Department of Physics and Astronomy

University of Heidelberg

Master thesis

in Physics

submitted by

Panagiota Chatzidaki

born in Heraklion, Greece

2020

Common-mode effect and ion-tail analysis of the GEM upgrade of the ALICE TPC at CERN

This Master thesis has been carried out by Panagiota Chatzidaki

at the

Physikalisches Institut Heidelberg

under the supervision of

Priv.-Doz. Dr. Yvonne Pachmayer

“Common Mode” und “Ion Tail” Effekt Analyse des GEM Upgrades der ALICE TPC am CERN

Um die ALICE Zeitprojektionskammer (TPC) kontinuierlich auszulesen, d.h. die komplette minimum bias Interaktionsrate von 50 kHz in Pb-Pb Kollisionen, so wie in LHC Run 3 und 4 antizipiert, aufzunehmen, wurden ihre Auslesekammern, basierend auf Vieldraht-Proportionalkammern (MWPC), durch Gas Electron Multipliers (GEMs) ersetzt. Bevor die TPC im Experiment installiert wurde, wurde die vorläufige Inbetriebnahme im Reinraum am LHC Point 2 von November 2019 bis August 2020 durchgeführt.
Das Ansprechverhalten der Pads zeigt eine charakteristische Unterschwingung gleichzeitig mit den Signalen, die von Tracks geladener Teilchen stammen, hervorgerufen durch das kapazitive Verhalten zwischen der Padebene und den GEM-Folien. Nach solchen Signalen beobachtet man eine Überschwingung, ausgelöst durch langsam von der Padebene wegdriftende Ionen. Diese zwei Effekte, genannt der "Common-Mode"-Effekt und "Ion-Tail", benötigen eine Online-Korrektur, und zwar in den FPGA-basierten Common-Readout-Units (CRU's), um die Performance des Detektors bzgl. Teilchenidentifikation und Spurrekonstruktion nicht zu beeinträchtigen.
Diese Arbeit beschreibt die Eigenschaften und Korrektur des Common-Mode-Effekts und Ion-Tails in der GEM-basierten TPC, mit Daten, die während der vorläufigen Inbetriebnahme mit einem Laserkalibrationssystem aufgenommen wurden. Die Abhängigkeiten der Effekte werden mit Machine-Learning-Methoden untersucht und verstanden, und Online-Korrektur-Modelle werden entwickelt und getestet.

Common-mode effect and ion-tail analysis of the GEM upgrade of the ALICE TPC at CERN

To operate the ALICE Time Projection Chamber (TPC) in continuous mode, recording the full minimum bias interaction rate of 50 kHz in Pb–Pb collisions as anticipated in LHC Run 3 and 4, its gated Multi-Wire Proportional Chamber-based readout (MWPC) was replaced by Gas Electron Multipliers (GEMs). Before installation of the TPC underground, pre-commissioning took place in the clean-room at LHC Point 2 during November 2019–August 2020.
The pad responses in the TPC show a characteristic undershoot simultaneously with the signals caused by charged particle tracks, due to capacitive coupling between the pad-plane and the GEM foils. At the same time, after the signal pulse, a long overshoot is observed, caused by the slow movement of ions away from the pad-plane. These two effects, referred to as the “common-mode” and “ion tail”, respectively, require an online correction, namely in the FPGA-based Common Readout Units (CRUs), in order to maintain the Particle IDentification (PID) and tracking performance of the detector.
This work describes the properties and correction of the common-mode effect and ion tail in the GEM-based TPC, using data collected with its laser calibration system during the pre-commissioning phase. The dependencies are investigated and understood using Machine Learning techniques, and online correction models are developed and tested.

Contents

1 Introduction

The *Large Hadron Collider* (LHC) is the final acceleration ring of the CERN accelerator complex. It is designed to collide counter-rotating beams of protons or heavy ions at the center of its four main experiments. To date, it holds the record for the highest beam energy in proton–proton collisions (6.5 TeV) as well as the highest instantaneous luminosity ($2 \times 10^{34}\,\mathrm{cm^{-2}s^{-1}}$) [1].

A Large Ion Collider Experiment (ALICE) [2] is dedicated to the study of heavy-ion collisions at extreme temperatures and energy densities. During such collisions, the *Quark–Gluon Plasma* (QGP) is formed for a short period of time, which is a very dense state of deconfined quarks and gluons [3]. The understanding of the QGP is essential in revealing the properties of the strong interaction, and consequently the conditions prevailing in the very early universe [4].

The *Time Projection Chamber* (TPC) [5] is one of the principal sub-detectors of ALICE. It is a cylindrical gaseous detector mainly providing tracking, momentum and $\mathrm{d}E/\mathrm{d}x$ measurements. The ionization electrons generated by incoming particles are detected at the readout chambers which are mounted on both its end-plates. Until the end of LHC Run 2 (2015–2018), these consisted of *Multi-Wire Proportional Chambers* (MWPCs). Ions produced during the signal amplification in the readout chambers were successfully blocked by the *gating grid*, thus preventing space-charge distortions in the drift region. The space-charge distortions lead to deterioration of the tracking and $\mathrm{d}E/\mathrm{d}x$ resolution of the TPC.

To facilitate its second major upgrade, the LHC is currently shut down until 2021. Starting from LHC Run 3, the luminosity in Pb–Pb collisions will be increased by a factor of about six, resulting in a minimum bias collision rate increase to 50 kHz, in accordance with the physics goals set by ALICE [6]. At the same time, the physics programme of ALICE requires recording of all minimum bias events, necessitating the integration from a triggered readout to a continuous one. For the above reasons, several components of ALICE required severe modifications or complete replacements.

The usage of a gating grid in the MWPC-based TPC limits its operation to a maximum collision rate of about 3 kHz. On the other hand, non-gated operation would result in prohibitively large space-charge distortions. Consequently, alternative readout technologies with intrinsic ion backflow capabilities were explored. The *Gas Electron Multiplier* (GEM) detectors [7], having intrinsic ion backflow and also high rate capabilities, were a natural choice. The final design of the ALICE TPC GEMs was a result of years of extensive R&D [8]. To accommodate the transition to a continuous readout mode, the electronics had to be completely re-designed as well.

After the replacement of all readout chambers and front-end electronics, and

before commissioning of the TPC in the cavern, pre-commissioning took place in the clean-room at LHC Point 2 during the months November 2019–August 2020, thus ensuring the functionality of all its components. During that time, dedicated measurements were essential in gaining understanding of the upgraded TPC and developing algorithms that will perform online and offline corrections during the normal operation.

Crucial are the corrections for the common-mode effect and the ion tail, which, if not accounted for, result in significant cluster losses, as well as deterioration of the tracking and $\mathrm{d}E/\mathrm{d}x$ resolution of the TPC. The common-mode effect, which is the result of capacitive coupling between the GEM foils and the pad plane, causes a characteristic undershoot simultaneously with the signals generated by charged particle tracks in the TPC volume. At the same time, an unforeseen long tail was observed after the signal pulse, which is caused by the motion of ions away from the pad-plane. Thus, to study both effects, non zero-suppressed laser data were collected and analysed within this thesis, using the laser calibration system of the TPC.

This thesis is organized as follows: A brief introduction to the physics of the strong interaction and a description of the ALICE sub-detectors is presented in Chapter 2. There, the TPC is discussed in further detail. In Chapter 3 the TPC upgrade is motivated, and the technical design in terms of readout chambers and front-end electronics is discussed. The main procedures that took place during the pre-commissioning of the TPC are described. Furthermore, the common-mode effect and ion tail in the GEM detectors are explained, and the importance in correcting for these is highlighted. Chapter 4 describes the data-taking campaign with the laser calibration system of the TPC, and the analysis steps from raw data to the final data structures that were then used for the analysis of the two effects. In Chapter 5, the common-mode effect analysis is presented, which is divided into two steps; firstly, understanding of the effect using Machine Learning techniques, and subsequently, developing online correction models that can be programmed in the FPGA-based CRUs. In Chapter 6, the origin of the ion tail in the GEMs is addressed, and its two sources are disentangled. Finally, the proposed online correction algorithm is tested. Summary and outlook constitute Chapter 7.

2 A Large Ion Collider Experiment

A Large Ion Collider Experiment (ALICE), located at Interaction Point 2 of the CERN *Large Hadron Collider* (LHC), is the only one of the collider's four main experiments that is dedicated to study the physics of ultra-relativistic heavy-ion collisions [2]. During normal operation, the LHC accelerates protons, but for about one month every year, heavy ions (usually lead or xenon, depending on the experimental requests) are accelerated and collide at the center of the ALICE detector. Such heavy-ion collisions are the main interest of the ALICE Collaboration, since, for a very short period of time, a high-temperature and high-density medium is created, which is called the *Quark–Gluon Plasma* (QGP). The study of the QGP is essential in understanding the properties of the strong interaction. In the universe, the QGP conditions were supposed to have only been fulfilled about $10\,\mu$s after the big bang [4]. Since astronomical observations cannot probe this early phase of the universe[1], the only possible means to study the QGP is via recreation of these conditions in the laboratory. Apart from ion–ion collisions, ALICE also collects proton–ion data, as well as proton–proton data that serve as reference for the heavy-ion programme.

In this chapter, the properties of the strong interaction that govern the QGP are firstly introduced (Section 2.1). In Section 2.2 the main components of the ALICE detector are briefly presented, while in Section 2.3 the *Time Projection Chamber* (TPC), of interest for this thesis, is described in more detail.

2.1 Quantum Chromodynamics

In the *Standard Model* (SM) of particle physics, all interactions between particles are governed by either the electromagnetic, weak or strong interaction. The SM has undergone extensive experimental tests, and in some cases has even demonstrated its predictive powers (for example, the Higgs boson, postulated in the sixties, was eventually discovered in 2012). Despite its many successes, the Standard Model is not a full theory of all fundamental interactions, since, among some caveats, it fails to incorporate gravity, and cannot describe dark matter and dark energy.

Quantum-Chromo Dynamics (QCD) is the part of the SM which describes the strong interaction of particles and is based on the SU(3) local gauge group. In analogy to the electromagnetic interaction, where particles interact via their electric charge, in the strong interaction particles interact via their *color charge*. Each quark therefore carries one color (red, green or blue for quarks, anti-red, anti-green and

[1] To date, the earliest detected signal is the cosmic microwave background, which dates back to a few thousand years after the big bang.

anti-blue for anti-quarks). A system of quarks must obey the Fermi–Dirac statistics, which dictates that the wave-function of the system must be anti-symmetric under the exchange of two quarks. Since some baryons consist of three identical quarks (for example the Δ^{++}=(uuu) baryon), the additional quantum number of color is required in order to ensure that the total wave-function is anti-symmetric. In the language of group theory, QCD is invariant under SU(3) transformations in the color space.

The strong interaction between quarks is mediated by eight massless gauge bosons, the *gluons*, which are the eight generators of the SU(3) local gauge symmetry. Since gluons carry the color charge, gluon–gluon interactions are also possible (unlike electromagnetism, where direct photon–photon interactions are forbidden, since the photon is chargeless). The leptons, as they do not participate in the strong interactions, do not carry a color charge.

In the Standard Model, quarks cannot exist freely, but are always bound to color-singlet states (either mesons or baryons). Mesons are composed of a quark and an anti-quark ($q\bar{q}$), while baryons consist of three quarks or three anti-quarks (qqq or $\bar{q}\bar{q}\bar{q}$). This property of *color confinement*, which is believed to originate by the gluon self-interactions, is indeed backed up by many fruitless experimental attempts to detect free quarks.

The theory of the SM predicts that, at extremely high energies, the coupling of the strong interaction α_S decreases, and quarks start to behave asymptotically as free particles. This behavior, known as *asymptotic freedom*, has been experimentally verified [9]. In such extremely high energies, and in particular for densities of at least 1 GeV/fm^3, the quark–gluon plasma is formed [3], which is classified as a different state of matter. Under these conditions, the valence quarks of a nucleon cannot any longer be associated with that particular nucleon, and the quarks can move freely over distances comparable to the system size.

2.2 The ALICE detector

The ALICE detector, sketched in Figure 2.1, with overall dimensions 16×16×26 m^3, consists of 18 different sub-detectors [2, 10]. ALICE is particularly designed to cope with the very high track multiplicities that occur in ion–ion collisions, which can be up to three orders of magnitude higher than in pp collisions. ALICE has to date collected heavy–ion collision data in LHC Run 1 (2009–2013) and Run 2 (2015–2018). Since in Run 3 (2021–2024) the interaction rates expected are considerably higher, the detector is currently undergoing a major upgrade programme. The motivation and consequences of the upgrade strategy are described in Chapter 3.

ALICE consists of a central part, which is placed inside a large solenoid magnet[2], and a forward muon spectrometer. The central part comprises of the *Inner Tracking*

[2]The magnet, which is re-used from the L3 experiment of LEP, provides a 0.5 T magnetic field along the beam direction.

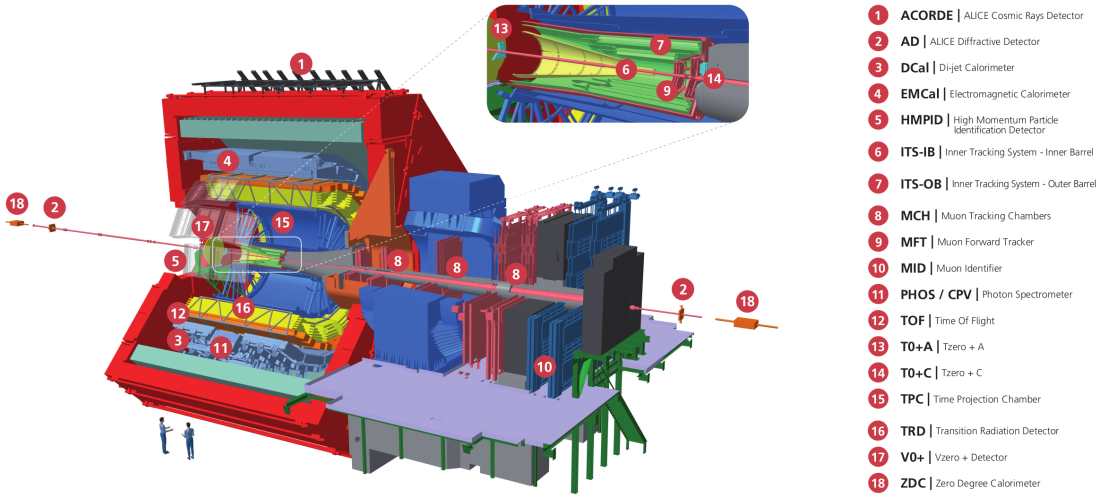


Figure 2.1: The Run 3 ALICE detector, with its components labeled. ALICE consists of a central part, placed inside a large solenoid magnet, drawn with red color, as well as the forward muon spectrometer. The sub-detectors are explained in the text. Figure taken from [11].

System (ITS), the *Time Projection Chamber* (TPC), the *Transition Radiation Detector* (TRD) and the *Time-Of-Flight* (TOF) detector, which cover at full acceptance the range of pseudorapidity $|\eta| < 0.9$, as well as single-arm detectors with a smaller azimuthal and polar acceptance. The forward muon spectrometer is located at backward pseudorapidity $-4.0 \leq \eta \leq -2.5$ and is equipped with a dipole magnet, providing a field integral of 3 T·m. Combining the ALICE detectors, all *Particle Identification* (PID) techniques are employed: $\mathrm{d}E/\mathrm{d}x$ measurement, electromagnetic calorimetry, Time-of-Flight (TOF) measurement, transition radiation, Cherenkov radiation, muon filters, as well as topological decay reconstruction. Below, each sub-detector, as will be used in the Run 3 setup, is shortly described.

Inner Tracking System

The *Inner Tracking System* (ITS) [12] is the sub-detector closest to the interaction point. It is a set of silicon detectors crucial for the determination of the primary vertex, as well as for the measurement of the impact parameter of the secondary tracks. For Run 3, the ITS will consist of a 7-layer detector employing the *Monolithic Active Pixel Sensor* (MAPS) technology and will be read out in a continuous mode. For the total of seven detector layers, the material budget is kept as low as 1.1% X_0.

Time Projection Chamber

The *Time Projection Chamber* (TPC) [5] is a gaseous detector with an active area of $\approx 88\,\mathrm{m}^3$. It is the main tracking and PID detector of ALICE, also complementing the vertex determination (together with the other central barrel sub-detectors) and providing momentum measurements with a good two-track separation. The TPC is described in further detail in Section 2.3.

Transition Radiation Detector

The *Transition Radiation Detector* (TRD) [13] is a gaseous detector that provides electron identification above 1 GeV/c. It is additionally used for global tracking. Each TRD module is divided into a 48 mm radiator and 30 mm Xe–CO_2 (85–15) drift region, and is read out via *Multi-Wire Proportional Chambers* (MWPCs).

Time-of-Flight Detector

The *Time-Of-Flight* (TOF) detector [14] encloses the TRD detector. By making use of the *Multi-gap Resistive-Plate Chamber* (MRPC) technology, it complements the identification of pions, kaons and protons in the momentum range between 0.5 and 2.5 GeV/c. The "start time" for the TOF measurement is provided by the *Fast Interaction Trigger* (FIT) detector, discussed below. With an overall time resolution of 80 ps, it is possible to separate pions from kaons and kaons from protons for track momenta up to 2.5 GeV/c and 4 GeV/c, respectively.

High-Momentum Particle Identification Detector

The *High-Momentum Particle Identification Detector* (HMPID) [15] is a *Ring Imaging CHerenkov* (RICH) detector, covering about 5% of the central barrel acceptance. It supplements the ITS, TPC and TOF in terms of PID in the momentum range of 1 to 5 GeV/c.

Calorimeters

EMCal: The *ElectroMagnetic Calorimeter* (EMCal) [16] is a sampling calorimeter which enhances the jet energy resolution, as well as the high-momentum photon and electron measurements. It is composed of alternating lead and polystyrene scintillator layers, and is placed about 5 m from the beam pipe, covering a pseudo-rapidity range of $|\eta| \leq 0.7$.

DCal: The *Di-jet Calorimeter* (DCal) [17], which will be employed from Run 3 onward, enhances the EMCal performance by allowing for back-to-back correlation measurements. The EMCal and DCal form a two-arm electromagnetic spectrometer.

PHOS: The *Photon Spectrometer* (PHOS) [18] is a lead–tungstate calorimeter dedicated to the measurement of central rapidity photons, π^0 and η mesons.

ZDC: The *Zero-Degree Calorimeter* (ZDC) [19] is a Quartz calorimeter providing the event-by-event determination of the collision centrality. In addition, it can be used as a fast trigger.

Multiplicity detectors

FIT: The *Fast Interaction Trigger* (FIT) [20] detector will replace the detectors *Time-0* (T0), *Vertex-0* (V0) and *Forward Multiplicity Detector* (FMD) that were employed during LHC Run 1 and 2 by ALICE, providing their original functionality with a simultaneous improvement in the performance. FIT will provide a minimum

bias trigger for the central barrel detectors with an efficiency higher than 90% for proton–proton collisions, as well as monitoring of the LHC luminosity (thus replacing the V0 detector). Additionally, it will determine the collision time with a resolution better than 50 ps, which will be used as the "start time" signal for the TOF measurement (thus replacing the T0 detector). Finally, it will be used to measure the event multiplicity (thus replacing the FMD). FIT consists of two arrays of Cherenkov radiators and a single large-diameter scintillator ring.

Di-muon Forward Spectrometer
The *Di-muon Forward Spectrometer* [21] is a detector covering the pseudorapidity range of $2.4 \leq \eta \leq 4$, dedicated to identify the heavy quark vector mesons (J/Ψ, Ψ', Υ, Υ', Υ'') via their di-muon decay channel.

2.3 The ALICE Time Projection Chamber

The TPC is the main tracking device of the central barrel, which, in combination with the other central barrel detectors, can provide momentum measurements with a good two-track separation, particle identification and vertex determination. It is cylindrical in shape, with its active volume ranging from an inner radius of 85 cm to an outer radius of 250 cm, and an overall length along the beam direction of 5 m. Filled with $\approx 88\,\mathrm{m}^3$ of gas, it is the largest gas-filled TPC in the world to date. Both end-plates of the TPC are equipped with readout chambers, where the signal is detected. The TPC is divided into two halves by the *Central Electrode* (CE), made of a stretched aluminised mylar foil. A schematic drawing of the TPC is shown in Figure 2.2, where the basic parts are labeled. The TPC is designed to provide a $\mathrm{d}E/\mathrm{d}x$ resolution better than 8% in Pb–Pb collisions[3], and a capability of separating tracks with a momentum difference less than 5 MeV [5]. Additionally, it is designed to provide a relative transverse momentum resolution better than 1% for momenta below 2 GeV/c, and less than 2.5% for momenta of 4 GeV/c.

When a charged particle traverses the active volume of the TPC, it ionizes the gas along its track. The produced electrons drift towards the end-plates due to the electric field which is generated by applying a voltage difference along the beam direction. The drifting electrons are then detected at the readout chambers which are mounted on the end-plates, therefore providing the projection of the track along the global x–y plane (see A.1 of the Appendix for the definition of the global coordinate system). The z-coordinate is determined by measuring the drift time of the ionization electrons. Combining the above information, the origin of ionization can be determined, resulting in a three-dimensional reconstruction of the particle tracks. The physics governing the ionization process can be found in B.1 of the Appendix. The drift properties as well as the diffusion effect are explained in B.2.

A momentum measurement can be obtained by placing the TPC in a magnetic field parallel to the aforementioned electric field. Then, the bending radius of

[3] In pp collisions, even a 5% $\mathrm{d}E/\mathrm{d}x$ resolution is attainable [22].

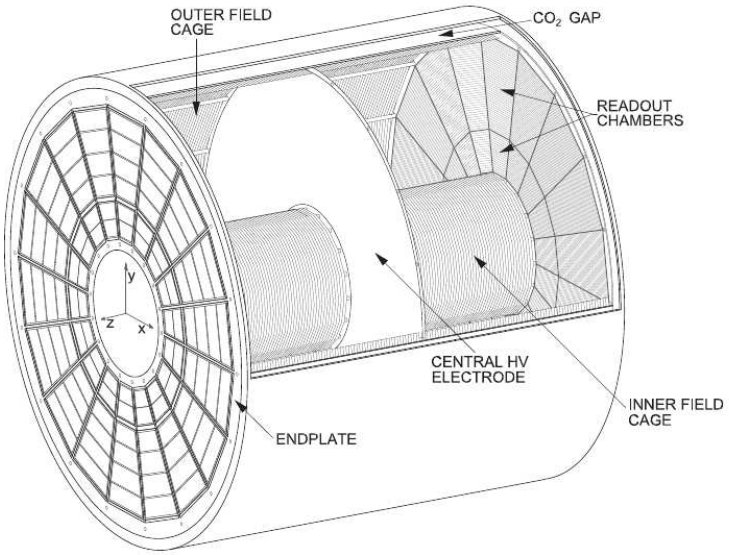


Figure 2.2: Layout of the ALICE TPC. Visible are the inner and outer field cage, the central electrode dividing the TPC into two halves as well as the end-plates, where the readout chambers are mounted on. The global reference system is also visible [23].

the particle trajectory in the x–y plane provides the transverse momentum of the particle. The radius is calculated by using the three-dimensional hit information with up to 159 (152) points in LHC Run 1 and 2 (in Run 3), equal to the number of pad rows (see Subsection 3.2.1 for the definition of a pad row). Overall, a dynamic range from tens of MeV/c to over 100 GeV/c is covered. At the same time, the mean energy loss per track length, $\langle \mathrm{d}E/\mathrm{d}x \rangle$, can be measured[4]. Combining the momentum information with the $\langle \mathrm{d}E/\mathrm{d}x \rangle$ allows for particle identification.

A uniform electric field in the TPC active volume (and therefore a homogeneous electrons drift velocity) is ensured by the usage of the *field cage*, which is a set of strips placed along the z-axis, as shown in Figure 2.2. In particular the central electrode, located at $z = 0$, is kept at a voltage of -100 kV, which results in a rather high drift field of ≈ 400 V/cm. The potential is degraded via the field strips in such a way that at the center of each strip its value is equal to the nominal one.

2.3.1 Readout in LHC Run 1 and 2

On each side of the TPC, 18 inner (IROCs) and 18 outer (OROCs) readout chambers are mounted. These, during LHC Run 1 and 2, consisted of *Multi-Wire Proportional Chambers* (MWPCs) with a segmented cathode pad.

The basic MWPC geometry is depicted in Figure 2.3. The chambers consist of

[4] In practice, the mean of 60% of the smallest signals is used as an estimator for the $\langle \mathrm{d}E/\mathrm{d}x \rangle$. Refer to Appendix B.1 for details.

a pad-plane and three wire planes, named *anode plane*, *cathode plane* and *gating grid*, the usage of the latter is justified below. The cathode wires and the pad-plane are at ground potential, while the anode wires at positive potential. The electrons produced in the TPC active volume by ionizing particles drift towards the end-plates. At the vicinity of an anode wire, the electric field is very high and therefore electron multiplication takes place. The signal induced by these avalanche electrons is however too fast to be detected by the readout electronics. The ions that are simultaneously produced will slowly[5] drift either towards the cathode wires or the pads, inducing a current signal on the pads. This ion-induced signal spreads on average over three adjacent pads, which increases the spatial resolution, since the center of gravity can be used to obtain the position.

The gas ionization results in equal amounts of produced ions and electrons, most of which are generated in the amplification region. The ions, due to the applied electric field, drift backwards towards the active volume and may distort the electric field lines. The so-called *space-charge distortions* would limit the position and $\mathrm{d}E/\mathrm{d}x$ resolution and, as a consequence, the performance parameters of the TPC would be significantly deteriorated. To quantify the effect of *ion backflow*, the ion backflow factor is introduced (IB), which is defined as the number of ions in the TPC active volume per ionization electron. The IB includes also contributions from ions produced in the active volume due to primary ionizations. Ideally, for perfectly

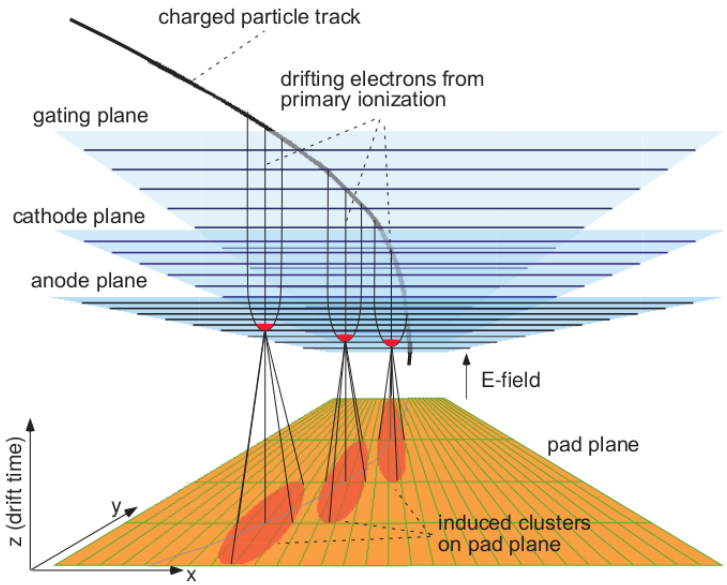


Figure 2.3: Illustration of the MWPC detection principle. Apart from the cathode plane, anode plane and the pad-plane, also the gating grid employed in the ALICE TPC is visible. The ionization clusters of the particle track are detected, after amplification, on the pad-plane [24].

[5]The ion drift velocity is roughly 2000 times smaller than the electron one [8].

working ion backflow suppression, only ions from primary ionization are found in the TPC drift volume.

For the MWPC-based TPC in Run 1 and 2, the ion backflow suppression is realized with a gating grid, which is a system of wires above the cathode plane. When the gating grid is "open", it does not influence the drift field; the grid is transparent. When it is "closed", the wires are biased with an alternating positive and negative voltage, thus preventing the ions from drifting into the drift volume. The gating grid is by default "closed", and only "opens" in case of an event trigger. The duration of the "open" mode is $\approx 90\,\mu$s, which is dictated by the time needed for an ionization electron to drift over the entire TPC length (the nominal electron drift velocity is $\approx 2.58\,\mathrm{cm}/\mu\mathrm{s}$). With this system, the resulting ion backflow suppression during Run 1 and Run 2 is 10^{-5}.

There is a fixed closure time of the gating grid of $200\,\mu$s in order to fully absorb the created ions, due to the small ion drift velocity. This total $\approx 300\,\mu$s defines the dead time of the detector, which limits the TPC readout rate in pp collisions to about 3.5 kHz.

2.3.2 Gas choice

The gas, apart from ionization and drift medium, determines the electron amplification parameters in the readout chambers. Therefore, a dedicated study was carried out in order to choose the gas mixture, since it is extremely critical for the detector performance.

As described in [25], the gas should provide a high electron drift velocity, since the electron drift time highly contributes to the dead time of the TPC. In addition, the drift velocity should ideally be saturated, i.e. it should not heavily depend on the electric field. At the same time, a small transverse and longitudinal diffusion is desired to maintain a good spatial resolution. A high Townsend coefficient with an absence of sparks will guarantee a good energy resolution and a stable operation of the TPC. A low electron attachment is needed to minimize electron losses in the drift region. A high ion mobility will quickly eliminate the space-charge in the TPC volume. A compromise in the ionization rate is necessary, since on the one hand a high number of electron–ion pairs guarantees a good spatial and $\mathrm{d}E/\mathrm{d}x$ resolution, but on the other hand ions contribute to the space-charge. Additional factors include ageing and stability, as well as inflammability, toxicity and radioactivity and of course, cost. Nowadays potent greenhouse gases are in general avoided (e.g. CF_4).

Via a process of elimination, the chosen gas mixtures for the MWPC-based TPC were either Ar–CO_2 (88−12) or Ne–CO_2–N_2 (90−10−5)[6], with the latter being used only in 2010 and 2017. Despite the fact that Ne is less stable in the LHC environment, for the upgraded TPC in Run 3 the chosen baseline mixture is Ne–CO_2–N_2 (90−10−5) since it results in a higher ion mobility compared to the Ar–

[6] Originally, a mixture of Ne–CO_2 (90−10) was used, but an addition of five parts of nitrogen was observed to contribute to the stability against discharges.

CO_2 mixture. This is essential in limiting the quite larger space-charge distortions expected in Run 3 (see Chapter 3). An unavoidable drawback of the chosen gas mixture is that the CO_2 component in the gas leads to a non-saturated drift velocity. Since neon is a quite expensive gas, a re-circulating system is employed in the TPC. As a consequence, external gases from the ambient air such as N_2, O_2 and H_2O can compromise the gas purity. To avoid the unwanted electron attachment caused by the oxygen molecules, the gas is routed to a purging station before being refilled in the TPC.

It is worth mentioning that due to the gravitational force, the gas pressure is lower and the temperature is higher at the top of the TPC. As a result, the mobility is slightly higher at the top, or equivalently the drift times are smaller.

2.3.3 The TPC laser calibration system

Pulsed UV laser beams were originally used by the STAR experiment TPC to simulate ionizing tracks inside the active volume of the detector [26]; subsequently, the method was adapted by ALICE. The obvious advantage of using laser tracks is that they are not affected by a magnetic field and are not subject to multiple scattering. During normal operation, the purpose of the laser system is to monitor the drift properties of the gas and the amplification parameters. During the pre-commissioning phase, it is additionally used to align the readout chambers, calibrate the electronics, study the field distortions and the $E \times B$ effects. In the framework of this thesis, it is used for studying the signal shapes and consequently for the correction of the common-mode effect and the ion tail (Chapter 5 and Chapter 6, respectively).

A detailed description of the laser system is given in [27]. The laser type used is a pulsed (5 ns pulse duration, 10 Hz repetition rate) Nd:YAG (1064 nm) laser, equipped with two frequency doublers, resulting in a wavelength of 266 nm. By making use of a telescope system, the beam diameter is expanded from 10 mm to about 25 mm. Close to the laser source, the laser beam has a constant intensity profile across the beam spot, while after 20−30 m it smoothly develops into a Gaussian profile. The laser is placed in a hut outside the solenoid magnet; from the hut, two beams exit, one for each TPC side. Via a system of mirrors, beam splitters and prisms, the two laser beams are guided to the entrance windows of the TPC field cage, separately on the so called A- and C-side. The beam travels along the end-plate perimeter, being reflected by a set of prisms. At certain positions along the perimeter, beam splitters direct part of the beam in the z–direction (along one of the six *laser rods*). The beam-splitter properties are chosen in such a way that in the end, the intensity of the individual beams along the laser rods are equal. *Micro-mirror bundles*, placed in each half of the TPC at $z \approx 130$, 850, 1690, 2470 mm for odd-numbered laser rods, and $z \approx 100$, 790, 1630, 2410 mm for even-numbered laser rods, reflect the beam into the TPC active volume and parallel to the x–y plane, as shown in Figure 2.4. Each of the micro-mirror bundles contains seven quartz fibres of 1 mm in diameter, cut at 45°. The effect of the bundle on the laser beam is to create a “fanning-out

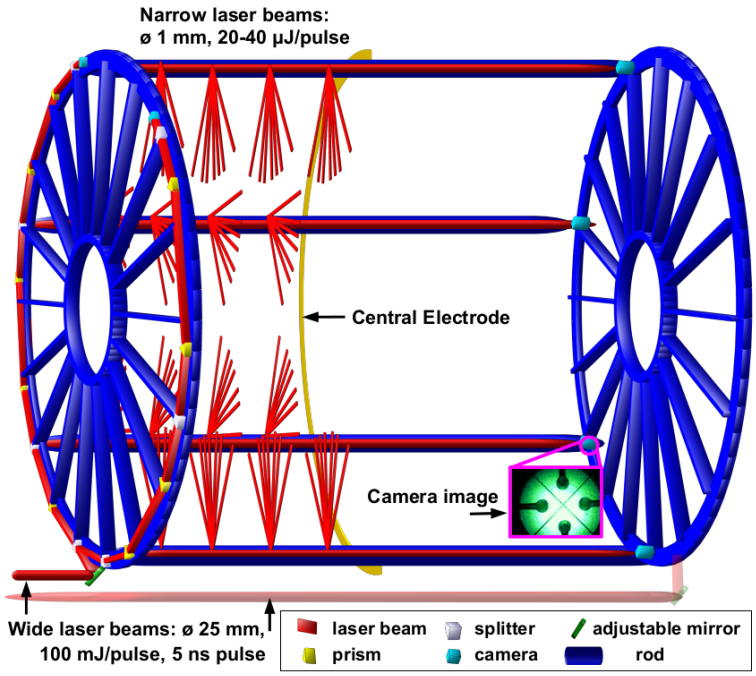


Figure 2.4: Three-dimensional drawing of the laser tracks inside the TPC active volume. Only some A-side tracks are drawn [28].

pattern" consisting of seven beams entering the TPC volume. With this system, 336 rays illuminate the TPC active volume at the same time. A projection of the laser tracks to the end-plates is shown in Figure 2.5.

The final laser wavelength of 266 nm corresponds to an energy of 4.66 eV. However, the ionization potential of the TPC gas components is much larger (Ne ≈22 eV, CO_2 ≈14 eV, N_2 ≈16 eV). This means that even two-photon ionization processes are unable to ionize the gas components, while higher-order processes are very improbable. The laser rather ionizes, with a two photon process, organic impurities in the gas with ionization potentials of 5−8 eV, which exist at a concentration of about 1 ppm. At the same time, many photons will reach the CE surface by scattering in the gas. Since the central electrode has work functions below the laser energy, these photons will generate photo-electrons simultaneously with the laser pulse.

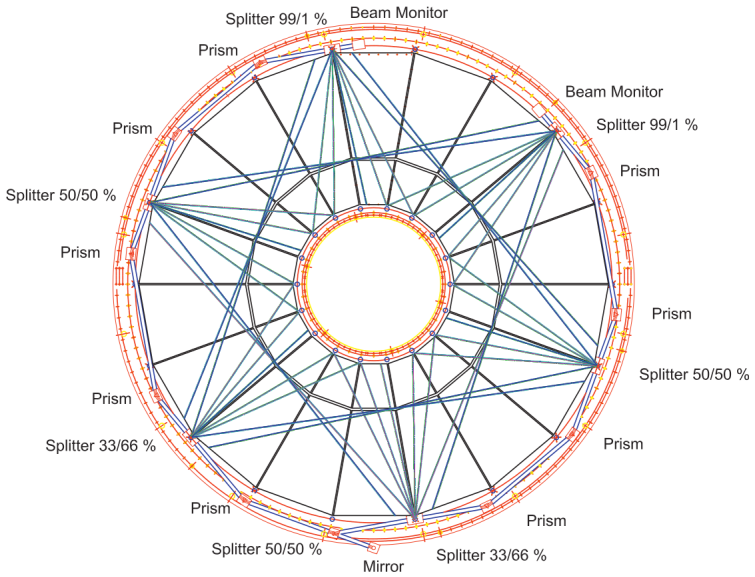


Figure 2.5: Projection of the laser tracks in the x–y plane. Each micro-mirror bundle results in seven separate laser beams entering the TPC volume [28].

3 The ALICE TPC upgrade

During the *Long Shutdown 2* (LS2), the ALICE experiment is undergoing a major upgrade programme. Starting from LHC Run 3, the minimum bias collision rate with Pb–Pb beams in ALICE will be increased to around 50 kHz, which is about a factor ten higher than the current one. This is motivated by the desired sensitivity to rare probes which are essential to the characterization of strongly interacting matter at high temperatures [29].

The rate limitations of the MWPC-based TPC at ≈ 3 kHz, mainly restricted by the gating grid closure time of 200 μs, are prohibitive for achieving the scientific goals set for Run 3. However, operation of the TPC without the gating grid would lead to excessive space-charge distortions caused by the ion backflow. At the same time, the focus of ALICE on rare probes at low momentum requires recording of all minimum bias events which will be written to tape only after compressing the data as much as possible. This implies that the TPC will be read out in a continuous mode, instead of a triggered one.

