Correction of the baseline fluctuations in the GEM-based ALICE TPC

The ALICE TPC collaboration

J. Alme,[4] T. Alt,[12] C. Andrei,[10] V. Anguelov,[19] H. Appelshäuser,[12] M. Arslandok,[30,*]
G.G. Barnaföldi,[29] E. Bartsch,[12] P. Becht,[19,21] R. Bellwied,[25] A. Berdnikova,[19] N. Bialas,[12]
P. Bialas,[12] S. Biswas,[1] B. Blidaru,[19,21] L. Boldizsár,[29] L. Bratrud,[12] P. Braun-Munzinger,[21]
M. Bregant,[24] C.L. Britton,[17] S. Brucker,[19] E.J. Brücken,[8] H. Büsching,[12] R. Soto
Camacho,[9] A.J. Castro,[26] P. Chatzidaki,[19] P. Christiansen,[14] L.G. Clonts,[17]
T.M. Cormier,[17,†] P. Dhankher,[11] S. Dittrich,[12] R. Ehlers,[30] M.N. Ericson,[17] N.B. Ezell,[17]
L. Fabbietti,[20] F. Flor,[25] J.J. Gaardhøje,[16] M.G. Munhoz,[24] C. Garabatos,[21] P. Gasik,[21]
T. Geiger,[12] Á. Gera,[29] P. Glässel,[19] D.J.Q. Goh,[15] A. Grein,[12] T. Gundem,[12] T. Gunji,[27]
M. Habib,[21] H. Hamagaki,[15] G. Hamar,[29] J.C. Hansen,[16] J.W. Harris,[30] P. Hauer,[7]
S. Hayashi,[27] J. Hehner,[21] J.K. Heino,[8] E. Hellbär,[12] H. Helstrup,[6] M. Hemmer,[12]
A. Herghelegiu,[10] R.A. Hernandez,[24] H.D. Hernandez Herrera,[24] T.E. Hilden,[8] C. Hughes,[26]
S. Hummel,[19] M. Ivanov,[21,*] J. Jäger,[12] J. Jung,[12] M. Jung,[12] E. Kangasaho,[8] B. Ketzer,[7]
S. Kirsch,[12] M. Kleiner,[12] T. Klemenz,[20] A.G. Knospe,[25] M. Kowalski,[22] L. Kreis,[21]
M. Krüger,[12] N. Kupfer,[19] R. Lang,[20] L. Lautner,[20] M. Lesch,[20] Y. Lesenechal,[5] F. Liebske,[12]
C. Lippmann,[21] R.D. Majka,[30,†] C. Markert,[23] T.A. Martins,[24] S. Masciocchi,[21]
O. Matonoha,[14] A. Matyja,[22] M. Meres,[2] D.L. Mihaylov,[20] D. Miśkowiec,[21] T. Mittelstaedt,[19]
C. Mordasini,[20] T. Morhardt,[21] S. Muley,[19] J. Mulligan,[30] R.H. Munzer,[12] H. Murakami,[27]
K. Münning,[7] A. Nassirpour,[14] C. Nattrass,[26] B.S. Nielsen,[16] W.A.V. Noije,[24] M. Ogino,[15]
A.C. Oliveira Da Silva,[26] A. Oskarsson,[14] K. Oyama,[15] A. Önnerstad,[14] L. Österman,[14]
Y. Pachmayer,[19] G. Paić,[13] R.N. Patra,[28] V. Peskov,[12] M. Petris,[10] M. Petrovici,[10]
M. Planinic,[18] L. Prodan,[10] A. Radu,[10] J. Rasson,[17] K.F. Read,[17] A. Rehman,[4] R. Renfordt,[12]
K. Røed,[3] D. Röhrich,[4] E. Rubio,[19] A. Rusu,[17] B.C.S. Sanches,[24] J. Schambach,[23]
S. Scheid,[12] C. Schmidt,[21] A. Schmier,[26] K. Schweda,[21] D. Sekihata,[27] S. Siebig,[19]
R.W.D. Silva,[24] D. Silvermyr,[14] B. Sitar,[2] N. Smirnov,[30] H.K. Soltveit,[19] S.P. Sorensen,[26]
J. Stachel,[19] L. Šerkšnytė,[20] G. Tambave,[4] K. Ullaland,[4] B. Ulukutlu,[20] D. Varga,[29]
O. Vazquez Rueda,[14] A. Velure,[4] S. Vergara Limón,[9] O. Vorbach,[19] B. Voss,[21] C. Weidlich,[12]
J. Wiechula,[12] B. Windelband[19] and S. Winkler[20]

[†]Deceased.
[*]Corresponding author.

[1]*Bose Institute, Department of Physics and Centre for Astroparticle Physics and Space Science (CAPSS),
Kolkata, India*

[2]*Comenius University Bratislava, Faculty of Mathematics, Physics and Informatics,
Bratislava, Slovakia*

[3]*Department of Physics, University of Oslo, Oslo, Norway*

[4]*Department of Physics and Technology, University of Bergen, Bergen, Norway*

[5]*European Organization for Nuclear Research (CERN), Geneva, Switzerland*

[6]*Faculty of Engineering and Science, Western Norway University of Applied Sciences,
Bergen, Norway*

[7]*Helmholtz-Institut für Strahlen- und Kernphysik,
Rheinische Friedrich-Wilhelms-Universität Bonn,
Bonn, Germany*

[8]*Helsinki Institute of Physics (HIP), Helsinki, Finland*

[9]*High Energy Physics Group, Universidad Autónoma de Puebla, Puebla, Mexico*

[10]*Horia Hulubei National Institute of Physics and Nuclear Engineering,
Bucharest, Romania*

[11]*Indian Institute of Technology Bombay (IIT), Mumbai, India*

[12]*Institut für Kernphysik, Johann Wolfgang Goethe-Universität Frankfurt,
Frankfurt, Germany*

[13]*Instituto de Ciencias Nucleares, Universidad Nacional Autónoma de México,
Mexico City, Mexico*

[14]*Lund University Department of Physics, Division of Particle Physics,
Lund, Sweden*

[15]*Nagasaki Institute of Applied Science, Nagasaki, Japan*

[16]*Niels Bohr Institute, University of Copenhagen, Copenhagen, Denmark*

[17]*Oak Ridge National Laboratory, Oak Ridge, Tennessee, U.S.A.*

[18]*Physics department, Faculty of science, University of Zagreb, Zagreb, Croatia*

[19]*Physikalisches Institut, Ruprecht-Karls-Universität Heidelberg,
Heidelberg, Germany*

[20]*Physik Department, Technische Universität München, Munich, Germany*

[21]*Research Division and ExtreMe Matter Institute EMMI,
GSI Helmholtzzentrum für Schwerionenforschung GmbH,
Darmstadt, Germany*

[22]*The Henryk Niewodniczanski Institute of Nuclear Physics, Polish Academy of Sciences,
Cracow, Poland*

[23]*The University of Texas at Austin, Austin, Texas, U.S.A.*

[24]*Universidade de São Paulo (USP), São Paulo, Brazil*

[25]*University of Houston, Houston, Texas, U.S.A.*

[26]*University of Tennessee, Knoxville, Tennessee, U.S.A.*

[27]*University of Tokyo, Tokyo, Japan*

[28]*Variable Energy Cyclotron Centre, Kolkata, India*

[29]*Wigner Research Centre for Physics, Budapest, Hungary*

[30]*Yale University, New Haven, Connecticut, U.S.A.*

E-mail: mesut.arslandok@cern.ch, marian.ivanov@cern.ch

ABSTRACT: To operate the ALICE Time Projection Chamber in continuous mode during the Run 3 and Run 4 data-taking periods of the Large Hadron Collider, the multi-wire proportional chamber-based readout was replaced with gas-electron multipliers. As expected, the detector performance is affected by the so-called common-mode effect, which leads to significant baseline fluctuations. A detailed study of the pulse shape with the new readout has revealed that it is also affected by ion tails. Since reconstruction and data compression are performed fully online, these effects must be corrected at the hardware level in the FPGA-based common readout units. The characteristics of the common-mode effect and of the ion tail, as well as the algorithms developed for their online correction, are described in this paper. The common-mode dependencies are studied using machine-learning techniques. Toy Monte Carlo simulations are performed to illustrate the importance of online corrections and to investigate the performance of the developed algorithms.

Contents

1 Introduction

Charged particles passing through the active volume of the Time Projection Chamber (TPC) ionize the gas along their path. The electrons generated in this process drift toward the end plates on which the readout chambers are mounted. By amplifying the signal in the readout chambers, the TPC provides a three-dimensional reconstruction of the charged-particle tracks. In the data-taking periods Run 1 (2009–2013) and Run 2 (2015–2018) of the Large Hadron Collider (LHC), the readout chambers of the ALICE TPC [1, 2] consisted of multiwire proportional chambers (MWPCs) [3]. The ions generated during the amplification process in the MWPCs were blocked by a "gating grid", a series of wires located between the cathode wires and the drift volume. However, the time requirements of the operation with gating grid limited the maximum readout rate of the TPC to around 3 kHz. On the other hand, operation of the MWPC-based TPC without ion gating would lead to intolerably large space-charge distortions in the drift region. Therefore, in order to operate the TPC with the expected minimum bias Pb-Pb collision rate of 50 kHz in Run 3 and Run 4

(2022–2030) [4], all readout chambers were replaced by gas electron multipliers (GEMs) during the Long Shutdown 2 (LS2) (December 2018 – March 2022) of the CERN LHC, since stacks of GEMs have intrinsic ion blocking capabilities [5]. To achieve the required gain while effectively suppressing the back-flow of ions, a stack of four GEM foils named from top to bottom (i.e. from drift to pad plane): GEM1, GEM2, GEM3, and GEM4, was employed for all chambers [6]. At the same time, due to the change from triggered to continuous operation, the front-end electronics (FEE) had to be replaced. A detailed description of the upgrade, from the design of the chambers and the new FEE to the pre-commissioning of the upgraded TPC, can be found in ref. [7].

