalibration des High-Range-Bereiches der ADCs,
Bachelor thesis: Rheinische Friedrich-Wilhelms-Universität Bonn, 2019
(cit. on p. 66).

[Die+15] S. Diehl, R. W. Novotny, B. Wohlfahrt, and R. Beck,
Readout concepts for the suppression of the slow component of BaF_2 for the upgrade of the TAPS spectrometer at ELSA,
Journal of Physics: Conference Series **587** (2015) 012044 (cit. on pp. 16, 27).

[Die08] J. Dielmann, *Entwicklung, Aufbau und Test eines Detektors zur Bestimmung des Photonenflusses an der Bonner Photonenmarkierungsanlage*,
Diploma thesis: Rheinische Friedrich-Wilhelms-Universität Bonn, 2008
(cit. on p. 27).

[Edw+11] R. G. Edwards, J. J. Dudek, D. G. Richards, and S. J. Wallace,
Excited state baryon spectroscopy from lattice QCD,
Phys. Rev. D **84** (7 2011) 074508 (cit. on p. 5).

[Ess89] K. Esslinger, *CB CsI-Shaper*, 1989 (cit. on p. 30).

[FH22] F. Frommberger and W. Hillert, *Elektronen-Stretcher-Anlage ELSA*,
[Online; 12.7.2022], 2022,
URL: http://www-elsa.physik.uni-bonn.de/Beschleuniger/bilder/elsaplan_en.pdf
(cit. on p. 21).

[FJ05] M. Frigo and S. G. Johnson, *The Design and Implementation of FFTW3*,
Proceedings of the IEEE **93** (2005) 216, Special issue on "Program Generation, Optimization, and Platform Adaptation" (cit. on p. 105).

[For10] K. Fornet-Ponse, *Die Photonenmarkierungsanlage für das Crystal-Barrel/TAPS-Experiment an ELSA*,
PhD thesis: Rheinische Friedrich-Wilhelms-Universität Bonn, 2010,
URL: https://hdl.handle.net/20.500.11811/4507 (cit. on p. 24).

[Fun08] C. Funke, *Analyse der Triggerfähigkeiten zur Selektion hadronischer Ereignisse und Entwicklung eines Hochgeschwindigkeits-Triggers für den Vorwärtskonus des Crystal-Barrel-Detektors*,
PhD thesis: Rheinische Friedrich-Wilhelms-Universität Bonn, 2008,
URL: https://hdl.handle.net/20.500.11811/3678 (cit. on p. 28).

[Gab+94] A. Gabler et al., *Response of TAPS to monochromatic photons with energies between 45 and 790 MeV*,
Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **346** (1994) 168,
ISSN: 0168-9002,
URL: https://www.sciencedirect.com/science/article/pii/0168900294907013
(cit. on p. 27).

[Gas+13] M. Gascón, S. Lam, S. Wang, S. Curtarolo, and R. Feigelson,
Characterization of light output and scintillation emission in CsI(Tl), NaI(Tl), and LaBr 3 (Ce) under isostatic pressure, Radiation Measurements **56** (2013) 70
(cit. on p. 15).

[GG93] A. Georgiev and W. Gast, *Digital pulse processing in high resolution, high throughput, gamma-ray spectroscopy*,
IEEE Transactions on Nuclear Science **40** (1993) 770 (cit. on p. 128).

[Gla61] S. L. Glashow, *Partial-symmetries of weak interactions*,
Nuclear Physics **22** (1961) 579, ISSN: 0029-5582 (cit. on p. 1).

[GM63] R. Gwin and R. B. Murray,
Scintillation Process in CsI(Tl). I. Comparison with Activator Saturation Model,
Phys. Rev. **131** (2 1963) 501 (cit. on pp. 12, 16, 18).

[Got+21] M. Gottschall et al.,
Measurement of the helicity asymmetry E for the reaction $\gamma p \to \pi^0 p$,
The European Physical Journal A **57** (2021) 40, ISSN: 1434-601X,
URL: https://doi.org/10.1140/epja/s10050-020-00334-2 (cit. on p. 29).

[Grä19] M. Gräf, *Teilchenidentifikation mit der neuen SADC-Auslese des Crystal-Barrel Kalorimeters*,
Bachelor thesis: Rheinische Friedrich-Wilhelms-Universität Bonn, 2019
(cit. on pp. 116, 118).

[Gri+14] S. Gridin, A. N. Vasil'ev, A. Belsky, N. Shiran, and A. Gektin, *Excitonic and activator recombination channels in binary halide scintillation crystals*,
physica status solidi (b) **251** (2014) 942 (cit. on p. 12).

[Gro+22] D. E. Groom et al., *Atomic and Nuclear Properties of Materials*, (2022),
URL: https://pdg.lbl.gov/2022/AtomicNuclearProperties/index.html
(cit. on pp. 12, 18, 27, 100).

[Gru16] M. Gruener, *Technical drawing*, 2016 (cit. on p. 28).

[Grü22] M. Grüner, Private communication, 2022 (cit. on p. 29).

[Ham09] C. Hammann, *Aufbau eines Flüssigwasserstofftargets zur Durchführung von Kalibrationsmessungen am Crystal-Barrel Experiment an ELSA*, Diploma thesis: Rheinische Friedrich-Wilhelms-Universität Bonn, 2009 (cit. on p. 24).

[Har08] J. Hartmann, *Zeitkalibrierung und Photonenflussbestimmung für das Crystal-Barrel-Experiment an ELSA*, PhD thesis: Rheinische Friedrich-Wilhelms-Universität Bonn, 2008 (cit. on p. 27).

[Har17] J. Hartmann, *Measurement of Double Polarization Observables in the Reactions $\gamma p \to p\pi^0$ and $\gamma p \to p\eta$ with the Crystal Barrel/TAPS Experiment at ELSA*, PhD thesis: Rheinische Friedrich-Wilhelms-Universität Bonn, 2017, URL: https://hdl.handle.net/20.500.11811/7258 (cit. on pp. 22, 23, 34).

[Har78] F. Harris, *On the Use of Windows for Harmonic Analysis With the Discrete Fourier Transform*, Proceedings of the IEEE **66** (1978) 51 (cit. on p. 106).

[Hil00] W. Hillert, *Erzeugung eines Nutzstrahls spinpolarisierter Elektronen an der Beschleunigeranlage ELSA*, Habilitationsschrift: Rheinische Friedrich-Wilhelms-Universität Bonn, 2000 (cit. on pp. 20, 21).

[Hon+23] C. Honisch et al., *The new APD-Based Readout of the Crystal Barrel Calorimeter – An Overview*, 2023, arXiv: 2212.12364 [physics.ins-det] (cit. on p. 29).

[Hon15] C. Honisch, *Design, Aufbau und Test einer neuen Ausleseelektronik für das Crystal-Barrel-Kalorimeter*, PhD thesis: Rheinische Friedrich-Wilhelms-Universität Bonn, 2015, URL: https://hdl.handle.net/20.500.11811/6529 (cit. on pp. 29, 32, 40).

[Kai07] D. Kaiser, *Aufbau und Test eines Gas-Cherenkov-Detektors für den Crystal-Barrel Aufbau an ELSA*, Diploma thesis: Rheinische Friedrich-Wilhelms-Universität Bonn, 2007 (cit. on p. 26).

[Kla23] P. Klassen, *Entwicklung eines neuen Cluster Finders für das Crystal-Barrel-Kalorimeter (unfinished)*, PhD thesis: Rheinische Friedrich-Wilhelms-Universität Bonn, 2023 (cit. on pp. 31, 32).

[Kna19] J. Knaust, *A Controller for the Crystal Barrel Sampling ADCs*, Master thesis: Rheinische Friedrich-Wilhelms-Universität Bonn, 2019 (cit. on p. 39).

[Kno00] G. F. Knoll, *Radiation Detection and Measurement*, 3rd edition, New York: John Wiley and Sons, 2000 (cit. on pp. 12, 13, 27).

[Kon+19] A. Koning et al., *TENDL: Complete Nuclear Data Library for Innovative Nuclear Science and Technology*, Nuclear Data Sheets **155** (2019) 1, Special Issue on Nuclear Reaction Data, ISSN: 0090-3752 (cit. on pp. 123, 124).

[Lan21] B. J. Land, 2021,
URL: https://ben.land/post/2021/06/24/deconvolving-with-convolution/
(cit. on p. 137).

[Lec20] P. Lecoq, "Scintillation Detectors for Charged Particles and Photons",
Cham: Springer International Publishing, 2020 45 (cit. on p. 13).

[Lu+17] X. Lu et al., *Energy-Dependent Scintillation Pulse Shape and Proportionality of Decay Components for CsI:Tl: Modeling with Transport and Rate Equations*,
Phys. Rev. Applied **7** (1 2017) 014007 (cit. on pp. 12–16).

[Mar12] P. Marciniewski,
A 32-channel, 14-bit, dual range, 80-125 MSPS ADC for the EMC-Endcap, 2012,
URL: https://indico.gsi.de/getFile.py/access?contribId=18&sessionId=7&resId=0&materialId=slides&confId=2126 (cit. on p. 37).