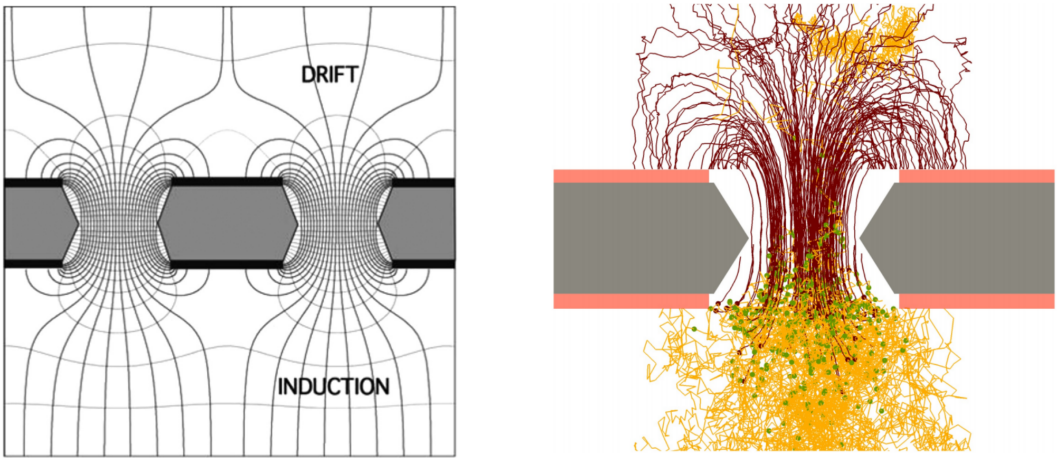
Consequently, an alternative readout technology with intrinsic ion backflow capabilities was necessary for the Run 3 TPC. Such requirements can be fulfilled by the *Gas Electron Multiplier* (GEM) detectors. To accommodate the continuous readout mode, the readout electronics had to be fully re-designed. In Section 3.1, the working principle of GEMs is explained, while in Section 3.2 the TPC technical design for Run 3, in terms of readout chambers and upgraded electronics, is described. Finally, in Subsection 3.4.1, the common-mode effect and ion tail in the GEMs are presented, and the significance in accounting for them is highlighted.

3.1 Working principle of Gas Electron Multipliers

The Gas Electron Multipliers (GEMs), invented by F. Sauli in 1996 [7], consist of a thin polyimide foil, coated on both sides with a thin metal layer (*GEM top* and *GEM bottom electrodes*) and perforated with a high density of holes. By applying a moderate potential difference between the two sides, a high electric field is created in the holes, which serve as the amplification region.

Figure 3.1a illustrates the electric field lines in the vicinity of a GEM foil. The electric field is very strong inside the holes, where electron multiplication takes place, therefore each hole acts as an individual proportional amplifier. By keeping the electric field low enough in the region designated *drift*, the electrons coming from the top simply drift towards the foil. After multiplication in the GEM holes, the amplified electrons induce signals on the pad-plane (not visible in the figure) by drifting towards it, in the *induction* region. In Figure 3.1b the amplification of two

primary electrons entering the GEM hole is shown, as simulated by Garfield [30]. The electron and ion paths are indicated with yellow and dark red, respectively, while green dots mark the ionization points. Due to much smaller diffusion, the ions follow very closely the electric field lines, therefore most of them are absorbed at the top of the GEM foil. Some ions will inevitably escape, drifting backwards towards the drift region, while at the same time some electrons will be collected at the bottom electrode, resulting in *extraction losses*. The fraction of ionization electrons that will be transferred through the GEM foil, or *transparency*, is highly affecting the energy resolution of the GEM.



(a) GEM electric field lines [31].

(b) GEM Garfield simulation [32].

Figure 3.1: (a) Electric field lines in the vicinity of a GEM foil. The multiplication occurs only inside the hole, where the field is very strong. The induction field is higher than the drift field. (b) Garfield simulation of two primary electrons entering the GEM hole. The electron and ion paths are shown with yellow and dark red, respectively, while with green dots the ionization points are shown.

The standard GEM design employs a 50 μm thick kapton foil, with a 5 μm metal coating on each side. The holes are double-conical with an inner radius of $\approx 50\,\mu$m and an outer radius of $\approx 70\,\mu$m. The distance between nearest neighbors is 140 μm.

As already mentioned, the clear advantage of the GEM technology is that most of the ions created inside the GEM holes during the avalanche process are directly absorbed at the top of the foil, thus do not drift back to the active volume. Therefore, by employing a GEM readout system for the TPC, a continuous readout without the usage of a gating grid is possible. Nonetheless, the GEM TPC is expected to have an ion backflow of $\leqslant 1\%$, which is significantly higher than the achieved 10^{-5} with the gated MWPC TPC[1]. Consequently, the space-charge distortions during Run 3 are expected to be considerably larger. In addition, another advantage of the GEMs is their high rate capability, in contrast to MWPCs, which would saturate in gain at the expected Run 3 rates.

[1] As a comparison, the MWPC without a gating grid would result in $\approx$ 8–10% ion backflow.

Another apparent difference between the two readout systems is that in the MWPCs, the ions that are produced during the amplification stage drift either towards the cathode wires or the pad-plane, inducing a mirror charge on the pads, which creates the (positive) pad signal. Since the electron multiplication occurs $\leqslant 20\,\mu$m away from the anode wires, the electrons are immediately absorbed by the anode wires and therefore only marginally contribute to the signal. In contrast to the above, in the GEMs only the (negative) signal of the electrons drifting for 2 mm from the GEM4 foil towards the pad-plane is measured, making the timing properties of the detector very appealing.

3.2 TPC technical design for LHC Run 3

3.2.1 Readout Chambers

An extensive R&D campaign was carried out in order to optimize the final geometry and applied voltages for the ALICE TPC GEMs. Stacks of four GEM foils were ultimately chosen, separated by ≈ 2 mm of transfer gap, as shown in Figure 3.2. The electrons coming from the active volume that drift towards GEM1 will multiply in the first GEM plane, and via the applied electric transfer field E_{T1} will be transferred to GEM2. After four subsequent stages of multiplication in total, the electrons exiting GEM4 will drift towards the pad-plane, where the signals are induced. GEM1 and GEM4 have a "standard" hole pitch of 140 μm, while GEM2 and GEM3 have a "large" pitch size of 280 μm. To maximize the blocking of ions produced in GEM4, the transfer field between GEM3 and GEM4, $\mathrm{E_{T3}}$, is kept as low as 0.1 kV/cm, while $\mathrm{E_{T1}}$, $\mathrm{E_{T2}}$ and $\mathrm{E_{ind}}$ are kept at 3.5 kV/cm.

The properties of the chosen GEM configuration are the following: a nominal gain[2] of 2000 for the Ne–CO_2–N_2 (90−10−5) gas mixture, an IBF of less than 1%

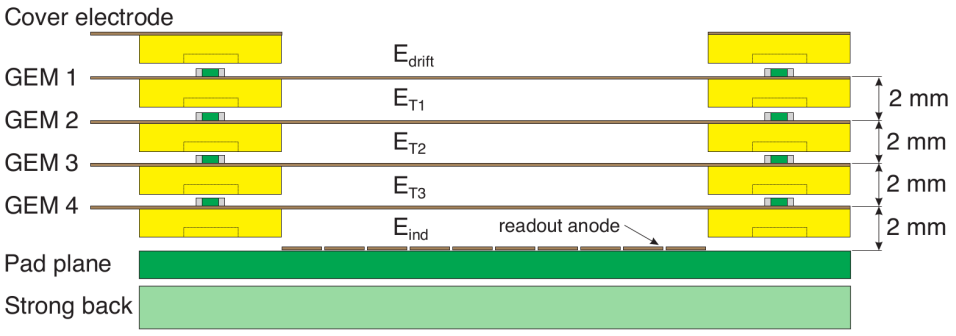


Figure 3.2: Cross section of an ALICE TPC GEM. Four GEM foils are stacked on top of each other, resulting in four subsequent stages of multiplication [6].

[2]The gain is defined as the number of electrons produced in the GEM structure per primary electron.

in order to maintain acceptable levels of space-charge distortions[3], and an energy resolution better than 12% for a Fe-55 source [8]. It should be emphasized that the main argument for choosing a neon-based gas mixture instead of an argon-based one was the significantly lower expected space-charge distortions in the neon case. Although the ion backflow is very similar for these two noble gases, the usage of argon, because of the much lower ion mobility, would result in larger space-charge distortions.

3.2.2 Mechanical structure and high-voltage system

Similarly to Run 1 and 2, each TPC end-plate is equipped with 18 inner readout chambers (IROCs) and 18 outer ones (OROCs). The dimensions of a sector, which consists of an IROC and an OROC, are shown in Figure 3.3. The dimensions and segmentation are the same as for the MWPC TPC. The two different types are used to account for the radial dependence of the track density, therefore the IROC pad sizes are smaller to maintain a good two-track separation. In the OROC, the pad sizes also moderately increase as the radius increases. The sector and individual readout chamber numbering scheme is explained in Appendix A.3.

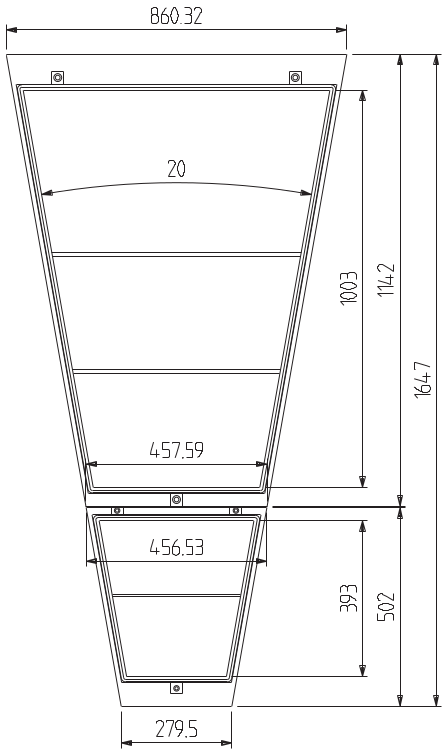


Figure 3.3: Dimensions of one ALICE TPC sector in mm. A sector consists of an inner and an outer readout chamber (IROC and OROC respectively) [6].

[3]Note that common triple GEM stacks leak up to 5%.

Since GEM foils having the dimensions of an outer readout chamber (≈ 1000 mm length, ≈ 900 mm width) are beyond today's technology, the OROCs are segmented into three independent modules or *stacks*, visible in Figure 3.3. The dimensions of each detector module are listed in Table 3.1. The pads have a size that varies from 7 to 15 mm in the row direction (which corresponds to the r-direction, or local x-direction) and from 4 to 6 mm in the pad direction (which corresponds to $r\phi$, or local y-direction). The pad sizes in the radial direction are chosen to optimize the sampling of the track, the two-hit resolution, the price of the electronics, as well as the charge per pad row created by a single track (commonly referred to as *cluster*). The number of charge measurements determines the $\mathrm{d}E/\mathrm{d}x$ resolution, as it is a statistical measurement. In this way, a maximum of 152 charge and position measurements (equal to the number of pad rows) can be obtained for a single track.

In order to reduce the capacitance between the two metallic sides of a GEM foil, therefore minimizing the discharge probability, the one side of each GEM foil is segmented into individually powered HV sectors soldered directly onto the foil. The potentials at the unsegmented side are supplied directly. With this HV scheme, sketched in Figure 3.4, the potential differences in the four GEM foils are $\Delta U_{\mathrm{GEM1}} = 270$ V, $\Delta U_{\mathrm{GEM2}} = 230$ V, $\Delta U_{\mathrm{GEM3}} = 320$ V and $\Delta U_{\mathrm{GEM4}} = 320$ V. In the MWPC TPC, the field cage was designed to degrade the drift field from the central electrode at -100 kV to the cathode wires kept at 0 kV potential. Since in the GEM TPC the GEM1 bottom electrode is kept at the high voltage of 3.15 kV, the field cage had to

Chamber Type (*stack*)	Stack area (mm^2)	Pad size ($r\phi \times r$, mm^2)	N. of pad rows	N. of pads
IROC	171154	4.16−4.36 x 7.5	63	5280
OROC1	174853	6 x 10	34	2880
OROC2	231284	5.88−6.08 x 12	30	3200
OROC3	294836	6.04−6.07 x 15	25	3200

Table 3.1: Dimensions and number of readout pads in one sector, for each stack [33].

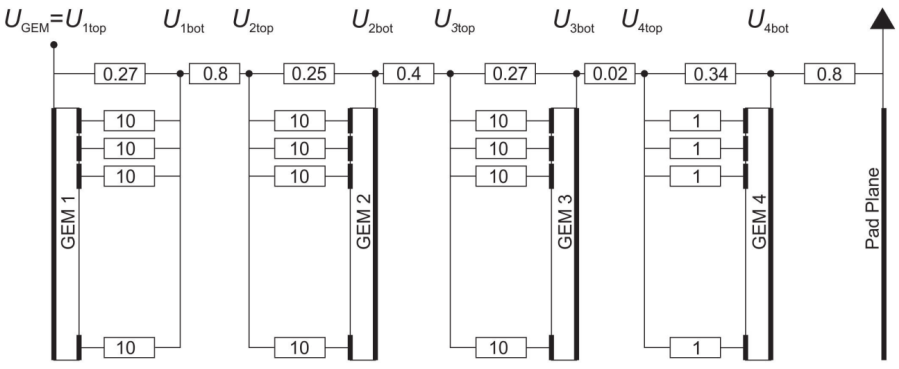


Figure 3.4: High-Voltage distribution scheme for a GEM module. The resistance values are in units of MΩ [6].

be modified accordingly.

An exploded view of a GEM OROC is shown in Figure 3.5. It consists of a trapezoidal aluminum frame (*alubody*), a support plate made of fiberglass (*strong back*), the pad-plane (multi-layer PCB[4]) and the GEM stacks. The three separate detector modules are visible, as well as the HV segmentation. An additional 1.5 mm wide *spacer cross* (one longitudinal and one transverse bar), not visible in the figure, is installed in each stack, to ensure mechanical stability of the GEM foils due to electrostatic forces [8]. Copper pipes, responsible for removing the residual heat propagating from the electronics through the kapton cables and into the chamber, are installed on the alubody.

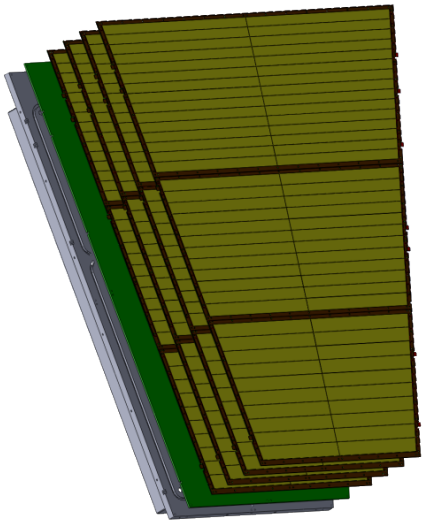


Figure 3.5: Exploded view of a GEM OROC, consisting of three independent detector modules [6].

3.2.3 New electronics and signal readout

During LHC Run 1 and 2, the current signals induced on the pads were transferred to the *Front-End Cards* (FECs) mounted on the end-plates, which were responsible for the signal amplification, shaping, baseline and tail correction and zero-suppression (see Subsection 3.3.1). However, the readout scheme had to be dramatically modified for the Run 3 set-up, to adapt to the new requirements stemming from the integration to a continuous readout. A continuous mode implies concurrently streaming data from the *Front-End Electronics* (FEE) to the CRUs[5]. Therefore, in the upgraded TPC, after digitization on the FEE, the data are passed to the back-end readout

[4]Printed Circuit Board.

[5]The *Common Readout Unit* is an FPGA-based readout card developed within ALICE. It is an interface to the online farm and detector control system.

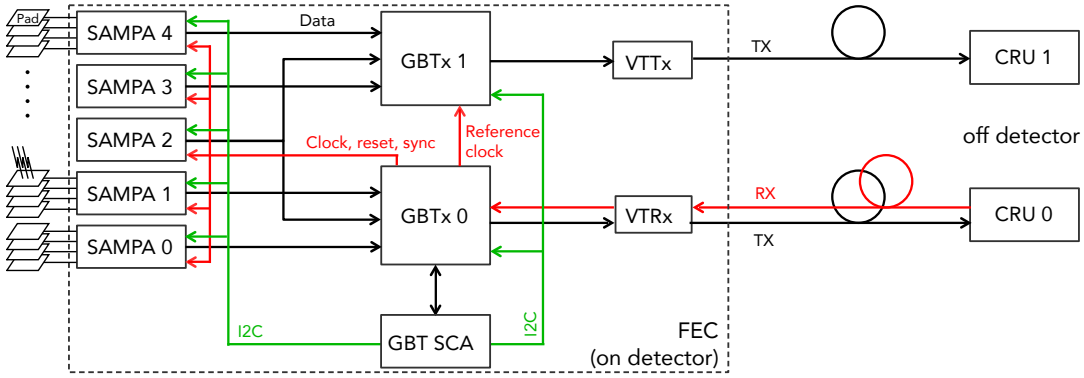


Figure 3.6: Readout scheme of a Run 3 FEC [8].

system without further front-end digital processing or compression. The data correction for the baseline fluctuations (see Subsection 3.3.1), common-mode effect (see Subsection 3.4.1) and ion tail (see Subsection 3.4.2) etc. is performed online, but off-detector (in the CRUs). Before written permanently to tape, the data are further compressed via cluster finding, online tracking and track selection.

The readout scheme of a Run 3 FEC is shown in Figure 3.6 and only briefly explained here. For further details, one can refer to [8, 34]. Each FEC consists of 5 SAMPA chips[6] which are responsible for the signal amplification, shaping and analog-to-digital conversion. Since each SAMPA can process 32 readout channels simultaneously, a FEC handles the data of 160 pads. After the analog-to-digital conversion, a 10–bit information for each channel (corresponding to $2^{10} = 1024$ ADC values, which is equivalent to 5 V) is transferred to either GBTx0 or GBTx1[7], which are responsible for the data multiplexing. GBTx0 also communicates with the GBT *Slow-Control ASIC* (SCA), and dictates the sampling clock. Each SAMPA has 10 serial electrical output eLinks (one for each bit), and an additional one for the sampling clock. The two GBTx are connected to the VTTx (versatile link twin transmitter) and VTRx (versatile link transceiver) as shown in the figure. From the VTTx and VTRx the digital information is transferred to the CRUs via optical fibers. Additionally, the clock is supplied from the CRU to the VTRx via trigger/control fibers. For each sector, 182 data fibers and 91 trigger/control fibers have to be connected. Each sector communicates with ten CRUs, therefore a total of 360 CRUs handle the TPC data stream.

The ADC sampling frequency is 5 MHz[8] therefore the ADC values are transferred from the FEE to the CRUs every 200 ns. The 3276 FECs of the TPC transfer 3.28 TB of information to the CRUs per second, of which, after intensive compression, only 100 GB/s are written to tape. A photograph of a FEC is shown in Figure 3.7.

[6] The SAMPA is a novel custom-made Application-Specific Integrated Circuit (ASIC).

[7] The *GigaBit Transceiver* (GBT) is a CERN-developed radiation hard readout architecture and transmission protocol.

[8] During Run 1 and 2, the frequency was 10 MHz, but as shown in [34], a factor two frequency reduction reduces the data size by half, without consequences for the physics performance.

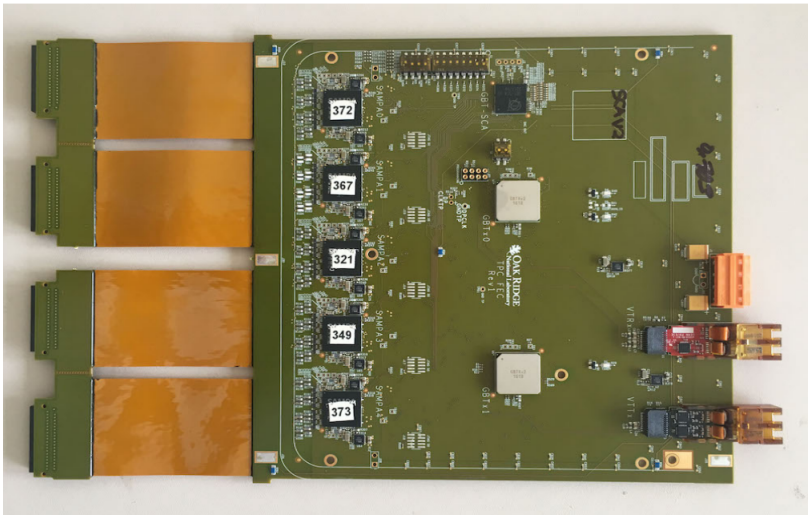


Figure 3.7: Photograph of a Run 3 FEC, to be compared to the sketch of Figure 3.6. The FEC is mounted on the TPC pad-plane and, via flexible kapton cables (orange), the pad currents are transferred to the SAMPA chips.

3.3 TPC upgrade procedures in SXL2

The three-year long LHC Run 2 ended on the 3rd of December 2018, signaling the start of the *Long Shutdown 2* (LS2). The last heavy-ion run in which ALICE collected data was during November 2018.

Immediately after the start of LS2, the disassembling procedures of the TPC began down in the SX2 Cavern; the shieldings, forward detectors and beam-pipe were removed from the A-side, and the un-cabling of the TPC began. At the same time, the SXL2 *Clean-Room* (CLR), at the surface of Point 2, was prepared, the *Yellow Platform*[9] (YP) was installed, and a test sector, where some tests would be performed at, was moved inside the CLR. After removing the Miniframe[10], the TPC was extracted from the detector and brought to the surface, next to the CLR.

During March 2019, a series of tasks were performed; over-pressure test of the TPC to ensure that no leaks are present, removal of distribution boxes, blue covers, cables, LV bundles, fibers, hoses, gas pipes, temperature sensors etc.. After removing all the *Front-End Cards* (FECs), the *Service Support Wheels*[11] (SSW) were unmounted as well. Finally, the ITS load test was performed, to ensure that the TPC can support the 450 kg Run 3 ITS, which will be heavier than in the previous setup.

By the end of March the TPC was cleaned and entered the CLR in SXL2, where

[9]The Yellow Platform is a hydraulically-controlled elevating platform which allows access to one TPC end-plate.

[10]The Miniframe is a removable structure installed on the A-side, serving as a support for all the services [10].

[11]The SSW is a support structure on which the FECs as well as the supply services are mounted, to avoid loading with additional weight the end-plates and simultaneously reducing the thermal coupling to the drift volume.

an X-ray irradiation test for the field cage took place to check its stability against discharges. The swapping of the readout chambers began at the end of April (replacement of the MWPCs with GEMs), which finished by July 2019. Access to the different levels of the TPC end-plates was possible via the YP on the C-side, and via an aluminum scaffolding on the A-side. After the installation, the new readout chambers were surveyed and aligned. The TPC schedule allowed that, if during the dedicated sector tests the performance of a chamber was not satisfactory, it could still be replaced. After the installation of all GEMs, the TPC was filled with dry air, in order to measure the leakage current of the GEM stacks.

Next, all 3276 FECs were installed on the TPC end-plates, together with the necessary services. Figure 3.8 shows a photograph of a fully equipped sector, with its most important parts labeled. One can see the *Service Support Wheel* aluminum structure, where the services are mounted on, and the *copper pipes* in which water, responsible for cooling the FECs, runs in. From the copper pipes, the water is directed through the FECs via the *hoses*, also visible in the photograph. The light-blue *fibers*, which connect each FEC to the CRUs, as well as the *low-voltage cables*, which supply the FECs with analog and digital voltage are visible. Finally, one can see the *resistor box*, which supplies the GEM stacks with the high-voltage.

In November 2019 the pre-commissioning tests began. The purpose of the tests was, first of all, to ensure the functionality of all readout chambers, front-end cards and connections to the CRUs, secondly to align the readout chambers, and finally to understand the upgraded TPC and develop algorithms that will perform online and

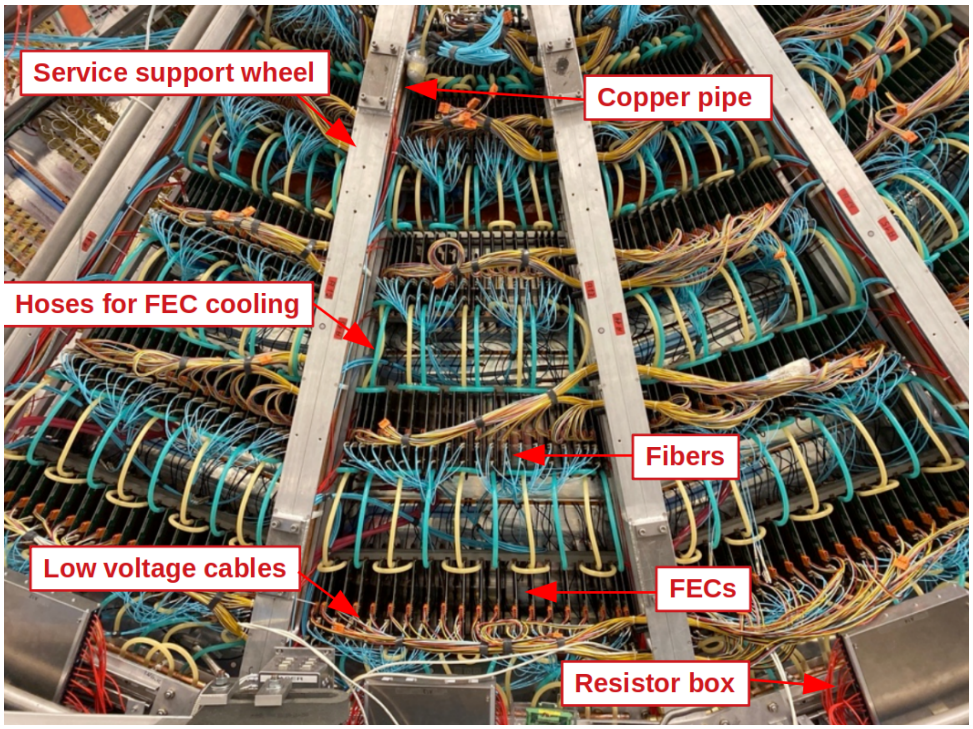


Figure 3.8: A fully equipped sector, with the most important parts labeled.

offline corrections during the normal operation. For the pre-commissioning tests, one of the two lasers from the cavern was installed in the clean-room, and placed on the A-side. By changing the configuration, the laser system could be adjusted to irradiate either the A- or the C-side.

The cooling system available in the SXL2 can only supply two sectors at the same time. Thus, sector pairs were tested for a certain time, and when a sector pair was fully tested, the cooling, LV supply and fiber connections to the CRUs were moved to the next sector pair, so that in the end all sectors were fully checked. The sectors tested together every time were selected either based on geometrical criteria (to better visualize cosmic tracks or laser rays) or because of convenience in moving the services.

After extensive irradiation tests with an X-ray source during July and August 2020, on the 14$^{\text{th}}$ of August the TPC was transported back to the cavern and inside the magnet to what is called the *parking position*. This position is a bit forward of where it will be finally installed for normal operation, thus allowing for free space to work behind the TPC, around the muon absorber.

The main measurements performed during the pre-commissioning phase were the following:

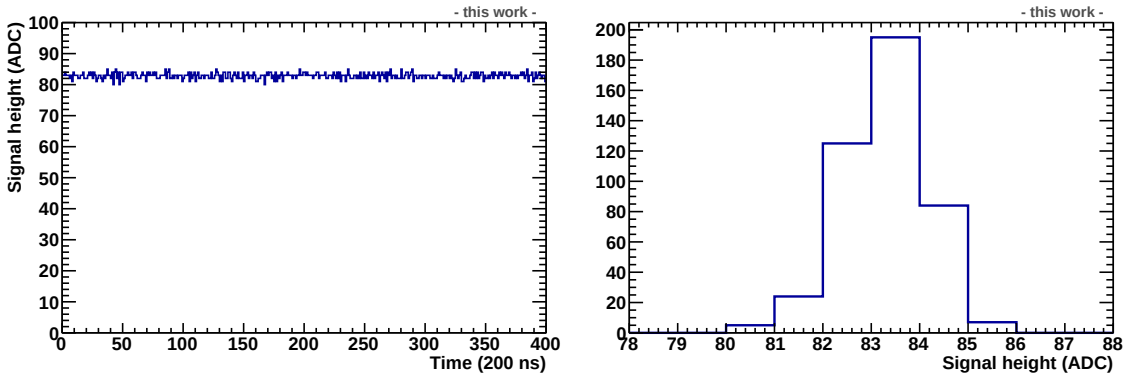
- *Pedestal and noise measurements*, necessary for the baseline correction and zero-suppression,
- *Calibration pulser measurements*, for studying the pulse-shaping electronics properties as well as the GEM4B–pad-plane capacitance,
- *Cosmic data-taking with a random trigger*, for studying the gas gain properties and for testing the track reconstruction algorithms,
- *X-ray irradiation tests with an Ag-source*, for studying the gas gain properties and the space-charge distortions, for performing pad-by-pad gain calibration, as well as for testing the stability of the field cage and chambers against discharges,
- *Laser measurements*, for alignment purposes and for studying the common-mode effect and ion-tail (see Chapter 5 and Chapter 6).

To ensure the functionality of the TPC over extended periods of time, to test the readout chain, as well as to collect enough cosmic events[12], two 8-hour long shifts took place every day after work hours [35]. In the following, the first two measurements, namely pedestal–noise and calibration pulser measurements are briefly described, since they are strongly connected with the laser data analysis. The classes of the O2 framework [36] used to analyze these data can be found in Appendix C.

[12]Since the triggering for the cosmic data-taking was random, most events do not contain a cosmic particle.

3.3.1 Pedestal and noise measurements

The *pedestal* or *baseline* is a systematic shift of the measured charge value due to the front-end electronics, while the *noise* describes the fluctuations around this mean value. Typically for the Run 3 setup the baseline shift is around 80 ADC, while the noise around 1 ADC. These values change from pad to pad, depending on its size, position in the chamber, minor manufacturing differences in the SAMPA chips etc.. The baseline and its fluctuations contain no physically interesting information, therefore to reduce the data volume, the *zero-suppression* is applied in the CRUs. ADC values below the mean value plus typically three standard deviations are considered as background and rejected.



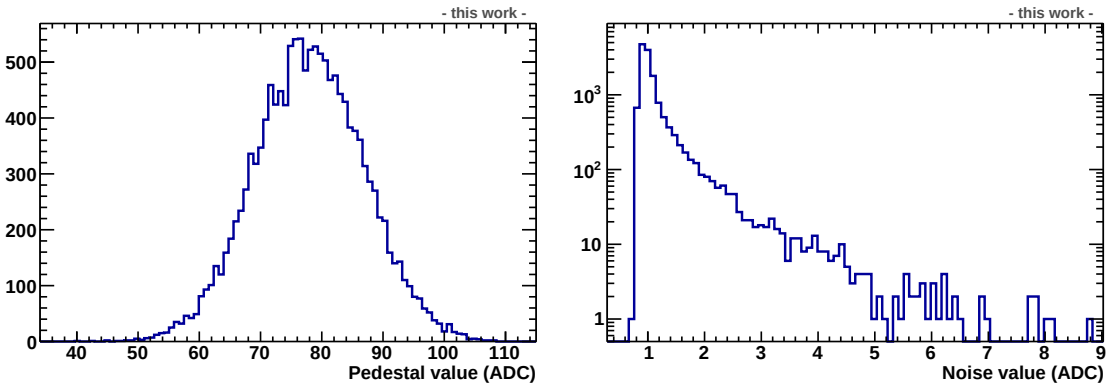
(a) (b)

Figure 3.9: (a) Baseline as a function of time for a pad of sector C04 for a single event, and (b) projection of the baseline to the ordinate.

It is therefore clear that, for a correct application of the zero-suppression, a precise knowledge of the pedestal and noise of each channel is necessary. To determine the pedestal and noise, data were collected for each pad, without external irradiation of the TPC. An example of the baseline as a function of time for a pad of sector C04 (for explanation of the sector numbering scheme refer to Appendix A.3) for a single *event* is shown in Figure 3.9a. In a single event, data are streamed for $\approx$ 480 timebins. During these measurements, the influence of external noise contributors present in the CLR (ventilation, electronics crates etc.) could not be entirely eliminated. For detailed noise studies with optimal conditions during the pre-commissioning phase, see [8].

To calculate the pedestal and noise of a single pad, the baseline is projected to the ordinate, as shown in Figure 3.9b. The mean value of the distribution is the pedestal value for the given pad, while the width is the noise. To get a statistically better value, many events are summed up.

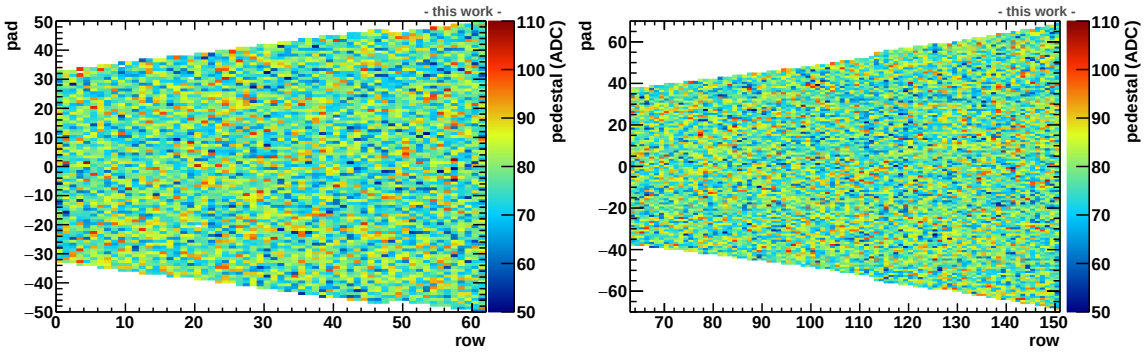
The pedestal and noise distributions for the 14560 pads of sector C04, as calculated during the pedestal runs, are shown in Figure 3.10a and Figure 3.10b, respectively. It can be seen that the pedestal varies quite substantially from pad to pad, with



(a) Pedestal distribution of C04 pads

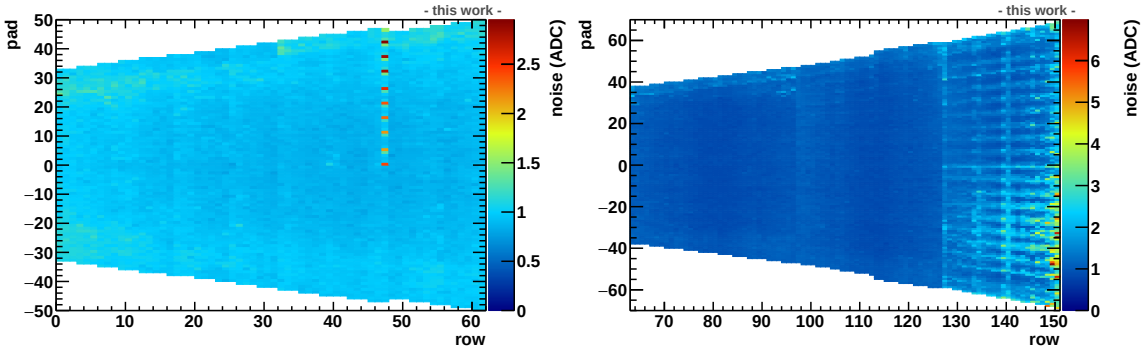
(b) Noise distribution of C04 pads

Figure 3.10: (a) Pedestal distribution of the 14560 pads of sector C04. The average pedestal value is ≈ 80 ADC, with an RMS of ≈ 9 ADC. (b) Corresponding noise distribution. The mean noise value is ≈ 1.14 ADC, however a number of high-noise outliers is observed.



(a) Pedestal values for C04 IROC.

(b) Pedestal values for C04 OROC.



(c) Noise values for C04 IROC.

(d) Noise values for C04 OROC.

Figure 3.11: 2D pedestal and noise distributions for the IROC and OROC of C04.

the average pedestal value being ≈ 80 ADC. The average noise value is ≈ 1.14 ADC, however a number of high-noise outliers are observed, which bias the mean. To identify the position of these outliers, in Figure 3.11, the pedestal and noise are plotted in the pad-row coordinate system separately for the IROC and OROC. It can be seen that the pedestal values are quite uniformly distributed in the chamber, however regions of high noise are observed. For the IROC, these correspond to the edges of the chamber, and some pads of row 47, where a temperature sensor has been installed. For the OROC, the pads in OROC3 and especially those at the edges have considerably larger noise values.

It should be noted that the pedestal and noise measurements taken during the pre-commissioning in SXL2 are not the final ones that will be used during the normal TPC operation, since these values heavily depend on the power supply configuration, noise from neighboring detectors, ventilation etc.. The pre-commissioning tests were merely a first round of measurements, aiming to test the functionality of all channels and to confirm that the noise and pedestal are within reasonable values. The pedestal and noise are discussed in further detail in Subsection 4.3.2, since they play an important role in the quality of the laser data analysis.

3.3.2 Calibration pulser measurements

The pulser measurements aim to study the shaping characteristics of the FEE. These characteristics in turn determine the timing and amplification of the signals in the SAMPA chips. Ideally, these properties should be universal for all channels, however due to manufacturing tolerances some differences are expected to be observed. The pulser measurements are described here since the nature of the pulser signals is very similar to the signals created due to the common-mode effect (discussed in Subsection 3.4.1). At the same time, they are probe to the GEM4 to pad-plane distance variations, which affect the effective field in the induction gap, which in turn influences the ion tail characteristics (discussed in Subsection 3.4.2). For the pulser measurements, a pulse is injected on GEM4B (bottom electrode of the GEM foil closest to the pad-plane) which induces a signal on the pad-plane due to capacitive coupling. A purely capacitive signal is measured, which is not caused by the motion of electrons, therefore the gas amplification properties are not coupled to the measurement.