The typical semi-Gaussian pulse shape of a single readout channel is shown in figure 1. The signals from the charged particles are influenced by various effects, such as diffusion and track inclination angle. They also exhibit a characteristic undershoot, due to capacitive coupling across the induction gap between the pad plane and the GEM foils, and a long overshoot after the signal pulse caused by the slow movement of the ions in the induction gap. These two effects are referred to as "common-mode" (CM) effect and "ion tail" (IT), respectively. They are more prominent for high multiplicity environments and, if not accounted for, lead to significant deterioration of particle identification (PID) and tracking performance of the detector. Detailed studies conducted during Run 1 and Run 2, where similar baseline fluctuations were observed, emphasize the crucial need to understand, simulate, and correct for these effects to maintain the performance of the TPC [8].
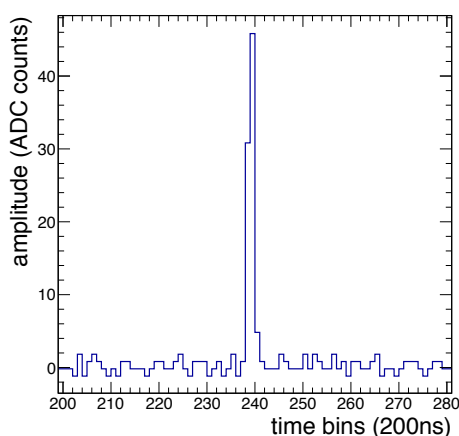
Figure 1. Typical electronics response of a single readout channel [6, 7]. Reproduced from [6]. © 2021 CERN for the benefit of the ALICE collaboration. Published by IOP Publishing Ltd on behalf of Sissa Medialab. CC BY 4.0.

This paper is organized as follows; section 2 describes the measurements obtained during the pre-commissioning phase during LS2. These data served as input for the initial analysis of the two effects. In section 3, the dependencies of the common-mode effect are investigated using machine learning techniques. In section 4, a detailed analysis of the ion-tail properties is performed. In section 5, the two online correction algorithms as implemented at the hardware level in the FPGA-based common readout units (CRUs) are described, and a proof of principle is provided. Toy Monte Carlo (MC) simulations are performed to demonstrate the impact of the two effects on the TPC signals and to investigate the performance of the two online correction algorithms.

2 Measurements

A quadruple GEM configuration with foils having different hole pitch was optimized for the ALICE TPC during an extensive R&D phase. The TPC volume is divided equally into two readout sides (A-side and C-side) by a central electrode. The TPC sectors are positioned on the endplates, each covering 20° in azimuth. They are radially segmented into an inner and outer readout chambers (IROC and OROC, respectively). The OROC is further subdivided into three individual GEM stacks, therefore a TPC sector consists of a total of four GEM stacks, labeled IROC (stack 0), OROC 1 (stack 1), OROC 2 (stack 2), and OROC 3 (stack 3) (see figure 2). The pad signals are read out by the FEE mounted on the TPC end plates at a sampling rate of 5 MHz, which corresponds to a time bin duration of 200 ns [7, 9]. Custom-made ASICs, called "SAMPA" chips [10], are responsible for the signal amplification, shaping, and analog-to-digital conversion. The digitized data are then streamed to the CRUs, where the baseline subtraction, common-mode effect correction, ion-tail filtering, and zero suppression (ZS) are performed.
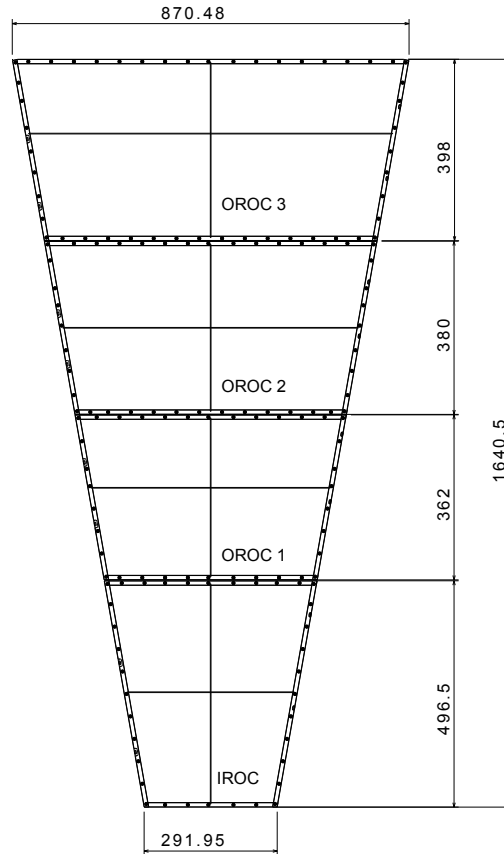
Figure 2. Dimensions (mm) of a sector of the TPC with four GEM stacks. The spacer cross, a structure that ensures the mechanical stability of the GEM foils against electrostatic forces, is shown as two 1.5 mm wide and 2 mm deep bars in the longitudinal and transverse directions for each stack [7]. Reproduced from [6]. © 2021 CERN for the benefit of the ALICE collaboration. Published by IOP Publishing Ltd on behalf of Sissa Medialab. CC BY 4.0.

2.1 Laser data

To study the common-mode and ion-tail effects, the laser calibration system of the TPC was used [11, 12]. Laser data were collected for all sectors during the pre-commissioning of the TPC in the clean room located at the LHC Point 2 [7]. The laser is a pulsed (5 ns pulse duration, 10 Hz repetition rate) Nd:YAG laser equipped with two frequency doublers, resulting in a final wavelength of 266 nm, which corresponds to an ionization energy of 4.66 eV. However, the ionization potential of the TPC gas components is much larger ($E_{Ne} \approx 22$ eV, $E_{CO_2} \approx 14$ eV, $E_{N_2} \approx 16$ eV). Therefore, with two-photon processes [13], the laser ionizes organic impurities (approximately 1 ppm) in the gas with ionization potentials of $5 - 8$ eV. Through a series of mirrors, beam splitters, prisms, and
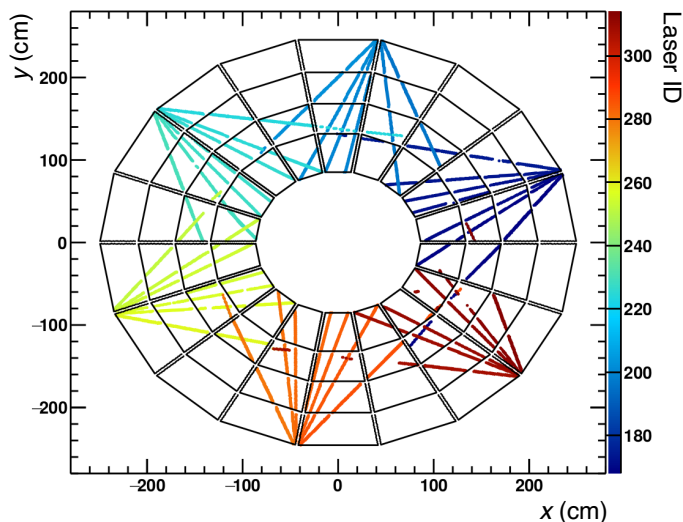

Figure 3. Reconstructed laser tracks of a particular bundle on the C-side of the TPC. The black lines indicate the sector boundaries. The *laser ID* is a unique number assigned to each of the 336 laser tracks.

micromirror bundles, 336 laser tracks simultaneously irradiate the TPC. In each TPC half, 6 wide laser beams illuminate 24 bundles of micromirrors located at four nearly equidistant positions along the LHC beam direction. For each bundle, 7 narrow reflected beams enter the TPC gas parallel to the end plates. Figure 3 shows the reconstructed laser tracks of the C-side bundle 0 (the bundle closest to the endplate), with the color scale indicating the *laser ID*.

In these measurements, for each laser event (corresponding to one laser pulse), raw data with a length of about 500 time bins (one time bin corresponds to 200 ns) were streamed from the FEE to the CRUs, corresponding to the full electron drift time. The data taking was triggered at 10 Hz by the laser system. For each TPC sector, a sufficient amount of laser events (approximately 400–1200) were collected. To increase the signal-to-noise ratio, the signals were averaged over all available events in order to reach better precision of the shapes of the common-mode signal and of the ion tail. Figure 4 shows three laser clusters[1] detected on one TPC pad row. The simultaneous common-mode signal is seen as an undershoot in the remaining pads, which are referred to as *non-signal* or *empty* pads. After the signal pulse, the long ion tail is also clearly visible in the signal pads.

---

[1]A cluster is defined as concentrated deposited charge detected within a search window of 3 bins in pad direction and 3 bins in time direction.
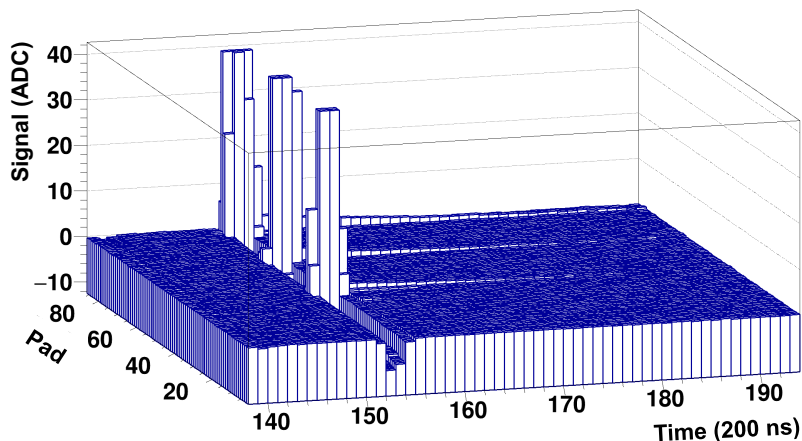
Figure 4. Laser signals and induced common-mode signals in the pads of a given pad row. The signal height axis is zoomed-in such that the signals are not to scale. The ion tail is also visible for the signal pads.