[McG08] W. R. McGehee, *The Gamma Intensity Monitor at the Crystal-Barrel-Experiment*,
Master thesis: Massachusetts Institute of Technology. Dept. of Physics, 2008
(cit. on p. 27).

[Mit19] B. Mitlasoczki,
Design and Test of a New Energy Sum Trigger for the CBELSA/TAPS Experiment,
Master thesis: Rheinische Friedrich-Wilhelms-Universität Bonn, 2019
(cit. on p. 33).

[MM61] R. B. Murray and A. Meyer, *Scintillation Response of Activated Inorganic Crystals to Various Charged Particles*, Phys. Rev. **122** (3 1961) 815 (cit. on pp. 12, 16, 18).

[Mos+12] W. W. Moses et al., *The Origins of Scintillator Non-Proportionality*,
IEEE Transactions on Nuclear Science **59** (2012) 2038 (cit. on p. 17).

[Mül18] J. Müller, *Bestimmung der Doppelpolarisationsobservablen E in der Reaktion $\gamma p \rightarrow p\eta$ am CBELSA/TAPS-Experiment in Bonn*,
PhD thesis: Rheinische Friedrich-Wilhelms-Universität Bonn, 2018
(cit. on pp. 29, 64).

[Mül19] J. S. Müllers, *An FPGA-based Sampling ADC for the Crystal Barrel Calorimeter*,
PhD thesis: Rheinische Friedrich-Wilhelms-Universität Bonn, 2019,
URL: https://hdl.handle.net/20.500.11811/8118
(cit. on pp. 29, 37, 39, 40, 42, 43, 45, 47, 55, 61, 63, 178).

[Mur+86] T. Murakami et al., *Properties of BaF_2 scintillators in charged particle detection*,
Nucl. Instrum. Methods Phys. Res. A **253** (1986) 163, ISSN: 0168-9002
(cit. on p. 27).

[NB13] O. Niemitalo and Bob K., *Window functions and frequency responses*,
Creative Commons CC0 1.0 license, 2013, URL: https://en.wikipedia.org/w/index.php?title=Window_function&oldid=1117292479
(cit. on p. 106).

[Nov91] R. Novotny, *The BaF_2 photon spectrometer TAPS*,
IEEE Transactions on Nuclear Science **38** (1991) 379 (cit. on p. 27).

[Nyq28] H. Nyquist, *Certain Topics in Telegraph Transmission Theory*, Transactions of the American Institute of Electrical Engineers **47** (1928) 617 (cit. on p. 38).

[OM59] H. Olsen and L. C. Maximon, *Photon and Electron Polarization in High-Energy Bremsstrahlung and Pair Production with Screening*, Phys. Rev. **114** (3 1959) 887, URL: https://link.aps.org/doi/10.1103/PhysRev.114.887 (cit. on p. 22).

[OPA06] OPAL collaboration, *Measurement of the running of the QED coupling in small-angle Bhabha scattering at LEP*, The European Physical Journal C **45** (2006) 1 (cit. on p. 3).

[OSB99] A. Oppenheim, R. Schafer, and J. Buck, *Discrete-time Signal Processing*, Prentice Hall international editions, Prentice Hall, 1999, ISBN: 9780137549207, URL: https://books.google.de/books?id=h_lSAAAAMAAJ (cit. on p. 105).

[Ott] B. Otto, *Upgrade des CBELSA/TAPS-Experimentes unter Vorwärtswinkeln (in preparation)*, PhD thesis: Rheinische Friedrich-Wilhelms-Universität Bonn (cit. on p. 163).

[Ott22] B. Otto, *Optimisation of the Sampling-ADC Readout of the Crystal Barrel Calorimeter*, Master thesis: Rheinische Friedrich-Wilhelms-Universität Bonn, 2022 (cit. on pp. 43, 44, 54–57, 59, 60, 67, 68, 70, 72, 75, 77, 96, 107, 108, 177).

[Ott24] B. Otto, Private communication, 2024 (cit. on p. 77).

[PDG22] R.L. Workman et al. (Particle Data Group), *Review of Particle Physics*, PTEP **2022** (2022) 083C01 (cit. on pp. 3, 8, 11, 19, 64, 100, 101).

[Pol16] T. Poller, *Simulation, Aufbau und Test einer Schaltung zur Baseline-Verschiebung der Crystal Barrel Kalorimetersignale für den Sampling-ADC*, Bachelor thesis: Rheinische Friedrich-Wilhelms-Universität Bonn, 2016 (cit. on pp. 39, 41).

[RM11] M. Ronniger and B. Metsch, *Effects of a spin-flavour dependent interaction on the baryon mass spectrum*, The European Physical Journal A **47** (2011) (cit. on p. 5).

[Roe20] T. Roelandts, *FIIIR! - Design FIR & IIR Filters*, Web-Applet, 2020, URL: https://fiiir.com/ (cit. on pp. 107, 109).

[Rön+22] D. Rönchen, M. Döring, U.-G. Meißner, and C.-W. Shen, *Light baryon resonances from a coupled-channel study including* $\mathbf{K\Sigma}$ *photoproduction*, Eur. Phys. J. A **58** (2022) 229, arXiv: 2208.00089 [nucl-th] (cit. on p. 6).

[Rot16] A. Roth, *Extraction of energy and time from pile-up pulses with fast sampling ADC analysis techniques*, Master thesis: Lunds Universitet, 2016 (cit. on p. 129).

[Run+16] S. Runkel, H. Dutz, S. Goertz, S. Reeve, and M. Bornstein, "CFD-Simulations of a 4π-continuous-mode dilution refrigerator for the CB-ELSA experiment", 2016 018 (cit. on p. 25).

[Sai22] Saint-Gobain Crystals, Company website, 2022, URL: https://www.crystals.saint-gobain.com/radiation-detection-scintillators/crystal-scintillators/csi-cesium-iodide# (cit. on p. 12).

[Sal16] B. W. Salisbury, *Entwicklung einer grafischen Benutzeroberfläche zur Energiekalibrierung des Crystal-Barrel-Kalorimeters*, Bachelor thesis: Rheinische Friedrich-Wilhelms-Universität Bonn, 2016 (cit. on p. 64).

[Sal18] B. W. Salisbury, *Results of SADC energy-calibration*, Internal Meeting, 2018 (cit. on p. 61).

[Sal68] A. Salam, "Weak and Electromagnetic Interactions", *Elementary particle theory. Relativistic groups and analyticity. Proceedings of the Eighth Nobel Symposium*, ed. by N. Svartholm, Stockholm: Almquist & Wiksell, 1968 367 (cit. on p. 1).

[Sch16] J. Schultes, *Test and Improvement of Feature-Extraction Methods for the new SADC-Readout of the Crystal Barrel Calorimeter*, Master thesis: Rheinische Friedrich-Wilhelms-Universität Bonn, 2016 (cit. on pp. 46, 49, 61, 66, 79, 80, 84, 87, 88, 178).

[SCI75] T. Stockham, T. Cannon, and R. Ingebretsen, *Blind deconvolution through digital signal processing*, Proceedings of the IEEE **63** (1975) 678 (cit. on p. 137).

[Sei09] T. Seifen, *Verbesserung der Rekonstruktion und Entwicklung eines First-Level-Triggerschemas für das Crystal-Barrel-Kalorimeter*, Diploma thesis: Rheinische Friedrich-Wilhelms-Universität Bonn, 2009 (cit. on p. 33).

[Sei21] T. Seifen, *Messung von Polarisationsobservablen in der $2\pi^0$-Photoproduktion mit dem Crystal-Barrel/TAPS-Experiment*, PhD thesis: Rheinische Friedrich-Wilhelms-Universität Bonn, 2021, URL: https://hdl.handle.net/20.500.11811/8984 (cit. on p. 34).

[Sha48] C. E. Shannon, *A mathematical theory of communication*, The Bell System Technical Journal **27** (1948) 379 (cit. on p. 38).

[SJW58] R. S. Storey, W. Jack, and A. Ward, *The Fluorescent Decay of CsI(Tl) for Particles of Different Ionization Density*, Proceedings of the Physical Society **72** (1958) 1 (cit. on pp. 12, 15).

[Smi97] S. W. Smith, *The Scientist and Engineer's Guide to Digital Signal Processing*, California Technical Publishing, 1997, ISBN: 0-9660176-3-3 (cit. on p. 137).

[Sok+15] V. Sokhoyan et al., *High-statistics study of the reaction $\gamma p \to p2\pi 0$*, The European Physical Journal A **51** (2015) 95, ISSN: 1434-601X, URL: https://doi.org/10.1140/epja/i2015-15095-x (cit. on p. 100).

[Spi08] H. Spieler, *Semiconductor Detector Systems*, Oxford University Press, 2008, ISBN: 978-0-19-852784-8 (cit. on p. 42).

[Sta19] N. Stausberg, *Time Calibration and Target Asymmetry Determination for Single Meson-Photoproduction at the CBELSA/TAPS Experiment*,
master thesis: Rheinische Friedrich-Wilhelms-Universität Bonn, 2019
(cit. on pp. 31, 50, 78).