In the top of Figure 3.12, the pulser charge for all C-side pads is plotted. Ideally, all pads belonging to the same stack and having the same dimensions should have the same pulser charge. However, as shown in the bottom of Figure 3.12, where the pulser charge normalized to the median pulser charge in the stack is plotted, the situation is more complicated. In Figure 3.13 the same quantity is shown for the pads of IROC C00, so that the structures are discernible. It is observed that the pulser charge (and therefore the pad capacitance) at the stack edges and under the spacer cross (see Subsection 3.2.2) is much higher. Despite the spacer, due to mechanical effects related to the foil stretching and electrostatic forces, the GEM foil "sinks" towards the pad-plane, thus reducing the distance and consequently increasing the

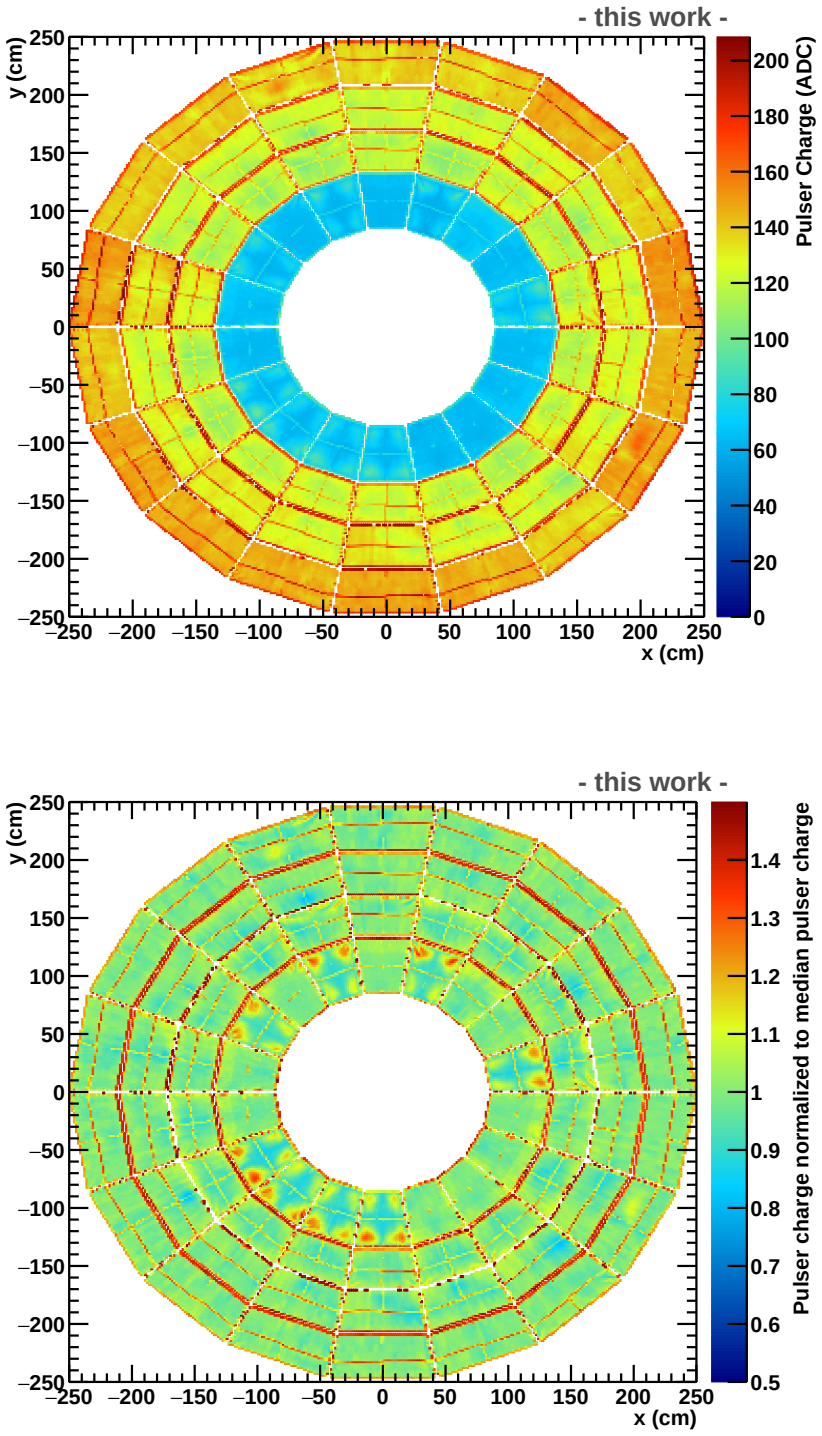


Figure 3.12: Pulser charge (top) and ratio of pulser charge to median pulser charge in the stack (bottom), for all C-side pads.

pad capacitance. This effect is referred to as *foil sagging*. The observed foil sagging varies significantly among the stack types due to geometrical differences, but also among the different chambers due to manufacturing differences during the stretching and framing procedures. The foil sagging is more pronounced in the IROC chambers. It was also observed in some OROC1 stacks, but very rarely in OROC2 and OROC3. Pulser data were collected for all the TPC sectors, and were used to understand and analyze the laser data.

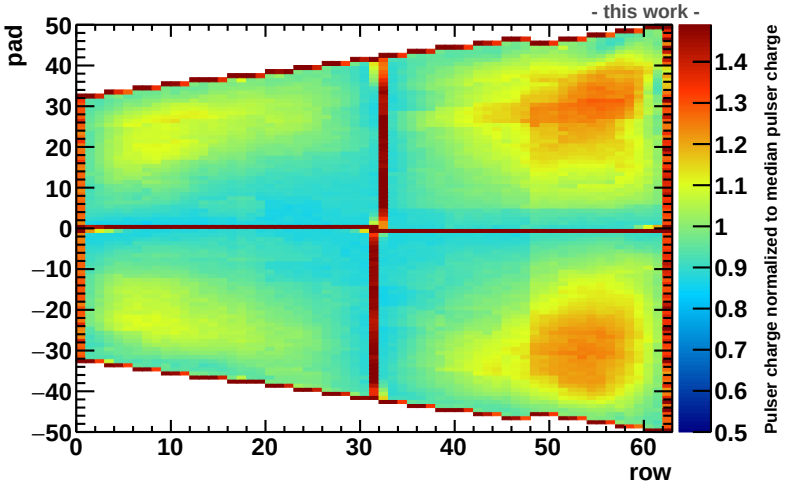


Figure 3.13: Ratio of pulser charge to median pulser charge in the stack for IROC C00.

3.4 Common-mode effect and ion tail in GEMs

3.4.1 Common-mode effect in GEMs

The common-mode effect is the result of capacitive coupling between the pad-plane and the GEM foils, which exists due to their common high-voltage supply via a resistor network. The common-mode effect is illustrated in Figure 3.14; when a real signal (due to the motion of electrons towards the pad-plane) is detected on a single pad, a capacitive signal of opposite polarity is created on all pads of the same stack. The amplitude of the common-mode signal in one pad is suppressed by a factor of N^{-1} with respect to the original signal, where N is the number of pads belonging to the same stack.

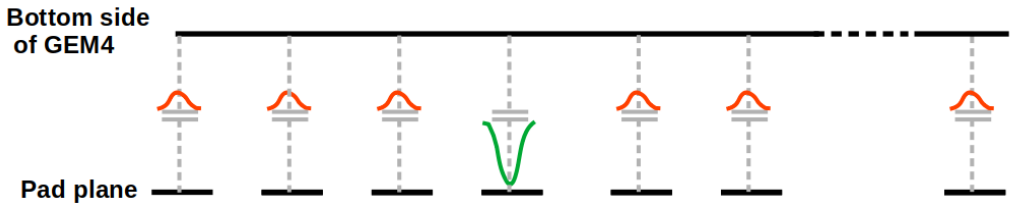


Figure 3.14: Illustration of the common mode effect. When a real signal (green) is detected on a single pad, a capacitive signal (orange) of opposite polarity is created in all pads of the same stack. The signal heights are not to scale.

In the case where coupling exists only between GEM4B and the pad-plane, the total positive signal in a given stack and the total induced negative charge[13] must sum up to zero, due to charge conservation. In that case, the pad-plane–GEM4B capacitance C can be calculated via the formula

$$C = \epsilon \frac{A}{d} \ , \tag{3.1}$$

where A is the stack area, d the pad-plane–GEM4B distance and ϵ the medium permittivity. However, as will be shown in Chapter 5, due to additional couplings between the pad-plane and the rest of the GEM foils (see Figure 3.4), as well as parasitic capacitance due to a ground cable connected to the pad-plane and a pulser cable connected to GEM4B, the total induced negative charge is in reality about 50% of the total positive signal in the stack. At the same time, either due to distance variations between GEM4B and the pad-plane, or due to the position of the pad in the chamber, pad capacitance variations are observed (see Subsection 3.3.2). This results in pad by pad common-mode charge variations that can be quantified using

[13] Although the polarity of the electron signal is negative (see Section 3.1), the ADC values sent to the CRUs are positive. Following this convention, the electron signals mentioned in the thesis are positive, while the common-mode signals negative.

the *common-mode fraction*, CF, which is the ratio between the (absolute) common-mode response of an individual pad normalized to the average positive signal in the stack. It will be shown that CF nominally takes values from 0.4 to 0.6, depending on the stack area and individual capacitance variations, however in some pads this number can reach up to ≈ 1.5.

As a comparison, in the case of the MWPCs, capacitive coupling of the anode wires to the readout pads results in uniform common mode responses with CF≈ 1. Therefore, for the same deposited charge, in the MWPC-based TPC the common-mode responses were on average about 50% larger compared to the GEM-based TPC. However, due to the interaction rate increase from $\approx$8 kHz in Run 2 to $\approx$50 kHz in Run 3 and therefore the consequent deposited charge increase, the common-mode effect is expected to have a similar or even larger magnitude in Run 3.

3.4.2 Ion tail in GEMs

Ions created during ionization processes in the vicinity of the pad-plane are expected to contribute to the measured signal in the pad-plane. These ions are either produced inside the GEM4 holes during electron multiplication or in the induction gap (see Chapter 6). Ions created in the previous multiplication stages are screened by the subsequent electrodes and therefore do not contribute to the signal. In both cases, the produced ions closely follow the electric field lines as shown in Figure 3.1, moving away from the pad-plane, with mobilities $\approx$ 2000 times smaller than the electrons [8]. The signal polarity of ions drifting away from the pad-plane is the same as that of electrons drifting towards the pad-plane. Therefore the measured signal is characterized by a fast rise due to the electron movement towards the pad-plane, followed by a slow positive tail due to the ion movement away from the pad-plane.

In the case of MWPCs, ions produced near the anode wires in the amplification stage drift towards the pad-plane, cathode wires or the gating grid wires. The signal is characterized by a fast rise due to the ions moving away from the anode wires in the high-field region, and a long tail of opposite polarity due to the slow motion of ions in the lower-field regions [37]. The tail integral in the MWPCs accounted for $\approx$50% of the total signal, while in GEMs this number was measured to be $\approx$9% (see Section 6.1). Again due to the foreseen rate increase in Run 3, the effect is expected to be quite significant.

3.4.3 Importance of the common-mode and ion-tail corrections

Extensive studies before and during LHC Run 1 and 2 were performed in order to understand, simulate and also correct for the two effects in the MWPC-based TPC [37]. The knowledge and the experience gained during that time can be partially transferred to the GEM-based TPC, however it should be kept in mind

that, although the common-mode effect characteristics are relatively similar, the characteristics of the ion tail differ remarkably from the old setup, primarily due to its opposite polarity. Therefore simulations for the common-mode effect in the MWPC-based TPC will be presented to highlight the importance of its correction, while a similar approach for the ion tail is avoided since the results would be misleading.

The impact of the common-mode effect in the performance of the TPC was studied by simulating central Pb–Pb events via the HIJING event generator with GEANT3. In these simulations, the common-mode effect with CF= 1 was included. As demonstrated in Figure 3.15, where the response of one TPC pad is shown, due to signals in neighboring pads the baseline is systematically shifted towards negative values. Applying the zero-suppression, all the physics-interesting structures below the threshold value would be lost. Therefore it is clear that the common-mode effect, if not accounted for, results in cluster losses as well as significant deterioration of the $\mathrm{d}E/\mathrm{d}x$ resolution, both of which undermine the tracking and the PID performance of the TPC.

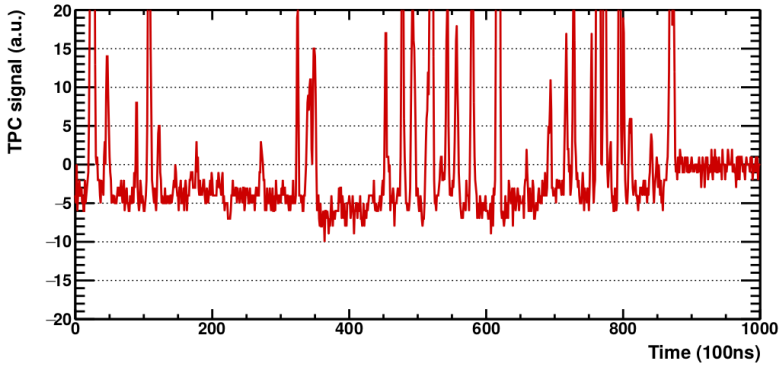


Figure 3.15: Simulation of a pad signal for a central Run 2 Pb–Pb event, where only the common-mode was implemented, with CF= 1. It can be seen that, due to the common-mode responses, the baseline is systematically shifted towards negative values [37].

Initially, for the MWPC-based TPC an online signal processing algorithm (*Moving Average Filter*) was developed to correct for the aforementioned baseline fluctuations on the hardware level. However, due to instabilities in the firmware, the online correction had to be abandoned and as an alternative, an offline software correction was used. HIJING simulations, shown in Figure 3.16, demonstrate the impact of the common-mode effect for the case of minimum ionizing pions and the achieved performance recovery using the offline correction. For the highest multiplicity events shown in the figure, a reduction of the mean $\mathrm{d}E/\mathrm{d}x$ position of $\approx 12\%$ is expected, of which $\approx 6\%$ can be recovered with the offline correction. The separation power is thus reduced by $\approx 18\%$, of which $\approx 12\%$ can be recovered.

Since the common-mode effect could not be corrected on the hardware level, its contribution to the baseline fluctuations inevitably requires a higher zero-suppression

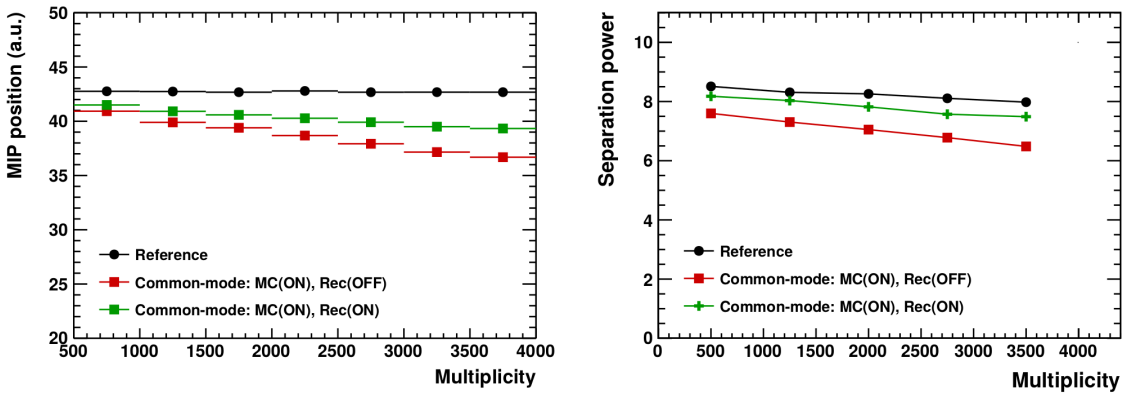
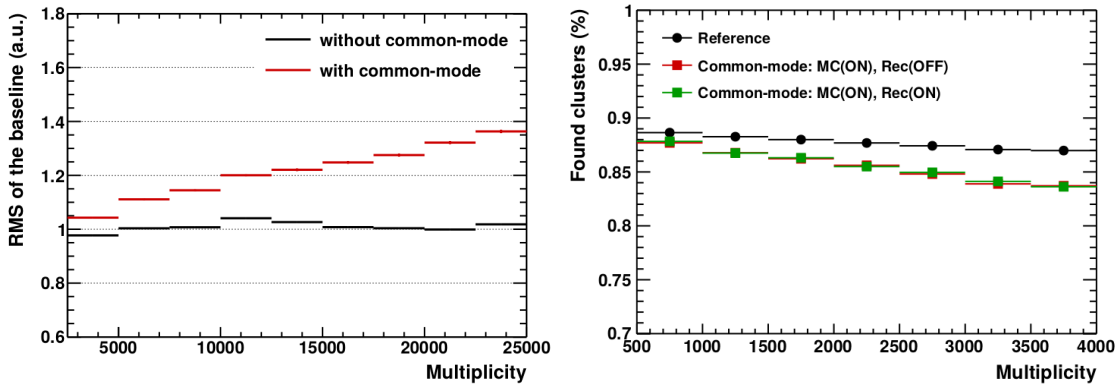


Figure 3.16: (Color Online) Mean dE/dx position (left) and separation power (right) results of a HIJING simulation for minimum ionizing pions as a function of the multiplicity where the common-mode was included. The reference data without common-mode (black), data with simulated common-mode (red), and data after offline correction (green) are plotted [37].

threshold, in turn resulting in more missing charge. In Figure 3.17a, the RMS of the baseline is plotted as a function of the multiplicity, with and without the common-mode effect. For the highest plotted multiplicity, its contribution to the baseline fluctuations is about ≈40%. Additionally, the common-mode effect unavoidably leads to missing charge clusters, which cannot be recovered with an offline correction, as shown in Figure 3.17b.



(a) (b)

Figure 3.17: (Color Online) (a)RMS of the baseline as a function of multiplicity, with or without simulated common-mode effect. (b) Found clusters as a function of multiplicity, without simulating the common-mode effect (black), simulating the common-mode effect (red) and correcting for it offline (green). The correction does not improve the fraction of missing clusters [37].

From the above, two main conclusions can be made; firstly, the common-mode effect has a significant impact on the performance of the TPC in terms of dE/dx

resolution (and, consequently, separation power), baseline fluctuations and missing clusters. Secondly, the performance can only be partially recovered if the common-mode effect is corrected offline. Therefore, for the Run 3 TPC, an online correction is aimed for, with which an almost full performance recovery can be achieved.

Due to its positive sign, the ion tail in the GEM-based TPC will not have an impact on the number of missing clusters. However, if not accounted for, it will lead to a (more localized) baseline bias and fluctuations, $\mathrm{d}E/\mathrm{d}x$ and cluster position resolution deterioration. The cluster position resolution deterioration is a consequence of the smearing of the *pad response function* (PRF) due to the ion movement. Ideally, the ion tail should be corrected online, primarily to restore the baseline bias, but also to significantly save resources on bandwidth and computing time.

From the above it becomes clear that the study of the common-mode effect and ion tail in the Run 3 setup is extremely crucial. These studies must be conducted before operation of the TPC with ion beams, so that online correction models can be developed to fully restore the performance parameters of the TPC.

4 Analysis of the signal shape and developed tools

The first round of pre-commissioning of the TPC took place in the months of November 2019−February 2020. Every sector pair was tested for a few days as described in Section 3.3, and during that time data were collected using the laser calibration system. To study the common-mode effect and the ion tail, "black" laser data were analyzed, namely pad signals before zero suppression. In this chapter, the laser data-taking procedure is described and the nomenclature is introduced (Section 4.1). The observation and treatment of the saturated signals is described in Section 4.2. Section 4.3 and Section 4.4 sketch the necessary analysis steps to produce the data structures that were then used for the analysis of the common-mode effect and ion tail. Some technical aspects of the analysis are summarized in Appendix C.

4.1 Laser runs and data format

A *laser event* is defined as one laser pulse. The laser pulse is very short (5 ns) and its frequency is 10 Hz (corresponding to one laser pulse every 100 ms). The data-taking is triggered by the *laser trigger*. For each laser event, raw data for about 480 timebins are streamed to the CRUs. Since the sampling frequency is 5 MHz, each timebin (tb) is 200 ns long. In a *laser run*, many laser events (200 to 1200) are collected concurrently[1].

The raw data for each sector pair consist of the digitized and multiplexed charge (10-bit ADC information) for every event and every read-out FEC channel. The raw data for each laser run are then converted to the *digits* format, which consists of one ADC value per timebin for each pad. The information is stored in tabulated structures, the so-called `TTrees` of the `ROOT` data analysis framework [38]. The resulting `digits TTree` contains the following information:

`event index, sector, row, pad, timebin, ADC value`

Ideally, for a *signal-pad*[2], four laser peaks should be expected, one for each bundle along z (refer to Subsection 2.3.3 for the laser tracks geometry in the TPC volume). Assuming a nominal electron drift velocity $u_\mathrm{d} \approxeq 2.58\,\mathrm{cm}/\mu\mathrm{s}$ [8], a data-taking trigger

[1] During the testing of the first few sectors, the laser system was still being optimized in terms of intensity and alignment. These laser runs mainly served as a test for its functionality, therefore the number of events collected was relatively low.

[2] A signal-pad is considered a pad where the maximum charge in a single timebin exceeds a threshold value.

synchronous with the laser trigger, and taking into account the z-positions of the laser bundles[3], the peaks should ideally be located at:

1st laser peak (bundle = 0):

- $z \approx 2470$ mm $\rightarrow \approx 6$ tb for odd laser rods,
- $z \approx 2410$ mm $\rightarrow \approx 17$ tb for even laser rods.

2nd laser peak (bundle = 1):

- $z \approx 1690$ mm $\rightarrow \approx 157$ tb for odd laser rods,
- $z \approx 1630$ mm $\rightarrow \approx 169$ tb for even laser rods.

3rd laser peak (bundle = 2):

- $z \approx 850$ mm $\rightarrow \approx 320$ tb for odd laser rods,
- $z \approx 790$ mm $\rightarrow \approx 331$ tb for even laser rods.

4th laser peak (bundle = 3):

- $z \approx 130$ mm $\rightarrow \approx 459$ tb for odd laser rods,
- $z \approx 100$ mm $\rightarrow \approx 465$ tb for even laser rods.

Additionally, a central-electrode signal should be visible for all pads, corresponding to the full electron drift length of 2500 mm. Therefore, another peak should be observed at ≈ 484 tb.

However, the measured pad signals differ substantially from the ideal scenario. Firstly, it was observed that the drift velocity of the electrons was systematically higher than the nominal one. At the same time, since each sector pair was tested at different dates, variations in the gas temperature and composition, as well as the ambient pressure, resulted in variations in the drift velocity and the gas gain. Furthermore, it was seen that the laser and the data-acquisition triggers were not perfectly synchronous, but rather that the data-acquisition trigger is delayed by a few timebins. Moreover, imperfections of the micro-mirror bundles result in misalignments either along z or along the x–y plane.

The response of a signal-pad is shown as an example in Figure 4.1. Only two laser peaks are visible, corresponding to bundle 0 and bundle 2, at ≈ 5 tb and at ≈ 295 tb, respectively. The last, smaller peak at ≈ 450 tb corresponds to the central-electrode signal. The two "missing" peaks are indeed observed in neighboring pads, due to the slight misalignment of the tracks in the pad-plane.

[3]Note that the z–positions are given with respect to the central electrode ($z = 0$). The electron drift distance to the end-plates will then be given by $2500\,\text{mm} - z$.

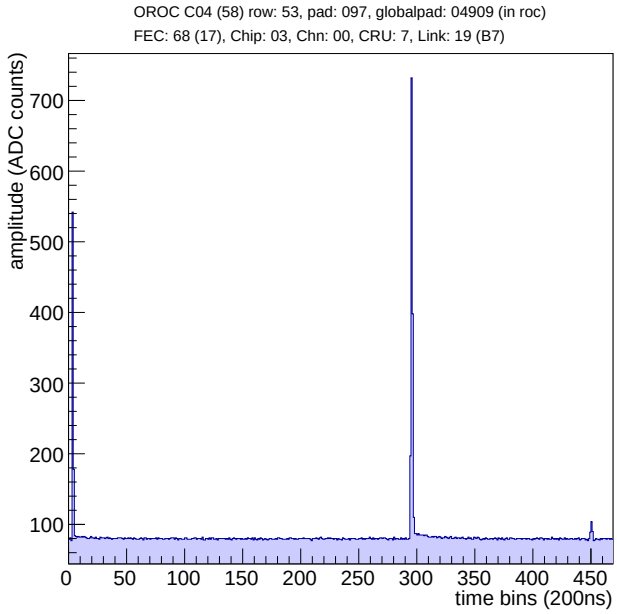


Figure 4.1: Single-event response of a pad of sector C04. Only two laser peaks are visible, corresponding to bundles 0 and 2. The smallest peak at ≈ 450 tb corresponds to the central-electrode signal. The two missing laser peaks are observed in neighboring pads.

4.2 Saturated laser signals

By analyzing the data of the first sector pair tested (A01, A16), it was observed that individual SAMPA channels temporarily malfunction if the pad charge for a single timebin exceeds the maximum value that can be streamed by the electronics (1024 ADC). Two examples of the effect of saturation are shown in Figure 4.2. The saturation shape differs significantly in the two shown cases, depending on the deposited charge. In all cases, the observed saturations last for less than ≈ 100 timebins, therefore the functionality of the channel is restored before the next laser signal.

It is underlined that this effect is an artifact of the electronics and therefore it does not affect the neighboring pads in any way. However, in such events the real signal height is not known, since the maximum is reached, and therefore a quantitative analysis cannot be conducted. Additionally, the long "tail" observed in the saturated chip would affect the ion-tail shape determination, which is the subject of Chapter 6. After this effect was discovered, the laser intensity was decreased, however the relatively small intensity fluctuations of the laser, combined with the stochastic nature of the charge deposition in the gas still resulted in some saturated events. Therefore, for the analysis, these saturated signals were excluded, as will be described in Subsection 4.3.1.

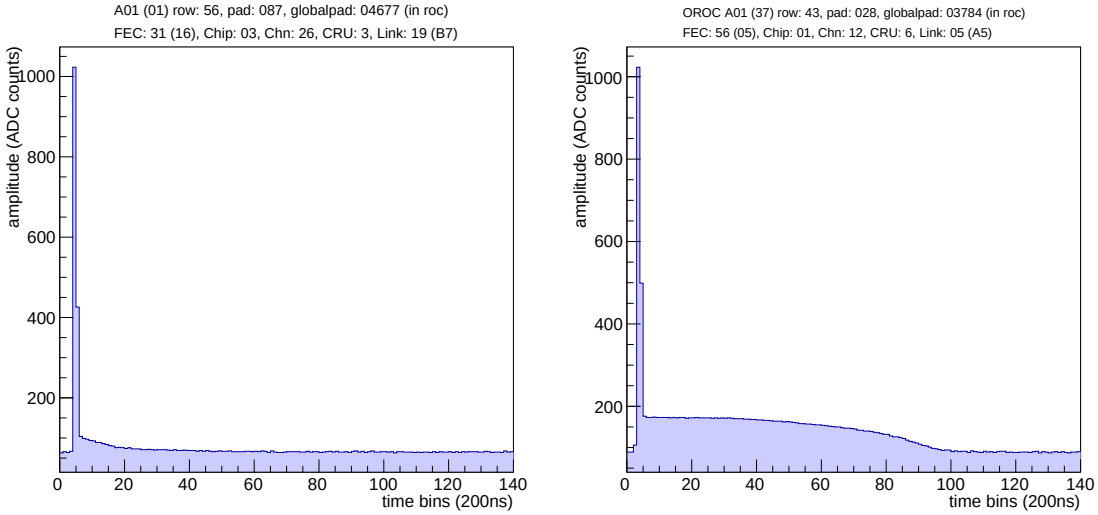


Figure 4.2: Examples of saturated chip signal shapes, for a single event. The saturation shape differs significantly in the two cases. In the right plot the deposited charge is much larger than in the left.

4.3 Analysis of the pad signals

A set of steps were followed to convert the `digits` format to the desired `signal TTree` that was used for the data analysis:

- Exclusion of saturated signals on an event-by-event, bundle-by-bundle and stack-by-stack basis (Subsection 4.3.1).
- Construction of the signal vs timebin graph for each pad (or *pad response*), after averaging over all non-saturated events. By averaging over all laser events, the *signal-to-noise ratio* (SNR) is significantly improved.
- Robust calculation of the baseline and noise for each TPC pad (Subsection 4.3.2).
- Inclusion of global and local pad coordinates.
- Inclusion of pulser data from the calibration pulser runs.

In the end, the `signal TTree` contains the information below for each TPC pad:

- sector, stack, row, pad, global & local pad coordinates,
- number of events (per stack and bundle),
- pad response,
- baseline & noise properties,
- pulser information.

4.3.1 Exclusion of saturated signals

Since the common-mode effect has to be studied for all pads of a given stack collectively (see Subsection 3.4.1), the exclusion of the saturated signals must be performed on a stack basis instead of a pad basis. At the same time, it was seen that for a given pad the peak maximum is not the same for all bundles, due to the micro-mirror bundle imperfections and diffusion (see for example Figure 4.1). Since the saturated signals do not last for more than ≈100 timebins, it is possible to perform the exclusion separately for each bundle.

Taking the above into account, the analysis procedure that was conducted was the following: the ≈480 timebins of each laser event were divided into four intervals, corresponding to each of the four bundles: [0,100), [100,250), [250,350), [350,480). Then, for each stack, if the signal of even one pad reaches the threshold value of 1000 ADC (before pedestal subtraction), the event is discarded for the entire stack, but only for the timebins belonging to this particular bundle. With this method, the number of events in one sector can be different for its four different stacks and for the four different bundles. In this way, the number of usable events was maximized.

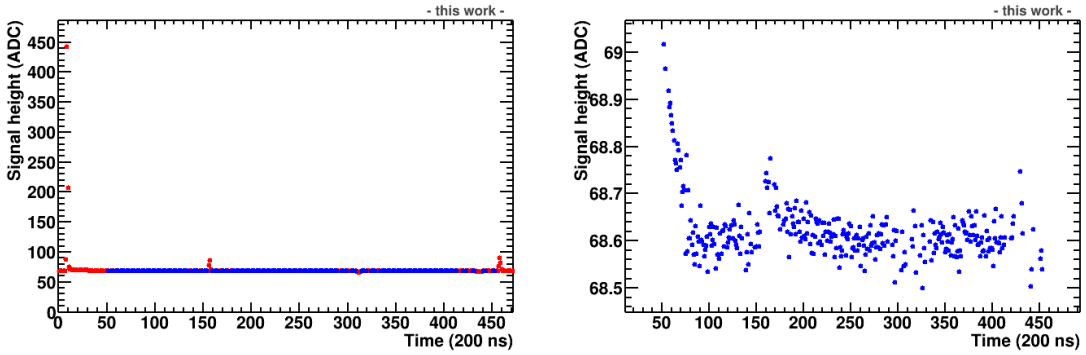
4.3.2 Baseline and noise calculations

As was already mentioned in Subsection 3.3.1, during normal operation of the TPC, the precise knowledge of the baseline and noise is essential for a correct application of the zero-suppression. At the same time, the common mode and the ion tail are both effects with small amplitudes (for the studied laser signals, the common mode responses range from less than 0.5 ADC to around 10 ADC, depending on the signal height, and typically the ion tail maximum is ≈0.5−1% of the total electron signal). Therefore, a well-defined baseline and low noise will ensure a high-quality analysis. Since the conditions during the pedestal calibration runs were often different compared to the laser runs, it was decided not to rely on the pedestal run results, but to calculate the baseline and noise directly from the laser data, using the pad responses averaged over all the available laser events. It is highlighted that the term *noise* during the course of this analysis mostly refers to the baseline fluctuations after averaging over all non-saturated events. Therefore the calculated noise not only reflects the intrinsic per event pad noise, but also the available number of events.

For each pad, a two-step process was performed in order to exclude the laser signals, the common-mode signals, as well as the observed ion tails. In the first iteration, if the charge `Q` at timebin `t`, `Q(t)`, fulfills the below criteria:

- `|Q(t - 1) - Q(t + 1)| < Q_T`
- `|Q(t) - mean(Q(t - 1), Q(t + 1))| < Q_T`
- `|Q(t) - Q(t ± 10)| < Q_T`
- `|Q(t) - Q(t ± 20)| < Q_T`

where Q_T is a threshold value, then the charge Q(t) will be included in the 1st subset. The first two conditions guarantee that positive and negative peaks (corresponding to laser and common-mode signals, respectively) are excluded, while the other two account for the expected tails after the laser peaks. The threshold value for the 1st iteration is fixed to Q_T = 0.4 ADC for all the TPC pads, based on the noise value of the full sample. In particular, the number of non-saturated events per bundle ranges from ≈100−1200. The expected per-event noise for a pad is roughly 1 ADC (see Figure 3.10b). Since the noise scales with the square root of the number of events, for the minimum number of 100 events, the average noise is expected to be around 0.1 ADC, therefore for 4-sigma, the threshold would be 0.4 ADC. In reality, in most cases the number of events is higher, while the per-event noise can be smaller or larger, depending on the pad size, position in the readout chamber etc.. These deems the application of a second iteration necessary, where the threshold is set individually for each pad, as will be explained below.



(a) (b)

Figure 4.3: (Color Online) Demonstration of the 1st iteration subset selection. In (a), the signal as a function of time for a single pad is plotted. With red markers, all points are shown, while with blue only the points selected during the 1st iteration. (b) Selected points during the 1st iteration for the same pad. Note the y-axis range.

An example of the 1st iteration subset selection for a single pad is shown in Figure 4.3. It can be seen that the signal peaks are successfully cut, and so are the common-mode signals. At the same time, most of the long ion tail in the first bundle is removed. However, part of the tail, as well as some signal timebins before and after the peaks are still selected.

From the subset selected via the first iteration, a first noise estimation can be conducted. To further avoid some outliers that might bias the calculations - as the ones observed in Figure 4.3b - the *Least Trimmed Squares* (LTS) method was used. The LTS method is a robust statistical method in which, to estimate the statistical properties of a sample, a specified fraction (e.g. 90%) of the sample is selected which results in the smallest chi-square. Therefore, the 90% noise is calculated

based on the subset selected in the first iteration (referred to as `noise1I90`). Since `noise1I90` might still not be an accurate estimator of the noise, a second iteration is performed on the initial data, where the same algorithm is applied, but this time with a threshold based on `noise1I90`, and in particular `thr = 4 · noise1I90`. From the subset of data in the second iteration, again using the LTS method, the 90% and 75% pedestal and noise values were calculated (`pedestal2I90`, `noise2I90`, `pedestal2I75`, `noise2I75`). These values were then used in the common-mode and ion-tail analysis. The effectiveness of the 2nd iteration is demonstrated in Figure 4.4, where apart from all points (red) and the 1st iteration subset (blue), the 2nd iteration subset is shown in black. Most of the tail remaining after the first iteration is now successfully removed, as well as some additional low-signal points close to the central electrode signal.

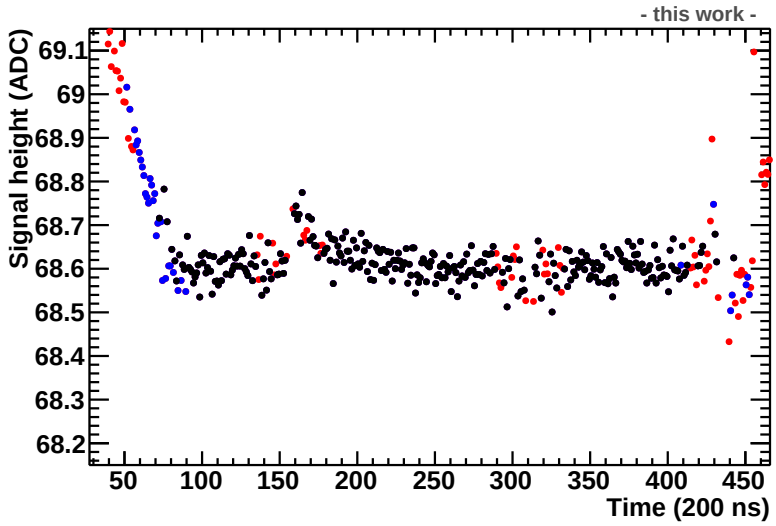
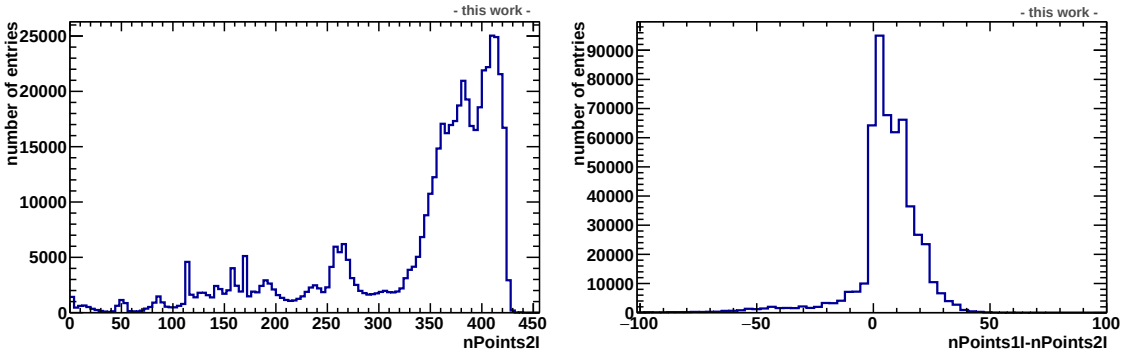


Figure 4.4: (Color Online) Demonstration of subset selection during the 1st and the 2nd iteration. With red markers all points are shown, while overlaid with blue and black only the points selected during the 1st and the 2nd iteration, respectively.