Very large signals saturating the dynamic range of the SAMPA chip were discarded from the analysis, since their actual amplitude is not known. Moreover, very large charges fed into the input of the SAMPA chip may lead to loss of sensitivity of this channel for a short time. This can be seen as a saturation of the SAMPA response at a constant value of about 100 ADC, as shown in the right panel of figure 5 with red and brown solid lines. Note that this effect is an artifact of the electronics and therefore does not affect the neighboring pads.
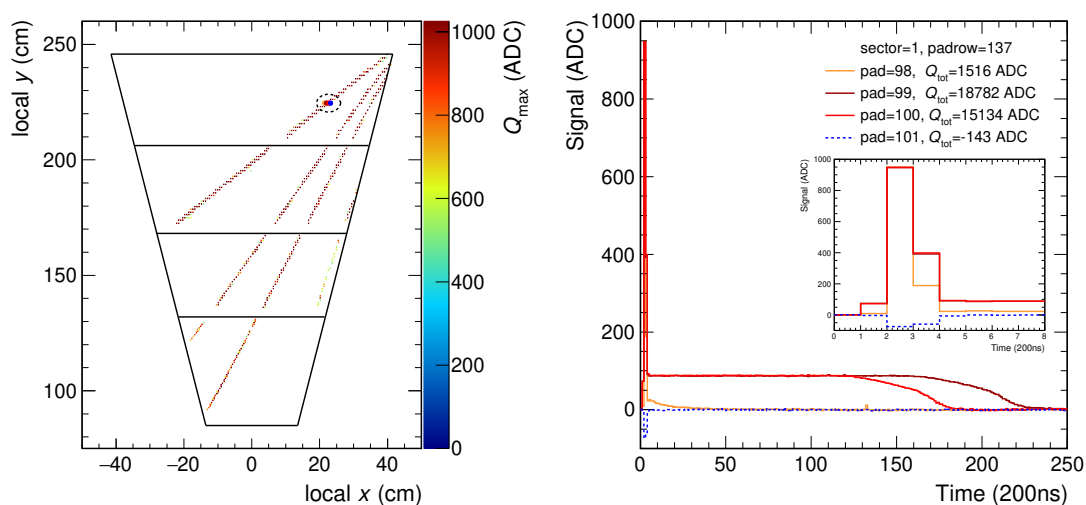
Figure 5. Left: four laser tracks projected on to the local x-y plane. The dashed oval highlights saturated signals. Right: saturated laser signals of the cluster highlighted in the left panel.

2.2 Calibration pulser data

Measurements with a calibration pulser system were performed independently during the pre-commissioning phase in order to investigate the shaping characteristics of the FEE. These measurements involve injecting a pulse into the bottom electrode of the GEM4 foil (GEM4B), which

induces a signal at the pad plane due to capacitive coupling. Ideally, the same charge should be measured in all pads of the same stack since they have the same dimensions. However, as shown in figure 6, significant variations are observed from pad to pad due to sagging of the GEM foil. The figure shows the normalized pulser charge, i.e., the charge normalized to the median charge in the stack, for an IROC. For pads positioned at the stack edges and under the spacer cross, the measured charge is significantly higher due to the presence of a dielectric material at these positions. Despite the spacer cross, the GEM4B bends slightly in the direction of the pad plane due to the stretching of the foil and electrostatic forces. This increases the capacitance and thus the pulser charge. As shown in figure 6, the relative change in the capacitance for a given pad reaches up to 50%.
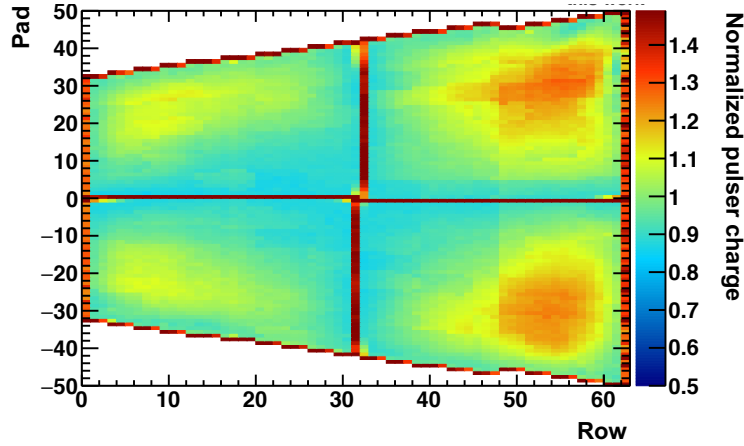
Figure 6. Ratio of pulser charge to median pulser charge in the stack for an IROC.

The measured pulser charge is proportional to the pad capacitance, which is an essential parameter in the analysis of the common-mode effect. Since both, the pulser charge and the effective field between the GEM4B and the pad plane, are inversely proportional to the distance between the GEM4B and the pad plane (except for pads in the spacer cross and at the edges), the pulser charge is also a relevant parameter for the ion-tail studies.

3 Common-mode effect analysis

3.1 The common-mode effect

The common-mode effect results from the capacitive coupling between the GEM foils and the pad plane. When electrons moving from GEM4 to the pad plane induce a signal on one pad, a simultaneous signal of opposite polarity (also called *undershoot*) is induced on all pads of the same stack, as shown in figure 4. This is caused by a voltage drop across the GEM electrode due to the currents caused by the charges drifting through the induction gap. The magnitude of the undershoot in each pad for a given time bin, $Q_{\text{pad}}^{\text{CM}}(t)$, is proportional to the sum of the positive signal in the stack at the same time bin. This pad- and time-independent proportionality factor will be referred to as *common-mode fraction (CF) factor*, $k_{\text{CF,pad}}$, and is defined as

$$k_{\text{CF,pad}} = \frac{Q_{\text{pad}}^{\text{CM}}(t)}{\langle Q_{\text{pos}}(t) \rangle_{\text{stack}}}, \tag{3.1}$$

where

$$\left\langle Q_{\text{pos}}(t) \right\rangle_{\text{stack}} = \sum_{\substack{Q(t)>0 \\ \text{instack}}} Q(t) \Big/ N \tag{3.2}$$

is the average positive signal in the stack and N is the number of pads in the stack. The main objective of the common-mode effect analysis is to investigate the dependencies of the $k_{\text{CF,pad}}$. The laser signals usually span three time bins (see figure 4), so for the analysis of the effect, the common-mode charge and the average positive signal in the stack were summed over three time bins around the laser signal. Since the common-mode effect influences all pads in a given stack, the common-mode signals are also present in the laser signal pads, i.e., the true laser signal is slightly larger than the measured signal. This is taken into account in the analysis.

## 3.2 Dependencies of the common-mode effect

All possible dependencies of $k_{\text{CF,pad}}$ were explored using the Random Forest (RF) machine learning algorithm [14] as implemented in the ROOTInteractive framework [15]. To estimate the common-mode signal, 33% of the non-signal pads were randomly selected as training data. Then, the selected sample was randomly subdivided into 200 estimators, and a decision tree with a depth of 12 was generated for each estimator. The dependencies of the $k_{\text{CF,pad}}$ on all available variables were tested. The importance of each variable is listed in table 1.

Table 1. Variable importance for $k_{\text{CF,pad}}$, reflecting how many times the decision tree was divided because of that specific variable.

Variable	Variable importance (%)
normalized pulser charge ($Q_{\text{pulser,pad}}^{\text{norm}}$)	61.1
stack type	36.1
average positive signal in stack	1.0
fraction of signal pads in stack	0.8
remaining dependencies	1.0

The normalized pulser charge, $Q_{\text{pulser,pad}}^{\text{norm}}$ (i.e., the pulser charge measured in a given pad normalized to the mean pulser charge in the stack), measured with the calibration pulser system and the stack type account for approximately 97% of the dependencies. The $Q_{\text{pulser,pad}}^{\text{norm}}$ is responsible for the pad-by-pad capacitance variations, while the stack type (IROC, OROC1, OROC2, OROC3) accounts for the absolute stack capacitance due to the different stack dimensions. Note that the amplitude of the measured laser signal is reduced due to the underlying common-mode signal. This is referred to as missing charge, which is responsible for the second-order effects mentioned in table 1; the *average positive signal in the stack* and the *fraction of signal pads in the stack*. The contribution of track-related properties such as the bundle (i.e. diffusion) and the beam (i.e. track inclination) was investigated, but no significant dependence was found.

In figure 7, the $k_{\text{CF,pad}}$ data, the RF prediction, and the difference between the two for each pad are plotted for the training data, where very good agreement between the data and the prediction is observed. The peak value of $k_{\text{CF,pad}}$ moves towards larger (absolute) values as the stack area and
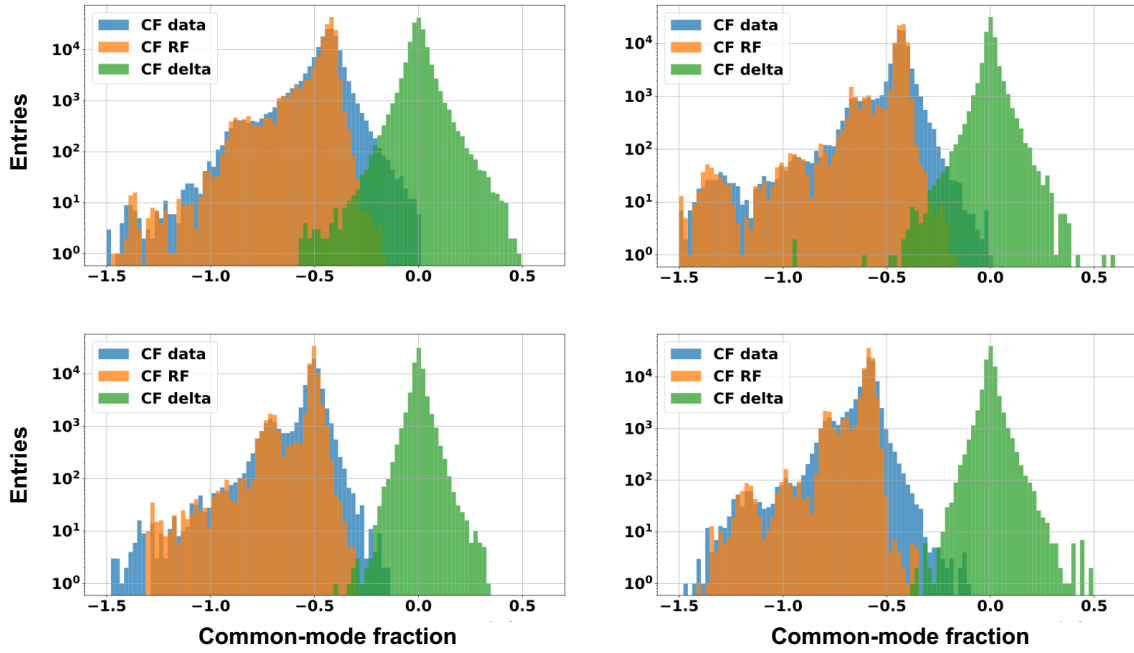
Figure 7. The common-mode fraction obtained from data (blue), random-forest prediction (orange) and difference between the two (green), for IROC (top left), OROC1 (top right), OROC2 (bottom left) and OROC3 (bottom right). The data shown are those used for the random forest training.

Table 2. Position of the maximum of the CF distributions in figure 7 for each stack type.