[Sta21] N. Stausberg, *Time in the Crystal Barrel Calorimeter*,
Presentation held at internal Meeting, 2021, URL: https://agthoma.hiskp.uni-bonn.de/intern/wiki/uploads/Main/GroupMeeting/time_in_cb_part_2.pdf
(cit. on pp. 86, 91).

[Sta23] P. S. Stahl, *Pulsformanalyse zur Teilchenidentifikation am MiniTAPS-Detektor des CBELSA/TAPS-Experiments*,
Rheinische Friedrich-Wilhelms-Universität Bonn, 2023 (cit. on p. 164).

[Sta24] N. Stausberg, *Measurement of Double Polarization Observables in the Reaction $\gamma p \to p\pi^0\pi^0$ with the Crystal Barrel/TAPS Experiment at ELSA (in preparation)*,
PhD thesis: Rheinische Friedrich-Wilhelms-Universität Bonn, 2024
(cit. on pp. 34, 68, 76, 82, 86, 87, 96, 154, 159, 163).

[Sto12] V. I. Stoica,
Digital pulse-shape analysis and controls for advanced detector systems,
PhD thesis: University of Groningen, 2012,
URL: https://pure.rug.nl/ws/portalfiles/portal/14645846/proefschriftStoica.pdf
(cit. on p. 128).

[Suf+05] G. Suft et al.,
A scintillating fibre detector for the Crystal Barrel experiment at ELSA,
Nucl. Instrum. Methods Phys. Res. A **538** (2005) 416, ISSN: 0168-9002
(cit. on p. 25).

[Syn+08] A. Syntfeld-Kazuch, M. Moszynski, L. Swiderski, W. Klamra, and A. Nassalski,
Light Pulse Shape Dependence on γ-Ray Energy in CsI(Tl),
IEEE Transactions on Nuclear Science **55** (2008) 1246 (cit. on pp. 12, 15, 16).

[TAW22] A. Thiel, F. Afzal, and Y. Wunderlich, *Light Baryon Spectroscopy*,
Progress in Particle and Nuclear Physics **125** (2022) 103949 (cit. on p. 7).

[Thi12] A. Thiel,
Bestimmung der Doppelpolarisationsobservablen G in π^0-Photoproduktion,
PhD thesis: Rheinische Friedrich-Wilhelms-Universität Bonn, 2012,
URL: https://hdl.handle.net/20.500.11811/5344 (cit. on p. 23).

[Tho+13] A. Thomas, N. S. Borisov, H. .-. Arends, A. N. Fedorov, G. M. Gurevich, et al.,
The new frozen spin target at MAMI, Physics of Particles and Nuclei **44** (2013),
ISSN: 1063-7796 (cit. on p. 25).

[Tia+18] L. Tiator et al., *Eta and etaprime photoproduction on the nucleon with the isobar model EtaMAID2018*, The European Physical Journal A **54** (2018),
ISSN: 1434-601X, URL: http://dx.doi.org/10.1140/epja/i2018-12643-x (cit. on p. 6).

[Urb18] M. Urban, *Design eines neuen Lichtpulsersystems sowie Aufbau und Inbetriebnahme der neuen APD Auslese für das Crystal-Barrel-Kalorimeter*, PhD thesis: Rheinische Friedrich-Wilhelms-Universität Bonn, 2018, URL: https://hdl.handle.net/20.500.11811/7512 (cit. on p. 29).

[Val+93] J. D. Valentine, W. W. Moses, S. E. Derenzo, D. K. Wehe, and G. F. Knoll, *Temperature dependence of CsI(Tl) gamma-ray excited scintillation characteristics*, Nucl. Instrum. Methods Phys. Res. A **325** (1993) 147, ISSN: 0168-9002 (cit. on p. 15).

[Wal07] D. Walther, *Technical drawings*, Private communication, 2007 (cit. on pp. 20, 22, 25, 26).

[Wan18] Y.-C. Wang, *Development of a Quality Assurance Procedure for the Crystal Barrel SADC*, Master thesis: Rheinische Friedrich-Wilhelms-Universität Bonn, 2018 (cit. on p. 39).

[Wei67] S. Weinberg, *A Model of Leptons*, Phys. Rev. Lett. **19** (21 1967) 1264, URL: https://link.aps.org/doi/10.1103/PhysRevLett.19.1264 (cit. on p. 1).

[Wen08] C. Wendel, *Design und Aufbau eines Szintillationsdetektors zur Identifizierung geladener Teilchen im Crystal-Barrel-Vorwärtsdetektor*, PhD thesis: Rheinische Friedrich-Wilhelms-Universität Bonn, 2008, URL: https://hdl.handle.net/20.500.11811/3640 (cit. on p. 26).

[Wik22] Wikipedia contributors, *Wiener deconvolution — Wikipedia, The Free Encyclopedia*, [Online; accessed 1-November-2022], 2022, URL: https://en.wikipedia.org/w/index.php?title=Wiener_deconvolution&oldid=1116471126 (cit. on p. 137).

[Wik24] Wikipedia contributors, *Standard Model — Wikipedia, The Free Encyclopedia*, [Online; accessed 13-May-2024], 2024, URL: https://en.wikipedia.org/w/index.php?title=Standard_Model&oldid=1222953765 (cit. on p. 2).

[Wor+12] R. L. Workman, M. W. Paris, W. J. Briscoe, and I. I. Strakovsky, *Unified Chew-Mandelstam SAID analysis of pion photoproduction data*, Phys. Rev. C **86** (2012) 015202, arXiv: 1202.0845 [hep-ph] (cit. on p. 6).

[Zaz01] S. Zazubovich, *Physics of halide scintillators*, Radiation Measurements **33** (2001) 699, Proceedings of the International Symposium on Luminescent Detecto rs and Transformers of Ionizing Radiation, ISSN: 1350-4487 (cit. on pp. 12, 14).

[Zyw19] L. Zywietz Rolon, *Optimierung der Baseline-Bestimmung fur die Crystal Barrel Sampling ADCs*, Bachelor thesis: Rheinische Friedrich-Wilhelms-Universität Bonn, 2019 (cit. on pp. 57–59).

APPENDIX A

Appendix

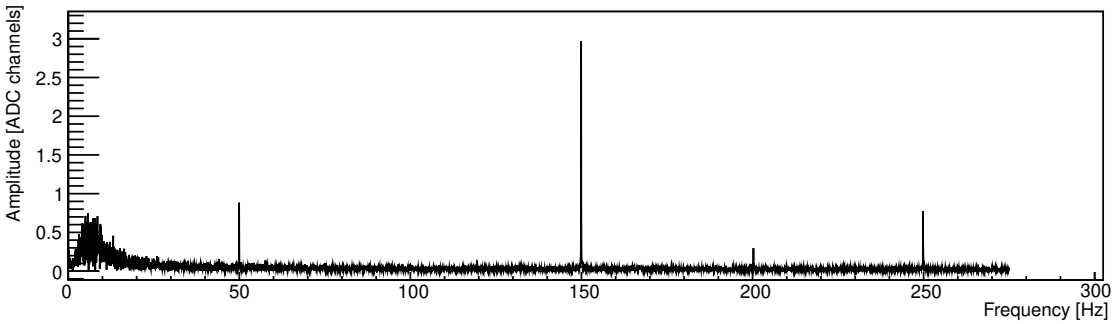


Figure A.1: FFT decomposition of baseline values from a single channel; recorded over several seconds. In addition to a relatively flat background with slightly stronger contributions at lower frequencies, multiples (harmonics) of the 50 Hz frequency of the power grid can be clearly identified.

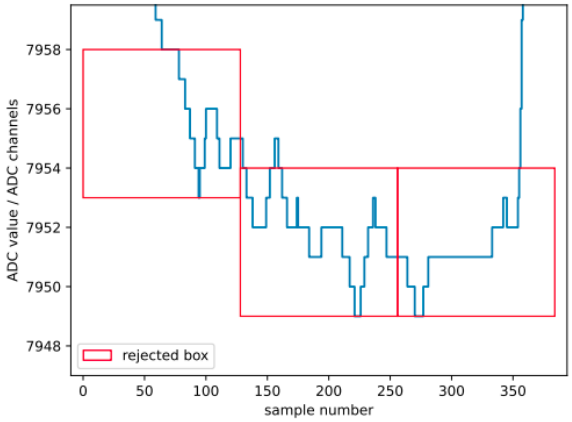


Figure A.2: Situation in which a sufficiently large area of flatness was not recognized by the baseline finding algorithm due to unfortunate alignment of the boxes [Ott22].