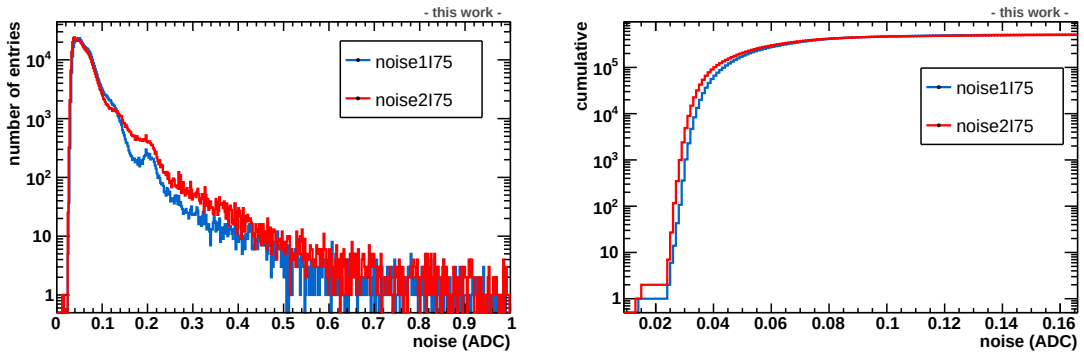
Figure 4.5a shows the number of points selected in the 2nd iteration. For the majority of pads at least 350 points (out of the ≈ 480) are selected. These correspond to either non-signal or low-signal pads with small ion tail signals. In ≈1% of the cases the number of points is below 50. These correspond to high-signal pads with very long tails, or pads were most events had to be excluded due to saturated signals, leading to large baseline fluctuations. Figure 4.5b shows the difference between the number of points selected during the 1st and 2nd iteration. In about 77% of the cases, fewer points are selected in the 2nd iteration; this indicates low-noise (intrinsic low noise and/or high number of events) pads where the 0.4 ADC threshold of the 1st iteration is an overestimation. In the rest of the cases, the 0.4 ADC threshold is an underestimation; this corresponds to higher noise pads that are mostly observed at the edges of the chambers.



(a) (b)

Figure 4.5: (a) Number of points selected for the baseline and noise determination during the 2nd iteration. (b) Difference between the number of points selected during the 1st and the 2nd iteration. Positive values indicate that the 0.4 ADC threshold is an overestimation, while negative values an underestimation.

The baseline and noise determination procedure described in Subsection 4.3.2 can be used to select a subset of the data with baseline and noise properties close to the ideal ones. Such a selection will improve the quality of the common-mode and ion-tail analysis. Figure 4.6a shows the 75% LTS noise as calculated during the 1st and the 2nd iteration. It can be seen that the 1st iteration threshold of 0.4 ADC is for some pads an underestimation, resulting in an increase in the number of high-noise pads in the 2nd iteration. Figure 4.6b shows the cumulative distribution of the 75% LTS noise, for small noise values. The number of low-noise pads is higher during the 2nd iteration, corresponding to overestimated noise values during the 1st iteration.



(a) (b)

Figure 4.6: (Color Online) (a) 75% LTS noise as calculated during 1st (blue) and the 2nd (red) iteration. The corresponding mean values are 0.067 and 0.068 ADC respectively, while the RMS of the distributions is 0.037 and 0.047 ADC. (b) Cumulative distribution of (a), for small noise values.

Similarly, the robustness of the baseline determination can be studied by plotting the difference between the 90% and 75% LTS baseline, as shown in Figure 4.7. The mean of the two distributions is ≈0.01 ADC, due to a slight biasing of the baseline of the signal-pads towards larger values (see Figure 4.4). In particular, in the case of the 2nd iteration, 99.5% of the pads have positive differences. By applying a 3-sigma cut (sigma = 0.01 ADC), namely selecting pads with `|Δ(pedestal)|` ≡ `|pedestal2I90-pedestal2I75|<0.03 ADC`, 4% of the points are excluded.

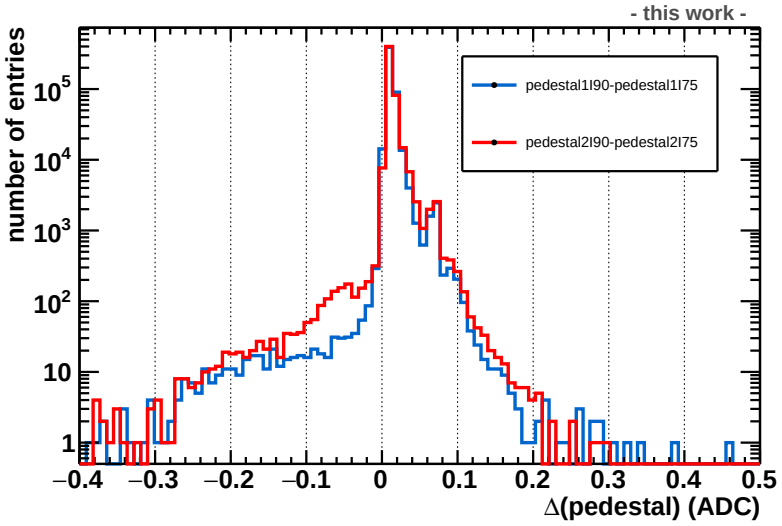
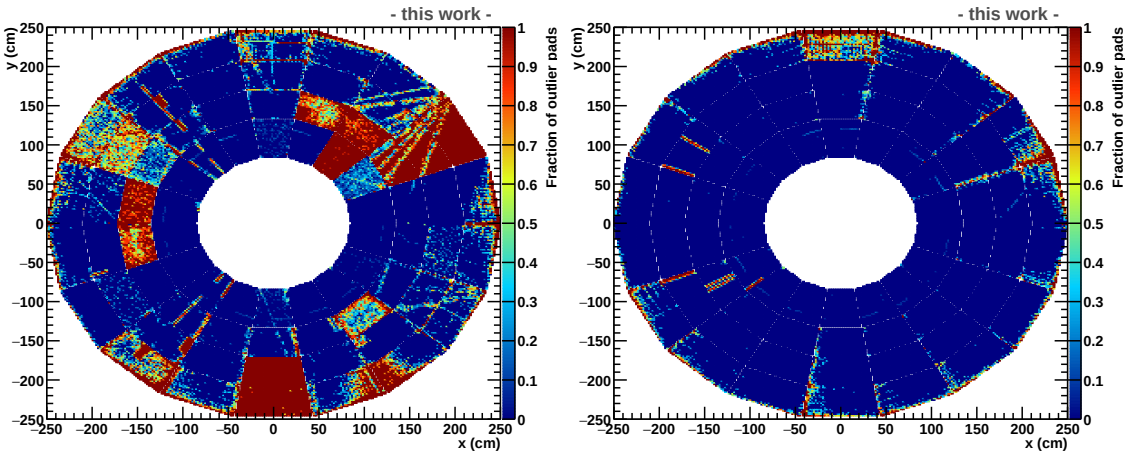


Figure 4.7: (Color Online) Difference between the 90% and 75% LTS baseline estimation for the 1st (blue) and the 2nd (red) iteration.

The above distributions were used to exclude particular *stack indices*[4] with a high number of noise or baseline outliers. This is demonstrated in Figure 4.8 where the pads which are identified as noise or baseline outliers (`noise2I75>0.1 ADC or |Δ(pedestal)|>0.03 ADC`, respectively) are plotted. A significantly larger number of outliers is observed on the A-side. Since the pre-commissioning started from the A-side sectors, in the first few runs a lower number of events was collected. At the same time, the laser intensity was not yet tuned, therefore a high number of events had to be excluded due to saturation.

In Figure 4.9 the fraction of (a) noise and (b) baseline outlier pads as a function of the stack index is shown. Cutting on stacks with more than 10% noise and 1% baseline outliers results in 33 (46%) and 15 (20%) stacks on the A- and C-side being excluded, respectively. In total about 30% of the stacks had to be excluded for the common-mode and ion-tail analysis.

[4]The `stack` corresponds to the *stack ID*, which can take the values 0 (IROC) to 3 (OROC3). The `stack index` corresponds to the *unique stack ID*: `stack index = sector + stack/4`, therefore it can take the values 0 (A00 IROC) to 36 (C17 OROC3), in steps of 0.25. Each TPC half consists of 72 ($4 \cdot 18$) stack indices.



(a) (b)

Figure 4.8: Fraction of baseline or noise outliers on the (a) A-side and (b) C-side.

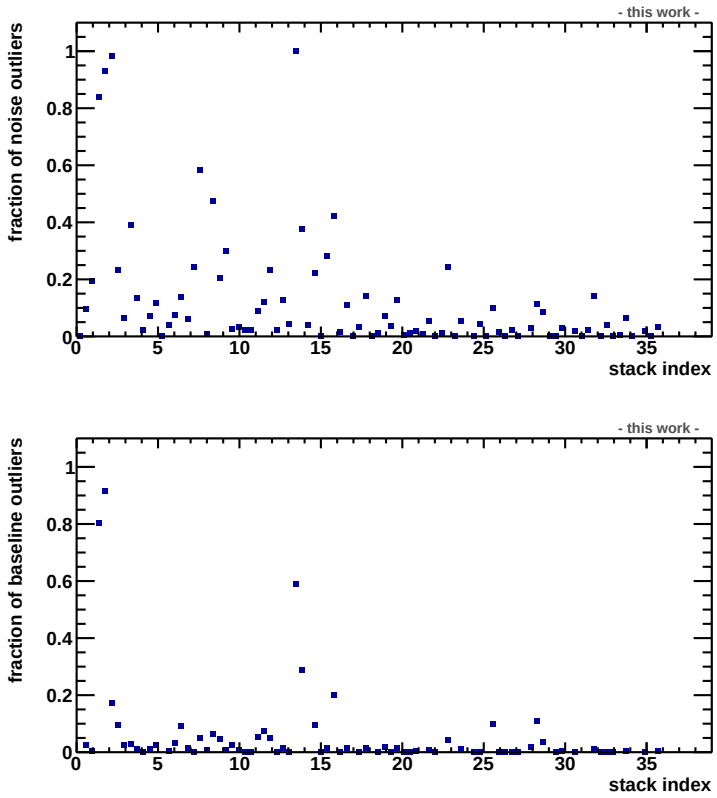


Figure 4.9: Fraction of pads identified as noise (top) and baseline (bottom) outliers.

4.4 Analysis of the laser tracks

Signal-pad to laser track association is a necessary part of the analysis in order to determine some unknown parameters such as track inclination[5], track misalignment (along the x–y plane as well as z), pad-plane misalignment etc.. This is especially useful for the ion-tail analysis, since the tail properties heavily depend on the distance of the signal-pad from the *center of gravity* (COG) of the cluster, the track inclination, diffusion etc.. Moreover, it is important to verify whether the common-mode effect is independent of such parameters.

The `track TTree` contains the full information that will be used for the common-mode and ion-tail analysis. In each entry, the following information, which will be explained below, is stored:

- **general information:**
  sector, stack, laser ID, bundle, beam, rod, ideal track coordinates, track fits of associated pads on the $x - y$ plane
- **information for each associated signal-pad (for ion-tail analysis):**
  row, pad, local and global coordinates, baseline and noise properties, pulser information, COG in time direction, maximum and total charge under signal peak, exponential fit parameters to the ion tail
- **information for each pad in the stack (for common-mode analysis):**
  row, pad, local and global coordinates, baseline and noise properties, pulser information, maximum and total charge under signal peak, template fit results

4.4.1 Laser track association

Each laser track is identified via a unique laser ID (0−335), or a set of properties:

- **TPC side:** A-side (0) or C-side (1)
- **Laser rod:** one of the six rods in which the beam travels along z (0−5)
- **Mirror bundle:** one of the four bundles along z (0−3)
- **Laser beam:** one of the seven beams of each micro-mirror bundle (0−6)

The laser track coordinates (x, y, z) are stored in the *Offline Calibration DataBase* (OCDB). The coordinates registered in the database correspond to the measured ones during LHC Run 2.

To associate the signal-pads[6] with the laser tracks, the following algorithm was developed: for each sector and laser bundle, the signal-pads were associated to the

[5] The track inclination describes the angle of the laser track in the local x - local y coordinate system. This parameter affects the distribution of the charge in the pads of a cluster.

[6] Following the noise estimation described in Subsection 4.3.2, the threshold value was set to 0.5 ADC, corresponding to 7-sigma of the average noise, Figure 4.6a.

laser tracks based on a 5 cm cut on the x–y distance between them. Due to the drift velocity variations, the z-coordinate matching was deemed unreliable. The 5 cm distance is a very loose cut to allow for possible deviations from the registered laser track coordinates. Robust[7] *sector-wise* (or *global*, i.e. for the entire sector) and *stack-wise* (or *local*, i.e. for each stack separately) fits on the local x–y pad coordinates are performed. The difference between the local and the global fits can account for stack-by-stack misalignments. Before performing the local fits, a 1.5 cm cut on the distance of each pad from the robust 80% global fit was applied. Since the pad distance from the laser track is an important parameter for the ion-tail analysis, several fit options were explored for the local fits, as will be discussed below.

An example of the pad-to-track association is shown in Figure 4.10, where the laser IDs for all associated pads for the C-side for bundle 0 are plotted. For visualization, the edges of each stack are overlaid in black.

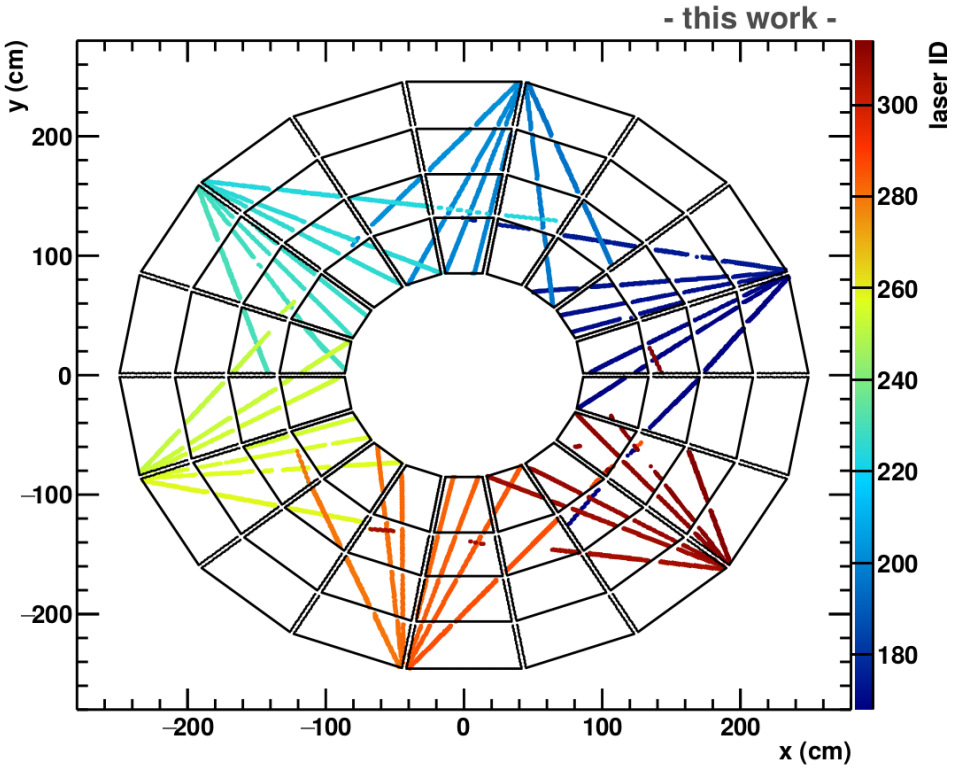


Figure 4.10: Laser IDs for all associated C-side pads of bundle 0. The stack edges are plotted in black.

For each associated signal-pad, the timebin position of the signal maximum in a given bundle is known, $t_{\max}$ ($Q_{\max} = Q(t_{\max})$). Thus, the center of gravity in time

[7] Using a pre-selected fraction of the data with the smallest chi-square.

can be calculated as

$$\mathrm{COG}_{\mathrm{time}} = \sum_{i=-1}^{i=+1} \Big[Q(t_{\max}+i) \cdot (t_{\max}+i) \Big] \Big/ \sum_{i=-1}^{i=+1} Q(t_{\max}+i) \quad . \tag{4.1}$$

Note that $t_{\max}$ is always an integer number, while $\mathrm{COG}_{\mathrm{time}}$ can take any value. In Figure 4.11 the COG in time is plotted for the associated pads of bundle 0, rod 1 on the C-side. It can be seen that the COG in time differs for each beam, and at the same time it is increasing with increasing track length. This is mostly due to micro-mirror bundle imperfections that result in the laser beams not being perfectly parallel to the end-plates, and the drift velocity variation along the y-axis [39] (due to the pressure derivative inside the TPC active volume). Furthermore, it can be seen that the COG in time is higher at the stack edges, caused by local distortions of the electric field. In particular, the electric field at the edges is not exactly perpendicular to the foil and the pad-plane, resulting in longer drift paths and therefore longer drift times for the electrons [40].

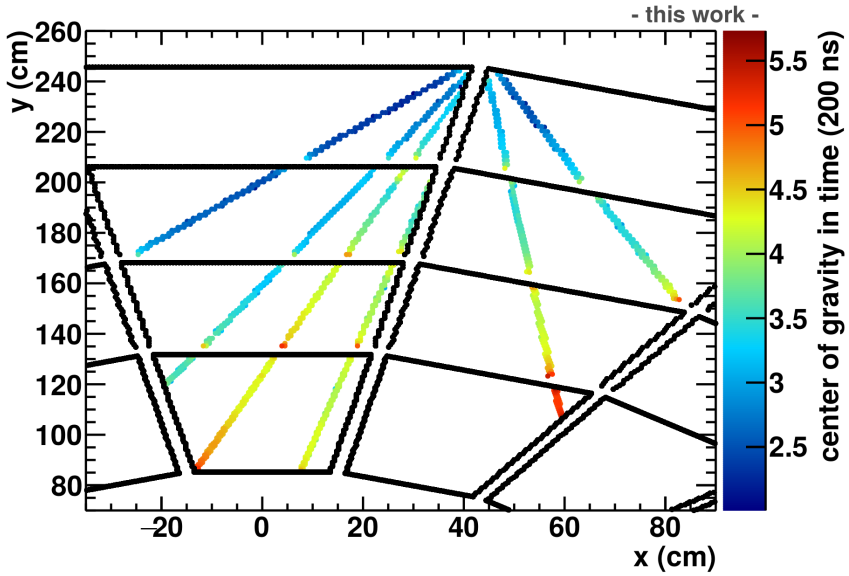


Figure 4.11: Center of gravity in time for the associated pads of bundle 0, rod 1 on the C-side. Two effects are convoluted: the track misalignment with respect to the pad-plane, and the drift velocity derivative due to the pressure derivative inside the TPC active volume.

4.4.2 Cluster finding and local fits

An alternative method to quantify the distance of the pad from the laser track is using the cluster information. The cluster is a set of neighboring signal-pads

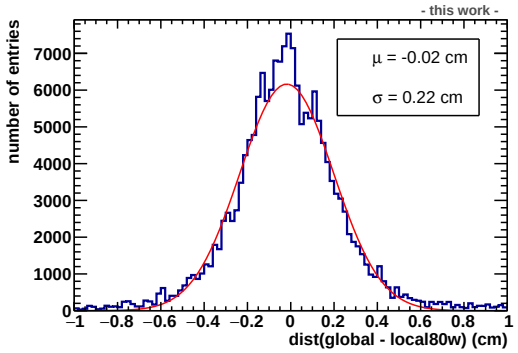
belonging to the same row, whose measured charge is assumed to be generated by the energy loss of a single particle – or beam – in the TPC volume. The laser clusters observed in the TPC usually contain 2–5 pads (depending on other parameters such as diffusion and track inclination). Once the pads belonging to the same cluster are identified, the center of gravity in the pad (local y) direction can be calculated,

$$\mathrm{COG}_y = \sum_{\mathrm{pad}\in\mathrm{cluster}} \Big[Q_{\mathrm{tot}}(\mathrm{pad}) \cdot y_{\mathrm{pad}} \Big] \Big/ \sum_{\mathrm{pad}\in\mathrm{cluster}} Q_{\mathrm{tot}}(\mathrm{pad}) \quad , \tag{4.2}$$

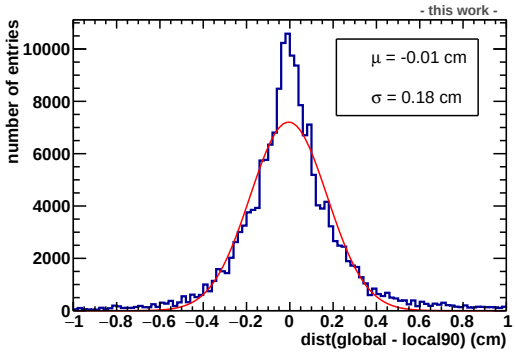
where Q_{tot} is the pad signal summed for three timebins around the peak maximum,

$$Q_{\mathrm{tot}} = Q(t_{\mathrm{max}} - 1) + Q(t_{\mathrm{max}}) + Q(t_{\mathrm{max}} + 1) \quad . \tag{4.3}$$

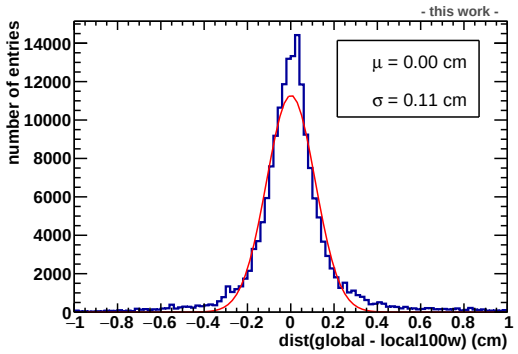
The advantage of the center of gravity is that, contrary to the local fits, it can account for field distortions which lead to deviations of the track linearity close to



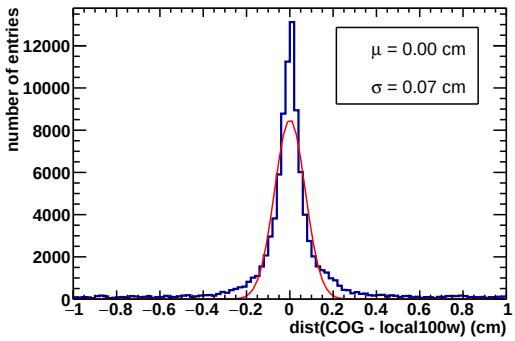
(a) Distance between global fit and local fit with 80% robust option and associated weights.



(b) Distance between global fit and local fit with 90% robust option without associated weights.



(c) Distance between global fit and non-robust local fit with associated weights.



(d) Distance between the center of gravity of the cluster and the non-robust local fit with associated weights.

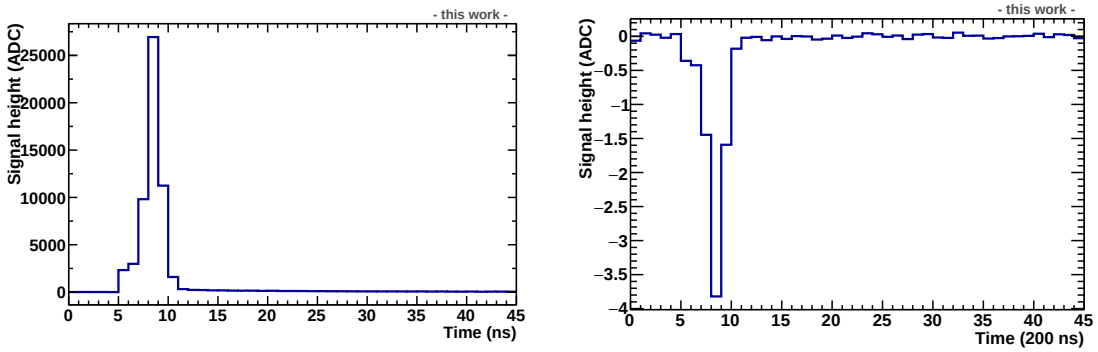
Figure 4.12: Difference between various methods of quantifying the distance from the laser track.

the stack edges. On the other hand, the center of gravity can be influenced by gain variations[8] between the pads of the same cluster.

Various options were explored for the local fits as well. A first method was to apply an 80% robust fit with associated uncertainties equal to $1/\sqrt{Q_{\max}}$. This method gives more weight to the pads with higher signal. However, it was observed that, in case of two-pad clusters – which are quite common in the laser tracks – this biases the fit by only accounting for the high-signal pad of each cluster. To avoid this biasing, a 90% fit without weights was also attempted. Finally, a non-robust fit with associated uncertainties forces the algorithm to also account for the low-signal pads. The difference between the various fit options is summarized in Figure 4.12, where the distance between them is shown for each signal-pad. It is seen that out of the three local fit methods, the non-robust local fit with associated weights results in the smallest RMS of the residuals from the global fit, Fig. (c). It is also observed that this local fit closely follows the center of gravity of the cluster, Fig. (d). In the end, these two methods, namely the distance from the center of gravity of the cluster, and the distance from the non-robust local fit with associated weights were used to quantify the pad distance to the laser track.

4.4.3 Template fits

For a given timebin, a few laser beams are simultaneously detected in a single stack (see Figure 4.11), all of which will contribute to the common-mode effect. In particular, the common-mode charge of a non-signal pad in a given timebin should scale linearly with the *summed* positive charge in the stack in the same timebin. An example is shown in Figure 4.13, where in (a) the summed laser signal in OROC2 of C01 is shown for bundle 0, while in (b) the common-mode response of a non-



(a) (b)

Figure 4.13: (a) Summed laser signal in OROC2 of C01. (b) Common-mode response of a non-signal pad of the same stack.

[8]The pad-by-pad gain calibration was performed during the pre-commissioning phase via an X-ray source. However, these gain maps were not yet available during the course of this analysis.

signal pad of the same stack is plotted for the same bundle. The common-mode charge of the non-signal pad as a function of the summed positive charge is shown in Figure 4.14 for 5 timebins around the peak. A linear fit is performed on these five points, which is called a *template fit*. Template fits were carried out for all pads, and the fit parameters together with the residuals were stored in the `track TTree`. These template fits are used to ensure that the timing information of the common-mode responses is consistent. By cutting on fits with large residuals, cases where either laser reflections in the inner field cage spoil the signal classification algorithm, or the pad baseline is biased, are excluded from the analysis.

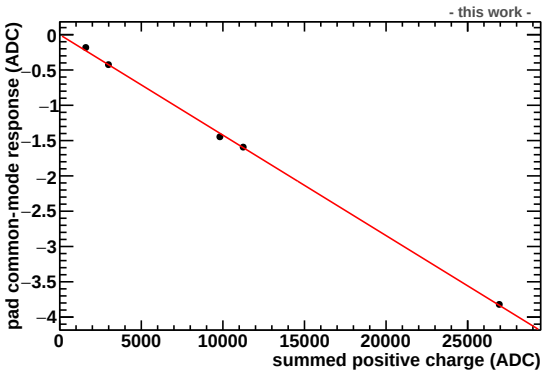


Figure 4.14: Example of a template fit for a single pad. The common-mode response of the pad for each timebin around the laser pulse is plotted as a function of the summed positive charge in the stack.

5 Common-mode effect analysis

The understanding and correction of the common-mode effect (see Subsection 3.4.1) in the GEM-based TPC using the laser data collected during the pre-commissioning phase is the subject of this chapter. In Section 5.1, the common-mode effect is demonstrated. Then, the analysis strategy is divided into two individual steps. Primarily, the physics behind the effect must be fully understood. For this, Machine Learning (ML) techniques are used to study and quantify all possible dependencies (Section 5.2). Subsequently, realistic correction models that can be programmed in the FPGA-based CRUs using only information available at the online stage are developed (Section 5.3).

5.1 Demonstration of the common-mode effect

In Figure 5.1 the laser tracks for bundle 1 of OROC C13 are shown in the pad-row coordinate system, after averaging over 344 events and subtracting the pedestal. The color axis indicates the maximum signal for each pad in the bundle 1 timebin range [100, 250). It is observed that, due to imperfections of the micro-mirror bundle, the intensity varies significantly from beam to beam. The two gaps visible along each track (at row 96 and 126) correspond to the dead regions at the stack edges. The three different track segments of each track appear slightly misplaced because the pad dimensions differ from stack to stack (see Table 3.1).

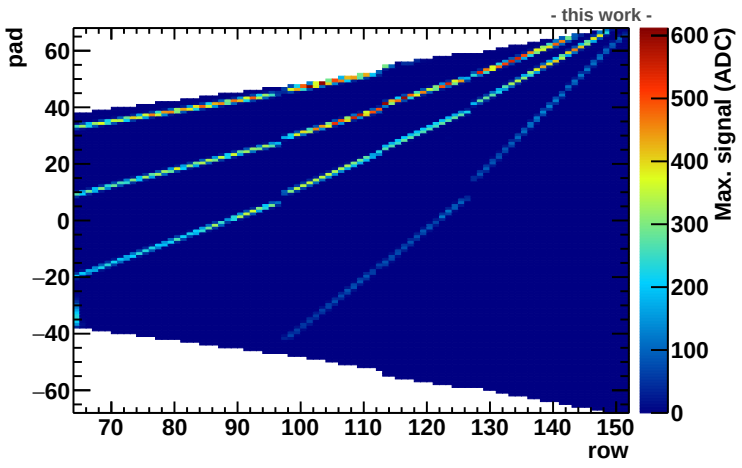


Figure 5.1: Laser tracks of OROC C13 for bundle 1 in the pad-row coordinate system. The color axis indicates the maximum signal in the timebin range [100, 250).

Figure 5.2 shows the responses for all pads of row 81 of sector C13 for bundle 1. The three laser tracks, also shown in Figure 5.1, create three charge clusters. In the rest of the pads the simultaneous common-mode signal is detected as an undershoot. After the signal pulse, the long ion tail, which is the subject of Chapter 6, is also clearly seen in the signal-pads.

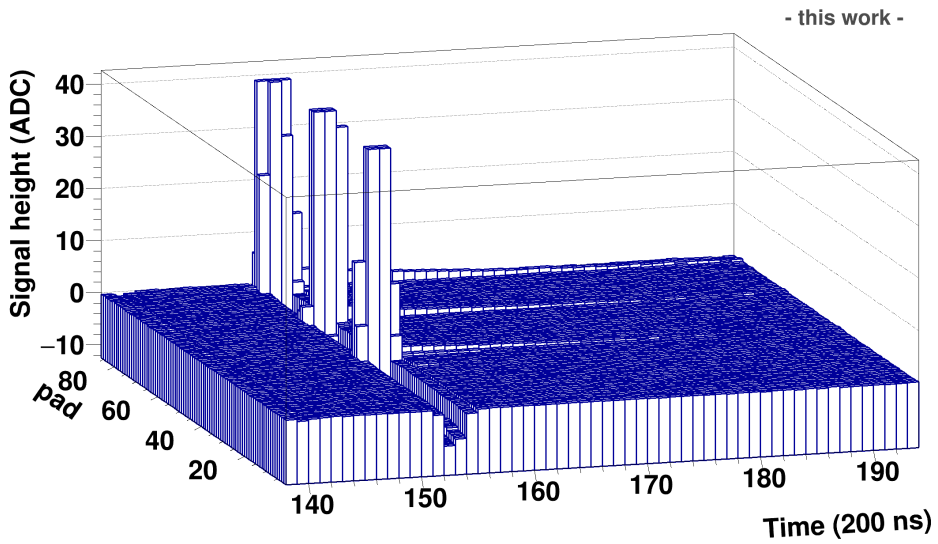
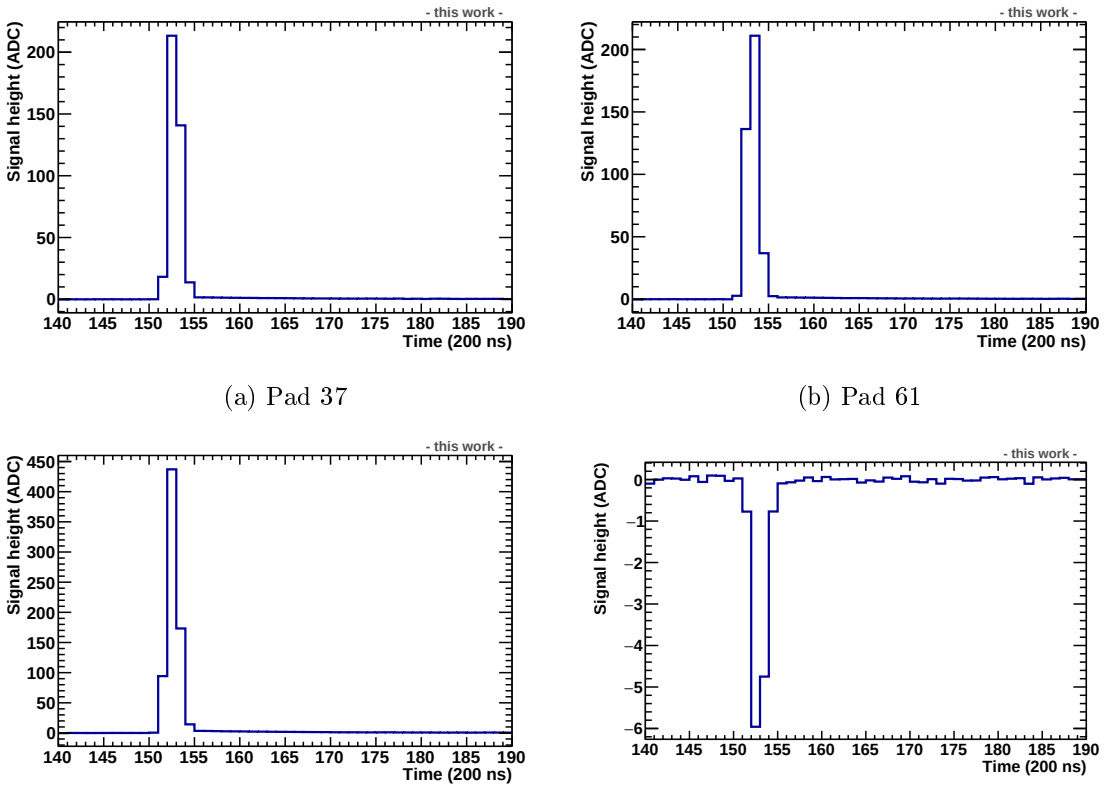


Figure 5.2: Laser signals and induced common-mode signals in the pads of row 81 of sector C13, for bundle 1. The signal height axis is zoomed-in (see Figure 5.1 for the actual signal heights).

The common-mode effect can be better visualized by plotting the responses of individual pads. In Figure 5.3, (a), (b) and (c) correspond to the pads at the center of the three laser clusters, while (d) corresponds to a common-mode (or *non-signal*, or *empty*) pad. It can be seen that the signal peaks extend over 3 timebins, however due to imperfections of the micro-mirror bundle, the individual beams are slightly misaligned along z and therefore have very different time profiles (see as an example the COG in time variations illustrated in Figure 4.11). By different time profiles it is meant that the position of the signal maximum and the distribution of the charge around the peak position differ from track to track. The undershoot in each pad is described by its response to the summed positive signal in the stack. Since the common-mode effect affects all pads of a given stack, common-mode signals are also superimposed to the laser pads, i.e. the real laser signal is slightly larger than the measured one.



(a) Pad 37

(b) Pad 61

(c) Pad 82

(d) Pad 40

Figure 5.3: Pad responses of four pads of row 81 of sector C13. (a), (b) and (c) correspond to the pads at the center of the charge clusters, while (d) corresponds to a non-signal pad. Imperfections of the micro-mirror bundle lead to different time profiles of the signal-pads. Note the different y-ranges.

5.2 Characterization of the common-mode effect

As a first step in the common-mode effect analysis, the machine learning [41] *Random Forest* (RF) algorithm [42] was used, as implemented in the ROOT interactive framework [43]. The main advantage of this approach is that, at this first stage, there is no need to provide a model which will describe the data. Additionally, the interactive aspect of the framework proved extremely helpful in the visualization and comprehension of the effect.

All possible dependencies of the common-mode fraction (CF) were explored. For a given timebin t, the CF is the ratio between the common-mode response of an individual non-signal pad, $Q(t)$, normalized to the average positive signal in the stack in the same timebin,

$$\mathrm{CF}(t) = \frac{Q(t)}{\langle \mathrm{signal}(t) \rangle_{\mathrm{stack}}} \ , \tag{5.1}$$

where

$$\langle \text{signal}(t) \rangle_{\text{stack}} = \sum_{\substack{Q(t)>0 \\ \text{in stack}}} Q(t) \Big/ N_{\text{pads in stack}} \ . \tag{5.2}$$

In the laser data, most of the signal is measured in three timebins around the peak maximum (see for example Figure 5.3). Therefore, for the CF determination, instead of merely using the charge maximum, $Q_{\text{max}} = Q(t_{\text{max}})$, it is preferred to use the charge summed in three timebins around the laser peak. This method, which results in a smaller relative uncertainty, is possible due to selection criteria on the quality of the template fits (see Subsection 4.4.3), which ensure that the timing information is consistent. Therefore for the analysis, the common-mode fraction is defined as

$$\text{CF} = \frac{Q_{\text{tot}}}{\langle \text{signal} \rangle_{\text{stack}}} \ , \tag{5.3}$$

where Q_{tot} is the charge of an individual pad summed for three timebins around the peak maximum,

$$Q_{\text{tot}} = Q(t_{\text{max}} - 1) + Q(t_{\text{max}}) + Q(t_{\text{max}} + 1) \ , \tag{5.4}$$

and

$$\langle \text{signal} \rangle_{\text{stack}} = \sum_{\substack{Q_{\text{tot}}>0 \\ \text{in stack}}} Q_{\text{tot}} \Big/ N_{\text{pads in stack}} \ . \tag{5.5}$$

The normalization in the CF with respect to $\langle \text{signal} \rangle_{\text{stack}}$ is performed since it is expected that the common-mode charge in a single pad is proportional to the average positive signal in the stack, at least to first order. Since in the laser data the total deposited charge in the stack varied significantly from stack index to stack index and from bundle to bundle, the normalization accounts for this proportionality.