	Stack area (mm^2)	Peak position of CF
IROC	171154	-0.42 ± 0.03
OROC1	174853	-0.43 ± 0.02
OROC2	231284	-0.50 ± 0.02
OROC3	294836	-0.58 ± 0.02

hence capacitance increases (see table 2). The spread within each stack results from the pad-by-pad capacitance variations, which can be seen as a linear dependence between $k_{\mathrm{CF,pad}}$ and $Q_{\mathrm{pulser,pad}}^{\mathrm{norm}}$ in figure 8. The proportionality also holds for pads with much larger capacitance located at the chamber edges and crosses. By performing linear fits on the data points, the common-mode fraction of a pad can be expressed as

$$k_{\mathrm{CF,pad}} = k_{\mathrm{stack}} \cdot Q_{\mathrm{pulser,pad}}^{\mathrm{norm}}, \tag{3.3}$$

where k_{stack} is the absolute value of the slope (0.44–0.58, depending on the stack type).

The following conclusion can be drawn from the studies: The $k_{\mathrm{CF,pad}}$ of a given pad depends mainly on the capacitance between the pad and the GEM stack (see table 1), which can be described by the stack type and the normalized pulser charge, the latter reflecting the pad-by-pad capacitance variations within a given stack. Note that the missing charge has a negligible effect on the common-mode correction, as discussed above. However, it is accounted for in the full MC simulations that use a GEANT3 implementation of the TPC detector setup [16].
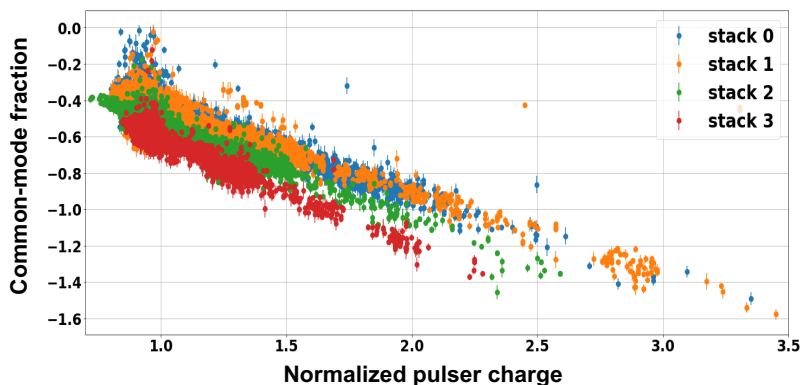
Figure 8. The common-mode fraction ($k_{\mathrm{CF,pad}}$) as a function of normalized pulser charge ($Q^{\mathrm{norm}}_{\mathrm{pulser,pad}}$) for each stack, only for the data with uncertainty of $k_{\mathrm{CF,pad}}$ less than 5%.

4 Ion-tail analysis

4.1 Ion tail in GEMs

The analysis of laser data for the common-mode effect revealed unanticipated signal tails. In figure 9 (left) the response of a single pad to a laser pulse is shown. On the right, the same response is zoomed in on the signal (y) axis. After the signal peak, a long tail is observed, caused by the ions generated in the amplification process, lasting about 16 μs. In this example, the maximum of the tail is about 0.7% of the maximum of the electron signal, while the integral of the tail, $Q^{\mathrm{tail}}_{\mathrm{tot}}$, corresponds to approximately 9% of the total electron signal, $Q^{\mathrm{signal}}_{\mathrm{tot}}$. The shape and duration of the tail depend on the distance between the GEM4B and the pad plane, the distance of the signal from the center of gravity (COG) of the cluster along the pad direction, the track inclination, and the diffusion. Consequently, the above ratios are strongly influenced by these parameters.
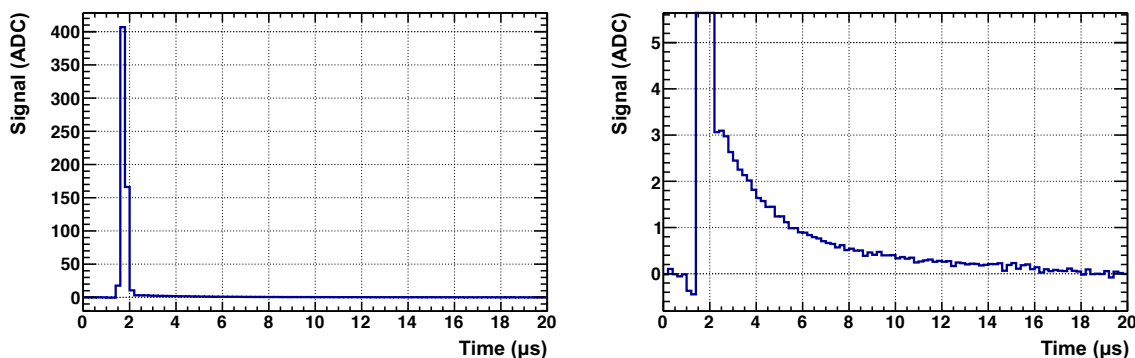
Figure 9. Left: Response of a pad to a laser signal. Right: The same signal zoomed in on the (y) axis. The undershoot observed before the signal pulse is the common-mode response due to signals in other pads of the same stack.

A long ion tail as a negative undershoot was also observed during Run 1 and Run 2, where the TPC was based on the MWPC technology. The integral of the ion tail accounted for approximately 50% of the total signal, resulting in significant degradation of the detector performance, especially

in the case of out-of-bunch pile-up events [8]. Despite its positive nature and smaller amplitude, the ion tail still requires a correction during the Run 3 data-taking period, where, on average, five pile-up events are expected within a full drift time.

Simulations were performed to understand the origin of the ion tail observed in GEMs. For these simulations, only the last GEM foil (GEM4) was modeled, which is sufficient to a first approximation due to the shielding of the ions generated in the previous amplification stages by the GEM4 electrodes. The electric field maps were calculated using the finite element method as implemented in ANSYS [17]. The transport properties of the charge carriers were determined using Magboltz [18], and their multiplication was simulated in Garfield++ [19]. The simulations show that two categories of ions contribute to the ion tail, which can be classified by their point of origin. They are created either in the GEM4 holes or in the induction gap, i.e., the region between GEM4B and the pad plane. While the first category can not be avoided and will be present in any GEM system, the latter is particular for the HV settings chosen for the ALICE TPC GEMs, where a high induction field (electric field between the GEM4B electrode and the pad plane, E_{ind}) plays an important role in minimising the ion backflow. A detailed description of the HV settings can be found in [7].
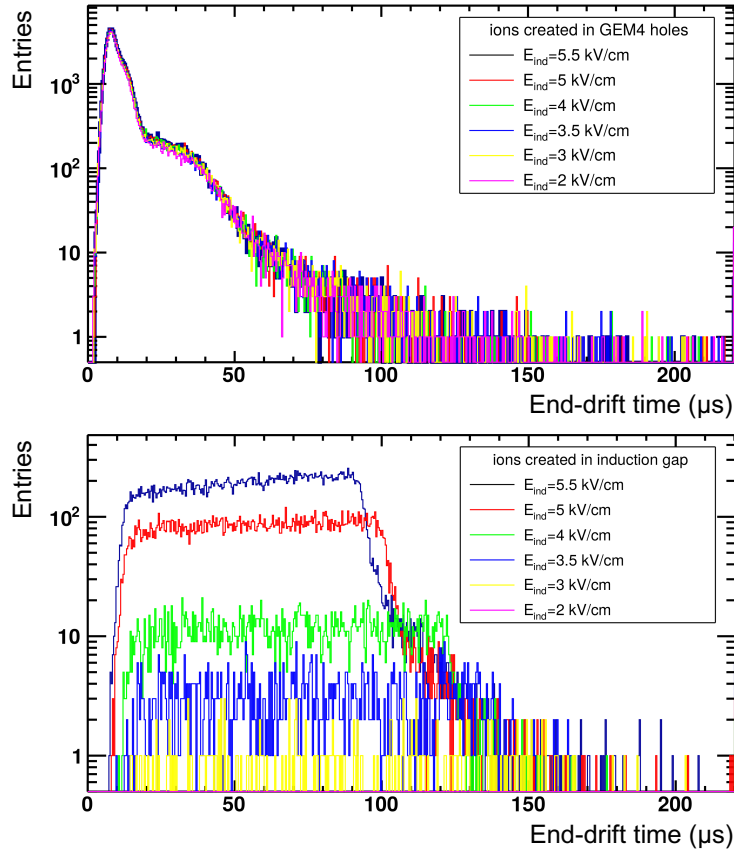
Figure 10. Simulated drift times of ions until they reach an electrode for two categories of ions: those generated in the GEM4 holes (top) and in the induction gap (bottom). Different induction field values are shown as different colors, with the nominal value corresponding to 3.5 kV/cm.

The simulation results are summarized in figure 10, where the time between the creation of each ion until it is collected at an electrode or leaving the amplification area (the *end-drift time*) is plotted separately for the two types of ions mentioned above. It can be seen that the number of ions generated in the amplification stage (top panel) is considerably larger than those generated in the induction gap (bottom panel). The former are referred to as the *fast component* of the ion tail due to their sharp distribution in small end-drift time values, while the latter are referred to as *slow component* due to their flat distribution. The distribution of the fast component does not depend on E_{ind}, however, in the case of the slow component, the distribution becomes narrower with increasing E_{ind} and acquires a slight slope. Since the probability of ionization depends on the induction field, the number of produced ions rapidly decreases with decreasing E_{ind}. As seen in the bottom panel of figure 10, for the low E_{ind} settings, the end-drift time distribution of the slow component ions is flat, indicating a uniform production of electron/ion pairs in the induction gap. On the other hand, for higher E_{ind} and in particular for $E_{\text{ind}} > 4$ kV/cm (see figure 2 of ref. [20]), the electron/ion pair creation probability is larger closer to the pad plane, due to avalanche effects. Note that the nominal induction field value, $E_{\text{ind}}^{100\%}$, is set to 3.5 kV/cm.

4.2 Induction field dependence

A dedicated set of measurements was added to the pre-commissioning program to study the tail properties and disentangle the contributions of the two different ion types. In these measurements, 5000 laser events were collected for two TPC sectors. The value of the induction field was set to 50% of the nominal value and then increased in 5% steps, from 75% to 100%. The high number of laser events compared to the standard laser calibration runs (with 1000 events) was chosen to ensure a good signal-to-noise ratio, since the tail magnitude is relatively small.