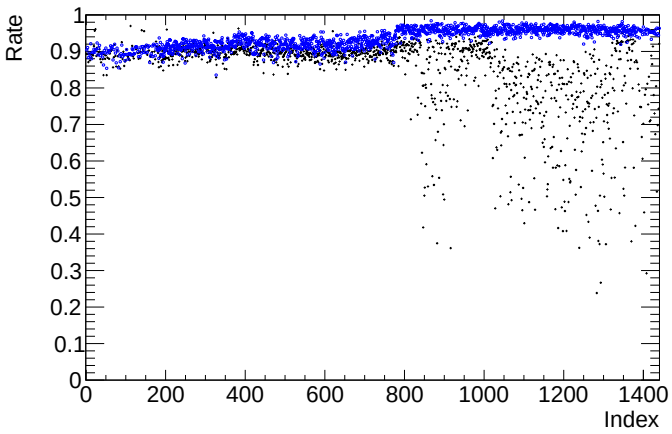


Figure A.3: Rate of baseline updates with normal parameter settings (box height of 5/10 ADC channels in DS/US [≈ 0.25 MeV] and box width of 256 samples [12.8 µs]) shown in blue and deliberately too tight settings (box height of 1 ADC channel everywhere, same box width) in black.

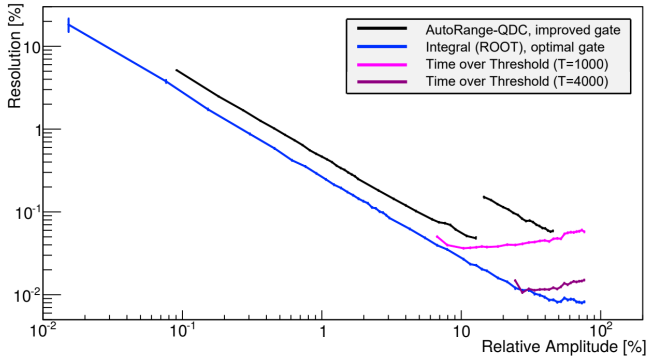


Figure A.4: Energy resolution of SADC (blue) and QDC (black) readout compared under laboratory conditions. 100% on the axis corresponds to the respective digitizers full dynamic range. The step around 12.5% is due to the QDC readout switching from low- to high-range. In addition to the integral (blue), the possibility of using Time-over-Threshold (magenta/violet) as an alternative energy measure was investigated, but ultimately not employed. Own work [Sch16], used by [Mül19].

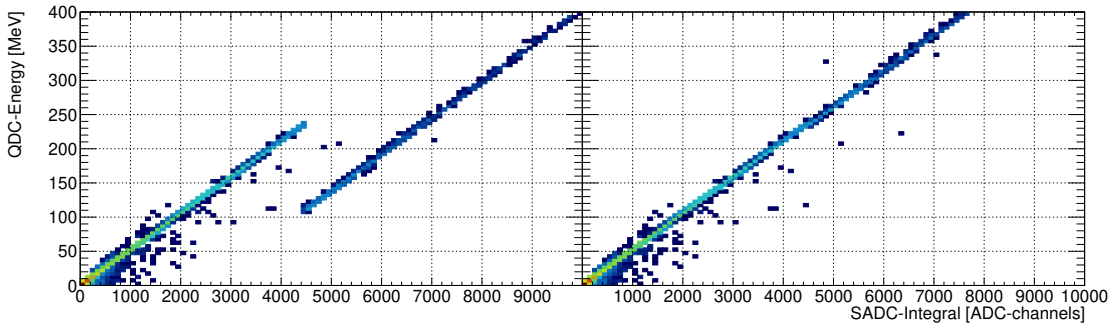


Figure A.5: Calibrated QDC energy versus uncalibrated SADC integral values for channel 40. Left: with original high-range factors. Right: After fixing.

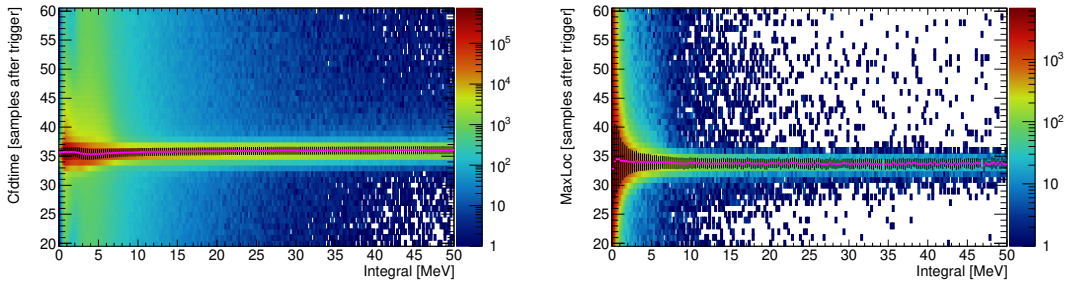


Figure A.6: Timewalk of CFD zero-crossing time (left) and location of the global maximum of FIR-filtered waveforms (from offline analysis; right) depending on the observed integral. Pink markers indicate the peakposition of a gaussian distribution; fitted in a 3σ-environment around the MPV for each vertical slice. Black bars indicate the σ-width of the fitted functions. Data from May 2018.

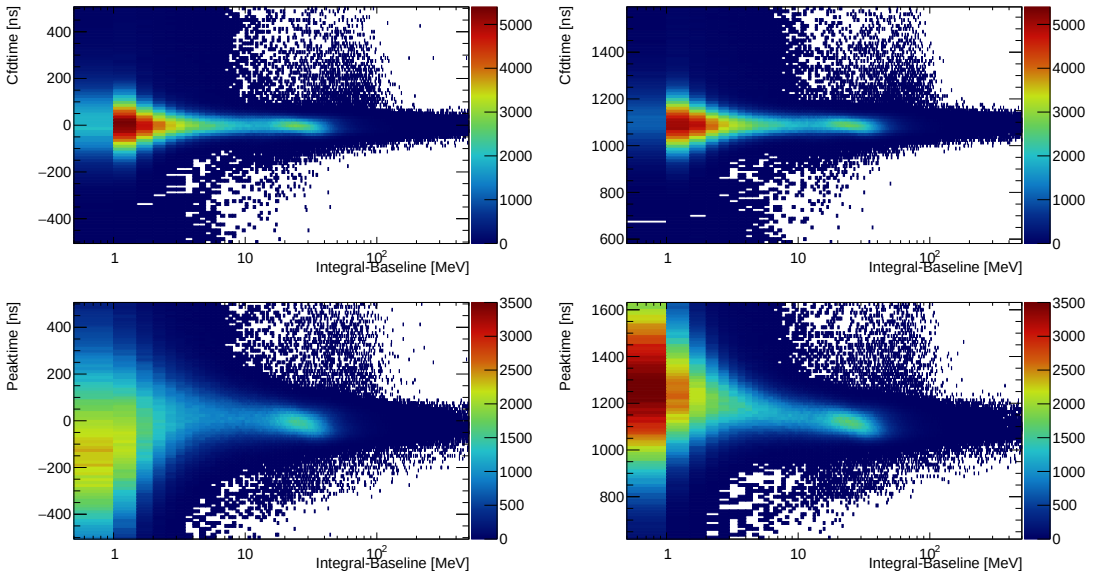


Figure A.7: Distribution of cfdtimes (top row) and peaktimes (bottom) in cosmic run from March 2022. Without FIR-filter on the left and with FIR-filter on the right (still with the previous, now incorrect, calibration).

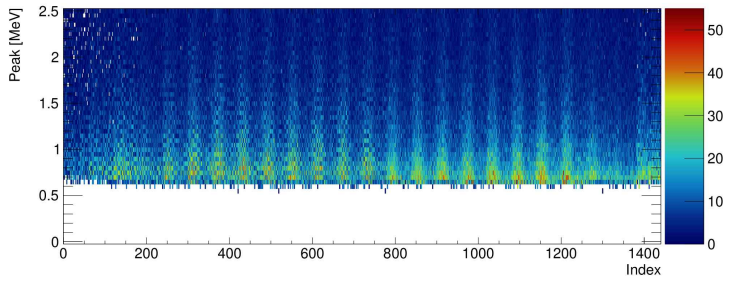


Figure A.8: Peak values (determined with an explicit height threshold) against crystal index. A number of 20 rising samples was required in both up- and downstream; in addition a minimum peak height of 20 (DS) or 40 (US) was required as well. The data were obtained from a cosmic run, which explains the seemingly oscillating structure along x.