Before the training of the algorithm[1], a subset of the data stored in the `track TTree` is selected based on the following criteria: stack indices with less than 10% noise and 1% pedestal outliers (see Figure 4.9) were selected, in order to guarantee baseline and noise properties close to the ideal ones. Additionally, since only the non-signal pads must be included in the analysis, pads with a distance from the ideal laser tracks of more than 2 cm and $Q_{\text{tot}} < 0$ ADC were selected. Finally, selection criteria based on the goodness of the template fits were applied, to reject entries with wrong timing information. From the data fulfilling these criteria only 33% were randomly selected as the training data.

[1] The *training* describes here the process during which the algorithm builds the model based on the provided data-set (known as the *training data*). Then, it can make predictions on other data-sets as well.

For the training, the chosen sample was randomly subdivided into 200 *estimators*[2], and for each estimator a decision tree of deepness 12 was generated. Dependencies of the CF on all available variables were tested. The advantage of this approach is that the algorithm calculates the *importance* of each variable that was used for the training, as listed in Table 5.1. The importance reflects how many times the RF decision tree was divided because of that specific variable. The *normalized pulser charge*, $Q_{\text{pulser}}^{\text{norm}}$ (meaning the pulser charge normalized to the mean pulser charge in the stack), as measured in the independent calibration pulser run described in Subsection 3.3.2, and the stack type account for $\approx 97\%$ of the dependencies. The variable $Q_{\text{pulser}}^{\text{norm}}$ accounts for pad-by-pad capacitance variations, while the stack type for the absolute stack capacitance, due to the different dimensions of each chamber. A second order correction is that of the average positive signal and the fraction of signal-pads in the stack, namely the *missing charge*. These two variables account for the fact that the measured positive signal (which enters in the denominator of the common-mode fraction) is smaller than the real one, due to the superimposition of the common-mode charge in the signal-pads. Since stacks with a high fraction of baseline and noise outliers are excluded from the data-set, the robustness of the baseline determination (difference between `pedestal2I90` and `pedestal2I75`, see Figure 4.7) is not significant[3]. The contribution of track-related properties such as the bundle (therefore the diffusion) and the beam (therefore the track inclination, see Subsection 4.4.1) was explored, but no notable dependence was discovered.

In Figure 5.4, the common-mode fraction data, the random forest prediction[4], and the difference between the two are plotted for the training data. A first observation is

Variable	Variable importance (%)
normalized pulser charge ($Q_{\text{pulser}}^{\text{norm}}$)	61.1
stack ID	36.1
average positive signal in stack	1.0
fraction of signal-pads in stack	0.8
robustness of baseline determination (`pedestal2I90-pedestal2I75`)	0.6
bundle	0.3
beam	0.1

Table 5.1: Variable importance for the CF, reflecting how many times the decision tree was divided because of that specific variable.

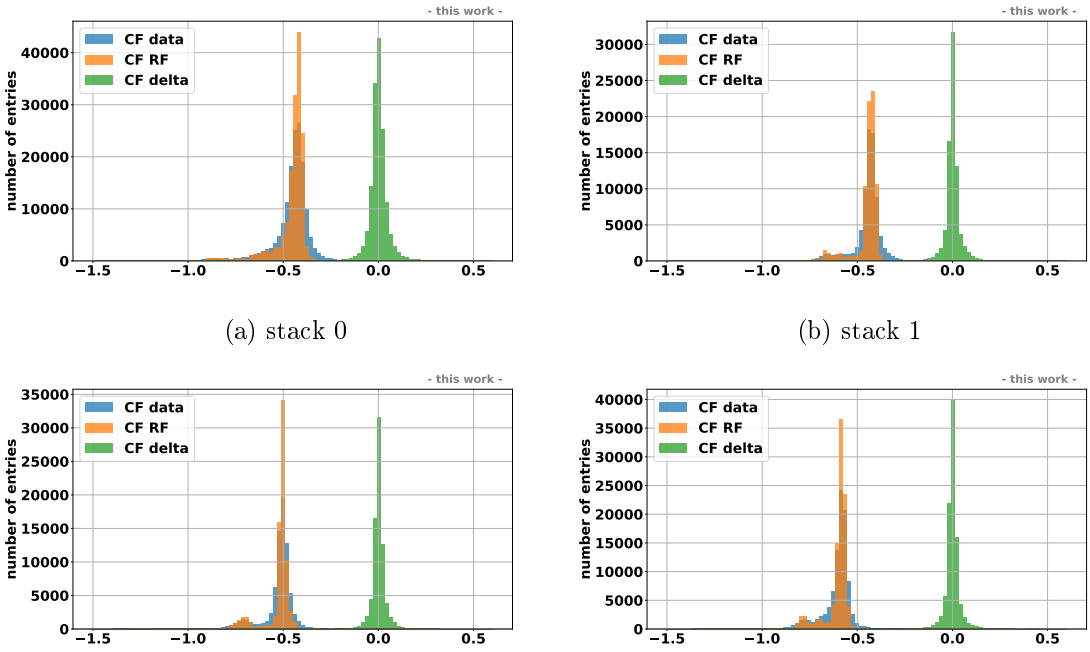
[2] See the *Bootstrap method* [44], a *random sampling with replacement* procedure which is used to draw inference on a population by studying the behavior of its sub-samples. In this method, an element may appear multiple times in the same sample.

[3] As a practical example, if one assumes that for a given pad the baseline was biased by 0.03 ADC, then Q_{tot} would be biased by 0.09 ADC. Taking a typical case, $Q_{\text{tot}} = 1.5$ ADC and $\langle \text{signal} \rangle_{\text{stack}} = 3$ ADC, the CF would be biased by 6%.

[4] The prediction is the output of the algorithm after it has been trained. The prediction can be applied to either the training data or to a new data-set.

that the peak of the CF moves towards larger (absolute) values as the stack area, and subsequently the capacitance, increases. In Table 5.2 the position of the maximum is listed for each stack type. The uncertainties correspond to the RMS value obtained by fitting each peak with a gaussian distribution. The CF spread observed within each stack is almost entirely attributable to the pad-by-pad capacitance variations, described by the pulser charge variations. In particular, a linear dependence between the CF and $Q_{\mathrm{pulser}}^{\mathrm{norm}}$ will be shown.

Apart from the main CF peak in each stack, smaller peaks are visible for all OROC stacks at higher (absolute) CF values. The CF distributions can be visually compared with Figure 5.5, where $Q_{\mathrm{pulser}}^{\mathrm{norm}}$ is shown for each stack. Because of the



(a) stack 0

(b) stack 1

(c) stack 2

(d) stack 3

Figure 5.4: (Color Online) Common-mode fraction data (blue), random forest prediction (orange) and difference between the two (green), for the data used for the random forest training.

stack	stack area (mm^2)	CF max.
0	171154	-0.42 ± 0.03
1	174853	-0.43 ± 0.02
2	231284	-0.50 ± 0.02
3	294836	-0.58 ± 0.02

Table 5.2: Position of the maximum of the CF distributions and for each stack type, as obtained from the training data.

normalization, the mean of each distribution is equal to one, however the shape and position of the peak maximum differ from stack to stack. The distribution for IROC is much broader due to substantial foil sagging. The continuous variation of the GEM4B to pad-plane distance because of the electrostatic forces results in only one main peak in case of IROC. In some cases foil sagging is also observed in OROC1, although less prominent compared to the IROC. This explains the moderately broader distribution of the OROC1 compared to the OROC2 and OROC3, where no foil sagging is seen (see Figure 3.12). The source of the outlier peaks in both Figure 5.5 and Figure 5.4 is pads at the edges of the chamber or below the spacer cross. Details on the impact of the normalized pulser charge distribution within a given stack index on the CF determination will be discussed below.

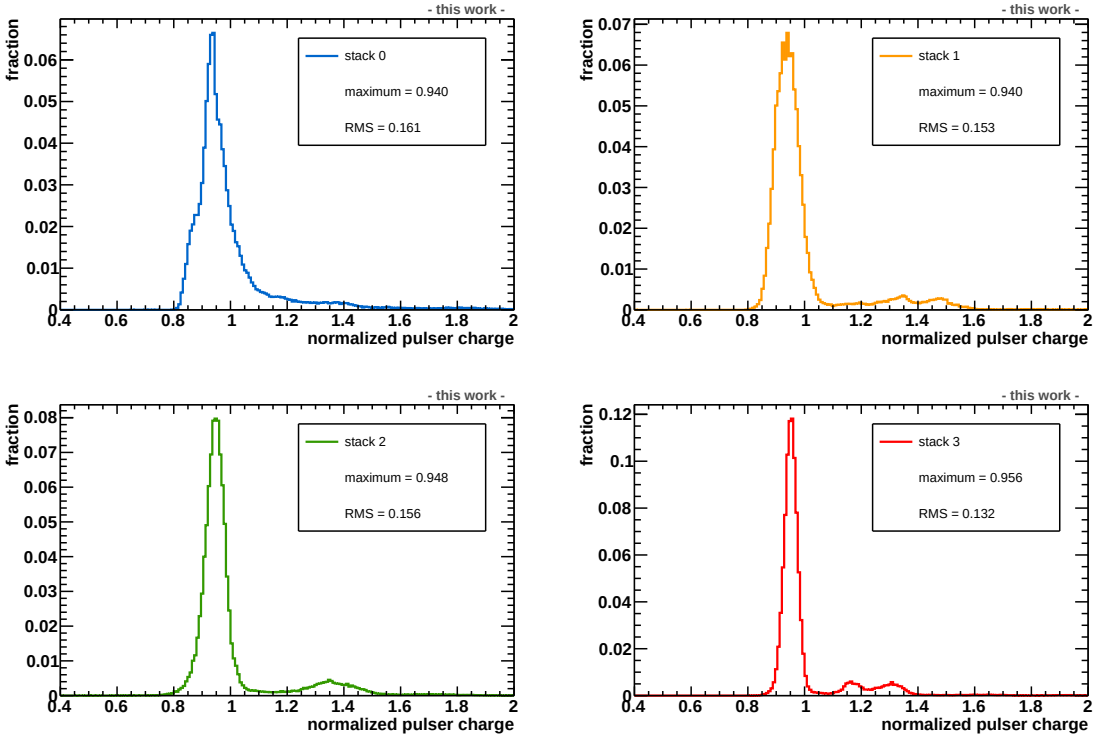


Figure 5.5: Pulser charge normalized to mean pulser charge in the stack, or $Q_{\mathrm{pulser}}^{\mathrm{norm}}$.

The CF values were compared with PSpice [45] simulations conducted by the TPC Collaboration [46]. In these simulations, the stack capacitance was approximated by that of a parallel-plate capacitor, where estimated parasitic capacitances of the pulser cable connected to GEM4B and the ground cable connected to the pad-plane were also added in parallel. The results of the simulations are listed in Table 5.3. Since the foil sagging was not accounted for, all pads have the same capacitance and therefore the CF is single-valued within a given stack. Firstly, it is observed that the position of the CF maximum follows the same trend both for the data

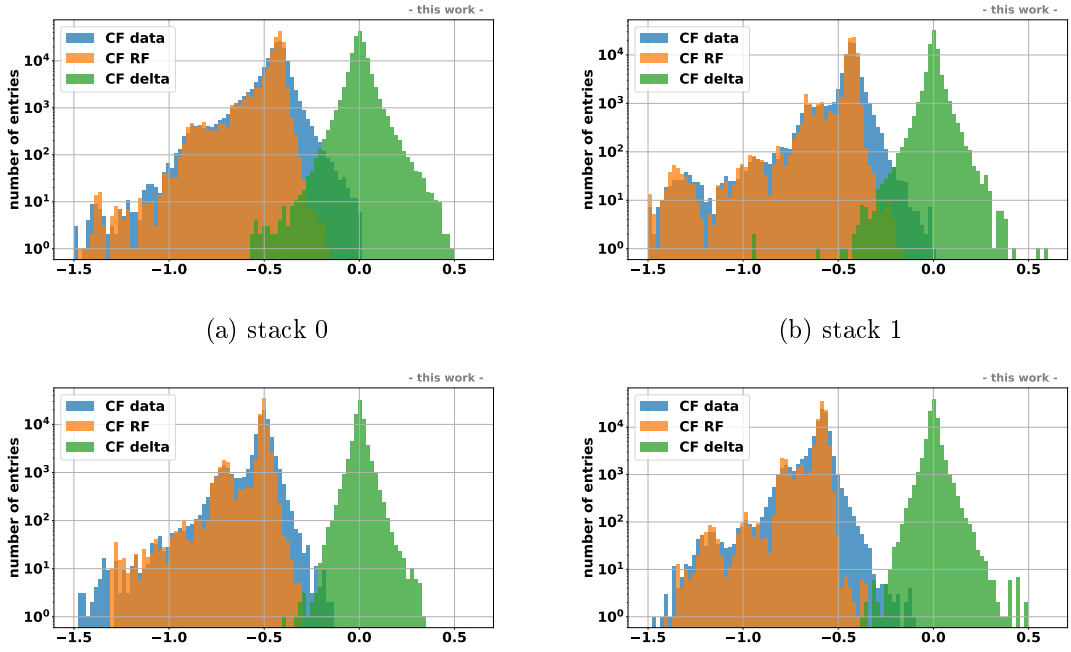
and the simulation. Systematic differences between the data and the simulation can be attributed to foil sagging, as well as additional capacitive couplings of the pad-plane to the rest of the GEM foils, which were not included in the simulations. Furthermore, it can be seen that the ratio between CF and area is not constant for both the data and the simulation, due to the aforementioned additional capacitances.

In Figure 5.6 the same quantities as in Figure 5.4 are plotted in a logarithmic scale, so that the differences between the data and the prediction are more visible. It is seen that the differences are centered around zero, but a number of data points not following the model are observed. The origin of these outliers is discussed below.

The trained RF algorithm was then used to predict the behavior of the data-set

stack	CF max. data	$\lvert CF \rvert$/area (m^{-2})	CF simulation	$\lvert CF \rvert$/area (m^{-2})	$\frac{\text{CF max. data}}{\text{CF simulation}}$
0	-0.42 ± 0.03	2.5	-0.50	2.9	0.84
1	-0.43 ± 0.02	2.4	-0.50	2.9	0.85
2	-0.50 ± 0.02	2.2	-0.56	2.4	0.89
3	-0.58 ± 0.02	2.0	-0.61	2.1	0.95

Table 5.3: Comparison of the position of the CF maximum calculated using the laser data from Table 5.2 with the CF prediction calculated via PSpice simulations.



(a) stack 0

(b) stack 1

(c) stack 2

(d) stack 3

Figure 5.6: (Color Online) Common-mode fraction data (blue), random forest prediction (orange) and difference between the two (green), for the data used for the random forest training. Note the logarithmic scale.

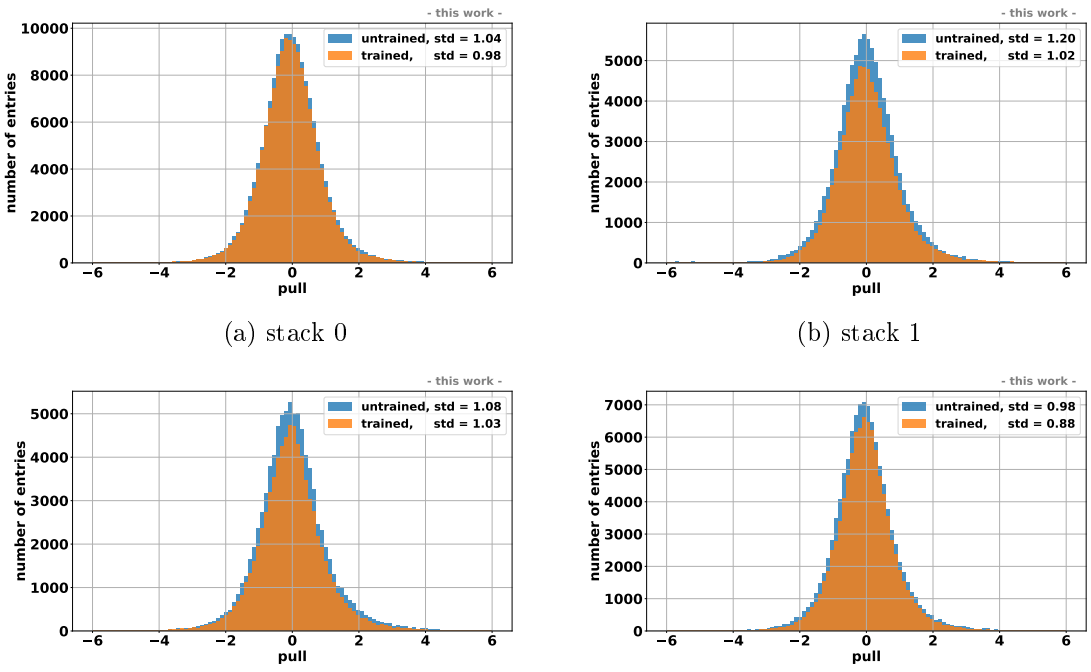
that was not used for the training, to confirm that the decision tree is not over-trained. Again, 33% of the data points fulfilling the selection criteria, excluding the training data, were randomly selected. In Figure 5.7, the pulls for the trained and the un-trained data are shown for each stack. The pull is defined as

$$\text{pull} = \frac{\text{CF}_{\text{data}} - \text{CF}_{\text{RF}}}{\text{error}\ (Q_{\text{tot}},\ \text{RF})}\ , \tag{5.6}$$

where CF_{data} and CF_{RF} are the measured common-mode fraction and the random forest prediction, respectively, while for the uncertainty both the Q_{tot} uncertainty[5]

$$\text{error}(Q_{\text{tot}}) = \sqrt{3} \cdot \text{noise}\ , \tag{5.7}$$

and the RF uncertainty[6] are summed quadratically. The pulls follow to a good approximation the expected behavior, namely a gaussian distribution with mean close to zero and standard deviation close to one. The standard deviation of each histogram, listed in the figure legend, is always larger for the un-trained data-set



(a) stack 0

(b) stack 1

(c) stack 2

(d) stack 3

Figure 5.7: (Color Online) Pulls for the trained (orange) and un-trained (blue) data.

[5]The total charge is summed for three timebins around the peak, and the uncertainty of each charge measurement is equal to the noise. Then, the uncertainty of the total charge is given by the error propagation formula.

[6]The RF uncertainty is the RMS of the distribution of predictions for each of the 200 decision trees.

compared to the trained one (5−18%, depending on the stack type). A slightly larger standard deviation could potentially hint to a small over-training of the data, or some minor neglected dependency of an effect. At this point it should be highlighted that the decision tree tries to describe the CF based only on the provided variables. If one assumes that there is an unknown variable which influences the CF but was not provided to the tree, then Table 5.1 would still look similar (i.e. the decision tree would be divided based on the – incomplete – input), however major differences would be observed between the CF data and the RF prediction. In case of over-training, only the pulls for the un-trained data-set would deviate significantly from the expected behavior. The above argument justifies the importance of the Machine Learning studies: to discover all dependencies of the CF and their importance, and to exclude all variables which the CF does not depend on.

In Figure 5.8 the dependence of the CF on $Q_{\mathrm{pulser}}^{\mathrm{norm}}$ is shown for each stack, only for data points with a CF uncertainty of less than 5%. A linear dependence is observed, which is expected as the pulser charge is proportional to the pad capacitance. The proportionality holds well even for pads with much larger capacitance (corresponding to the red regions of Figure 3.13). The capacitance differences result in slope differences between the stacks.

Some data points which do not follow the expected behavior are observed in Figure 5.8. The geometrical position of these outliers can be identified by plotting the pads with a pull value larger than e.g. 4·RMS, as shown in Figure 5.9 for the C-side. It can be seen that many of the outliers correspond to reflections of the laser beams in the inner field cage. These reflected tracks are not registered in the OCDB, consequently the applied selection on the distance of the pads to the ideal laser tracks cannot exclude these points. Another source of outliers is pads very close to the laser tracks, corresponding to the edges of the clusters. In both cases the pad signal is very weak, therefore due to the superimposed common-mode, the measured charge in these pads is still negative. The exclusion of weak signal pads with negative

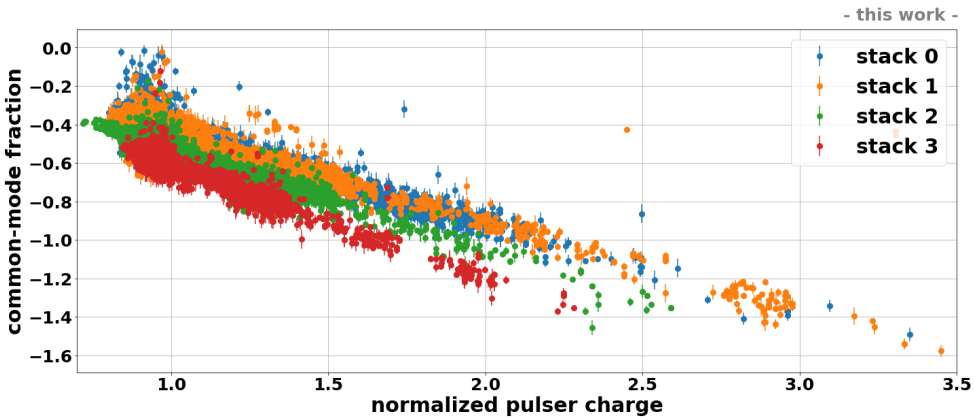


Figure 5.8: (Color Online) Common-mode fraction as a function of the normalized pulser charge for each stack, only for the data with a CF uncertainty of less than 5%.

total charge is discussed in Subsection 5.3.3. Finally, some outliers correspond to pads where the pulser calibration failed (discussed in Subsection 5.3.1).

From the Machine Learning studies, the following conclusions can be made: the common-mode fraction of a given pad mainly depends on the (time-independent) pad to GEM capacitance (see Table 5.1). The pad to GEM capacitance can be described by the normalized pulser charge as obtained from the calibration pulser runs (which reflects the pad-by-pad capacitance variations within a given stack) and the stack type (which accounts for the absolute capacitance). The dependence on the pad to GEM capacitance is almost perfectly linear (see Figure 5.8). A bias in the common-mode fraction estimation in the laser data comes from the missing charge, namely the average positive signal and the number of signal pads in the stack. Accounting for the above variables only is sufficient for the description of the observed data to a good degree (see Figure 5.6 and Figure 5.7).

The interactive visualization for the dependence of the CF on $Q_{\mathrm{pulser}}^{\mathrm{norm}}$ is available in an html file [47]. A video with explanation of the plotted data is also available [48].

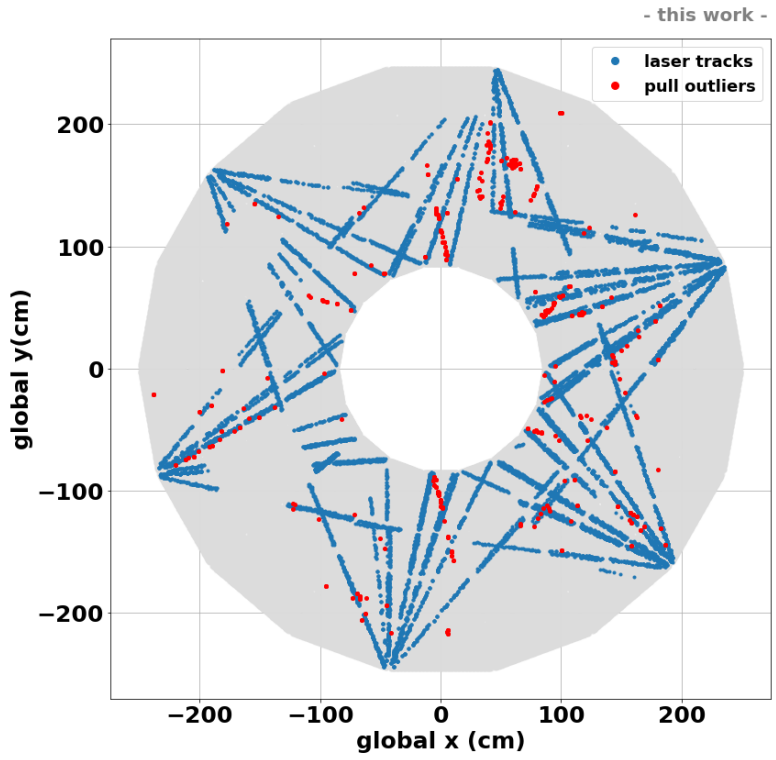


Figure 5.9: (Color Online) Position of pull outliers (|pull| $>$ 4·RMS) for all C-side pads derived from both the trained and un-trained data, in red. These points fulfill the selection criteria, namely $Q_{\mathrm{tot}} < 0$ and a distance from the ideal laser tracks of more than 2 cm. For visualization, the laser tracks are shown in blue.

5.3 Online correction in the CRUs

After the full understanding of the common-mode effect and its dependencies using the random forest algorithm, the next step is to develop correction models that can be programmed in the FPGA-based CRUs. It is highlighted that the actual hardware implementation is beyond the scope of this thesis.

As was discussed in Subsection 3.4.3, in order to recover the PID performance of the TPC, the common-mode effect requires a correction already at the online stage. The multiplexed data streamed from the FEE are decoded in the CRUs, and subsequently the pedestal subtraction and common-mode correction are applied, before the zero-suppression. For the common-mode correction in the CRUs, only simple mathematical operations are possible, using only the information that is available at this stage. A signal detection algorithm has been programmed in the CRUs, which identifies the peaks using first and second derivative cuts in the time direction. Therefore, for a given timebin, the different pads can be classified as either "signal" or "non-signal" pads. Consequently, within a CRU, the number of signal-pads and the average positive signal can be calculated on a timebin-by-timebin basis.

Figure 5.10 illustrates the CRU segmentation of one sector. The different stacks are shown with different colors. Due to the large number of pads, the IROC is read out by four different CRUs, while each OROC stack by two CRUs (see Table 3.1 for the chamber specifications). With this scheme, each CRU reads the data of

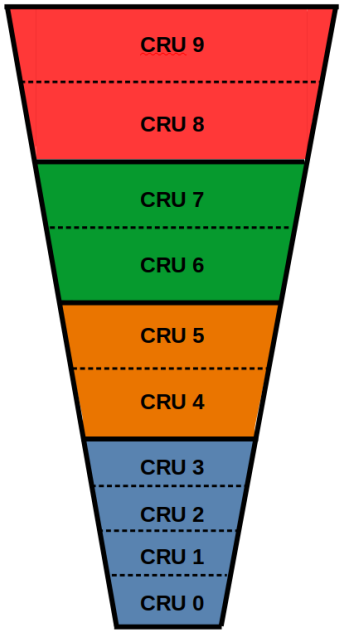


Figure 5.10: CRU segmentation of one TPC sector. The different stacks are shown with different colors. The IROC is read out by four different CRUs, while each OROC stack by two.

1300−1600 pads. With the current setup, information cannot be exchanged between different CRUs, therefore the additional CRU segmentation of each stack has severe implications on the common-mode correction algorithm. In particular, for a given timebin, the average signal and the number of signal-pads cannot be determined for an entire stack, since this would require combining information from different CRUs, which in turn would introduce undesirable latencies. In order to demonstrate the impact of the above, two models have been developed.

In *Model 1*, it is assumed that combining information from different CRUs is possible. Apart from the pulser charge, which can be provided as a static map to the CRUs[7], the common-mode fraction determination can also account for the missing charge contributions in a given timebin, namely the average positive signal in the stack and the number of signal-pads,

$$\mathrm{CF}_1(t) = a + b \cdot Q_{\mathrm{pulser}}^{\mathrm{norm}} + c \cdot R_{\text{signal pads}}(t) + d \cdot \langle \text{signal}(t) \rangle_{\text{stack}} \quad , \tag{5.8}$$

where $R_{\text{signal pads}}(t)$ is the fraction of signal-pads in the stack for a given timebin and $\langle \text{signal}(t) \rangle_{\text{stack}}$ is given by Equation 5.2. Since an estimation of the absolute capacitance is not possible (due to additional couplings to the rest of the GEM foils, as well as to the ground cable and pulser cable), the parameter b must be fitted separately for the different stack types. Then, to calculate the common-mode charge for a given pad and timebin, following Equation 5.1, the CF model prediction is multiplied by the average laser signal in the stack, which is assumed to be available:

$$Q_{\mathrm{CM},1}(t) = \mathrm{CF}_1(t) \cdot \langle \text{signal}(t) \rangle_{\text{stack}} \quad . \tag{5.9}$$

In *Model 2*, the common-mode fraction determination can only account for the static pulser charge,

$$\mathrm{CF}_2 = a + b \cdot Q_{\mathrm{pulser}}^{\mathrm{norm}} \quad . \tag{5.10}$$

Note that the CF_2 for a given pad is time-independent, while $\mathrm{CF}_1(t)$ accounts for the (time-dependent) missing charge information. As expected also from the random forest parameter importance (Table 5.1), the missing charge should have a relatively small impact on the CF determination. However, the real challenge in Model 2 is the calculation of the common-mode charge in a given pad, since in Model 2 the average laser signal in the stack is not available (compare to Model 1, Equation 5.9). In that case, the baseline has to be estimated using the *empty pads* as identified by the signal detection algorithm. In particular, accounting for the pad-by-pad baseline variations predicted by CF_2, the common-mode response of a given pad in a given timebin can be estimated by

[7] Once the TPC is re-installed in the cavern, another set of calibration pulser measurements will take place. The capacitance is constant over time, therefore the pulser charge can be provided as a static map to the CRUs.

$$Q_{\mathrm{CM},2}(t) = \mathrm{CF}_2 \cdot \left\langle \frac{Q_{\mathrm{CM}}^{\mathrm{meas}}(t)}{\mathrm{CF}_2} \right\rangle_{\text{empty pads}} , \tag{5.11}$$

where $Q_{\mathrm{CM}}^{\mathrm{meas}}(t)$ is the measured common-mode charge in an empty pad for timebin t. The averaging is ideally conducted over all available empty pads.

The scope of this study is to determine the parameters of Model 1 and Model 2 using the laser data, and to verify whether these linear models are sufficient to describe the measured common-mode fraction.

5.3.1 Outlier removal

The machine learning studies, through a comprehension of the origin of the outliers, allowed for a refinement of the selection criteria applied to the laser data that were then used for determining the parameters of Model 1 and Model 2. Apart from the selection discussed in Section 5.2, it was decided to increase the distance cut from the ideal laser tracks to 3 cm, in order to avoid some weak signal pads close to the laser tracks (see Figure 5.9). Secondly, only pads with err(CF) $< 5\%$ were used.

It was observed that another source of outliers is pads where the pulser calibration failed. In such cases, either the front-end electronics were not properly configured, or the CRU readout cards were not properly initialized [49]. These pads were identified as the outliers in the weighted mean (t_0) distribution of the pulser signal. The weighted mean is calculated in ± 2 timebins around the maximum, similarly to Equation 4.1. Figure 5.11 shows the t_0 outliers on the A- and C-side, namely the pads where the difference between t_0 and its average in the stack exceeds 3-sigma ≈ 0.1 timebins. This additional cut rejects 2% of the pads which passed all the other

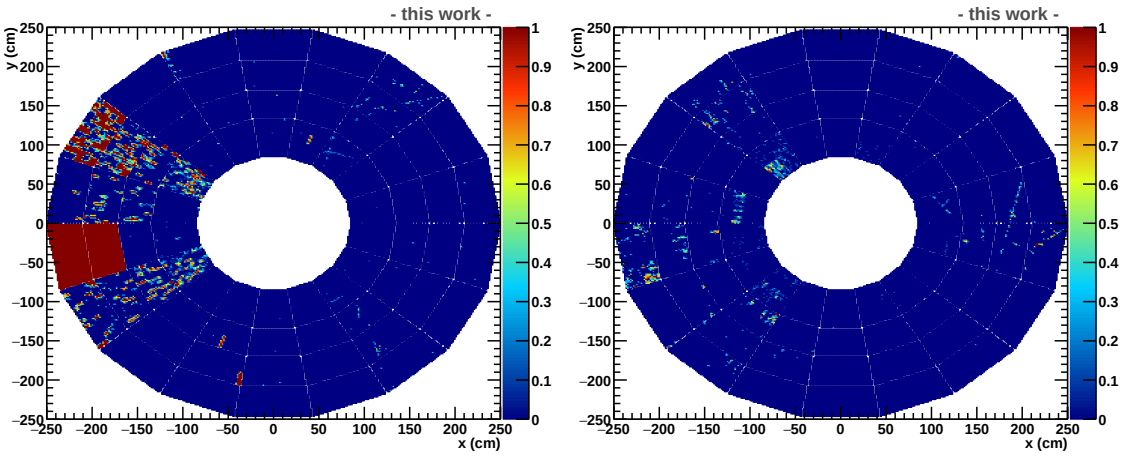


Figure 5.11: Pulser t_0 outliers on the A-side (left) and C-side (right). The color axis indicates the fraction of outliers within a bin.

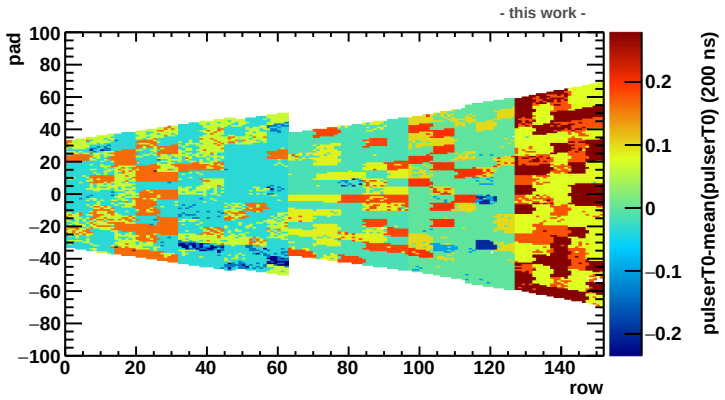


Figure 5.12: Difference between pulser t_0 and its average over the entire stack, for sector A07 in the pad-row coordinate system.

selection criteria. Figure 5.12 shows the aforementioned time differences for sector A07, where a high number of t_0 outliers is observed. The structures correspond to individual SAMPA chips.

5.3.2 Performance of common-mode fraction models

Using the selected data-set, the parameters of Model 1 and Model 2 are determined. Similarly to the machine learning studies, for the calculation of the CF via the laser data the charge summed in three timebins around the laser peak is used (Equation 5.3), which is possible due to selection criteria on the quality of the template fits. However, it should again be emphasized that in the CRUs the common-mode effect will be corrected on a timebin-by-timebin basis (Equation 5.1). The fit results for the two models are listed in Table 5.4 and Table 5.5. The typical range of each corresponding variable is also listed in the tables. In particular, $Q_{\mathrm{pulser}}^{\mathrm{norm}}$ mostly takes values in the range (0.9, 1.1) ADC (see Figure 5.5). In the laser data, the fraction of signal pads is typically $5-15\%$ of the total pads in the stack, while the average positive signal in the stack is $\approx 3-30$ ADC. A rough estimation of the resulting (absolute) contribution of each variable to the CF is also shown. This range is for simplicity calculated by taking the minimum parameter of the four stacks and multiplying by the lower bound of the corresponding variable range, and then taking the maximum parameter and multiplying by the upper bound of the variable range[8]. As was also expected from the random forest prediction, most of the dependency in the CF is due to the normalized pulser charge and the absolute capacitance. The reduced chi-square values[9], χ^2_ν, are slightly larger for Model 2, indicating that the missing charge is a second order correction to the model. Fits with a constrained constant ($a = 0$) were also explored, where it was observed that

[8] As an example, for the Model 2 $Q_{\mathrm{pulser}}^{\mathrm{norm}}$ contribution to the CF, the calculation is $(|-0.4369| \cdot 0.9,\ |-0.5847| \cdot 1.1) = (0.39, 0.64)$.