Figure 11 shows the average normalized laser signals of all central pads. The normalization is with respect to the maximum value of a given pad signal. It demonstrates that the magnitude of the tail decreases with decreasing induction value and the shape also varies depending on the induction field value. Apart from the magnitude, the shape of the tail is also different. In particular, the tail decays faster for lower values of the induction field. Note that, due to the steeply rising nature of the tail, as illustrated in figure 9, and the fluctuations in the laser signal positions, the bins near the peak exhibit greater fluctuations, leading to increased statistical errors.

4.3 Estimation of the contribution from the two categories of ions

Based on the simulation results, an estimate of the contributions from the two categories of ions described in section 4.1 can be made. It can be assumed that for $E_{\text{ind}}^{50\%} = 1.75$ kV/cm the contribution of the slow component is negligible with respect to higher field settings as see in figure 10. Since the fast component is practically independent of the induction field, the slow component contribution for a given E_{ind} value can be estimated as the difference between the value with this induction field and the one with $E_{\text{ind}}^{50\%}$, i.e., one can write

$$Q_{\text{fast}}(E_{\text{ind}}) \approx Q_{\text{meas}}(E_{\text{ind}}^{50\%}) \tag{4.1}$$

$$Q_{\text{slow}}(E_{\text{ind}}) \approx Q_{\text{meas}}(E_{\text{ind}}) - Q_{\text{meas}}(E_{\text{ind}}^{50\%}). \tag{4.2}$$
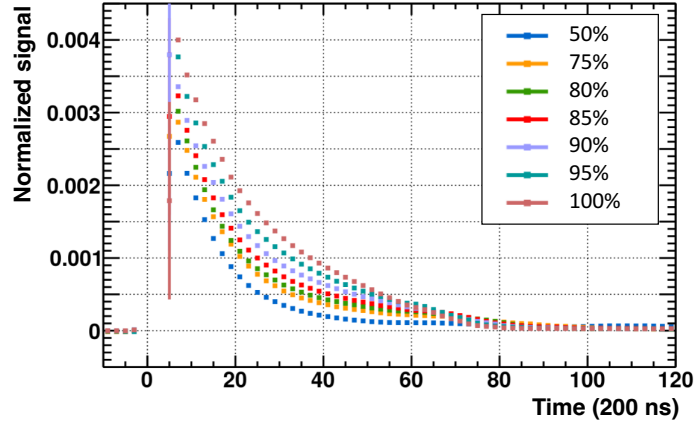
Figure 11. Normalized signals as a function of time for the different induction field settings, averaged over all central pads (within 0.2 cm from cluster COG). Numbers listed in the legend correspond to percentages of $E_{\rm ind}^{100\%}$. The error bars (standard error of the mean) are smaller than the marker size.

The results of the above procedure are shown for all central pads of OROC3 in figure 12 and figure 13. In figure 12, the average normalized ion tail is shown for the $E_{\rm ind}^{50\%}$ setting, which is dominated by the fast component. An exponential shape is observed. In figure 13, the normalized ion tail (left) and its difference from the $E_{\rm ind}^{50\%}$ setting (right) are shown for different values of the induction field. The difference approximates the slow component. A nearly linear shape of the slow component is observed in the right column of figure 13, consistent with ions uniformly produced in the induction gap.
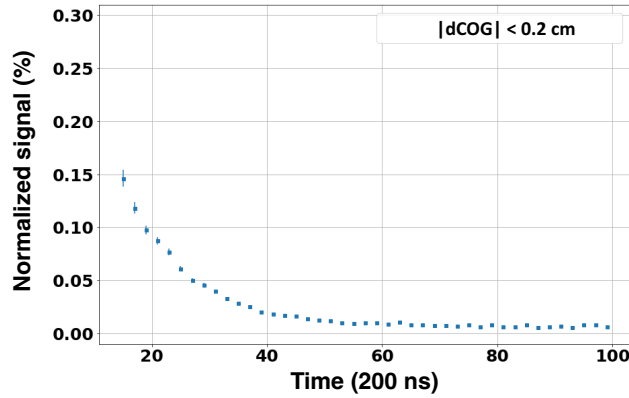
Figure 12. Average normalized ion tail for $E_{\rm ind}^{50\%}$, which is assumed to describe the fast component independent of the induction field setting. Only the data from the central pads of OROC3 are shown. The error bars correspond to the RMS of entries in each time bin.

4.4 Ion tail parametrization

The ion tail can be corrected online using an exponential filter algorithm (see section 5). However, the additional linear component must be taken into account to avoid any bias. For this, two external parameters, the *ion-tail slope* and the *ion-tail fraction* ($Q_{\rm tot}^{\rm tail}/Q_{\rm tot}^{\rm signal}$), are used as input during online processing. To optimize these parameters, each pad signal was fitted with the convolution of a
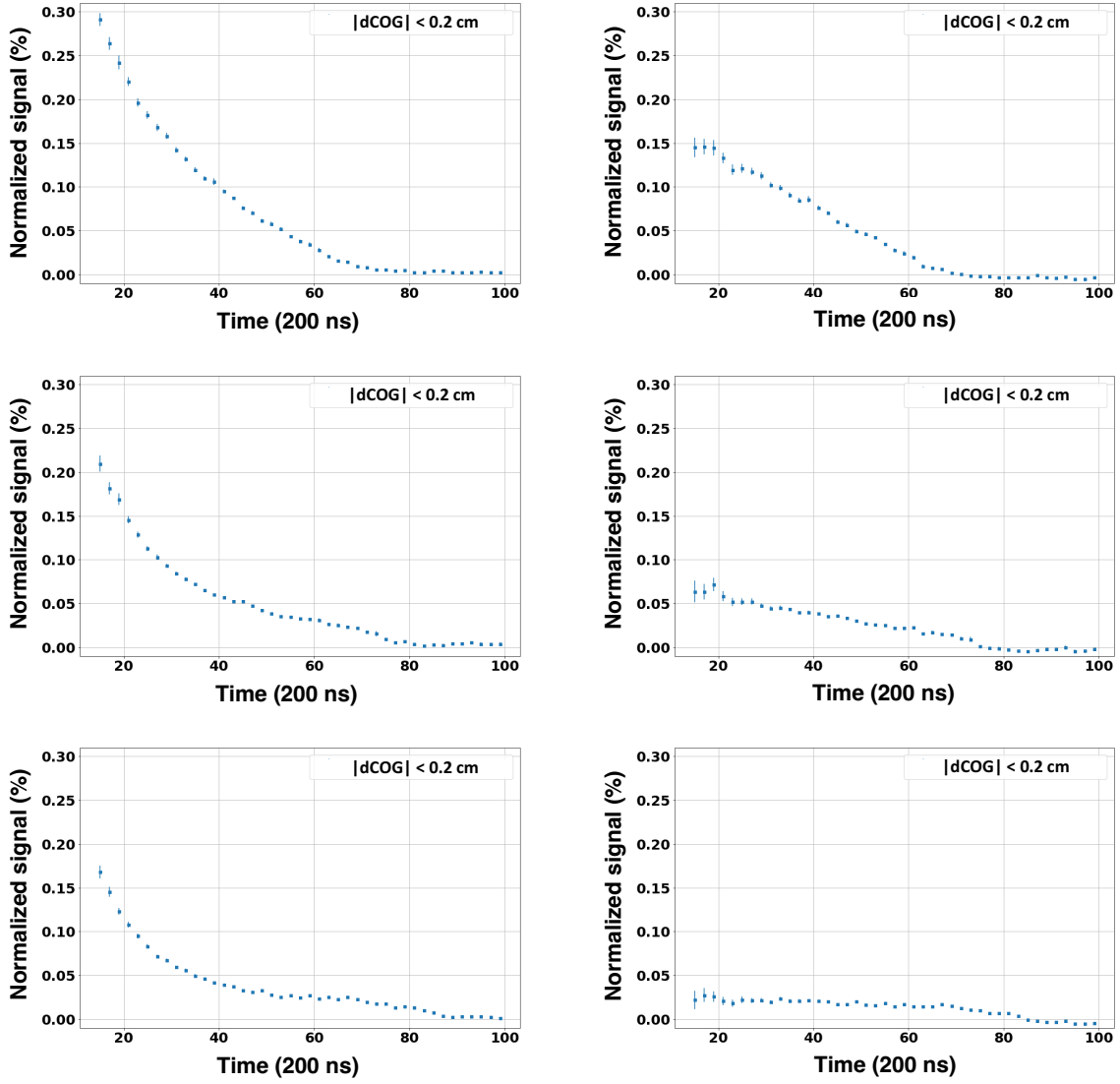
Figure 13. Average normalized ion tail (left) and difference from the $E_{\mathrm{ind}}^{50\%}$ (right), for $E_{\mathrm{ind}}^{100\%}$ (top), $E_{\mathrm{ind}}^{85\%}$ (middle), $E_{\mathrm{ind}}^{75\%}$ (bottom). Only the data from the central pads of OROC3 are shown. The error bars correspond to the RMS of the entries in each time bin. The slow component of the ion tail (right) is nearly linear for all induction field settings, with its contribution decreasing as E_{ind} decreases.