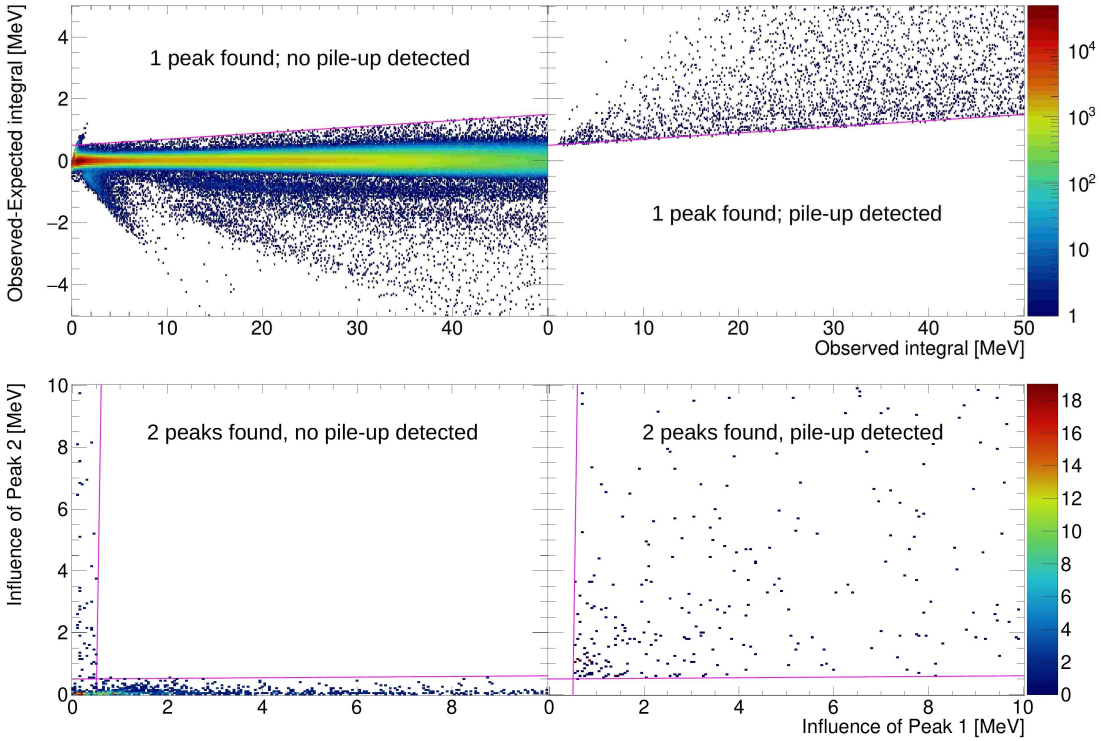


Figure A.9: Test of the new pile-up algorithm implemented in firmware. Data is from cosmic particles with inherently low pile-up probability. Thresholds are depicted in magenta.
Bottom row: If 2 peaks were detected and both have an expected integral above threshold, the waveform is classified as pile-up. If both are insignificant, it is not. If only one of the two peaks is significant, the waveform is treated as if it had just one peak to begin with (see top row).
Top row: If exactly 1 peak was significant (because only one was found or a second peak was insignificant), the difference between observed and expected integral is checked. Above a certain threshold, the waveform is classified as pile-up.
Note: The top row depicts cases in which 1 peak was found; regardless of whether it was actually significant. Insignificant peaks lead to the small area above the threshold in the top left plot, since the integral difference is explicitly not checked in such cases. Most of the observed integral (typically below 1 to 2 MeV in such cases will be due to pulses which were not detected (or other, unknown reasons).

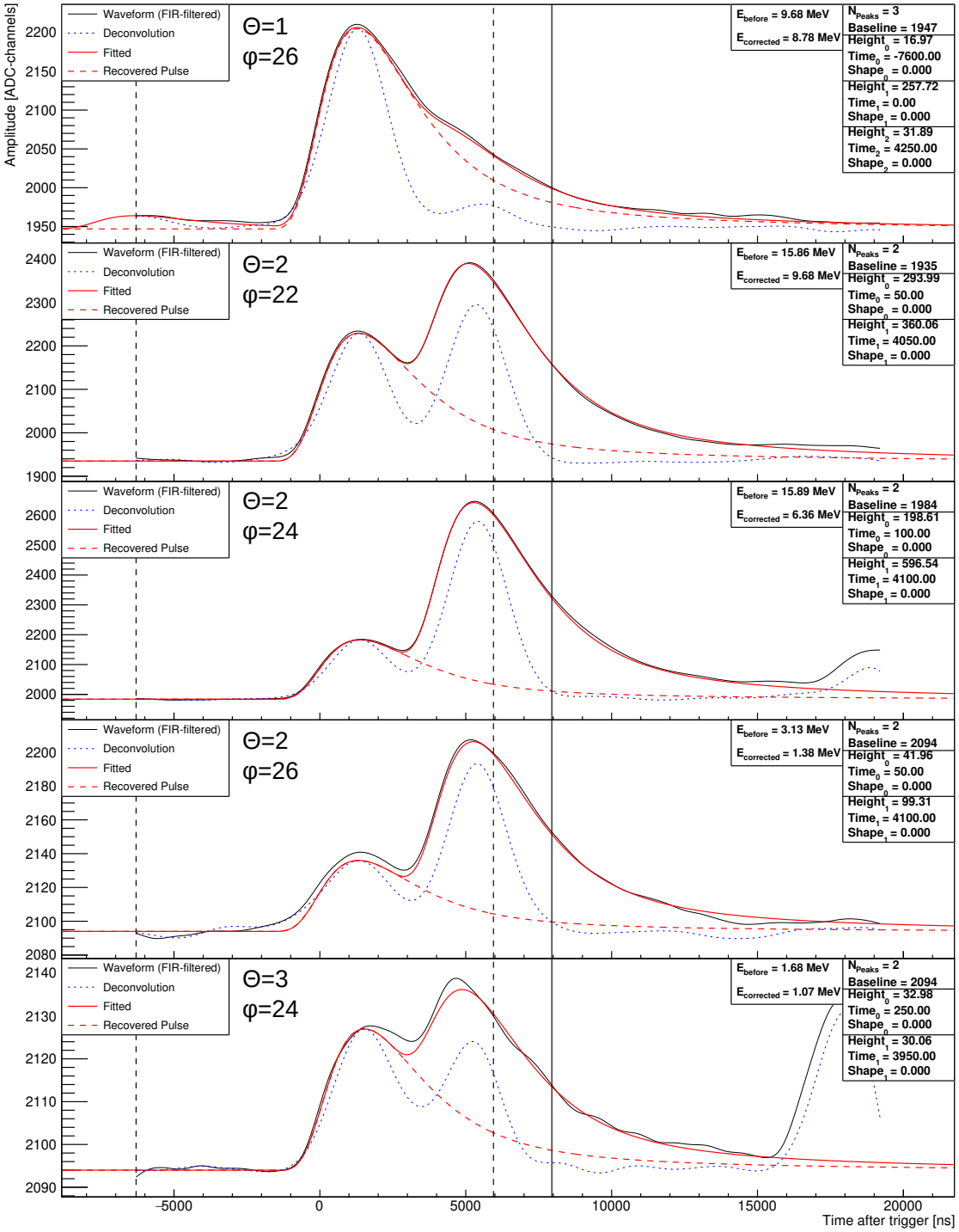


Figure A.10: Collection of (detected) pile-up waveforms in adjacent crystals from one particular event. The FIR-filtered waveforms are shown in black, blue is the deconvolution result (see section 6.6.4) and red is the result of the recovery process (see section 7). Θ and Φ refer to the particular counting by which a crystals polar and azimuthal angles are indexed. Data from June 2021 beamtime.

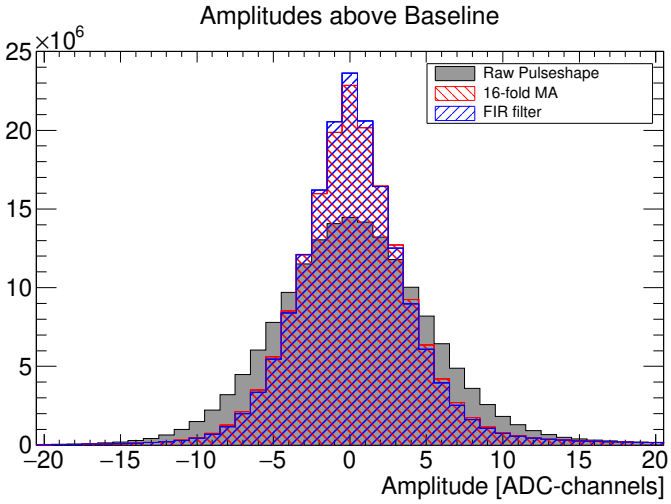


Figure A.11: Comparison of the spread of samples around the baseline. Raw values are shown in black, MA_{16} in red, FIR-filtered values in blue. Pulses of all heights were taken into account, therefore the distributions are slightly skewed towards the right. This figure is effectively a projection of the left-hand side of Fig. 6.2 onto the y-axis.