[9] The reduced χ^2 is defined as the χ^2 value per degree of freedom.

stack	a	b $\left(Q_{\text{pulser}}^{\text{norm}}\right)$	c $(R_{\text{signal pads}})$	d $(<\text{signal}>)$	χ^2_ν
0	0.0485	-0.4808	-0.2552	-0.0008	1.4611
1	0.0125	-0.4375	-0.3964	0.0001	1.5865
2	0.0035	-0.5117	-0.2991	0.0004	1.2583
3	-0.0024	-0.5842	-0.3737	0.0005	1.0768
Variable range		$0.9-1.1$	$0.05-0.15$	$3-30$	
Contribution to CF		$0.39-0.64$	$0.01-0.06$	<0.02	

Table 5.4: Model 1 fit parameters. See text for details.

stack	a	b $\left(Q_{\text{pulser}}^{\text{norm}}\right)$	χ^2_ν
0	0.0178	-0.4793	1.6856
1	-0.0210	-0.4369	1.9145
2	-0.0203	-0.5108	1.4257
3	-0.0242	-0.5847	1.1714
Variable range		$0.9-1.1$	
Contribution to CF		$0.39-0.64$	

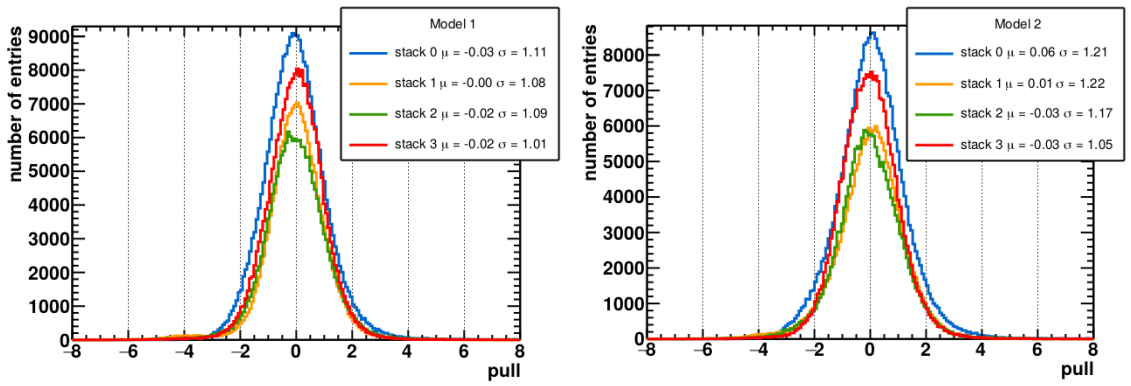
Table 5.5: Model 2 fit parameters. See text for details.

the reduced chi-square and residuals are larger. The interpretation of a non-zero constant a is discussed in Subsection 5.3.3. It is highlighted that IROC and OROC1, despite their almost identical area (see Table 5.2) have quite different slopes. This is also addressed in Subsection 5.3.3.

The pulls for each stack, defined as

$$\text{pull} = \frac{\text{CF}_{\text{data}} - \text{CF}_{\text{model}}}{\text{error}\ (Q_{\text{tot}})}\ , \tag{5.12}$$

where error (Q_{tot}) is given by Equation 5.7, are shown in Figure 5.13 for the two different models. The mean and standard deviation of the histograms are listed in the legend. Good agreement with the expected behavior is observed (compare with the pulls obtained via the random forest algorithm, Figure 5.7). Although Model 1 results in standard deviations closer to unity since it can account for the slight biasing of the CF due to the missing charge, Model 2, with standard deviations $\approx 4-15\%$ larger than Model 1 (depending on the stack type), is also sufficient in describing the data. Deviations of the distributions from the ideal behavior are a result of insufficient exclusion of the weak laser signals, as well as stack index by stack index capacitance variations, both of which are addressed below.



(a) (b)

Figure 5.13: (Color Online) Pulls for Model 1 (a) and Model 2 (b).

5.3.3 Stack index–by–stack index and laser–by–laser fits

Small capacitance variations between the different chambers can be expected due to manufacturing imperfections. These can be identified by calculating the fit parameters on a stack index by stack index basis.

Figure 5.14 shows as an example the data points for IROC C04 and laser ID 213. With red all data points (including the signal-pads) are shown, while overlaid in black are only the points passing the data-selection as described in Subsection 5.3.1. All points with CF > 0 (therefore $Q_{\mathrm{tot}} > 0\,\mathrm{ADC}$) are excluded as signal-pads, and so are pads close to the ideal laser tracks (weak signal-pads with negative charge due to the superimposed common-mode). However, it can be seen that the selection criteria still fail to discard some weak signal pads due to the reflected laser tracks in the inner field cage. These points have a CF value above the expected nominal

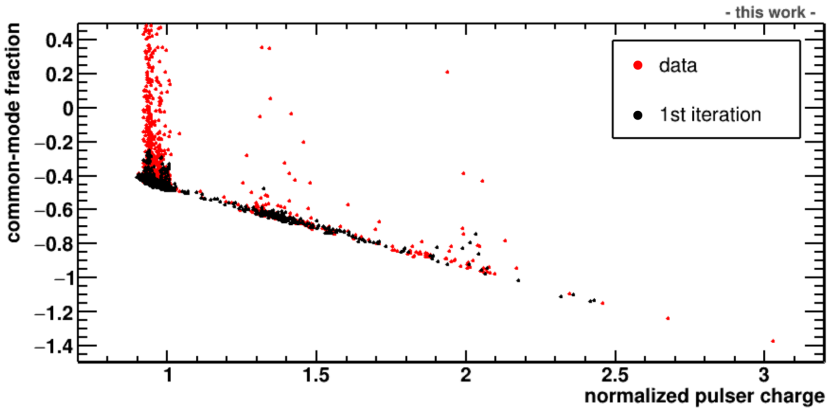


Figure 5.14: (Color Online) IROC C04 data points for laser ID 213 (red) and overlaid data points used for the 1$^{\mathrm{st}}$ iteration fitting (black).

(linear) behavior, mostly observed in the region $Q_{\text{pulser}}^{\text{norm}} \approx 1$[10]. These data points can bias the fitting results by shifting the slope and introducing a larger fit constant.

To further improve the fit quality, a first option is to vary the robustness of the fit. In the robust fits, a pre-defined fraction of the data-set is selected which minimizes the chi-square. Since the weak laser signal pads are outliers, they will not be included in the fitting process. However, depending on the data points available for a given stack index and laser ID, the same robust option might sometimes also exclude data points with large normalized pulser charge. Note that the high-capacitance points are "outliers" of the pulser charge distribution, however in these linear fits they define the lever arm of the fit and consequently are essential in describing the high-capacitance range. Alternatively, another option is to perform a second iteration fit, cutting on data points with large residuals[11] from the first iteration fit. This is demonstrated in Figure 5.15, where the blue points correspond to the selected data points after a 3-sigma cut on the residuals from the 95% robust fit. It is observed that, in the data points selected for the 2$^{\text{nd}}$ iteration fitting, most of the weak signal pads are successfully discarded. In most cases, the second iteration rejects less than 2% of the data points and improves the quality of the fit.

Different fit options ($80\% - 100\%$ robust fit, 1$^{\text{st}}$ and 2$^{\text{nd}}$ iteration, constrained or non-constrained linear fit constant) were compared with each other. It was concluded that a 2$^{\text{nd}}$ iteration 95% robust fit is usually ideal to describe the data. With these settings, most of the weak signal pads are removed, without rejecting the high capacitance data points. It was additionally seen that the fit with a non-zero constant describes the data points better. This will be shown in the following examples, which motivate the differential studies.

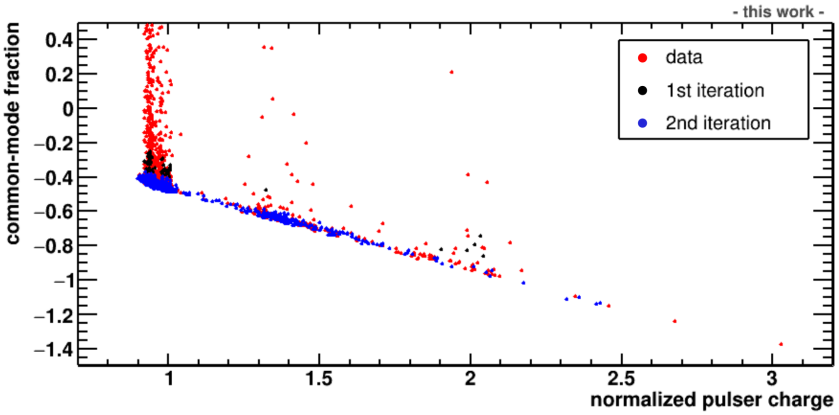


Figure 5.15: (Color Online) IROC C04 data points for laser ID 213 (red), overlaid data points used for the 1$^{\text{st}}$ iteration fitting (black), and overlaid data points used for the 2$^{\text{nd}}$ iteration fitting (blue), as described in the text.

[10] These outliers are concentrated in the $Q_{\text{pulser}}^{\text{norm}} \approx 1$ region simply because most pads have a $Q_{\text{pulser}}^{\text{norm}}$ value close to unity. In this example, one can also see some outliers at $Q_{\text{pulser}}^{\text{norm}} \approx 2$.

[11] The residual for a given pad is the difference between the CF value calculated from the data and the linear model prediction.

Stack index by stack index fits

Figure 5.16 shows the $Q_{\text{pulser}}^{\text{norm}}$ distributions in the pad-row coordinate system for IROCs C04 and C05. In C05 significant foil sagging is observed, which is absent in C04. In the latter, only the stack edges and pads below the spacer cross exhibit a larger capacitance. In Figure 5.17, the selected data points used for fitting the data from two different laser IDs are shown for the two chambers. In the case of IROC C04 most data points are concentrated in the region $Q_{\text{pulser}}^{\text{norm}} \approx 0.95$, while in IROC C05, due to foil sagging, the lever arm is larger, extending over the interval $Q_{\text{pulser}}^{\text{norm}} \approx 0.85 - 1.20$. In the ideal case, both chambers should have very similar fit parameters, however, as can be seen in Table 5.6, a difference in the slopes is observed. In the table, the RMS of the residuals is shown, which is almost identical for the two cases. The column "slope+const" will be explained below.

It is therefore demonstrated that, within the same stack type, due to differences in the pulser charge distributions between the different stack indices, stack index by stack index fits describe the data better than fitting all stack indices simultaneously, as was done in Subsection 5.3.2.

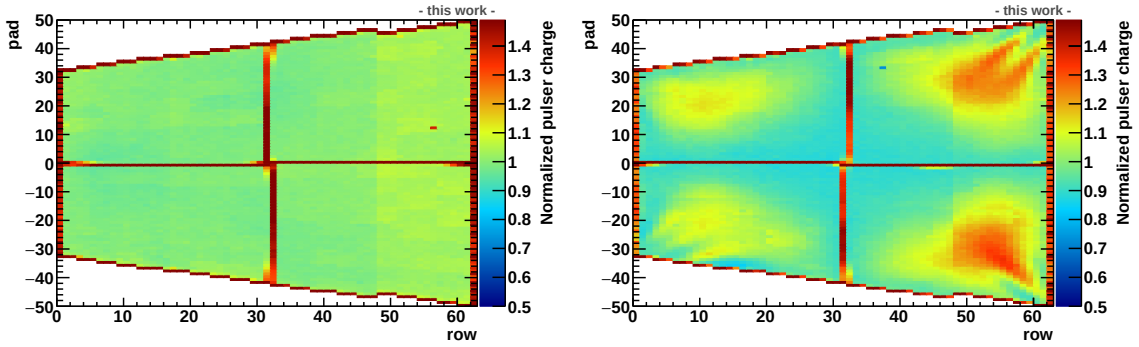


Figure 5.16: Normalized pulser charge in the pad-row coordinate system for IROCs C04 (left) and C05 (right). The two outlier pads are caused by a malfunction of the SAMPA analog front-end.

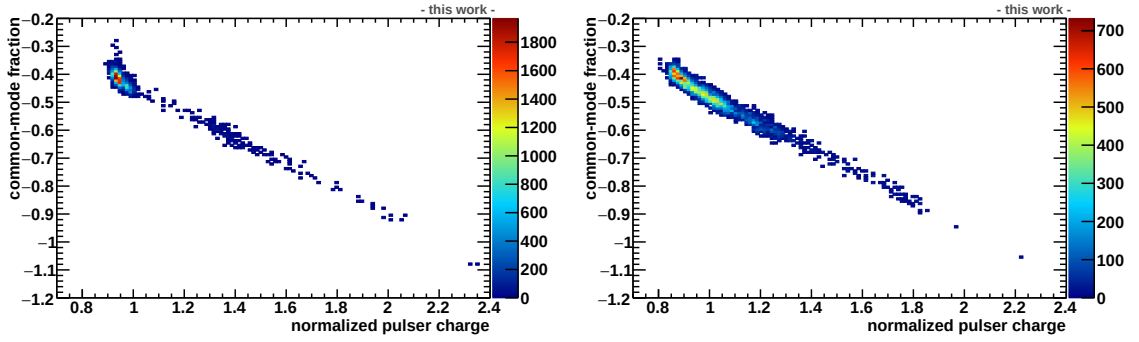


Figure 5.17: Data points used for fitting of data from sector IROC C04 laser ID 199 (left) and IROC C05 laser ID 232 (right).

sector	laser ID	slope	const	slope+const	RMS	$< \text{signal} >_{\text{stack}}$
C04	199	-0.4730	0.0308	-0.4423	0.0171	7.3854
C05	232	-0.5037	−0.0295	-0.4741	0.0172	5.0000

Table 5.6: Fit results for two laser IDs of sectors C04 and C05. See text for details.

Laser by laser fits

Another example, demonstrating the laser ID by laser ID fit differences for the same chamber is shown in Figure 5.18, where the data points used for fitting data from two laser IDs of IROC C13 are plotted. The reflected laser tracks in the inner field cage (see Figure 5.9) result in weak laser signals observed in the data of laser ID 284, which are absent in the case of laser ID 291. These data points, despite the robust fitting used, influence the results of the fits. More importantly, they bias the estimation of the average positive signal in the stack, which defines the CF (Equation 5.3). This biasing results in smearing of the data points around the nominal behavior. The RMS of the residuals, listed in Table 5.7, is more than two times larger for the case of laser ID 284.

It is therefore clear that the quality of the fit is heavily determined by the available data points for a given laser ID. In cases where a few laser IDs are available for a given stack index, the fit with the smallest RMS can be selected as the one best describing the data.

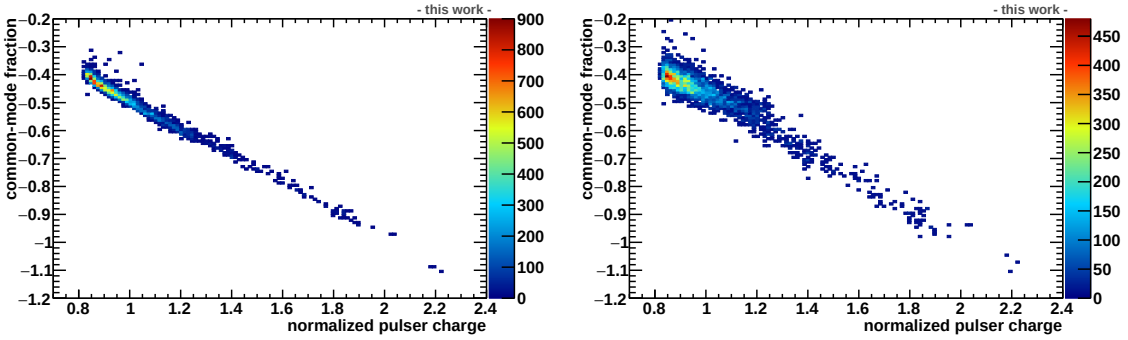


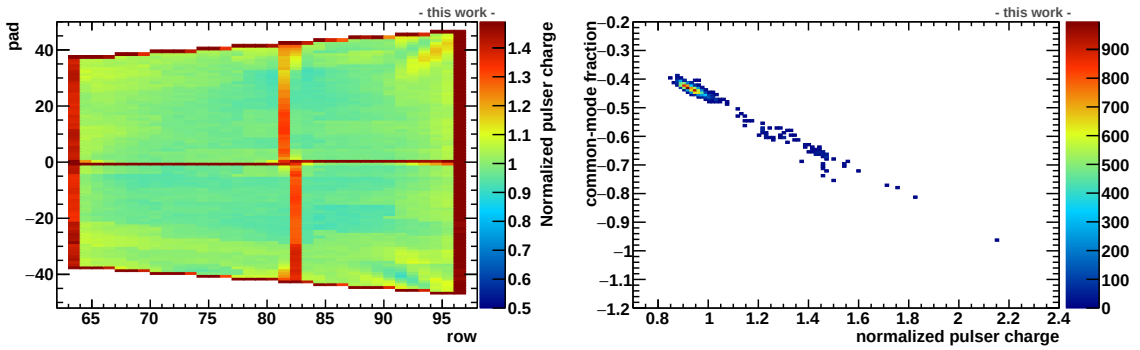
Figure 5.18: Data points used for the fitting for sector C13 laser ID 291 (left) and 284 (right).

stack	laser ID	slope	const	slope+const	RMS	$< \text{signal} >_{\text{stack}}$
0	284	-0.4641	-0.002	-0.4664	0.0329	2.5293
0	291	-0.5050	0.0134	-0.4850	0.0143	3.1783
1	291	-0.4357	-0.0246	-0.4603	0.0093	7.4768

Table 5.7: Fit results for laser IDs of C13 IROC (stack 0) and OROC1 (stack 1). See text for details.

IROC and OROC1 fit differences

Additionally, an explanation of the slope differences observed between IROC and OROC1, despite their almost identical area, is demonstrated below. Figure 5.19a shows the normalized pulser charge distribution for OROC1 of C13. In contrast to the IROCs, in all OROC1 stacks little or no foil sagging is observed (see also Figure 3.12). Again, most data points are concentrated in the region $Q_{\text{pulser}}^{\text{norm}} \approx 0.95$, which defines the lever arm of the fit. The fit results are also listed in Table 5.7. Additional differences between the IROC and OROC1 fit parameters could be explained by a small systematic bias of the calibration pulser measurement. The bias is larger for the IROC, where the measured pulser charge is about 50% smaller than in the OROC1.



(a) (b)

Figure 5.19: (a) Normalized pulser charge distribution for OROC1 of sector C13 in the pad-row coordinate system. (b) Data points used for the fitting of laser ID 291.

Since for all stack indices the majority of the pads have a $Q_{\text{pulser}}^{\text{norm}} \approx 0.9 - 1.1$, the constant term in the model can be interpreted as a "correction" to the slope variations, determined by the lever arm available for each fit. This is demonstrated in Figure 5.20 where in the top the slopes for each fit are shown as a function of the stack index, while in the bottom, the sum of the constant and the slope. Only fit parameters with RMS$<$ 0.02 are included. Within a given stack type, there is a large stack index by stack index and laser ID by laser ID slope variation due to the aforementioned effects, which is partially compensated by the constant term. In Figure 5.21 the distribution of the sum of the constant and slope for each fit is shown separately for each stack. The ordering of the values is in agreement with the stack areas, as listed for example in Table 5.2.

As a conclusion, it was demonstrated that, due to differences in the pad-by-pad capacitance distributions within individual stack indices, it is optimal to determine the fit parameters separately for each stack index. Within a given stack index, in case where multiple laser IDs are available, the fit parameters slightly vary. Weak

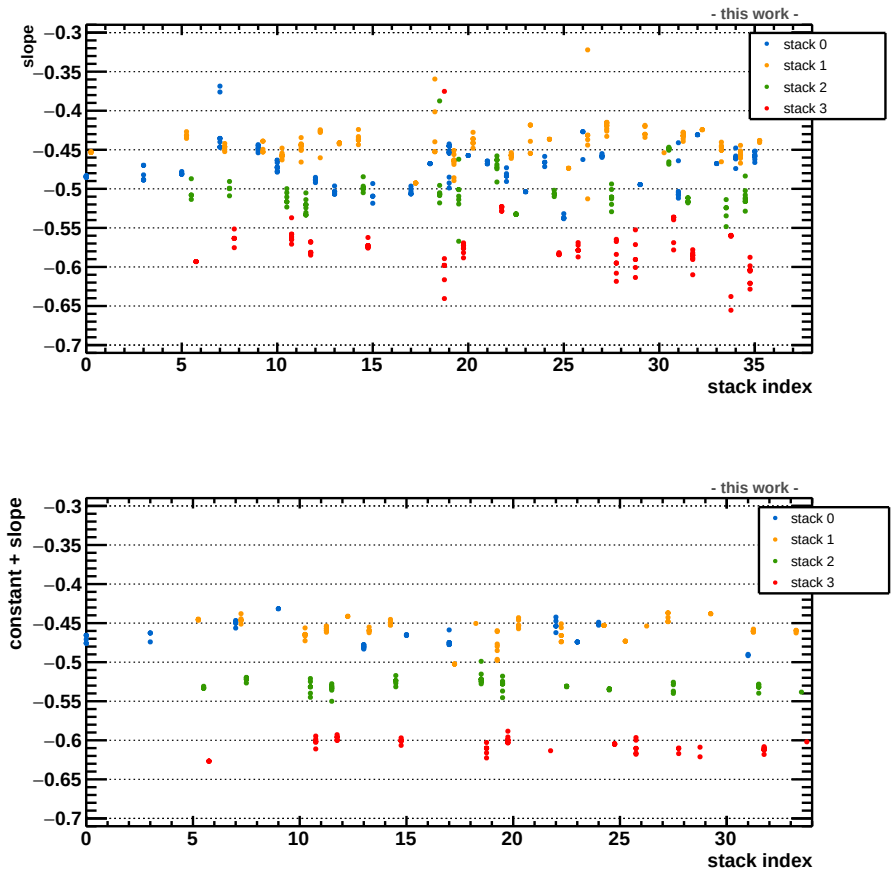


Figure 5.20: (Color Online) Slopes as a function of the stack index (top), and sum of constant and slope (bottom). Only fit parameters with RMS$<$ 0.02 are shown.

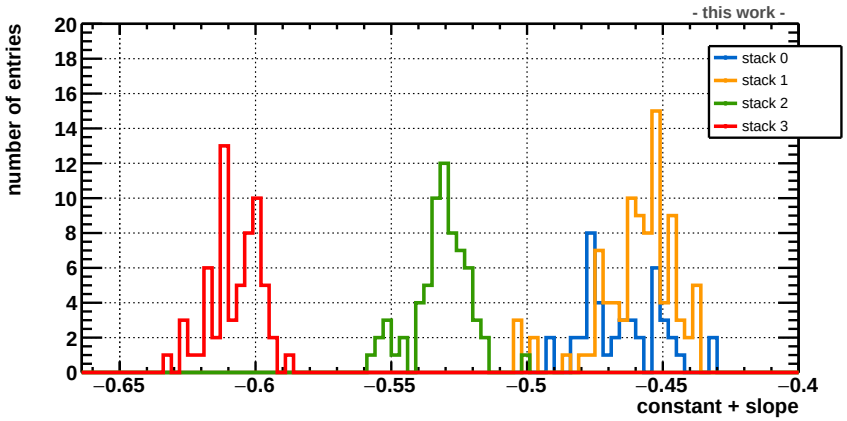


Figure 5.21: (Color Online) Distribution of the sum of the constant and slope, for each stack. Only fit parameters with RMS$<$ 0.02 are shown.

laser signals due to reflected laser tracks smear the data points around the nominal behavior due to underestimation of the average positive signal in the stack. Ideally, for the determination of the fit parameters, laser IDs with small RMS of the residuals should be opted.

Once the TPC is installed in the cavern a new set of pulser and laser measurements will take place, this time collecting a high number of events for all sectors. In general, the noise is expected to be slightly better compared to the one measured in the clean-room due to the ventilation and the power supply configuration [49]. However, for a few stack indices, due to the geometry of the tracks, no laser data can be collected (see for example Figure 4.10). In these cases, the model parameters can be estimated by those of stack indices with similar pulser charge distributions where laser data are available.

5.3.4 Common-mode signal estimation

The baseline estimation model described by Equation 5.11 was in the past simulated for the MWPC-based TPC [37], however in the old setup the common-mode responses were uniform, namely CF ≈ 1 for all pads of the same stack. A CF value close to unity for all pads in the stack practically means that, for every timebin, summing up all the positive signal in the stack (for all signal-pads) and the negative common-mode undershoots (for all empty pads) equals to zero. In that case, the (uniform) common-mode signals were simply estimated by

$$Q_{\mathrm{CM,\ MWPC}}(t) = \left\langle Q_{\mathrm{CM}}^{\mathrm{meas}}(t) \right\rangle_{\mathrm{empty\ pads}} . \tag{5.13}$$

A simulation of the baseline estimation by averaging over a different number of empty pads is shown in Figure 5.22 for the MWPCs. As the number of pads increases, the baseline estimation converges towards the reference baseline. In Figure 5.23, the contribution of the common-mode effect correction to the baseline fluctuations as a function of the number of used pads is plotted. By including a sufficient number of pads, the baseline can be restored with a negligible bias and noise contribution. Still, a remaining bias can be seen, both in Figure 5.22 (bottom right plot) and in Figure 5.23. This stems from imperfections in the signal detection algorithm, for example in cases where pads with remaining ion-tails from previous signals are mis-classified as "empty-pads" [50].

The same method is proposed for the GEM-based TPC, however the following differences must be highlighted:

(i) The common-mode fraction variations have to be accounted for before the averaging, using the constructed linear model, Equation 5.10.

(ii) The common-mode fraction in the GEM-based TPC is on average (depending on the stack type) $40-65\%$ smaller compared to the one of the MWPC-based TPC.

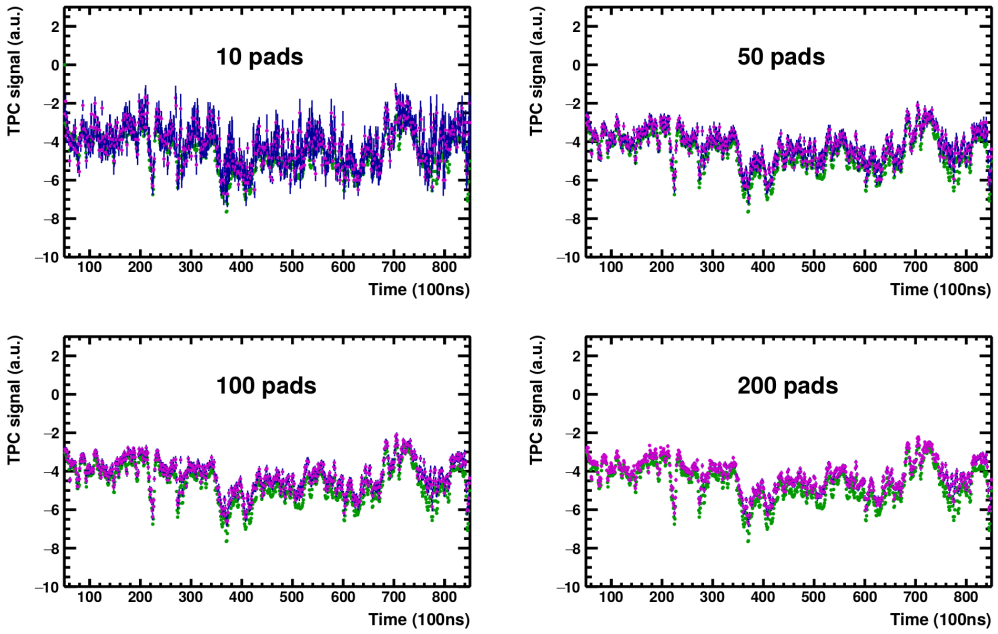


Figure 5.22: (Color Online) Simulation of baseline estimation in the MWPC-based TPC, using a different number of pads. With green, the reference baseline is shown, while in magenta the estimated baselines, with the corresponding uncertainties (blue lines). Figure taken from [37]

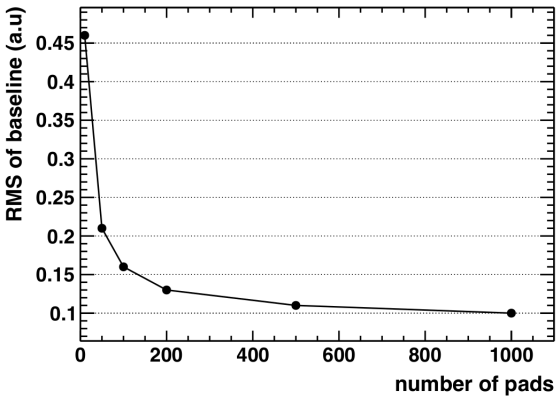


Figure 5.23: Contribution of the common-mode effect correction method to the baseline fluctuations in the MWPC-based TPC as a function of the number of pads used in the averaging, as simulated in [37].

(iii) The average pad noise in the GEM-based TPC is $\approx 0.9 - 1.0$ ADC, depending on the stack type [8], which is larger than that in the MWPC-based TPC, ≈ 0.7 ADC [24].

(iv) The higher occupancies in Run 3 result in larger deposited charges in the stacks. This in turn results in larger common-mode signals (see Equation 5.9).

(v) On the other hand, the higher occupancies imply that the number of available empty pads for the baseline estimation might be insufficient.

From the listed differences, (ii) and (iii) negatively affect the signal-to-noise ratio in the measurement of the common-mode charge in the empty pads, while (iv) improves it.

The acquired laser data are not suitable for testing the proposed algorithm for the complete Run 3 setup; due to the low laser track density, the common-mode responses are comparable to the per-event noise ($\approx 1\,\mathrm{ADC}$, see Figure 3.10b). On the other hand, correcting for the effect over the full event sample (averaged pad responses) results in a very quick convergence of the RMS of the baseline, due to the much better signal-to-noise ratio. This is demonstrated in Figure 5.24, where the contribution of the common-mode effect correction to the RMS of the baseline as a function of the number of pads used is plotted. It should be noted that for this correction an "ideal" signal detection algorithm is used, which differs from the realistic hardware implementation. The aforementioned arguments necessitate detailed Monte Carlo simulations, taking all the above into account, to test the effectiveness of the algorithm for the GEM-based TPC.

It is highlighted that the laser data are the only way to study the common-mode effect correction without zero-suppression. Taking non zero-suppressed Pb–Pb data is impossible due to the data size. Therefore testing with Pb–Pb events without zero-suppression is only possible in Monte Carlo simulations.

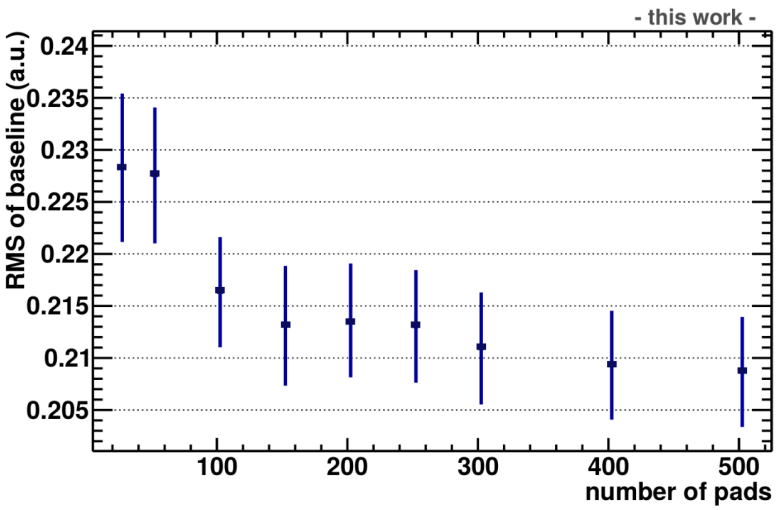


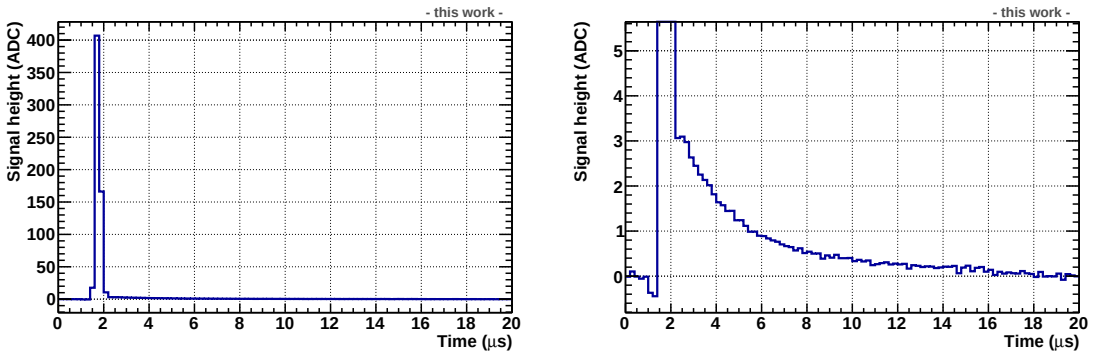
Figure 5.24: Contribution of common-mode effect correction to the RMS of the baseline for the GEM-based TPC, using the averaged laser data.

6 Ion-tail analysis

The subject of this chapter is the ion tail in the GEM-based TPC, as analyzed using the laser data. A demonstration of the ion tail is shown in Section 6.1, where its origin is also discussed. In Section 6.2, an estimation of the two ion type contributions is presented. Finally, in Section 6.3, the online correction in the CRUs is addressed.

6.1 Demonstration and origin of the ion tail

A known advantage of the GEM detectors, contrary to the MWPCs, is the absence of an ion-tail. However, during the analysis of the laser data for the common-mode effect in the scope of this thesis, an unforeseen tail was discovered. In Figure 6.1a, the response of a signal-pad for a few timebins of laser bundle 0 is plotted, while Figure 6.1b shows the same response zoomed in on the signal (y) axis. After the signal peak, a long tail is observed, lasting for $\approx 14\,\mu$s. In the particular example, the tail maximum (≈ 3 ADC) is $\approx 0.7\%$ of the signal maximum ($Q_{\text{max}} \approx 410$ ADC), while the tail integral corresponds to $\approx 9\%$ of the total electron signal ($Q_{\text{tot}} \approx 600$ ADC). It should be highlighted that the tail shape and duration depend on the distance of the pad from the center of gravity of the cluster, the track inclination, diffusion, GEM4B to pad-plane distance etc.. Consequently, the aforementioned numbers are heavily affected by such parameters. The ion-tail in the MWPC-based



(a) (b)

Figure 6.1: (a) Response of a signal-pad for a few timebins of laser bundle 0. (b) Response of the same pad, zoomed in on the signal (y) axis. The undershoot observed before the signal pulse is the common-mode response due to signals in other pads of the same stack. Note that the time range is in μs.

TPC corresponded to $\approx 50\%$ of the total signal. Despite having a much smaller magnitude in comparison to the MWPC-based TPC, the ion tail in the Run 3 setup still requires a correction.

Simulations were conducted by the TPC collaboration in order to understand the origin of the unforeseen ion tail [51]. For these simulations, the quadruple GEM configuration, depicted in Figure 3.2, was modeled in detail, and the electric field maps were calculated using the finite elements method as implemented in ANSYS [52]. The transport properties of the charge carriers were obtained by Magbolz [53] and then their multiplication was simulated in Garfield++ [54]. The simulations showed that the ions contributing to the tail are either created in the final amplification stage (GEM4 holes) or in the induction gap. Ions created in the

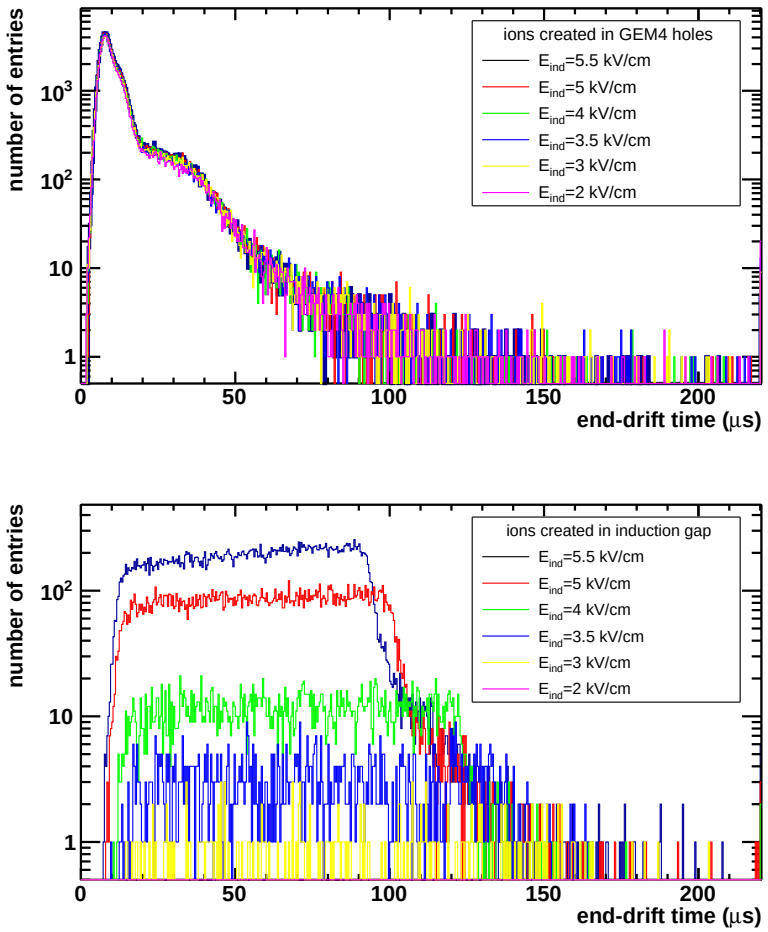
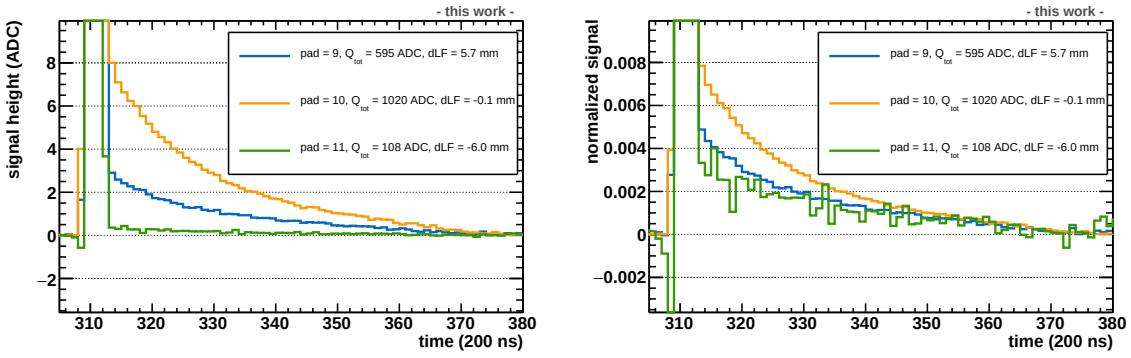


Figure 6.2: (Color Online) End-drift times of ions created in the GEM4 holes (top) and in the induction gap (bottom). Different induction field values are shown with different colors, with the nominal corresponding to 3.5 kV/cm. Note the logarithmic scale in y, and the different y-axis range in the two cases. Figure taken from [51].

previous amplification stages are screened by the GEM4 electrodes, thus they do not contribute to the measured signal.