Gaussian and an exponential function to simultaneously describe the signal and tail as shown in the left and right panel of figure 9, respectively. In figure 14, the ion-tail fraction averaged over all pads is shown as a function of the induction-field setting. The points are fitted with a function of the following form:

$$f_{\mathrm{IT}}(E_{\mathrm{ind}}) = A + B \cdot e^{C \cdot E_{\mathrm{ind}}} \, . \tag{4.3}$$

The first term corresponds to the contribution from the GEM4 holes that is independent of the value of the induction field, while the second term corresponds to the contribution from the induction gap. By reducing the induction field value from 100% to 95%, the average ion-tail fraction is

reduced by 10%. This indicates that the ion-tail fraction near the nominal setting is very sensitive to small fluctuations in the effective induction field. From the fit, one can also obtain $f_{\text{IT}}(0) \approx 0.045$ and $f_{\text{IT}}(50) \approx 0.05$, indicating that for $E_{\text{ind}} < 50\%$ there is a residual contribution of about 10% from the slow component. This negligible residual is consistent with the assumption used in eq. (4.2). Moreover, the data shown here are consistent with the observation of figure 13, where for the nominal induction field the contributions of the fast and slow components are almost equal, $f_{\text{IT}}(50)/f_{\text{IT}}(100) \approx 0.5$.
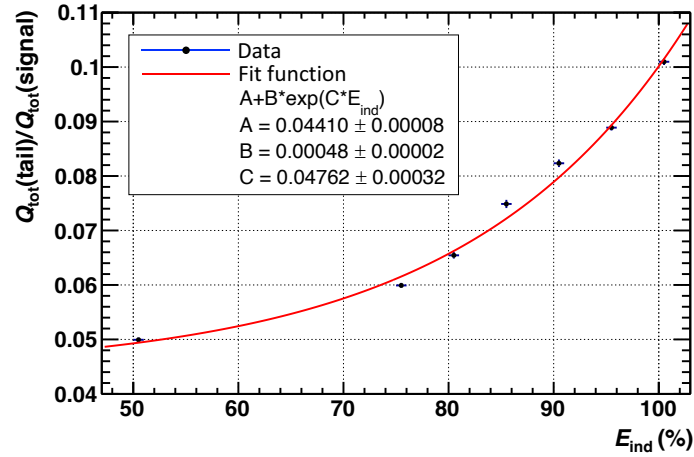
Figure 14. Dependence of the (average) ion-tail fraction on the induction field value. The data points are fit with the function defined in eq. (4.3).
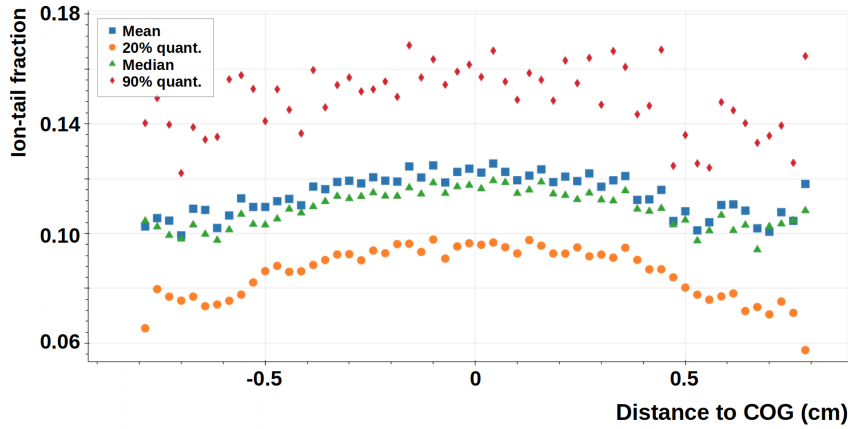
Figure 15 shows the dependence of the ion-tail fraction and the ion-tail slope on the normalized pulser charge. The pads located in edge regions or close to the spacer cross of each chamber were considered separately in the analysis, since the dielectric material placed in these regions influences the signal shapes. In the upper panels of the figure, these pads (*cross/edge*) are shown in red, while the rest of the pads (*bulk*) are shown in blue. A linear correlation is observed between the ion-tail fraction and the normalized pulser charge for the bulk pads. This is explained by the fact that the pulser charge and the value of the effective induction field are inversely proportional to the distance between the pad and GEM4B, resulting in a wide range of the ion-tail fraction, from about 5 to 20%. The proportionality does not apply to the pads in the cross/edge region due to the additional dielectric material. In the bottom panels, the bulk region is shown for pads with a distance to the COG of the cluster (dCOG) of less than 0.8 cm and for the four different GEM stacks.

Figure 16 shows the dependence of the ion-tail fraction on the distance from the COG of the cluster for the bulk pads. For the mean, about a 20% difference is observed between the center of the cluster and the center of the neighboring pad. The large spread of the two distributions shown in figure 15, mainly due to the pad-to-GEM4B distance, does not allow an accurate pad-by-pad calibration of the ion-tail parameters. This is because the foil sagging not only affects the capacitive coupling, but also increases the effective induction field, which leads to a stronger amplification and thus to a change in the ion-tail shape. Moreover, only a small fraction of about 10% of the TPC pads "see" the laser signals. Since accurate knowledge of the parameters is important for the restoration of the baseline bias and its fluctuations, krypton calibration data were used to disentangle the ion-tail
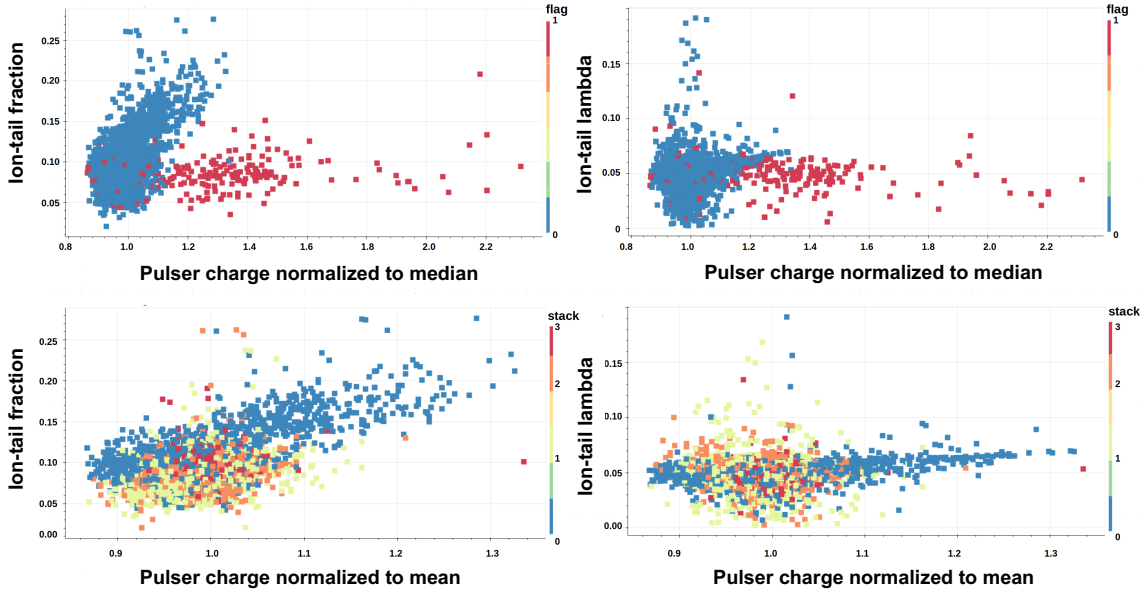
Figure 15. Ion-tail fraction (left) and ion-tail slope (right) as a function of $Q_{\text{pulser,pad}}^{\text{norm}}$. In the top panels, the color indicates the pad position in the chamber, with blue representing the bulk and red the regions influenced by the presence of the spacer cross or edges. In the bottom panels, the bulk region for pads with $|\text{dCOG}| \leq$ 0.8 cm are shown, where the color scale indicates the stack number.

dependencies (see section 5.1).

Figure 16. Dependence of the (average) ion-tail fraction on the distance of the pad to the COG of the cluster for the bulk pads. Only pads with $|\text{dCOG}| \leq 0.8$ cm are shown. The different quantiles are shown with different colors.

5 Common-mode and ion-tail corrections

Correcting the common-mode effect and ion tail online, before applying the ZS, is critical for maintaining the PID and tracking performance of the TPC and for limiting cluster losses. Moreover, the correction of the ion tail also helps with minimizing the data volume produced by the TPC. The

multiplexed data streamed from the FEE are decoded in the CRUs, where subsequently the pedestal subtraction, common-mode correction, ion-tail correction, and ZS are performed. Due to the large number of pads, the data from each TPC stack are read out by either two (for OROC) or four (for IROC) CRUs.

With the current configuration, information cannot be exchanged between different CRUs. This additional CRU segmentation implies that calculation of the common-mode charge for a pad using eq. (3.1) cannot be applied. The calculation of the average positive signal in the stack $\langle Q_{\mathrm{pos}}(t) \rangle_{\mathrm{stack}}$ would require combining information from different CRUs. Instead, a baseline estimation is performed using the *empty* (or *non-signal*) pads in the CRU, by using eq. (3.1) and eq. (3.3) [9]:

$$Q_{\mathrm{pad}}^{\mathrm{CM}}(t) = Q_{\mathrm{pulser,pad}}^{\mathrm{norm}} \left\langle \frac{Q_{\mathrm{pad}}(t)}{Q_{\mathrm{pulser,pad}}^{\mathrm{norm}}} \right\rangle, \tag{5.1}$$

where $\langle \rangle$ denotes averaging over empty pads in a given CRU. The algorithm consists of two parts: First, the empty pads are selected and the mean baseline is calculated for a given time bin. Second, the common-mode correction is applied to all pads, scaling accordingly using the normalized pulser charge. Since each pad has a different pulser charge, a static map must be provided to the CRUs. An advantage of this method is that the stack-dependent parameters (see figure 8) are not needed, since the correction is calculated at the CRU level. The efficiency of the algorithm depends on the correct selection of the empty pads on a time-bin basis. For this, apart from a simple threshold cut ($Q_{\mathrm{pad}}(t) \leq n \cdot \langle \mathrm{noise} \rangle$), an additional check is performed by comparing the pad charge to that of a number of randomly selected pads in the CRU. This ensures that pads measuring the ion tail of earlier signals (superimposed with common-mode) are excluded. A pseudo-code for the common-mode correction is given in appendix A.

Online correction of the ion tail is performed on a pad-by-pad basis prior to ZS in the CRUs. Since the polarity of the ion tail is positive, ZS does not result in missing charge and thus missing clusters. Consequently, the ion tail can be corrected online to first order and the remaining second-order deficiencies can be handled during track reconstruction, if necessary. An exponential correction of the form

$$Q_{\mathrm{out}}(t) = Q_{\mathrm{in}}(t) - \sum_i A_i \cdot e^{-\lambda_i (t - t_{i,\mathrm{max}})} \tag{5.2}$$

is quite difficult to implement directly in the CRU FPGAs. In the equation, $Q_{\mathrm{in}}(t)$ and $Q_{\mathrm{out}}(t)$ are the pad signals for the time bin t before and after the correction is applied, respectively. The sum runs over all previous signal peaks i for the given pad. The parameters A_i and λ_i are the maximum and slope of the tail corresponding to the signal peak i, respectively, while $t_{i,\mathrm{max}}$ is the position of the peak maximum. The complexity of such a correction stems from two factors: the number of resources required and the time needed to perform the calculations. First, applying the above correction would mean that the entire peak history for each of the roughly 1600 pads read out by a CRU would need to be stored. However, FPGAs are not typically designed to store large amounts of data. Second, while it is possible for the FPGAs to compute an exponential (e.g., using the CORDIC[2] functions), FPGAs would need to perform multiple exponential calculations in parallel.