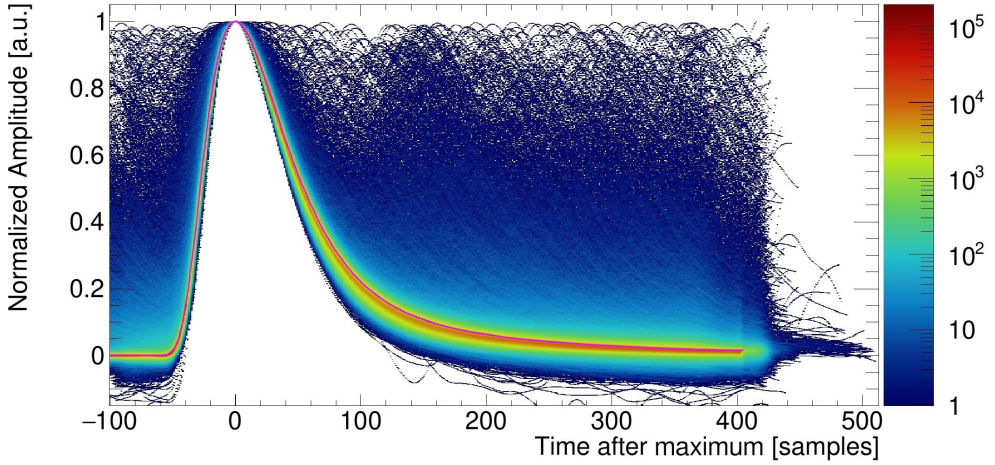
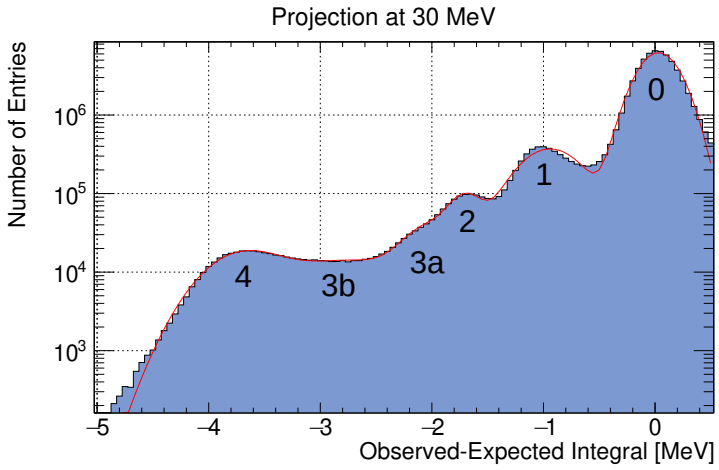


Figure A.12: Overlay of pulses from the lower area of Fig. 6.10 (FIR-filtered, normalized and shifted). The mean pulseshape extracted from all pulses (see section 6.3) is shown in magenta.



Region	Amplitude	μ [MeV]	σ [MeV]	Integrated number of entries
0	6.33×10^6	0.03	0.179	5.68×10^7
1	3.68×10^5	-0.94	0.288	5.31×10^6
2	7.36×10^4	-1.69	0.132	4.87×10^5
3a	3.63×10^4	-1.98	0.222	4.04×10^5
3b	1.38×10^4	-2.67	0.453	3.13×10^5
4	1.75×10^4	-3.66	0.353	3.10×10^5

Figure A.13: Difference of observed and expected integral at an observed energy of 29 to 31 MeV. A function consisting of 6 independent gaussians was fitted to the data in order to roughly estimate the respective positions and number of entries. The respective peaks are described by the parameters given in the table. Although barely visible as enhancement in the data, a sixth peak (3b) was necessary to get a satisfactory description of the data.

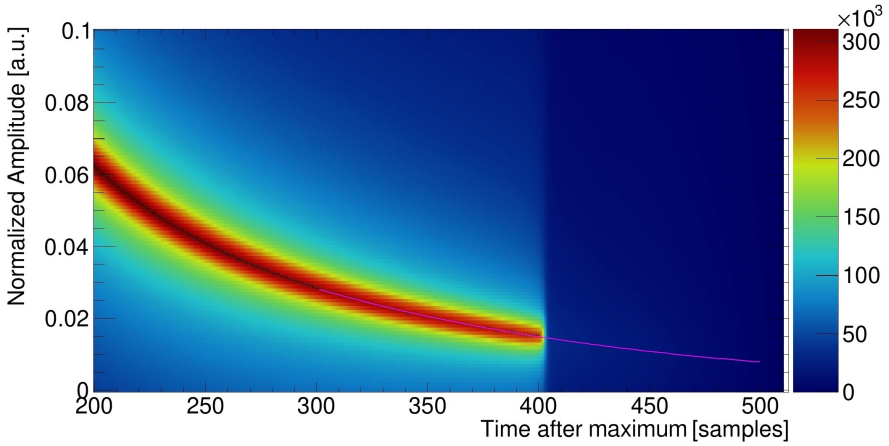


Figure A.14: Enhanced view of the fitted region in Fig. 6.5. The fit ranges from x = 300 to 400 samples, but was drawn up until sample no. 500.

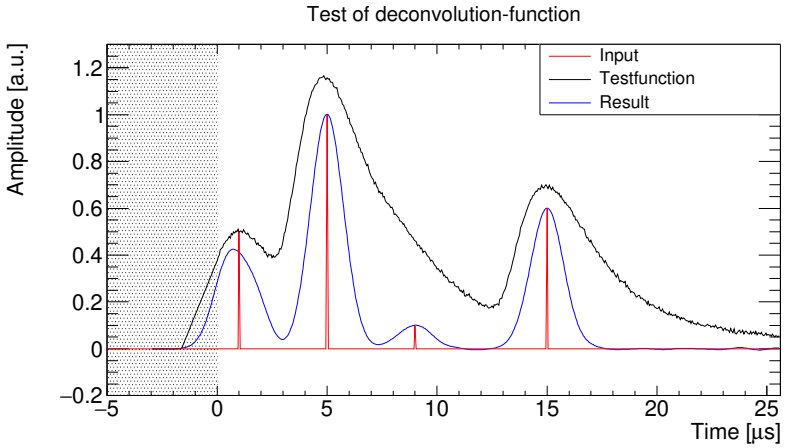


Figure A.15: Performance of the custom deconvolution in a modeled pile-up situation with an extrapolated extension in front of the waveform (shadowed area).

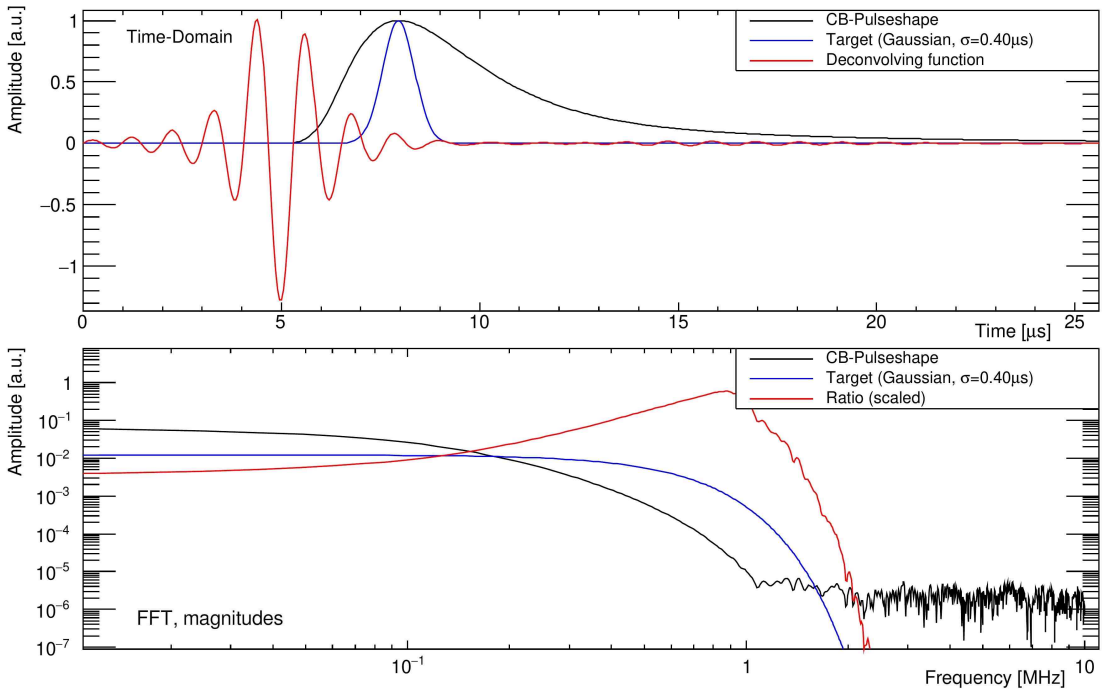


Figure A.16: Results of the calculation of a deconvolving function (similar to Fig. 6.21) if the width of the target gaussian is chosen too narrow.