The simulation results are summarized in Figure 6.2, where the end-drift time, namely the time since the creation of each ion until its absorption, is plotted separately for the two aforementioned ion types. Different induction field values (electric field between the GEM4B electrode and the pad-plane) are shown with different colors. A first observation is that the number of ions created in the amplification stage is considerably larger than the number of those created in the induction gap. The former, due to their sharp distribution in small end-drift time values, are referred to as the *fast component* of the ion tail, while the latter, due to their flat distribution, are referred to as the *slow component*. Additionally, while the distribution of the fast component does not depend on the induction field value, the distribution of the slow component is highly affected by it. In particular, since the probability for ionization depends on the induction field value, fewer and fewer ions are produced as the induction field value decreases. The almost flat distribution of the slow component indicates that these ions are uniformly produced in the induction gap. It is noted that the nominal induction field value, $E_{\text{ind}}^{100\%}$, is set to 3.5 kV/cm.

The ion-tail properties depend on various parameters such as the distance of the pad to the center of gravity of the cluster, the track inclination, diffusion etc.. Figure 6.3 shows as an example the tail dependency on the distance of the pad to the center of gravity of the cluster. The ion tail is shown for three pads of OROC2 which belong to the same cluster, and in particular in (a) the actual signal height is plotted, while in (b) the signal is normalized to the total charge under the peak of each pad (Q_{tot}). In the figure legend, dLF refers to the distance of the center of the pad from the non-robust local (i.e. stack-wise) track fit with associated uncertainties. This fit, as was shown in Subsection 4.4.2, describes quite consistently the center of



(a) (b)

Figure 6.3: (Color Online) Ion tail for three pads of the same cluster. In (a) the actual signal height is shown, while in (b), the signal is normalized to the total charge under the peak. dLF refers to the distance from the non-robust local track fit with associated uncertainties.

gravity of the cluster. In the normalized responses, the tail magnitude and shape differ for the pad closest to the COG of the cluster (pad 10, $d\mathrm{LF} = -0.1\,\mathrm{mm}$). It is reminded that the pad size along the pad-direction is $\approx 6\,\mathrm{mm}$ for the OROC2 pads (Table 3.1).

This demonstrates the significance of the *pad response function* (PRF) in the ion-tail analysis. The PRF describes the signal distribution on neighboring pads induced by a point-like avalanche. The measured cluster size on the pad-plane is determined by the intrinsic beam spread, the diffusion, and the PRF. The electron PRF is almost a delta-function; the electrons induce a signal explicitly on the pad where they will be collected at. On the other hand, due to the much smaller mobilities involved, the ion PRF extends over a few pads. This results in additional smearing, which is manifested as a dependence of the ion tail properties on the pad distance with respect to the center of gravity of the cluster.

To investigate the tail properties and to disentangle the contributions of the two different ion types, a dedicated set of measurements was added to the pre-commissioning programme. In these measurements, 5000-event laser runs were taken for the sector pair A01–A02, in which the induction field value was varied from 50% to 100% of the nominal one, in steps of 5%. The high number of events compared to the standard laser runs was chosen to guarantee a good signal-to-noise ratio, since the tail magnitude is quite small.

6.2 Estimation of ion type contributions

6.2.1 Data selection

Since the ion tail properties depend on many parameters, for the estimation of the two ion type contributions shown in Figure 6.2, the information was aggregated in multi-dimensional histograms. In many cases, tracks originating from even and odd laser rods irradiate the same stack (see Figure 4.10). The time difference between the two is $\approx$ 10 timebins (Section 4.1). Therefore very often a pad laser signal is followed by a common-mode signal, as shown in Figure 6.4. The undershoot is superimposed on the ion tail. Since this scenario occurs quite often, excluding these pad signals would significantly reduce the available size of the data sample. To avoid biasing the results, instead of using the measured charge in a given timebin (tb), $Q(t)$, the median charge in $t \pm 3\,\mathrm{tb}$ was used instead, referred to as $Q_{\mathrm{med}}(t)$. The median is not so sensitive to the outliers due to the common-mode.

To eliminate any dependency on possible track misalignments along z and pressure variations (see Figure 4.11), the time variable is redefined as

$$t_R = t - \mathrm{COG}_{\mathrm{time}} \ . \tag{6.1}$$

where for a given signal-pad the $\mathrm{COG}_{\mathrm{time}}$ is the center of gravity in the time direction for a particular laser bundle, given by Equation 4.1. In this way, $t_R = 0$

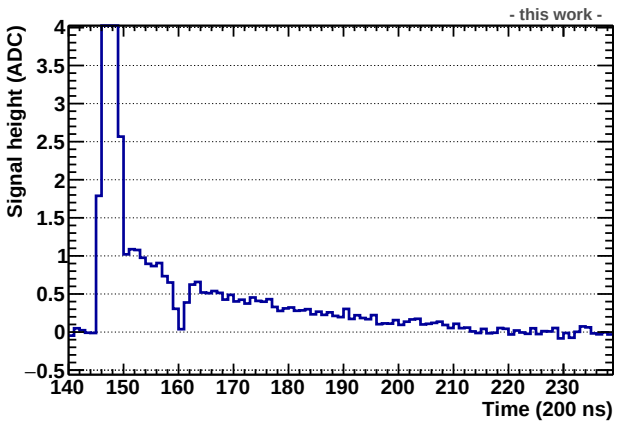


Figure 6.4: Example where a laser signal (tb≈ 146 – 149) is followed by a common-mode signal (tb≈ 159 – 161). The common-mode signal can affect the determination of the ion-tail properties.

always corresponds to the signal maximum irrespective of the bundle and pad-by-pad $\mathrm{COG_{time}}$ variations. For the dependent variable, the normalized signal for each pad is used, namely

$$Q_{\mathrm{norm}}(t_R) = \frac{Q_{\mathrm{med}}(t_R)}{Q_{\mathrm{tot}}} \quad . \tag{6.2}$$

The normalization is performed with respect to the total charge in three timebins around the maximum, given by Equation 5.4. The maximum charge, $Q_{\mathrm{max}} = Q(t_{\mathrm{max}})$, heavily depends on the parameter

$$\mathrm{phase}(\mathrm{COG_{time}}) = \mathrm{COG_{time}} - \mathrm{int}(\mathrm{COG_{time}}) \quad , \tag{6.3}$$

namely the relative position of the center of gravity in time with respect to the sampling timebin[1]. In the equation, "int" denotes the nearest integer. By using the total charge in three timebins around the center of gravity in time, this dependence on the so-called *cluster clock phase* is almost entirely eliminated.

6.2.2 Results

Figure 6.5 shows the normalized signal (defined in Equation 6.2) as a function of time for $E_{\mathrm{ind}}^{100\%}$, bundles 0–2, all beams (and therefore track inclinations) and for all central pads (distance from the center of gravity of the cluster of less than 0.2 cm). Since the peak of bundle 3 is very close to the central-electrode signal, it was excluded

[1] If the $\mathrm{COG_{time}}$ coincides with the center of the sampling timebin, then Q_{max} will be maximum. If the $\mathrm{COG_{time}}$ coincides with the edge of the sampling timebin, then the charge will be "shared" between the two neighboring timebins, resulting in a smaller Q_{max}.

from the data-set. By taking the *profile* with respect to the y-axis[2] and repeating the procedure for the other induction field values, Figure 6.6 is constructed. The listed numbers in the figure legend correspond to percentages of $E_{\text{ind}}^{100\%}$. The y error-bars are plotted, but they are smaller than the marker size. In Table 6.1 the maximum

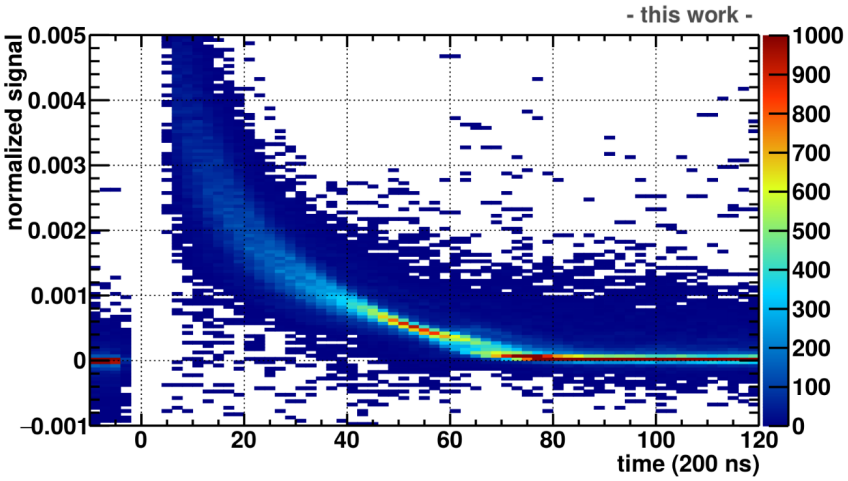


Figure 6.5: Normalized signal as a function of time for $E_{\text{ind}}^{100\%}$, bundles 0–2, all beams and for all central pads.

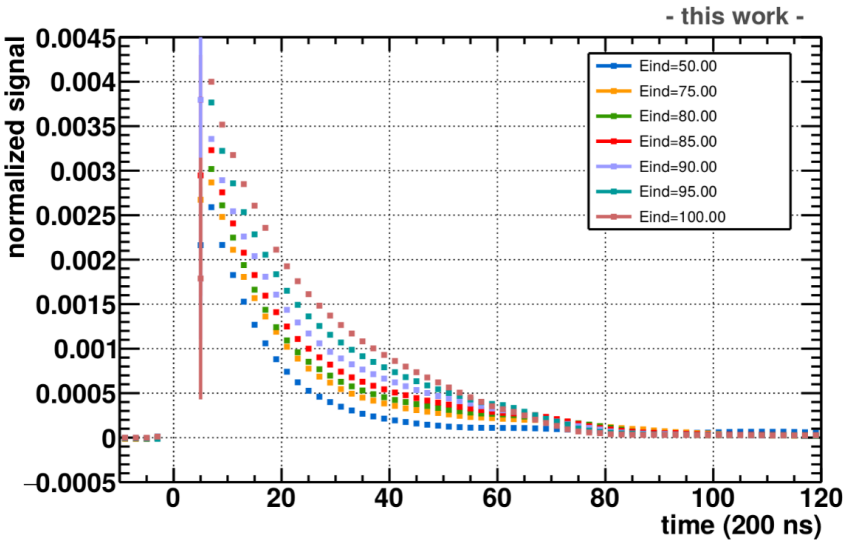


Figure 6.6: (Color Online) (Mean) normalized signal (Equation 6.2) as a function of time for the different induction field value settings. The listed numbers in the legend correspond to percentages of the nominal induction field value, which corresponds to 3.5 kV/cm. Signals from laser bundles 0–2 and only central pads are used.

[2] In the profile histogram, the mean value of the normalized signal (y) is shown for every timebin (x). The y error-bars correspond to the RMS of the entries in each x bin.

	E_{ind} ($\%E_{\text{ind}}^{100\%}$)						
	50	75	80	85	90	95	100
(tail maximum)/(Q_{tot}) (%)	0.26	0.29	0.30	0.32	0.34	0.38	0.40
(tail integral)/(Q_{tot}) (%)	3.8	5.1	5.5	6.1	6.7	7.6	8.5

Table 6.1: Tail maximum and integral, both normalized to the total electron signal, for different induction field values. Signals from laser bundles 0–2 and central pads are used.

and integral of the tail is listed for the different induction value settings. It can be seen that the magnitude of the tail decreases with decreasing induction field value. The difference between the ion tail of two subsequent E_{ind} values gets smaller as E_{ind} decreases (compare for example $E_{\text{ind}}^{100\%}$ to $E_{\text{ind}} = 95\% E_{\text{ind}}^{100\%} \equiv E_{\text{ind}}^{95\%}$, and $E_{\text{ind}}^{80\%}$ to $E_{\text{ind}}^{75\%}$), also in agreement with Figure 6.2. Apart from the magnitude, the shape of the exponential-like tails is also different.

An estimation of the two ion type contributions can be achieved by relying on the simulation results. It can be assumed that at $E_{\text{ind}}^{50\%} = 1.75\,\text{kV/cm}$, the contribution of the slow component is negligible. Since the fast component is practically independent of the induction field value, the slow component contribution for a E_{ind} value can be estimated as the difference of the ion tail between that induction field value and $E_{\text{ind}}^{50\%}$. Equivalently,

$$Q_{\text{fast}}(E_{\text{ind}}) \approx Q_{\text{meas}}(E_{\text{ind}}^{50\%}) \tag{6.4}$$

$$Q_{\text{slow}}(E_{\text{ind}}) \approx Q_{\text{meas}}(E_{\text{ind}}) - Q_{\text{meas}}(E_{\text{ind}}^{50\%}) \tag{6.5}$$

The results of the above procedure for OROC3 are summarized in Figure 6.7 and Figure 6.8. In Figure 6.7, the normalized ion-tail (Equation 6.2) is shown for the

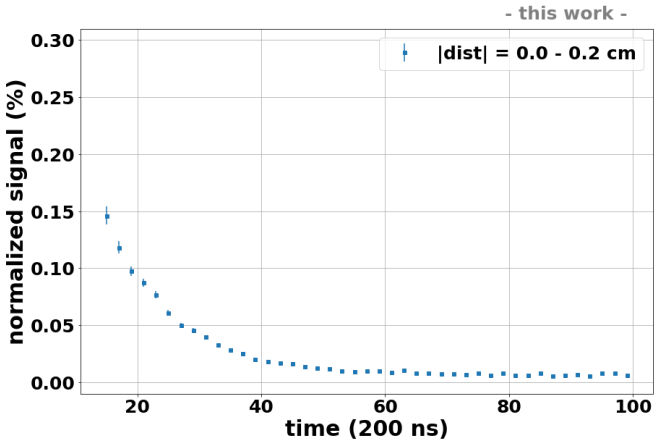
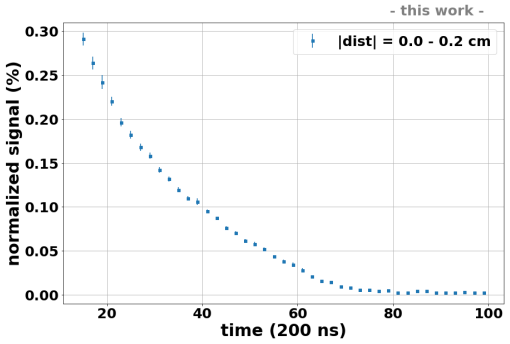
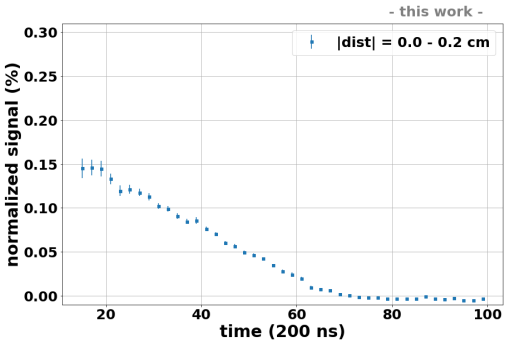


Figure 6.7: Normalized ion-tail for $E_{\text{ind}}^{50\%}$, which is assumed to describe the fast component independently of the E_{ind} setting. Only data from central pads of OROC3 are shown. The error bars correspond to the RMS of the entries in each timebin.

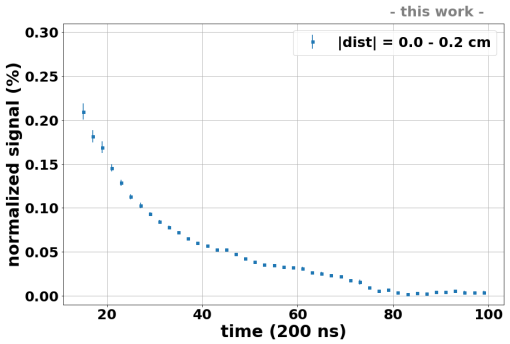
$E_{\text{ind}}^{50\%}$ setting, which is dominated by the fast component. An exponential-like shape is observed. In Figure 6.8 the normalized ion-tail as well as its difference from the $E_{\text{ind}}^{50\%}$ setting are shown for different values of the induction field. The difference



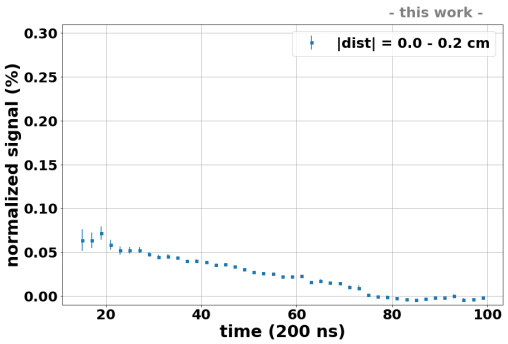
(a) Normalized ion-tail for $E_{\text{ind}}^{100\%}$.



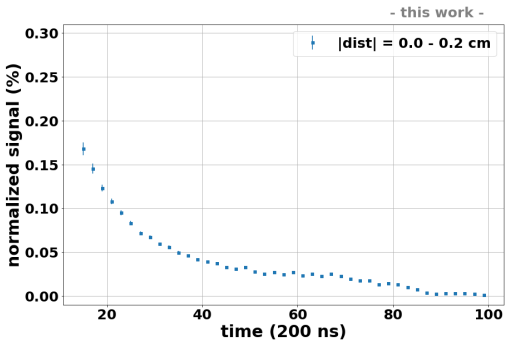
(b) Difference between normalized ion-tail for $E_{\text{ind}}^{100\%}$ and $E_{\text{ind}}^{50\%}$.



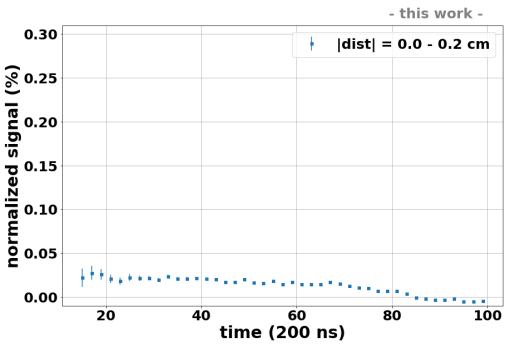
(c) Normalized ion-tail for $E_{\text{ind}}^{85\%}$.



(d) Difference between normalized ion-tail for $E_{\text{ind}}^{85\%}$ and $E_{\text{ind}}^{50\%}$.



(e) Normalized ion-tail for $E_{\text{ind}}^{75\%}$.



(f) Difference between normalized ion-tail for $E_{\text{ind}}^{75\%}$ and $E_{\text{ind}}^{50\%}$.

Figure 6.8: Normalized ion-tail (left) and difference from the $E_{\text{ind}}^{50\%}$ (right), for different E_{ind} settings. Only data from central pads of OROC3 are shown. The error bars correspond to the RMS of the entries in each timebin.

approximates the slow component. An almost linear shape of the slow component is observed, which is in agreement with uniformly produced ions in the induction gap (see also Figure 6.2). In the case of $E_{\mathrm{ind}}^{100\%}$, the slow component contribution to the signal is $\approx 50\%$, with the percentage decreasing with decreasing E_{ind}.

Since only two chambers were tested during the induction field scan, it was seen that a full differential analysis using Machine Learning techniques (i.e. an investigation of the tail dependencies on the track inclination, diffusion, distance to the center of gravity of the cluster, cluster clock phase etc.) could not be conclusive. For this, dedicated laser runs with a high number of events (5000 or higher) for all chambers would be necessary.

6.3 Online correction in the CRUs

An online correction of the ion tail on a channel-by-channel basis before the data compression is vital for significantly reducing the streamed data size, and also for decreasing the computing power necessary at the offline stage for its correction. However, the dependence of the ion tail shape and magnitude on cluster-related properties implies that a full online correction is impossible. Since the polarity of the ion tail is positive, the zero-suppression does not lead to missing charge and subsequently missing clusters. Consequently, the ion-tail can be partially corrected online, and the remaining contribution, if needed, can be treated offline, after the track reconstruction.

Applying an exponential correction directly in the CRUs, i.e. a correction of the form

$$Q_{\mathrm{corr}}(t) = Q_{\mathrm{meas}}(t) - \sum_i A_i \cdot e^{-(t-t_{\mathrm{i,max}})/\lambda_i} \quad , \tag{6.6}$$

is quite challenging. In the equation, $Q_{\mathrm{meas}}(t)$ and $Q_{\mathrm{corr}}(t)$ is the pad signal for timebin t before and after applying the correction, respectively. The sum runs over all previous signal peaks i for the given pad. A_i and λ_i is the maximum and slope of the tail corresponding to signal peak i, while $t_{\mathrm{i,max}}$ is the position of the peak maximum. The complexity of such a correction stems from two factors; the number of required resources and the time needed to perform the calculations. Firstly, applying the aforementioned correction would imply storing the full peak history for each of the $\approx$1600 pads read out by a CRU. However, the FPGAs are normally not suited for storing large amounts of data. Secondly, while it is possible for the FPGAs to calculate an exponential (using, for example, CORDIC[3] functions), the FPGAs would need to perform multiple exponential calculations in parallel. Based on the number of available DSPs (Digital-Signal-Processors) in the CRU, this would introduce some latencies [55]. To avoid the aforementioned complications,

[3]The *CORDIC* (Coordinate Rotation Digital Computer) algorithm is an iterative method for the calculation of elementary functions in the electronics using rotations.

the proposed correction model that can be programmed in the CRUs requires only simple mathematical operations,

$$Q_{\text{exp}}(t) = k_{\text{time}} \cdot [Q_{\text{exp}}(t-1) + Q_{\text{meas}}(t)] \quad , \tag{6.7}$$

$$Q_{\text{corr}}(t) = Q_{\text{meas}}(t) - k_{\text{amp}} \cdot Q_{\text{exp}}(t-1) \quad , \tag{6.8}$$

namely the corrected charge for timebin t is equal to the measured one minus a correction factor, $k_{\text{amp}} \cdot Q_{\text{exp}}(t-1)$. The correction factor depends on two static parameters, k_{time} and k_{amp}, and a buffered value calculated in the previous timebin (compare to the number of buffered values required in case of a correction of the form Equation 6.6). The parameters k_{time} and k_{amp} describe the tail slope and integral, respectively. It should be highlighted that the above filter can be applied continuously even in the absence of signals. In that case, the filter has no impact on the measured charge.

Ideally, the parameters k_{time} and k_{amp} would depend on all aforementioned track-related properties, as well as pad-related properties, such as the pad-plane to GEM4B distance (which in turn affects the effective induction field value), and gain variations. Given that all track-related properties are not available at the online stage, two options remain; either to provide a map for k_{time} and k_{amp} to the CRUs, which accounts for pad-by-pad distance and gain variations, or to universally fix the two parameters for all pads. In the former case, since only a small ($\approx$10%) fraction of the pads is irradiated in the laser data, a separate calibration measurement would be necessary to determine the two parameters for all pads (using for example an X-ray or a krypton source).

Below, the proof of principle for the exponential filter described by Equation 6.8 is demonstrated using the averaged pad responses. It is highlighted that for the demonstration, reasonable values for the parameters k_{time} and k_{amp} were selected. A comprehensive analysis is however required in order to determine these parameters with the goal to optimize the performance of the TPC. In Figure 6.9 the ion tail before and after the correction is plotted for 80 randomly selected pads of bundle 2 and rod 1, where the induction field value is 50% of the nominal one. Since the correction is applied on a pad and timebin basis, the binning described in Subsection 6.2.1 was not applied (here the measured charge $Q(t)$ was used instead of the median charge). All beams (and therefore track inclinations) and pad distances to the center of gravity of the cluster are included. The parameters are fixed to $k_{\text{time}} = 0.875$ and $k_{\text{amp}} = 0.008$ for all pads. The tail cannot be entirely eliminated due to the remaining cluster-related dependencies, but on average (bottom figure, where the profile histogram is shown), most of the effect disappears. It is also confirmed that the correction filter has no impact on the measured charge before the laser pulse (the measured charge before and after correction for timebins < 288 is identical). In Figure 6.10 the same is repeated for the nominal induction field value. This time, the parameters are adjusted to $k_{\text{time}} = 0.910$ and $k_{\text{amp}} = 0.009$. Finally, in Figure 6.11, the pad signals of the nominal induction field setting are corrected using the parameters selected for the $E_{\text{ind}}^{50\%}$ case. It is seen that the correction is

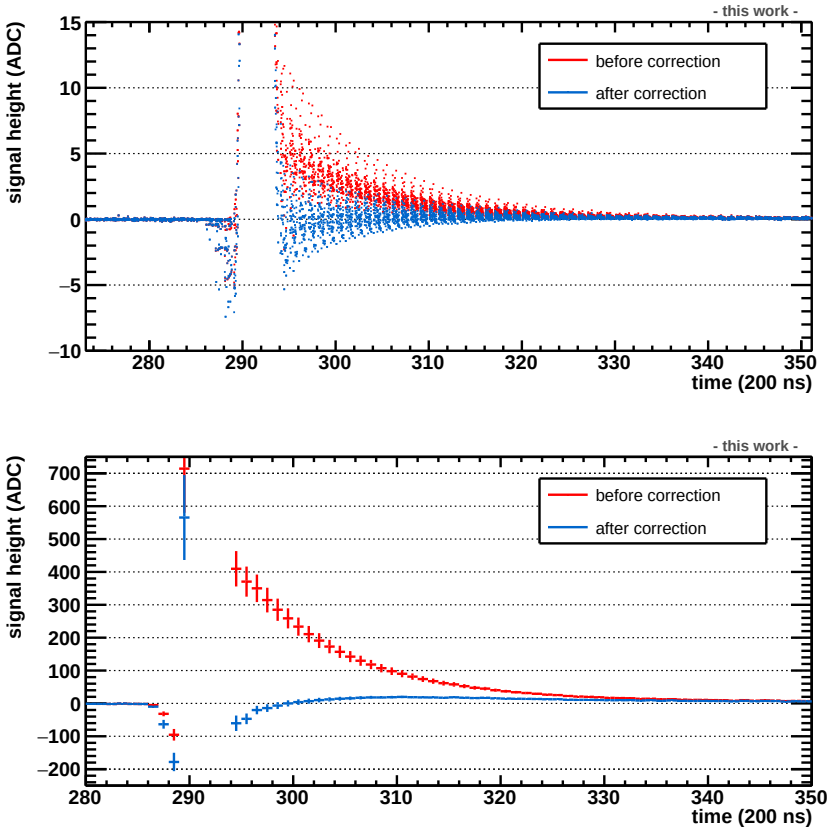


Figure 6.9: (Color Online) Ion tail before (red) and after (blue) correction for $E_{\mathrm{ind}}^{50\%}$ (top) and profile histogram of the above (bottom). A random selection of 80 pads of bundle 2 belonging to different beams and with varying distances to the COG of the cluster is performed. The parameters are fixed to $k_{\mathrm{time}} = 0.875$ and $k_{\mathrm{amp}} = 0.008$. The error bars in the bottom plot correspond to the RMS of the entries for each timebin.

quite sensitive to the chosen parameters (compare the corrected profile histogram of Figure 6.10 to the one of Figure 6.11).

It is therefore shown that the proposed exponential filter, despite its simplicity, can account for most of the ion tail. A detailed analysis is needed to select the optimal values for k_{time} and k_{amp}. Whether these are fixed for all pads or not, the optimization should rather focus on the physics performance of the TPC, and not on the minimization of the residual signal. The reasoning behind this is that it is preferable to under-correct the tail (i.e. so that the residual signal is positive), since over-correcting leads to shifting the baseline below the zero-suppression threshold, therefore results in loss of information that cannot be recovered [40].

Similarly to the common-mode correction, Monte Carlo simulations and further measurements are necessary to test the effectiveness of the proposed algorithm and its impact on the overall performance of the TPC on an event-by event basis.

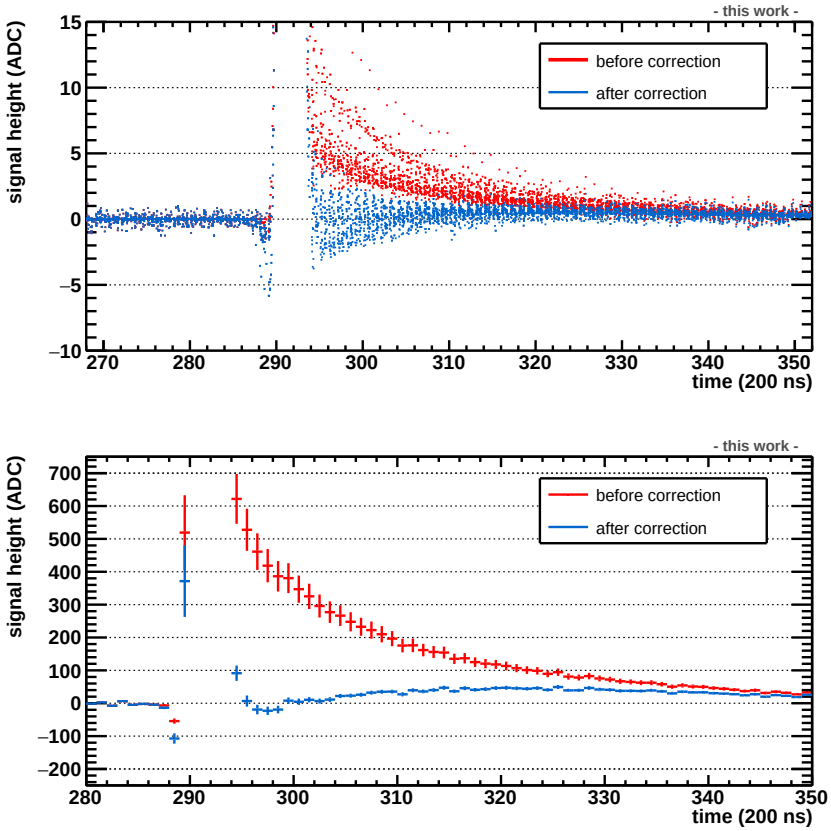


Figure 6.10: (Color Online) Ion tail before (red) and after (blue) correction for $E_{\mathrm{ind}}^{100\%}$ (top) and profile histogram of the above (bottom). A random selection of 80 pads of bundle 2 belonging to different beams and with varying distances to the COG of the cluster is performed The parameters are fixed to $k_{\mathrm{time}} = 0.910$ and $k_{\mathrm{amp}} = 0.009$.

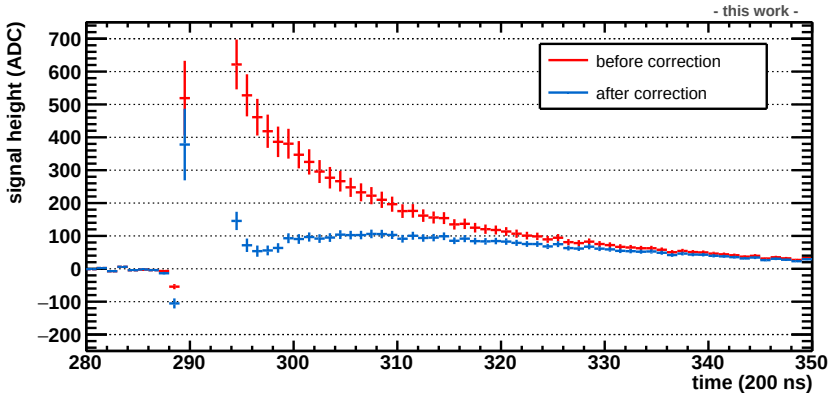


Figure 6.11: (Color Online) Profile histogram for the ion tail before (red) and after (blue) correction for $E_{\mathrm{ind}}^{100\%}$, using the parameters chosen for $E_{\mathrm{ind}}^{50\%}$.

7 Summary and outlook

In order to achieve the desired sensitivity to rare probes, in accordance with its physics goals, the ALICE detector is undergoing a major upgrade programme during the *Long Shutdown 2* (2018–2021). Starting from LHC Run 3, the Pb–Pb minimum bias collision rate at the ALICE detector will be increased to 50 kHz, which is about a factor ten higher than the current one. To overcome the rate limitations posed by the gated operation of the ALICE Time Projection Chamber (TPC) which are prohibitive for the anticipated interaction rate increase in LHC Run 3 and 4, its Multi-Wire Proportional Chamber-based (MWPC) readout was replaced by Gas Electron Multipliers (GEMs). Before commissioning the TPC in the cavern, a pre-commissioning took place in the clean room of LHC Point 2, in order to test the performance of all its components.

This thesis describes the analysis and development of online correction algorithms for the common-mode effect and ion tail in the GEM-based TPC, by analyzing data collected with its laser calibration system. The common-mode effect, which results from capacitive coupling between the pad-plane and the GEM foils, causes a characteristic undershoot in all pads simultaneously with signals generated by charged particle tracks in the TPC volume. The – initially unanticipated – long tail after the signal pulse is caused by the slow motion of ions away from the pad-plane. Correcting for the two effects online, and in particular in the FPGA-based Common Readout Units (CRUs), is essential in maintaining the $\mathrm{d}E/\mathrm{d}x$ and tracking resolution and limiting the cluster losses; if not, both effects would result in undermining of the Particle IDentification (PID) performance of the TPC. The common-mode effect and ion tail were also present in the MWPC-based TPC, however key differences were observed with the current setup.

Firstly, an analysis of the laser signal shape was presented. A detailed study of the baseline and noise using the averaged laser data via a two-iteration process and the Least Trimmed Squares (LTS) method was discussed. A well-defined baseline and low noise are essential to guarantee a high-quality data analysis. The signal-pad to ideal laser track association algorithm was developed. The track association is a necessary step to identify possible dependencies of the two effects on track-related parameters such as the pad distance to the center of gravity of the cluster, track inclination, diffusion, etc.. The cluster is defined as the charge collected in a TPC row, created by a single track in the TPC volume.

The connection of the calibration pulser measurements to the analysis of the two effects was also demonstrated. In the pulser measurements, a pulse is injected in the bottom of the lowermost GEM foil (GEM4), which induces a signal on the pad-plane due to capacitive coupling. The measured pulser charge in each pad is

proportional to the capacitance between the pads and the GEM foils. Regions of high capacitance are observed in cases when the GEM4 foil bends towards the pad-plane due to electrostatic forces. Pads positioned at the edges of the chamber and below the spacer cross (a support structure to ensure mechanical stability of the chamber) also exhibit higher capacitance.

The dependencies of the common-mode fraction, which is the parameter of interest for the common-mode effect, were understood using Machine Learning techniques. The common-mode fraction (CF) is the common-mode charge in a given pad for a given timebin normalized to the average laser signal in the stack. It was found that CF mostly (96%) depends on the stack type and the *normalized pulser charge* (meaning the pulser charge normalized to the mean pulser charge in the stack). The stack type accounts for the absolute capacitance of the stack, while the normalized pulser charge for the pad-by-pad capacitance variations within a given stack. In particular, there is a linear dependence of the CF on the normalized pulser charge. A second-order correction comes from the *missing charge*, namely the average positive signal in the stack and the number of signal-pads. The missing charge accounts for the biasing of the positive signal in the stack due to the superimposed common-mode in the signal-pads. No dependence on any track-related properties was observed.

Once the common-mode effect was understood, the implementation of its online correction in the FPGA-based Common Readout Units (CRUs) was addressed. Due to the additional CRU segmentation within each stack, the average positive signal and the number of signal-pads in the stack cannot be available at the online stage. Two linear models for the CF were compared, one accounting only for the (time-independent) pulser charge dependency (*Model 2*), and one also accounting for the (time-dependent) missing charge contributions (*Model 1*). While Model 1 results in slightly smaller χ^2 values, both models are sufficient to describe the CF data. It was therefore confirmed that the linear dependence on the pulser charge can describe the CF data to a very good degree. Since the average positive signal in the stack is not available online, to calculate the pad common-mode charge in a given timebin, an estimation of the baseline must be conducted (Equation 5.11 making use of the Model 2 CF calculation. For this estimation, the available *empty* or *non-signal* pads are used, as identified by the signal-detection algorithm implemented in the CRUs. However, the laser data are not suitable to test the effectiveness of the proposed algorithm on a per-event basis, because, due to the small deposited charge in the stack, the common-mode responses are comparable with the pad noise. The baseline estimation method as a way to correct for the common-mode effect online was in the past successfully demonstrated in the MWPC-based TPC (CF≈ 1) via detailed simulations. However, significant differences are expected in the Run 3 setup; in particular, in the GEM-based TPC there is an additional dependence of the CF on the pad-by-pad capacitance variations, described by the pulser charge, as well as a systematically smaller CF ($\approx 0.4-0.6$) and larger pad noise. For these reasons, full Monte Carlo Pb–Pb simulations with the Run 3 conditions are needed to guarantee that the proposed algorithm is suitable to account for the effect online. It is noted that, due to the enormous data size, it is not possible to test the algorithm on non

zero-suppressed Pb–Pb measurements.