---

[2]The *CORDIC* (Coordinate Rotation Digital Computer) algorithm is an iterative method for computing elementary functions in the electronics using rotations.
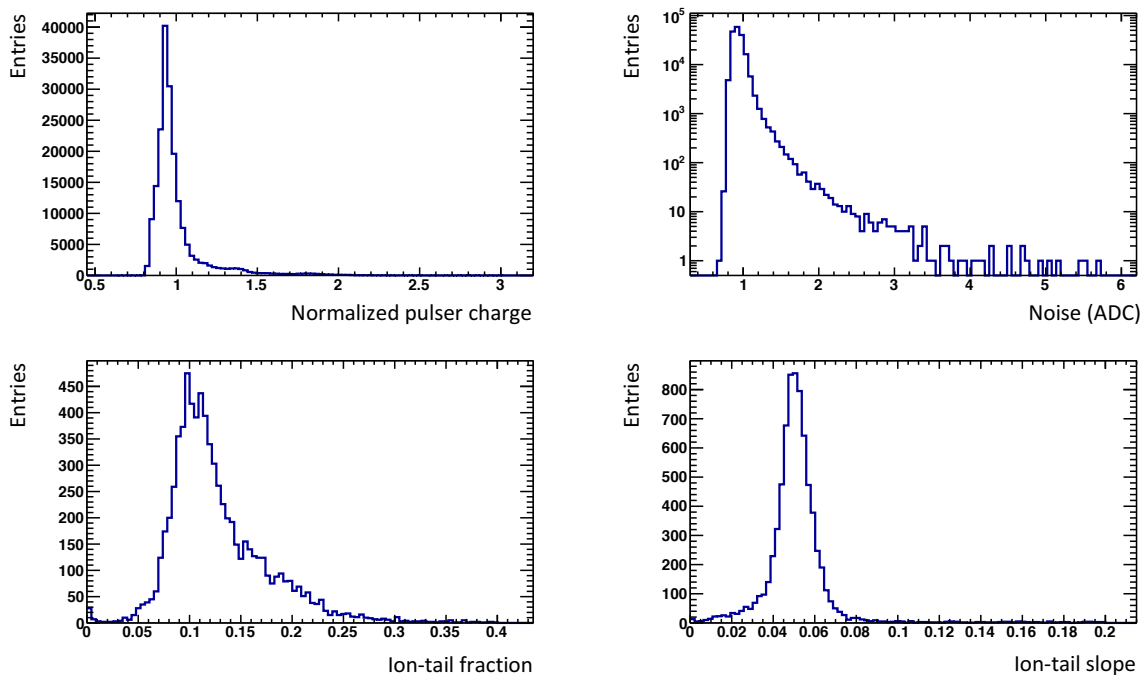
Figure 17. Parameter distributions used as input for the toy MC simulation. Only IROC pads are included. Top left: distribution of normalized pulser charge from a pulser calibration run. Top right: noise distribution from a pedestal/noise run. Bottom left (right): ion-tail fraction (slope) obtained from the laser data. For simplicity, no correlations between parameters were assumed.

Based on the number of digital-signal processors available in the CRU FPGA, these calculations would introduce some latency.

To avoid the aforementioned complications, an exponential filter has been developed for online correction of the ion tail. The filter requires only simple mathematical operations. Note that this correction assumes a perfectly exponential ion tail, which is not entirely realistic (see section 4.3). Moreover, the input parameters are tuned using the laser tracks (see section 4.4), so different track topologies are not taken into account. Namely, residual biases, expected to be on the per-mill level, are inevitable. However, they may still have an impact on the tracking efficiency and PID performance. More details on the implementation of the exponential filter are given in appendix B.

5.1 Toy MC simulations

To quantify the effects of common-mode and ion tail on the baseline, and to investigate the performance of the two online correction algorithms, toy MC simulations were performed. The stepwise procedure is as follows:

- 250 Pb-Pb events were simulated, with particle composition derived from particle spectra measured with the ALICE detector at 5.02 TeV (Run 2 conditions). An area of 1000 pads (data from between 1200 and 1600 pads are transferred to a single CRU) and 1000 time bins were considered for the simulation of the common-mode effect.
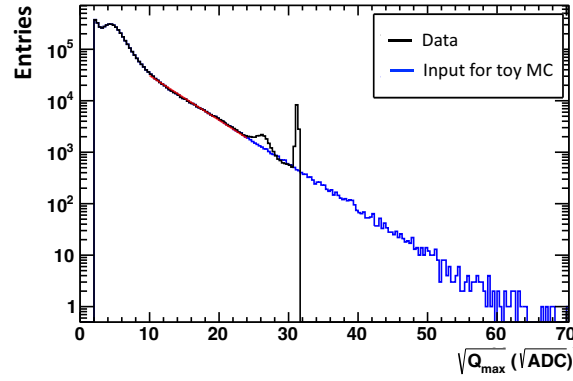
Figure 18. Distribution of Q_{max} as obtained from a data sample of Pb-Pb collisions (black). The two peaks around 26 and 31 are due to saturated signals. The distribution was fitted with an exponential function (red) in the range (10, 24). The input distribution for the MC (blue) is constructed using the data distribution for $\sqrt{Q_{max}} < 20$, and the fit function for $\sqrt{Q_{max}} > 20$.

- Random values for the normalized pulser charge, noise, ion-tail fraction and slope were assigned to each pad according to the distributions obtained from the calibration data (see figure 17). For the pedestal, a value in the range (0,1) was assigned, since the integer part of the pedestal can be subtracted trivially.

- A number of clusters were generated, with the number of clusters per event randomly distributed such that the variations in detector occupancy resembled the conditions of Run 3.

- For each cluster, a value for Q_{max} (maximum value among all digits in a cluster) was generated using the distribution obtained from the Pb-Pb collision data (see figure 18). Each cluster spreads over three pads and three time bins, following a Gaussian distribution in the pad- and time-space.

- The ion tail was simulated as a perfect exponential, and for each pad, the simulated parameters for the ion-tail slope and fraction were used.

- The common-mode was simulated using eq. (3.3). A value of 0.5 was used for the slope, namely $k_{CF,pad} = 0.5 \cdot Q^{norm}_{pulser,pad}$.

- Noise, pedestal and rounding were added.

- The pedestals were subtracted.

- The common-mode effect was corrected using the algorithm described in appendix A.

- The ion tail was corrected using the algorithm described in appendix B. The order in which the ion-tail and common-mode corrections are applied is important because the ion tail also produces small common-mode undershoots in the other pads.

Figure 19 shows a simulated pad signal for an event with multiplicity corresponding to approximately 30% occupancy in TPC. Both the common-mode effect and the ion tail are included.
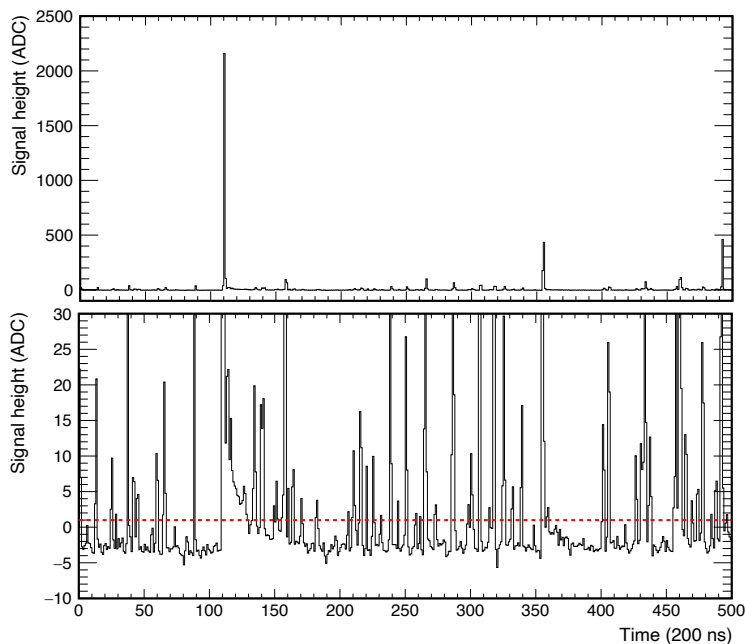
Figure 19. Top: pad signal for an event with about 30% occupancy, without noise. Bottom: pad signal zoomed in on the *y*-axis. The red dashed line shows the ZS threshold.

The baseline is systematically shifted to negative values due to the common-mode effect. Next, a series of simulations were performed to optimize the correction parameters, including simulated noise. For each event, 500 random settings were simulated, allowing for the following options:

- Common-mode simulation: ON or OFF.

- Common-mode simulation ON and correction of the common-mode effect either OFF, or using a mean correction, median correction, mean 2nd iteration, or median 2nd iteration.

- Random values (in reasonable ranges) for the parameters `nPadsRandom`, `nPadsMin`, `Q_thr1`, `Q_thr2`.

- Ion-tail simulation: ON or OFF.

- Ion-tail simulation ON and ion-tail filter either OFF, or correction using the pad-by-pad ion-tail parameters (*pad-by-pad*), or correction using the median value of the parameter distributions shown in the bottom of figure 18 (*fixed-to-median*).

- Random value for `k0` in the range 70–100%.