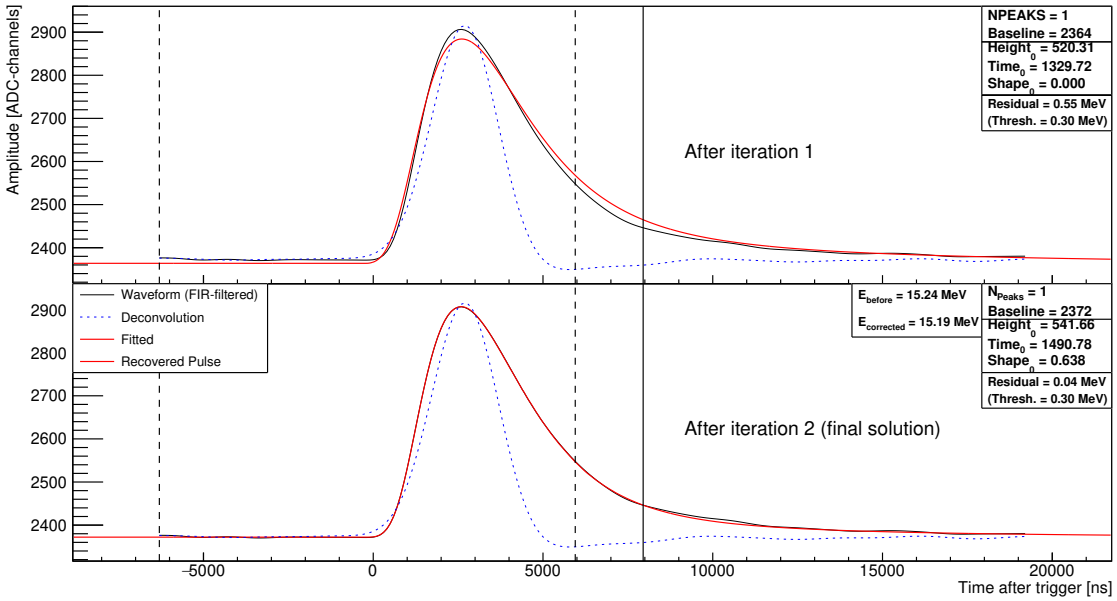


Figure A.17: Waveform which was not satisfactorily described after iteration 1 and required the shape parameter to be released in order to be well described. Since only one pulse was present, the recovered energy for this "pile-up" ($E_{\text{corrected}}$) is very close to the initial value; the small difference comes from a slightly more accurate baseline estimate after fitting.

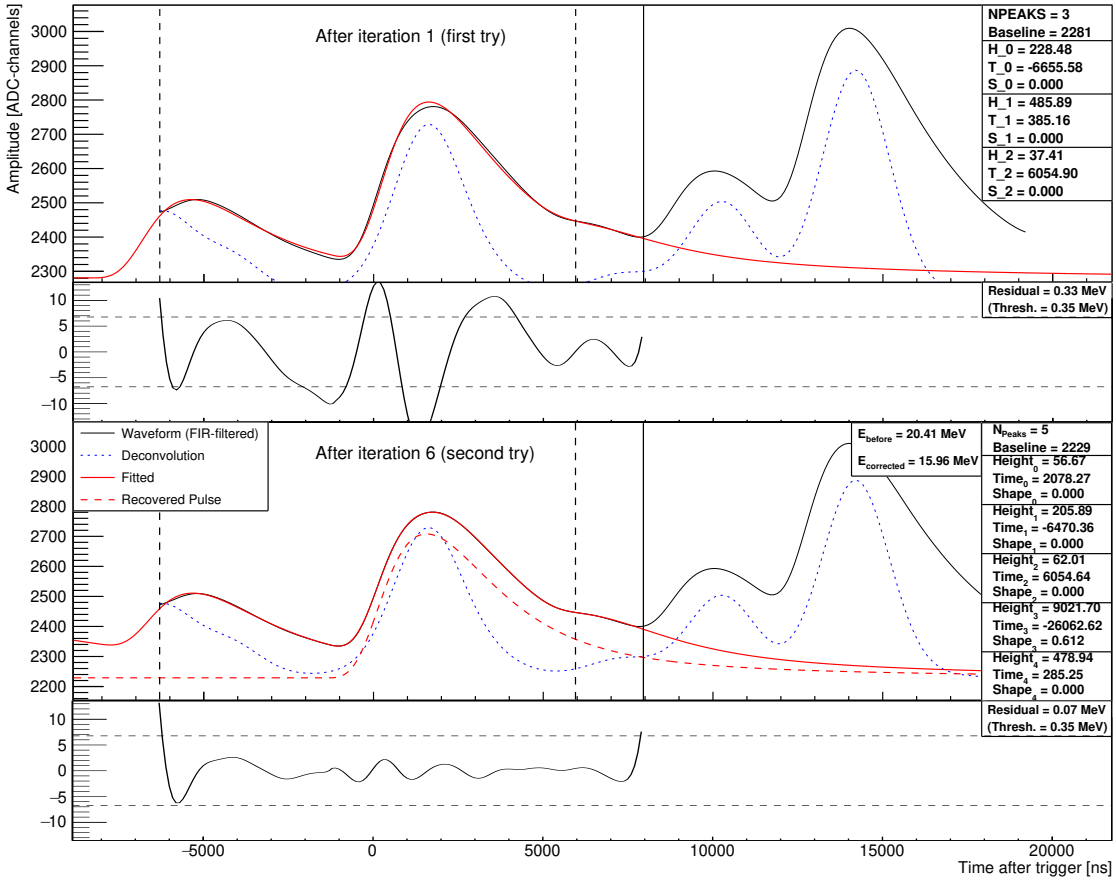


Figure A.18: Top: Example of a "suspicious" result of the recovery process. Bottom: Result of the second attempt at recovery, which yielded a much smaller residual value. Notice how the fit failed to resolve the two separate pulses in the middle peak in the first case and gave a significantly higher baseline value as in the latter.

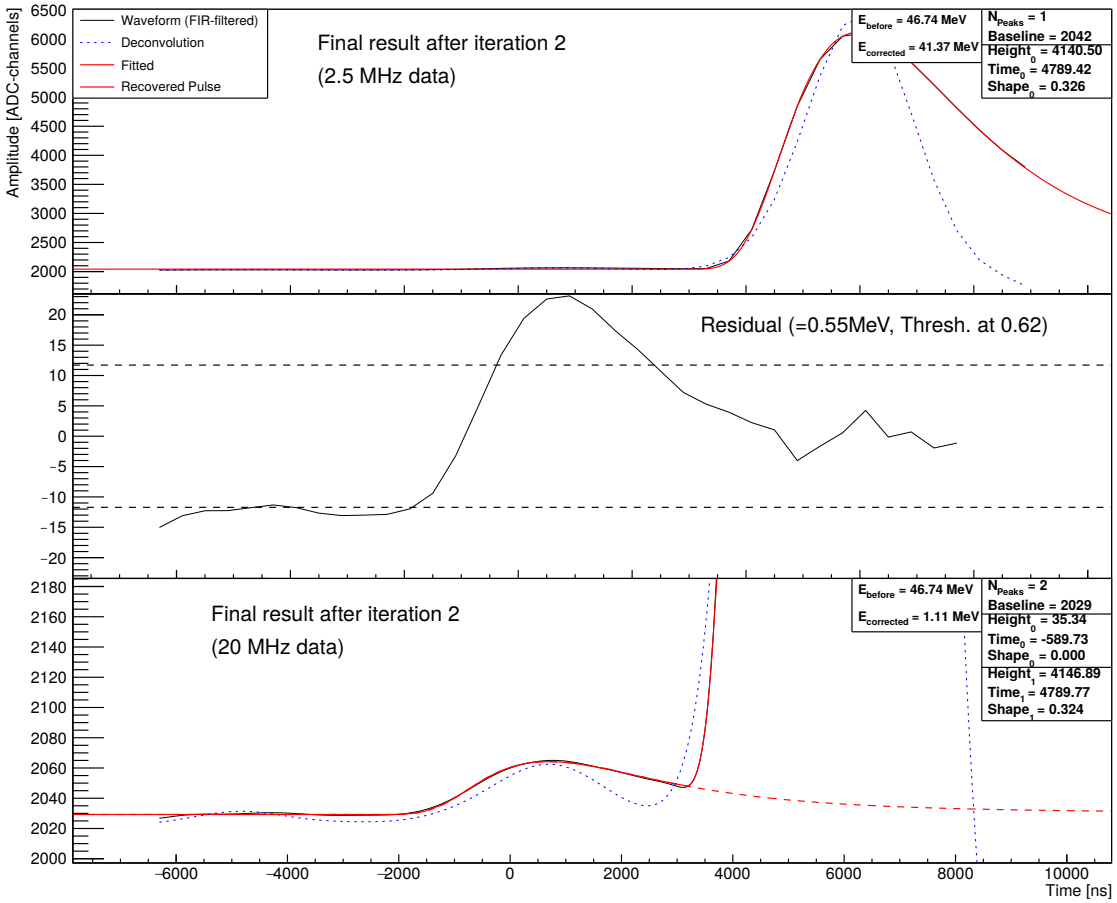


Figure A.19: (Rare) example in which using every 8th sample (top and middle) yielded a substantially different result than if every sample was used (bottom; residual of 0.05 MeV not shown for space reasons). Since the deconvolution was still the same, two peaks were initially assumed for the fit in both cases. However, since the second peak is narrower, the fit needed two iterations. In case of the reduced sampling rate, the fit in the first iteration (with S still fixed to zero), apparently favored a solution in which the small peak in the beginning was shifted into the larger peak and consequently removed. To compensate, the fit pulled up the baseline value to reach a (barely) sufficient residual (see middle). This was not the case with the 20 MHz data, where the small peak stayed at approximately the same position throughout both iterations. It should be noted, that since the "lost" peak is clearly visible in the residual, it would have been re-inserted in iteration no. 3, had the residual not been sufficient after iteration 2 already. As a consequence, the much larger, albeit clearly displaced right peak was identified as "main peak" (since it is the only one) in the first case. This causes a very large discrepancy between the features of both "main peaks" (e.g. in Fig. 7.7), although the overall compatibility of the fit with the data is not that different.
Note: If the time of the energy deposition is used in the analysis (e.g. by *time-clustering*), the first case is clearly identifiable as uncorrelated to the trigger. Since the main peak in the bottom could (depending on the cutoff) be considered to be correlated to the trigger, the final discrepancy would effectively only be 0 MeV (instead of 41.37 MeV) vs. 1.11 MeV.