An ion tail was observed in the laser data. The tail maximum corresponds to on average $\approx 0.4\%$ of the peak maximum, while its integral corresponds to $\approx 9\%$ of the total electron signal. As a comparison, in the MWPC-based TPC the ion-tail integral corresponded to 50% of the total signal. Simulations conducted by the TPC collaboration revealed the origin of the observed ion tail in the GEMs; the signal is generated by ions produced either in the last multiplication stage (GEM4 holes) or in the induction gap. To understand and quantify the two contributions, a dedicated set of measurements was performed during which the induction field value was varied from $50\% - 100\%$ in steps of 5%. The investigation showed that, for the nominal induction field setting, the two ion types contribute almost equally to the generated signal. The contribution of ions produced in the GEM4 holes is almost independent of the induction field value, and results in a exponentially-shaped tail. The contribution of ions produced in the induction gap heavily decreases with decreasing induction field value, and leads to an almost linearly-shaped tail. For the nominal induction field setting, the magnitude of the two contributions is very similar.

Unlike the common-mode effect, the ion tail vastly depends on cluster-related properties, which are not available at the online stage. These are, for example, the distance of the pad to the center of gravity of the cluster, the track inclination, diffusion etc.. Consequently, the ion tail can only be partially (on average) corrected for online, while the rest of the dependencies can be treated offline, if needed, after the track reconstruction. An online correction algorithm was proposed, which requires the buffering of one value calculated in the previous timebin for each pad, and depends on two parameters which describe the slope and integral of the tail. The two parameters can either be fixed for all the TPC pads, or can account for gain and pad to GEM4B distance variations (which in turn affect the effective field seen by the ions). For the latter, a dedicated campaign using either an X-ray or a krypton source would be required. Nonetheless, it was shown that fixing the two parameters for all TPC pads can correct for most of the effect. A comprehensive analysis is needed to select the optimal values for the two parameters. Similarly to the common-mode effect, further simulations and measurements are necessary to test the effectiveness of the algorithm of an event-by-event basis.

A ALICE Coordinate Systems and Conventions

A.1 Global Coordinate System

The global coordinate system of ALICE, sketched in Figure A.1, is a right-handed Cartesian system, where the point of origin is at the nominal beam *Interaction Point* (IP) [56].

The ***z*–axis** is parallel to the beam direction. The positive direction is pointing towards the shaft-side (away from the muon arm). For simplicity, the positive z direction is the designated *A-side* (*Shaft Side*), while the negative z is the designated *C-side* (*Muon Side*).

The ***x*-axis** is perpendicular to z, aligned with the local horizontal, pointing towards the LHC center. The positive direction is from the IP towards the accelerator center.

The ***y*-axis** is perpendicular to x and z, forming a right-handed system. Therefore, positive y is pointing upward.

The **azimuthal angle ϕ** increases counter-clock wise from x ($\phi = 0$) to y ($\phi = \pi/2$), as shown in the figure.

The **polar angle θ** increases from the z-axis towards the x–y plane.

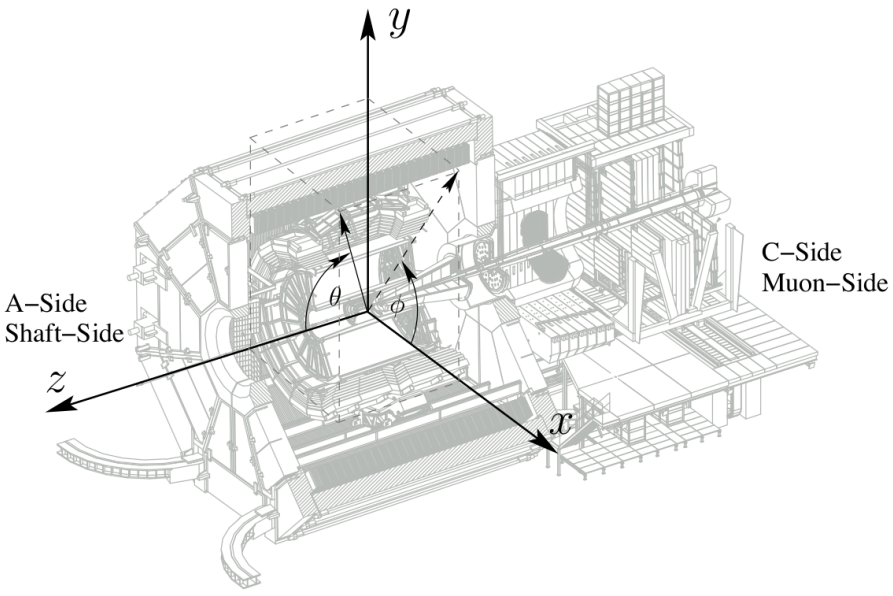


Figure A.1: ALICE detector, together with the Global Coordinate System [24, 56].

A.2 Local Coordinate System

The ALICE local coordinate system is a right-handed Cartesian system, having the same origin and same z-axis as the global coordinate system. It accounts for the azimuthal segmentation of the TPC in sectors. The x-axis of the local coordinate system always points radially outwards, and is perpendicular to the TPC pad row plane, as shown in Figure A.2. The y-axis is perpendicular to x.

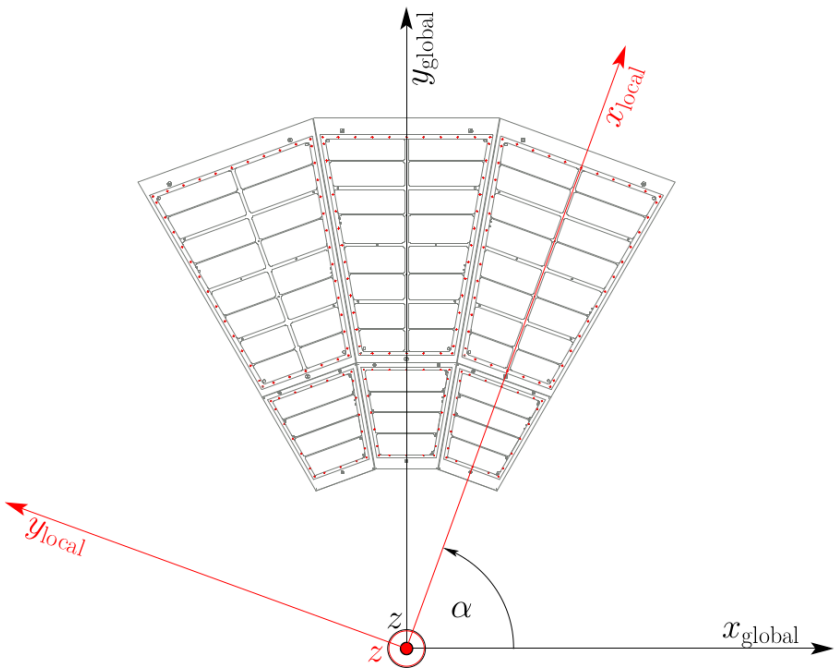


Figure A.2: Definition of the local Coordinate System for a particular sector. The global Coordinate System is also visible in black [24, 56].

One can transform between global and local coordinate system by rotating via the angle $\alpha = \text{sector} \times 20°$:

$$\begin{aligned} x_l &= x_g \cdot \cos(\alpha) + y_g \cdot \sin(\alpha) \\ y_l &= x_g \cdot \sin(\alpha) - y_g \cdot \cos(\alpha) \end{aligned} \tag{A.1}$$

A.3 Sector numbering conventions

As already described in Subsection 3.2.1, the TPC readout is divided into two sides, the A-side and C-side; each side is then divided into 18 sectors, and each sector is further subdivided along the radial direction into one Inner Readout Chamber (IROC) and one Outer Readout Chamber (OROC).

The standard numbering runs first over all IROCs on the A-side (0–17) and then on the C-side (18–35), and continues with all OROCs on the A-side (36–53) and then on the C-side (54–71).

An alternative labeling does not distinguish between IROC and OROC and treats the sectors as a whole; for example sector A00 corresponds to IROC 0 and OROC 36 together, while sector C00 corresponds to IROC 18 and OROC 54. With this numbering, mirror sectors have the same number, for example A00 and C00.

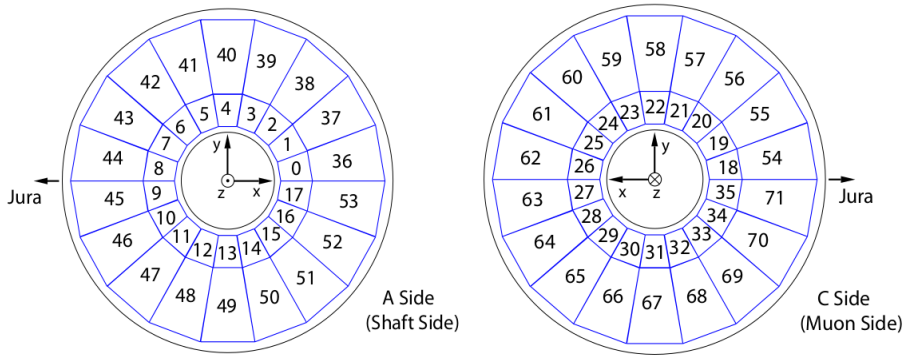


Figure A.3: Numbering of the 72 readout chambers of the TPC [57].

B Relevant physics processes

B.1 $\mathrm{d}E/\mathrm{d}x$ of charged particles in gases

When charged relativistic particles pass through the TPC active volume, they lose part of their kinetic energy by electromagnetic processes (through ionization, atomic or collective excitation). The mean energy loss is described by the "Bethe-Bloch" formula, for the region $0.1 \leqslant \beta\gamma \leqslant 1000$ [58]:

$$\left\langle -\frac{\mathrm{d}E}{\mathrm{d}x} \right\rangle = Kz^2 \frac{Z}{A}\frac{1}{\beta^2}\left[\frac{1}{2}\ln\frac{2m_e c^2\beta^2\gamma^2 W_{\max}}{I^2} - \beta^2 - \frac{\delta(\beta\gamma)}{2}\right] \quad , \tag{B.1}$$

where $K = 4\pi N_A r_e^2 m_e c^2 = 0.307075\,\mathrm{MeV\,mol^{-1}\,cm^2}$, Z the atomic number of the absorber, A its atomic mass, I the mean excitation energy (in eV) and $W_{\max}$ the maximum energy transfer in a single collision. For the energy regime of interest, $W_{\max}$ is commonly approximated with the formula $W_{\max} = 2m_e c^2\beta^2\gamma^2$, where β is the velocity of the incoming particle in units of c and γ its Lorentz factor, while $m_e c^2$ is the electron rest energy in MeV. Finally, δ is a correction term accounting for the density effect, discussed below.

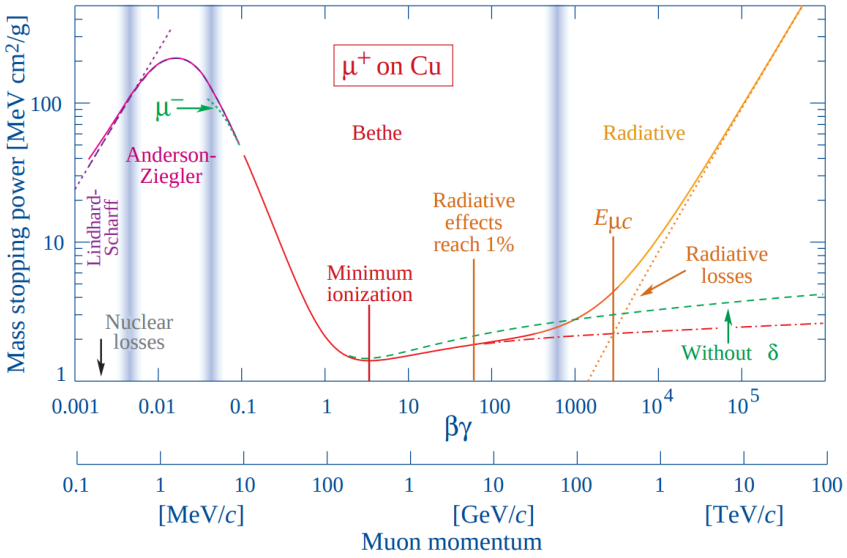


Figure B.1: Mass stopping power for positive muons in copper as a function of $\beta\gamma = p/Mc$. The Bethe-Bloch theory is applicable within the $\beta\gamma$ region defined by the vertical bands at around ≈ 0.1 to ≈ 1000 [58].

The mean rate energy loss discussed above, or *mass stopping power*, is expressed in units of $\mathrm{MeVg^{-1}cm^2}$. The mass stopping power of positive muons in copper as a function of $\beta\gamma = p/Mc$ can be seen in Figure B.1. The Bethe-Bloch region is indicated with vertical bands at $\beta\gamma$ around ≈ 0.1 to ≈ 1000. For low velocities in the Bethe-Bloch region, the mean energy loss is steeply decreasing as $\beta\gamma$ increases, being dominated by the $1/\beta^2$ term, until it reaches a minimum at around $\beta\gamma \approx 3.6$. In many cases, the particles of interest have a mean energy loss close to this minimum; these particles are often called *Minimum Ionizing Particles* (MIPs). Towards higher $\beta\gamma$ values (but still in the Bethe-Bloch regime) the $\ln\beta^2$ term starts to contribute, resulting in the observed *relativistic rise*. This rise is caused because for higher velocities, the Lorentz contraction of the electric field results in interactions with more distant molecules. In such a scenario, the mean energy loss would increase infinitely towards higher $\beta\gamma$ values (green dashed line in Figure B.1). However, since real media become polarized, the field extension is limited and thus so is the mean energy loss. This so-called *density effect* results in the Fermi Plateau (red dot–dashed line). Since polarization is more important in dense absorbers, it can mostly be neglected for gases under normal pressure conditions.

For higher $\beta\gamma$ values outside the Bethe-Bloch regime, radiative effects start to dominate (*Bremsstrahlung effect*), while for lower $\beta\gamma$ other effects need to be taken into account, such as the binding energy of the atomic electrons, as well as their

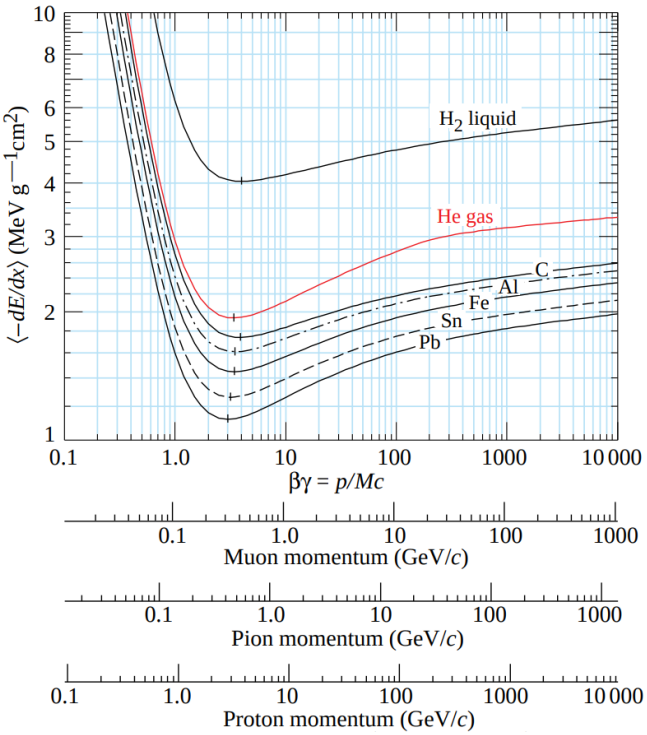


Figure B.2: Mass stopping power in some materials as function of $\beta\gamma$. The momentum scale is also shown for muons, pions and protons, in units of GeV/c [58].

velocity. The Bethe-Bloch region is the only mass-independent region (see form of Equation B.1, together with the $W_{\max}$ expression). The *linear stopping power*, in units of MeV/cm, can be obtained via $\langle -\mathrm{d}E/\mathrm{d}x \rangle \rho$. The advantage of using the mass stopping power is that it is mainly material-independent, as shown in Figure B.2. It can be seen that the mean energy loss minimum is located at around the same $\beta\gamma$ value, irrespective of the target material, and the curve shape is quite consistent as well. The significantly steeper relativistic rise that is observed for the helium gas is a result of the much lower density effect correction, due to the considerably smaller material polarization compared to denser absorbers.

The $\langle -\mathrm{d}E/\mathrm{d}x \rangle$ depends on the quantity $\beta\gamma = p/Mc$, i.e. to both the momentum and the mass of the particle. Consequently, by measuring the momentum of the particle (via the track curvature in a magnetic field) and the $\langle -\mathrm{d}E/\mathrm{d}x \rangle$, the mass of the particle can be determined. However, since the Bethe-Bloch formula describes the mean energy loss, for which rare events with very large single-collision energy losses are accounted for, the mean energy loss is not a good estimator. For the ALICE TPC, the maximum number of charge (and therefore energy) measurements is 159 (152) in LHC Run 1 and 2 (Run 3) - one per pad row. In that case, the mean of 60% of the smallest signals is used as an estimator (*truncated mean*) which is also called the *TPC signal* [22]. The TPC signal for different particle tracks, measured in pp collisions at $\sqrt{s} = 13\,\mathrm{TeV}$, as a function of their momentum, is shown in Figure B.3, illustrating the PID capabilities of the ALICE TPC. Solid curves correspond to the Bethe-Bloch formula. For the regions where the curves overlap, other PID techniques must be combined to successfully identify the particle species on a track-by-track basis.

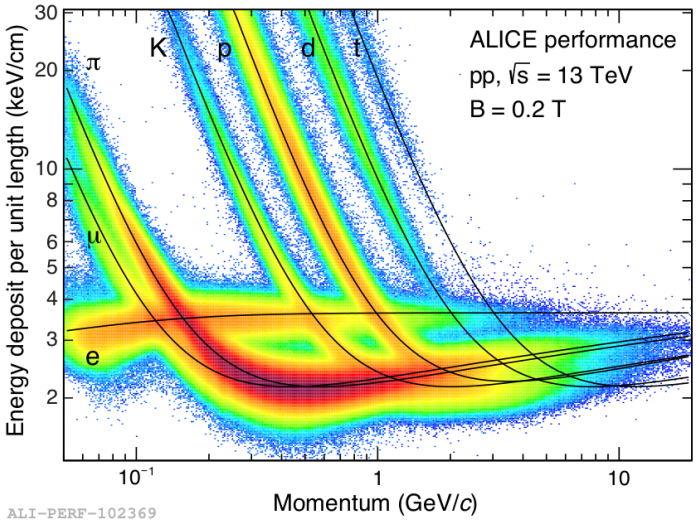


Figure B.3: Energy deposit per unit length as a function of the particle momentum in pp collisions at $\sqrt{s} = 13\,\mathrm{TeV}$. The black solid lines indicate the Bethe-Bloch parametrization [58].

B.2 Movement of electrons in gases

In the presence of an electric or a magnetic field, free electrons and ions gain energy from the field and move in the direction dictated by it; this movement is called *drift*. Additionally, due to the thermal energy in a gas, the charge carrier will also exhibit random motion due to collisions with the gas atoms or molecules; this movement is called *diffusion*.

B.2.1 Drift

Macroscopically, the equation describing the motion of a charged particle in a medium in which an electric and a magnetic field is applied is given by [59]:

$$m\frac{\mathrm{d}\boldsymbol{u}}{\mathrm{d}t} = e\boldsymbol{E} + e[\boldsymbol{u}\times\boldsymbol{B}] - K\boldsymbol{u} \ , \tag{B.2}$$

where m and e are the mass and charge of the electron respectively, while $\boldsymbol{u}$ is the velocity vector. The term $e\boldsymbol{E} + e[\boldsymbol{u}\times\boldsymbol{B}]$ is merely the Lorentz force exerted to the particle due to the electric and magnetic fields. The term $K\boldsymbol{u}$ describes a resistive force, proportional to the particle velocity, which is added to account for charge carrier collisions with the gas molecules during its acceleration.

The ratio m/K has time dimension and can be interpreted as the average time between collisions:

$$\tau = \frac{m}{K} \ . \tag{B.3}$$

At times $t \gg \tau$, a steady state ($\mathrm{d}\boldsymbol{u}/\mathrm{d}t = 0$) is reached, therefore replacing to Equation B.2 and solving for the velocity, one obtains:

$$\boldsymbol{u} = \frac{e}{m}\tau|\boldsymbol{E}|\frac{1}{1+\omega^2\tau^2}\left(\hat{\boldsymbol{e}} + \omega\tau[\hat{\boldsymbol{e}}\times\hat{\boldsymbol{b}}] + \omega^2\tau^2(\hat{\boldsymbol{e}}\cdot\hat{\boldsymbol{b}})\cdot\hat{\boldsymbol{b}}\right) \ , \tag{B.4}$$

where $\hat{\boldsymbol{e}}$ and $\hat{\boldsymbol{b}}$ denote the unit vectors in the direction of the electric and magnetic field and $\omega = (e/m)|\boldsymbol{B}|$ is the cyclotron frequency. Defined this way, the quantity $\omega\tau$ is dimensionless.

Setting the magnetic field to zero (therefore $\omega = 0$), one obtains the drift velocity $\boldsymbol{u}_\mathrm{d}$, which is parallel to the electric field:

$$\boldsymbol{u}_\mathrm{d} = \frac{e}{m}\tau\boldsymbol{E} = \mu\boldsymbol{E} \ , \tag{B.5}$$

where

$$\mu = \frac{e}{m}\tau \tag{B.6}$$

is called the *mobility* of the charge carrier. As an example, for the upgraded ALICE TPC with a Ne–CO_2–N_2 (90−10−5) gas mixture and for a drift field of 400 V/cm, the electron drift velocity is 2.58 cm/μs, while the ion drift velocity is 1.168 cm/ms [8].

B.2.2 Diffusion

In the absence of electric and magnetic fields, the charge carrier movement is only thermal. Through collisions with the gas molecules, the carriers exchange energy and eventually their energy distribution approaches the Maxwell-Boltzmann limit. A point-like electron cloud starting to diffuse at $t = 0$ from the origin will, after time t, follow a Gaussian density distribution:

$$n(r,t) = \left(\frac{1}{\sqrt{4\pi Dt}}\right)^3 \exp\left(-\frac{r^2}{4Dt}\right) \ , \tag{B.7}$$

where r is the distance from the origin and D the diffusion constant.

By introducing an electric field, the diffusion in the drift direction (parallel to the electric field) and perpendicular to it is different, D_l and D_t, respectively. In such a case, the formula takes the form:

$$n(r,t) = \left(\frac{1}{\sqrt{4\pi D_l t}}\right)\left(\frac{1}{\sqrt{4\pi D_t t}}\right)^2 \exp\left(-\frac{x^2+y^2}{4D_t t} - \frac{(z-u_d t)^2}{4D_l t}\right) \ , \tag{B.8}$$

where u_d is the drift velocity. By comparing Equation B.8 with the typical form of the normal distribution, one sees that the width of the charge distribution in either direction is given by $\sigma = \sqrt{2Dt}$, which increases over time. Together with $L = u_d t$, the width can be expressed as

$$\sigma = \sqrt{\frac{2D}{u_d} L} \ , \tag{B.9}$$

and finally the drift length-independent quantities can be introduced:

$$D_{\mathrm{L}} = \frac{\sigma_l}{\sqrt{L}} = \sqrt{\frac{2D_l}{\mu E}} \ , \tag{B.10}$$

$$D_{\mathrm{T}} = \frac{\sigma_t}{\sqrt{L}} = \sqrt{\frac{2D_t}{\mu E}} \ . \tag{B.11}$$

In the presence of a magnetic field, the diffusion transversal to $\boldsymbol{B}$ is reduced:

$$D(B \neq 0) = \frac{D(B=0)}{1+\omega^2\tau^2} \ , \tag{B.12}$$

while the diffusion along the magnetic field remains the same.

For the upgraded ALICE TPC, the electron transverse and longitudinal diffusion coefficients are $D_{\mathrm{T}} = 209\,\mu\mathrm{m}/\sqrt{\mathrm{cm}}$ and $D_{\mathrm{L}} = 221\,\mu\mathrm{m}/\sqrt{\mathrm{cm}}$, respectively. As an example, for a magnetic field of 0.5 T the transversal diffusion is thus reduced by $\approx 10\%$ [6].

C References to external code

Conversion of raw data

The raw data to digits conversion is conducted via the O2 framework [36] shell script `dumpDigits.sh`, which makes use of the `CalibRawBase` class. The resulting `digits TTree` contains the following branches:

```
event index, sector, row, pad, timebin, ADC value
```

Event visualization

For a simple visualization of the data, the raw data decoder can be used, `startMonitor`.

Pedestal and noise calibration

The pedestal and noise during the standalone pedestal measurements can be calculated via the `runPedestal.sh` shell script, which uses the `CalibPedestal` class of the O2 framework.

Pulser calibration

The data collected during the pulser measurements can be analyzed via the `runPulser.sh` shell script, which uses the `CalibPulser` class of O2.

Local-Global pad coordinates

A set of local and global x and y coordinates are assigned to each TPC pad, corresponding to the position of its center. They can be retrieved using the `Mapper` class of O2.

Ideal laser track coordinates

The ideal laser track coordinates can be accessed via the `AliTPCLaserTrack` class of `AliRoot` [60].

D Lists

D.1 List of Figures

D.2 List of Tables

E Bibliography

[1] P. Collier, "Status and Plans for the CERN Accelerator Complex", *PoS* **LeptonPhoton2019** (2019) 002.

[2] ALICE Collaboration, *ALICE: Technical proposal for a Large Ion collider Experiment at the CERN LHC*. LHC Tech. Proposal. CERN, Geneva, 1995. http://cds.cern.ch/record/293391.

[3] B. B. Brandt, M. Cè, *et al.*, "Lattice QCD estimate of the quark-gluon plasma photon emission rate", arXiv:1912.00292 [hep-lat]. https://arxiv.org/abs/1912.00292.

[4] D. Boyanovsky, "Phase transitions in the early and the present universe: From the big bang to heavy ion collisions", in *International School of Astrophysics 'Daniel Chalonge' 8th Course: Opening the 3rd Millennium, Phase Transitions in the Early Unverse: Theory and Observations*, pp. 3–44. 2, 2001. arXiv:hep-ph/0102120.

[5] ALICE Collaboration, "Upgrade of the ALICE Time Projection Chamber", Tech. Rep. CERN-LHCC-2013-020. ALICE-TDR-016, Oct, 2013. https://cds.cern.ch/record/1622286.

[6] The ALICE Collaboration, "Upgrade of the ALICE Time Projection Chamber",. https://cds.cern.ch/record/1622286.

[7] F. Sauli, "GEM: A new concept for electron amplification in gas detectors", *Nucl. Instrum. Meth. A* **386** (1997) 531–534.

[8] ALICE TPC Collaboration, "The upgrade of the ALICE TPC with GEMs and continuous readout", *Journal of Instrumentation (JINST), to be published* (2020) .

[9] D. J. Gross, "Twenty five years of asymptotic freedom", *Nuclear Physics B - Proceedings Supplements* **74** no. 1-3, (Mar, 1999) 426–446. http://dx.doi.org/10.1016/S0920-5632(99)00208-X.

[10] ALICE Collaboration, "The ALICE experiment at the CERN LHC", *JINST* **3** (2008) S08002.

[11] A. Tauro, *ALICE Schematics*, May, 2017. http://cds.cern.ch/record/2263642. General Photo.

[12] ALICE Collaboration, "Technical Design Report for the Upgrade of the ALICE Inner Tracking System", *Journal of Physics G: Nuclear and Particle Physics* **41** no. 8, (Jul, 2014) 087002. https://doi.org/10.1088%2F0954-3899%2F41%2F8%2F087002.

[13] ALICE Collaboration, "The ALICE Transition Radiation Detector: Construction, operation, and performance", *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **881** (2018) 88 – 127. http://www.sciencedirect.com/science/article/pii/S0168900217309981.

[14] ALICE Collaboration, *ALICE Time-Of-Flight system (TOF): Technical Design Report*. Technical Design Report ALICE. CERN, Geneva, 2000. https://cds.cern.ch/record/430132.

[15] ALICE Collaboration, *ALICE high-momentum particle identification: Technical Design Report*. Technical Design Report ALICE. CERN, Geneva, 1998. https://cds.cern.ch/record/381431.

[16] ALICE Collaboration, "Performance of prototypes for the ALICE electromagnetic calorimeter", *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **615** no. 1, (2010) 6 – 13. http://www.sciencedirect.com/science/article/pii/S0168900209023985.

[17] ALICE Collaboration, "ALICE DCal: An Addendum to the EMCal Technical Design Report Di-Jet and Hadron-Jet correlation measurements in ALICE", Tech. Rep. CERN-LHCC-2010-011. ALICE-TDR-14-add-1, Jun, 2010. https://cds.cern.ch/record/1272952.

[18] ALICE Collaboration, *ALICE Photon Spectrometer (PHOS): Technical Design Report*. Technical Design Report ALICE. CERN, Geneva, 1999. https://cds.cern.ch/record/381432.

[19] ALICE Collaboration, *ALICE Zero-Degree Calorimeter (ZDC): Technical Design Report*. Technical Design Report ALICE. CERN, Geneva, 1999. https://cds.cern.ch/record/381433.

[20] A. I. Maevskaya for the ALICE Collaboration, "Fast Interaction Trigger for the upgrade of the ALICE experiment at CERN: design and performance. The Fast Interaction Trigger detector for the upgrade of the ALICE experiment at CERN: design and performance", *EPJ Web Conf.* **204** no. arXiv:1812.00594, (Dec, 2018) 11003. 6 p. https://cds.cern.ch/record/2649931.

[21] ALICE Collaboration, *ALICE dimuon forward spectrometer: addendum to the Technical Design Report*. Technical Design Report ALICE. CERN, Geneva, 2000. https://cds.cern.ch/record/494265.

[22] ALICE Collaboration, "Performance of the ALICE Experiment at the CERN LHC", *Int. J. Mod. Phys. A* **29** no. CERN-PH-EP-2014-031. CERN-PH-EP-2014-031, (Feb, 2014) 1430044. 120 p. https://cds.cern.ch/record/1648854.

[23] A. Deisting for the ALICE Collaboration, "Status of the R&D activities for the upgrade of the ALICE TPC", *EPJ Web of Conferences* **174** (01, 2016) .

[24] J. Wiechula, *Commissioning and Calibration of the ALICE-TPC*. Doctoral thesis, Johann Wolfgang Goethe - Universität, Frankfurt am Main, Germany, 2008. https://www.uni-frankfurt.de/46491136/Generic_46491136.pdf.

[25] R. Veenhof, "Choosing a gas mixture for the ALICE TPC",. https://inspirehep.net/literature/1236940.

[26] The STAR Collaboration, "The laser system for the STAR time projection chamber", *Nucl. Instrum. Meth. A* **499** (2003) 692–702.

[27] B. Nielsen *et al.*, "Design Note on the ALICE TPC laser calibration system", *ALICE-INT-2002-022* (2002) . https://edms.cern.ch/ui/#!master/navigator/document?D:1508928712:1508928712:subDocs.

[28] ALICE TPC Collaboration, "The ALICE TPC, a large 3-dimensional tracking device with fast readout for ultra-high multiplicity events", *ALICE-INT-2003-29* (2003) . https://edms.cern.ch/document/470838/1.

[29] ALICE Collaboration, "Upgrade of the ALICE experiment: Letter of intent", *Journal of Physics G: Nuclear and Particle Physics* **41** no. 8, (Jul, 2014) 087001. https://doi.org/10.1088%2F0954-3899%2F41%2F8%2F087001.

[30] *Garfield - simulation of gaseous detectors*. http://garfield.web.cern.ch/garfield/.

[31] F. Sauli, "The gas electron multiplier (GEM): Operating principles and applications", *Nucl. Instrum. Meth. A* **805** (2016) 2 – 24. http://www.sciencedirect.com/science/article/pii/S0168900215008980.

[32] F. V. Böhmer *et al.*, "Space-Charge Effects in an Ungated GEM-based TPC", 2012. https://arxiv.org/abs/1209.0482.

[33] D. Vranic, *Private Communication*.

[34] H. Appelshäuser, M. Arslandok, *et al.*, "Readout scheme of the upgraded ALICE TPC",. https://cds.cern.ch/record/2231785.

[35] A. Önnerstad and P. Chatzidaki, *ALICE TPC Pre-Commissioning Song*. https://www.youtube.com/watch?v=G0RS--32VvA&ab_channel=AnnaYiota.

[36] *ALICE O^2 project for LHC Run 3.* https://alice-o2-project.web.cern.ch/.

[37] M. Arslandok, *Event-by-Event Identified Particle Ratio Fluctuations in Pb–Pb Collisions with ALICE.* Doctoral thesis, Johann Wolfgang Goethe - Universität, Frankfurt am Main, Germany, 2017. https://www.uni-frankfurt.de/74207220/Arslandok_PhD_2017.pdf.

[38] *ROOT Data Analysis Framework.* https://root.cern/.

[39] A. Deisting, *Measurements of ion mobility and GEM discharge studies for the upgrade of the ALICE Time Projection Chamber.* Doctoral thesis, Ruperto-Carola-University of Heidelberg, Germany, 2018. https://archiv.ub.uni-heidelberg.de/volltextserver/24133/.

[40] M. Ivanov, *Private Communication.*

[41] M. Kubat, "An Introduction to Machine Learning", 2017. https://link.springer.com/book/10.1007/978-3-319-63913-0#about.

[42] Tin Kam Ho, "Random decision forests", 1995.

[43] *ROOT interactive framework.* https://github.com/miranov25/RootInteractive.

[44] B. Efron and R. J. Tibshirani, *An introduction to the bootstrap.* Mono. Stat. Appl. Probab. Chapman and Hall, London, 1993. http://cds.cern.ch/record/526679.

[45] *PSpice - Electronic Circuit Optimization & Simulation.* https://www.pspice.com/.

[46] P. Gasik, C. Garabatos, *Private Communication.*

[47] *Common-mode effect visualization using the ROOT interactive framework – html file.* ttps://indico.cern.ch/event/889369/contributions/4044542/attachments/2118635/3565055/laserCommonModeVisualization.html.

[48] *Common-mode effect visualization using the ROOT interactive framework – demonstration video.* https://indico.cern.ch/event/889369/contributions/4044542/attachments/2118635/3603417/laserCommonModeVisualization.webm.

[49] C. Lippmann, *Private Communication.*

[50] M. Arslandok, *Private Communication.*

[51] T. Gunji, *Private Communication.*

[52] *ANSYS Engineering Software*. https://www.ansys.com.

[53] *Magboltz - transport of electrons in gas mixtures*. http://magboltz.web.cern.ch/magboltz/.

[54] *Garfield++*. https://garfieldpp.web.cern.ch/garfieldpp/.

[55] T. Alt, *Private Communication*.

[56] L. Betev and P. Chochula, *Definition of the ALICE Coordinate System and Basic Rules for Sub-detector Components Numbering*, 2003. http://edms.cern.ch/document/406391. ALICE-INT-2003-038.

[57] C. Lippmann and D. Vranić, *Alice TPC Numbering Conventions*. https://subversion.gsi.de/alice-tpc-fee/numbering/num.pdf.

[58] C. Patrignani *et al.*, "Review of particle physics", *Chinese Physics C* **40** no. 10, (Oct, 2016) 100001. https://doi.org/10.1088%2F1674-1137%2F40%2F10%2F100001.

[59] L. Rolandi, W. Riegler, W. Blum, "Particle Detection with Drift Chambers", 2008. https://link.springer.com/book/10.1007/978-3-540-76684-1#toc.

[60] *Aliroot: ALICE offline framework for LHC Run 1 and 2*. http://alidoc.cern.ch/AliRoot/master/.

Acknowledgments

The past twelve months have been full of experiences and surprises. During this time, many people contributed to my scientific and personal progress.

Firstly I would like to thank my supervisor, Priv.-Doz. Dr. Yvonne Pachmayer, the lecture of whom attracted me to the topic of detector physics. She has been supportive – professionally and mentally – in every step of this thesis, but also gave me freedom to thrive. I thank her in particular for her valuable feedback and for being extremely meticulous with proof-reading the thesis.

I am very thankful to Prof. Dr. Johanna Stachel for giving me this wonderful opportunity to join the TPC group at CERN. She has been closely following my work, making helpful suggestions along the way, and she personally ensured my smooth transition during the pandemic.

I truly appreciate the guidance I received from Mesut Arslandok. His patience in my first steps in ROOT was admirable, and his tangible explanations enabled me to improve immensely both in terms of programming skills but also in the comprehension of the effects in the TPC. His recommendations on conveying the important messages really shaped this thesis. I would also like to express my gratitude to Marian Ivanov. I learnt a lot from his problem-solving approach and his deep understanding of the TPC. His high expectations of me helped me grow and consider myself a scientist. With his attention to the details, he made this a work I am truly proud of. Many thanks to Jens Wiechula, who very gladly introduced me to the O2 framework during my first days at CERN and has always been available to answer my questions or provide detailed instructions.

Special thanks to Christian Lippmann, who trusted me with important tasks in the clean-room, and to Chilo Garabatos, who was the first to show me around at Point 2, and made me feel very welcome. Our common work in the clean-room with both of them gave me many opportunities to ask questions about every aspect of the TPC, all of which they happily answered. Overall the TPC group was a warm, ideal environment that allowed me to be myself, but at the same time to grow and learn (and have fun!).

Keeping in touch with the members of the ALICE group in Heidelberg was a pleasure, and also enabled me to learn about the physics analysis of ALICE.

I would like to thank Christian, Chilo and Marten Ole Schmidt for taking the time to proof-read parts of the thesis. I also value the interesting discussions with Taku Gunji, Piotr Gasik, Robert Helmut Munzer, Rob Veenhof and Torsten Alt.

I feel so glad to have met Anna and Lisa, whose friendship turned into something unique during our time at CERN. I would have never imagined us bonding so much and connecting so deeply. Finally, thanks to Bernd for writing the German abstract, but mostly for helping me out during my stressful times, and always understanding.

Erklärung:

Ich versichere, dass ich diese Arbeit selbstständig verfasst habe und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.

Heidelberg, den (Datum) 17/12/2020 ...