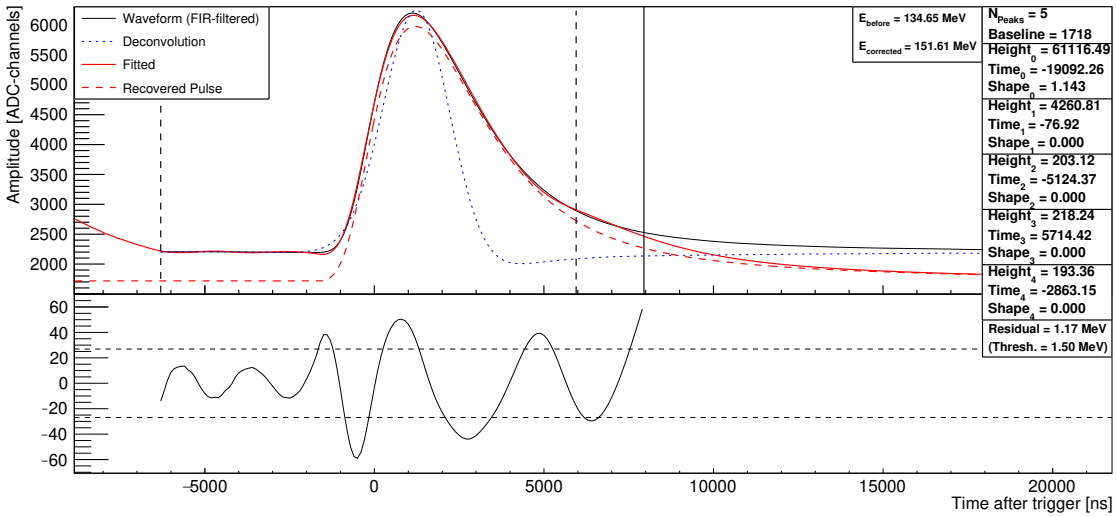


Figure A.20: Example of a (rare) recovery result which result in a fitted baseline value which is significantly too low. The originally extracted baseline (startvalue) was at 2196 ADC channels, which would have been a much better description. The reason why the fit diverged significantly from this value in the first place is a tiny peak in the deconvolution function directly at the start of the sampling window which led to the assumption of not one but two pulses from the start. Because a small pile-up pulse sits on the falling edge of the large pulse (hard to see by eye) the fit favored a solution in which the first pulse was shifted very far to the left and increased its height dramatically (almost to the maximum value of 65536) while the baseline decreased and the height of the other (main) pulse was increased as well. In addition to a third pulse which was eventually added to account for that small pile-up in the end, the falling nature of f(t) in front of the main pulse (due to the presumed pulse far in the past) required two more peaks to be inserted in front of the large pulse. Only this way could a somewhat "flat" area and a technically (barely) satisfactory residual be achieved.

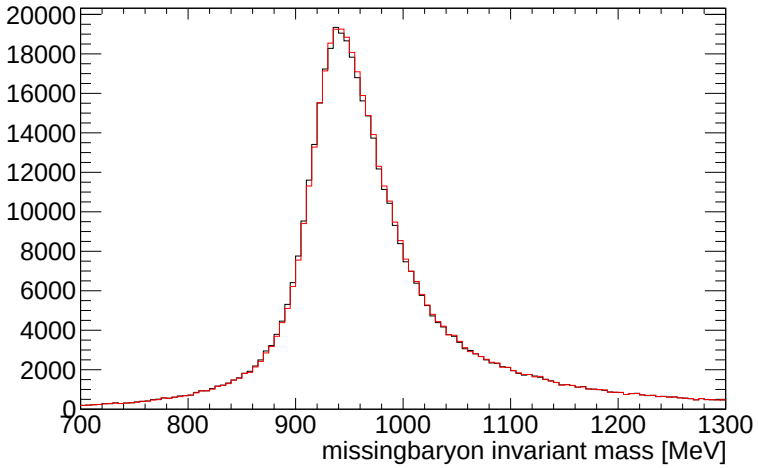


Figure A.21: Distribution of missing masses after all other cuts were applied. Black shows the result for uncorrected data, red for pile-up corrected data.

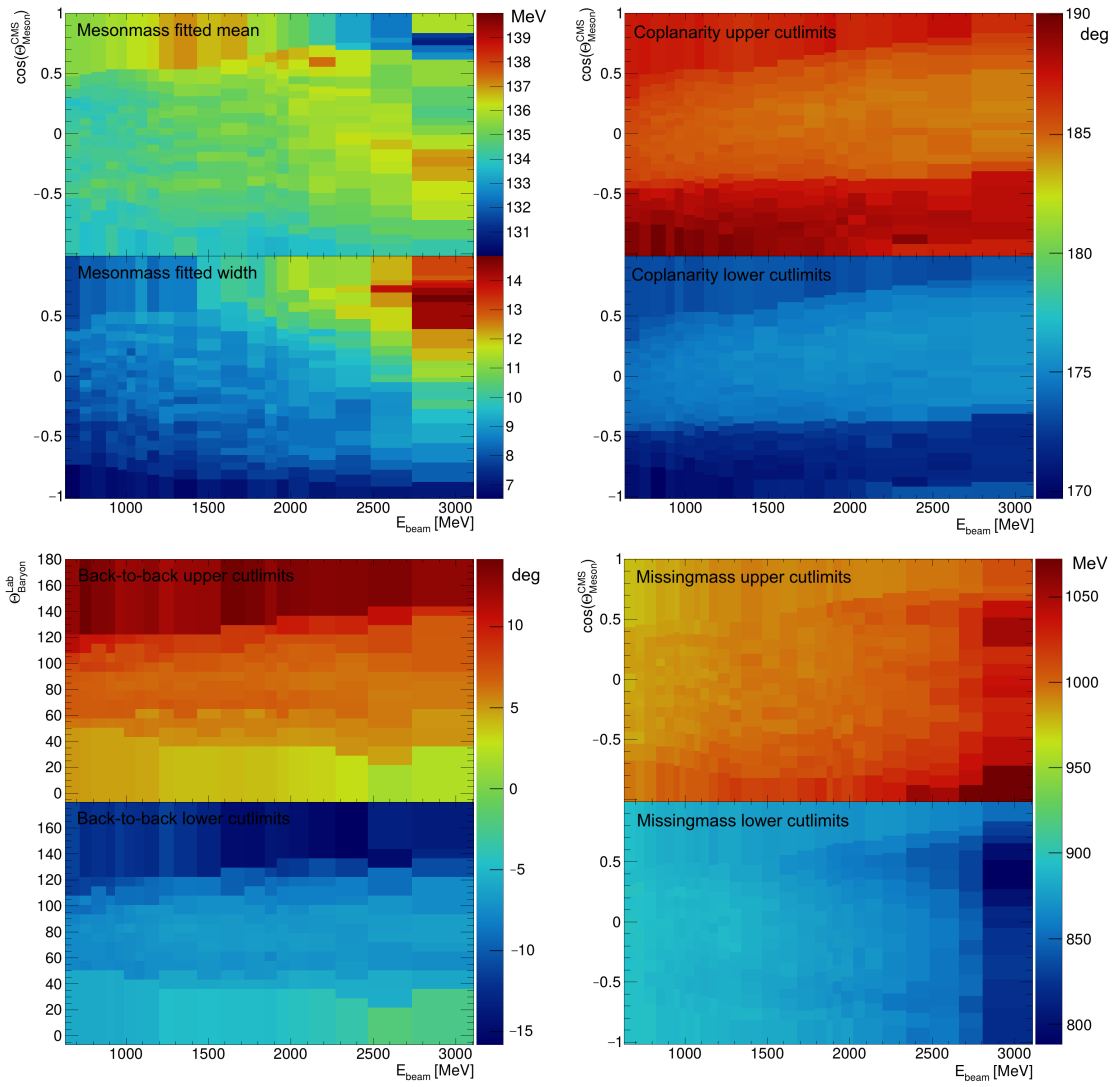


Figure A.22: Cutlimits used for the reconstruction of physical events in section 7.4.1.1. The diagrams were ordered corresponding to Fig. 7.10, with the exception of the top left, where the mean and width of the fitted peak (gaussian) for the π^0 spectrum is shown. The phase space was subdivided in E_{beam} and $\cos\left(\theta^{\text{CMS}}_{\text{Meson}}\right)$ (except for the back-to-back cuts in which $\Theta^{\text{Lab}}_{\text{Baryon}}$ was used as y-axis instead) with 20 bins each, resulting in 400 bins in total. Bin limits were chosen such that every bin contains (approximately) the same number of entries.
The cutlimits were fine-tuned over several iterations for a different analysis, but on exactly the same data set as the analysis in section 7.4.1.1 (where only a small subset of the available statistics was used).