Figure 20 shows the results of the parameter scan. On the left, the average baseline shift and on the right its RMS are shown as a function of occupancy. A 99% least-trimmed-squares method was used to exclude some extreme outliers. For the top panels, only the common-mode effect was simulated. The different correction methods (no correction, mean, median, and 2nd iterations) are shown with different colors. It can be seen that the average baseline bias can reach up to −5 ADC and its RMS up to 1.5 ADC if the common-mode effect is not corrected. Compared to the RMS in the absence of

the two effects (caused by the noise), this corresponds to roughly a 60% increase. For the bottom panels, only the ion tail was simulated. The average baseline bias reaches up to 1 ADC and its RMS 1.25 ADC, which corresponds to an increase of about 35% when the tail is not corrected.
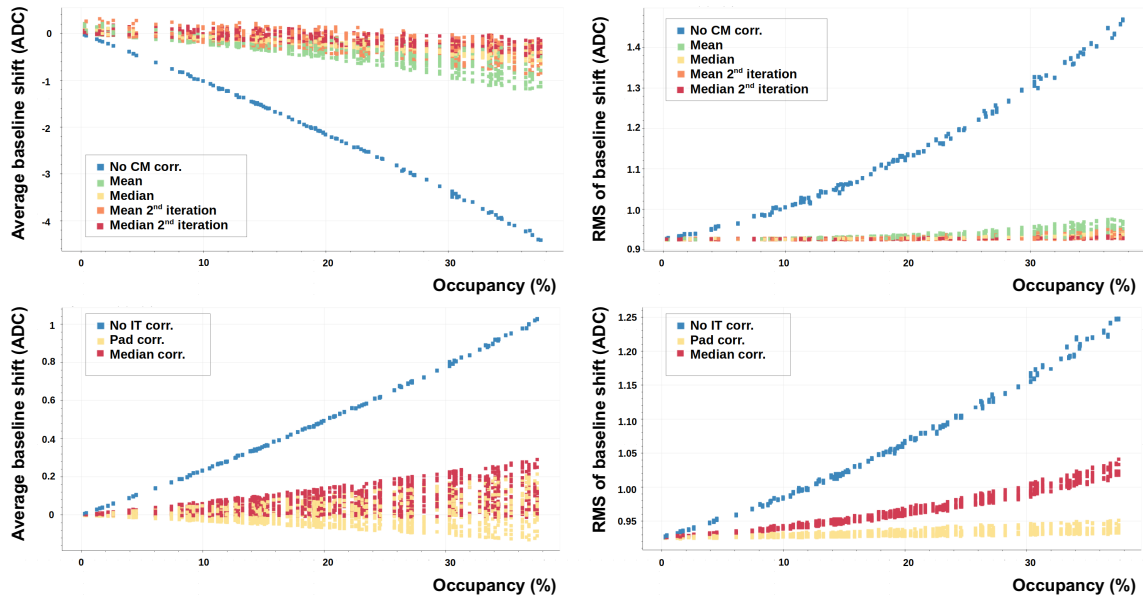
In figure 21, the average baseline shift is shown as a function of occupancy for the pad-by-pad (left) and the fixed-to-median (right) corrections of the ion tail. As mentioned in appendix B, a value of k0=100% overcorrects the data, which can be confirmed by the negative values of the baseline bias (red markers). It can be seen that the optimal value for the pad-by-pad method is k0 ≈ 85-90%, while for the fixed-to-median method it is k0 ≈ 90-95%. The baseline shift is slightly larger for the fixed-to-median method.

Figure 22 shows the average baseline shift for the different methods used in the common-mode (left) and ion-tail (right) corrections, where the occupancy is 28–32%, corresponding to the highest expected multiplicities in Run 3. All parameter settings were averaged. It can be seen that the common-mode correction can restore most of the baseline shift. Using the median 2nd iteration method instead of the mean results in roughly a 60% smaller average baseline bias. Similarly, correction of the ion tails restores most of the baseline shift. Using the pad-by-pad method instead of the fixed-to-median method results in an approximately 70% smaller average baseline shift.

The impact of the ion tail on the data volume produced by the TPC is shown in figure 23 as a function of occupancy. The *space saving* is defined as:

$$\text{space saving}(\%) = 100\% - \frac{\text{compressed data size}}{\text{uncompressed data size}} \quad , \tag{5.3}$$

where the uncompressed and compressed data size is the size before and after ZS, respectively. It can be seen that the ion tail significantly reduces the space saving compared to the original signal. As

Figure 20. Average baseline shift (left) and its RMS (right) as a function of occupancy. The simulation and correction of the "common-mode only" ("ion tail only") scenario is shown in the top (bottom) panel. The various marker colors corresponds to the different correction methods.
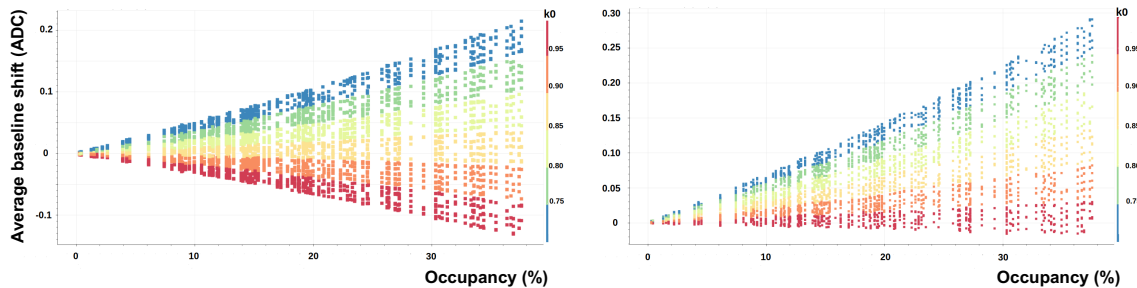
Figure 21. Average baseline shift as a function of occupancy for the pad-by-pad (left) and the fixed-to-median (right) methods of the ion-tail correction. The common-mode effect was not simulated. The color scale indicates the different correction fractions.
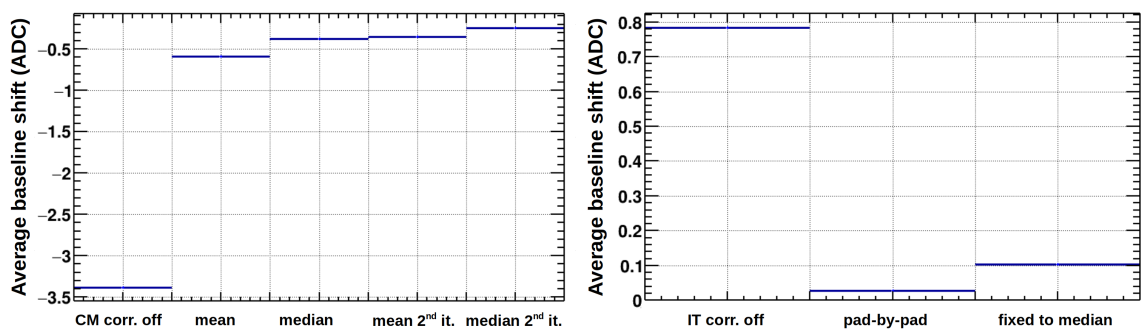
Figure 22. Average baseline shift for the different methods used in the common-mode (left) and ion-tail (right) corrections, where the occupancy is 28–32%. All parameter settings were averaged.

shown by the blue and magenta lines, simulating and correcting for the common-mode effect leads to a larger space saving than the original signal due to a remaining negative bias. The reduction in space saving due to the ion tail can be restored by applying its correction.
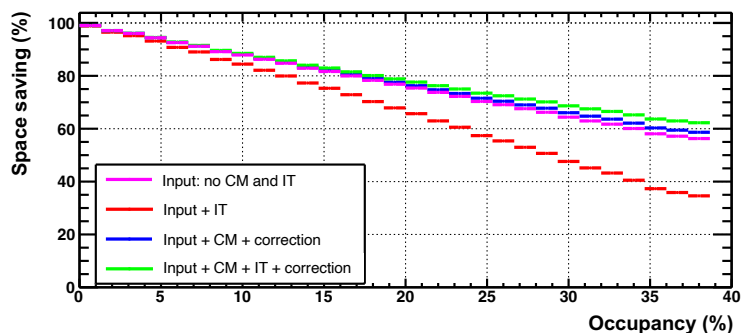
Figure 23. Impact of the ion tail on the space saving. Noise was not included, and a threshold cut of 1.2 ADC was applied.

Table 3 summarizes the importance of the ion-tail correction for space saving for minimum bias Pb-Pb events (occupancy $\leq$ 30%). Due to the ZS, an uncompressed data rate of 3.5 TB/s would be reduced to 650 GB/s in the absence of the ion tail and the common-mode. This number is consistent with the expected zero-suppressed rate of about 600 GB/s [21]. When adding the ion tail, the rate

increases by about 30%, to 888 GB/s. After correction for both effects, a given choice of parameters results in an inaccuracy of about 3 to 5% for each effect.

Table 3. Impact of the ion-tail correction on the space saving for minimum-bias events with occupancy ≤30%. Noise was not included, and a threshold cut of 1.2 ADC was applied. Note the realistic order of implementation of the effects and the corresponding corrections.

	Minimum bias space saving (%)	GB/s
input signal	81.4	650
input + ion tail	74.6	888
input + common-mode + correction	82.1	628
input + ion tail + common-mode + correction	83.1	594
current assumption	82.9	600

Although this toy MC simulation study is based on a data-driven approach, detector performance studies using the full MC simulations should be carried out in order to account for different track topologies and cluster shapes, diffusion, and other detector effects that have not been considered.

6 Conclusions

The signal response of the GEM-based ALICE TPC was studied in detail using the data collected with the laser calibration system. The dependencies of the common-mode effect were understood using machine learning techniques. It was found that the common-mode signal depends largely (96%) on the stack type and the normalized charge detected in runs with the dedicated calibration pulser system. The stack type accounts for the absolute capacitance of the stack, while the normalized pulser charge is responsible for the pad-by-pad capacitance variations. An unpredicted ion tail was observed in data recorded with the TPC laser system. The measurements have shown that ions from two categories contribute almost equally to the generated signal at the nominal induction field setting. The contribution of the ions generated in the GEM4 holes is practically independent of the value of the induction field and results in an exponentially shaped tail. Ions generated in the induction gap of the GEM stack result in an additional contribution, which depends on the electric field applied to the induction gap.

The performance of the two online correction algorithms was studied in detail using a toy MC with input parameters determined in a data-driven way. The common-mode correction algorithm correlates all pads of a given CRU for each time bin, while the ion tail is corrected on a per-channel basis using an exponential filter. Both effects are efficiently corrected, despite a residual bias in the baseline, comparable to the noise. Since the ion-tail parameters used in the exponential filter are obtained from the laser tracks, different track topologies were not considered. Therefore, possible imperfections in the ion-tail correction are to be expected. These should be considered when repeating these studies using a full MC simulation. Furthermore, the ion tail has a significant impact on online data compression. During Run 3 and Run 4 data taking periods, the raw data readout rates are estimated to be about 3.5 TB/s. The toy MC results obtained in this study show that in the presence of the ion tail, the final data rate after baseline subtraction is about 890 GB/s instead

of 650 GB/s, which is the estimated value given in the technical design report [6]. The proposed ion-tail correction algorithm fully covers this increased data rate. The two correction algorithms will be commissioned using pp and Pb-Pb collisions at record energy and luminosities as part of the TPC readout system from 2023 onward.

Acknowledgments

A Common-mode correction algorithm

The following pseudo-code for the common-mode correction is tested in section 5.1 using a toy MC simulation and implemented in the CRU firmware:

```
1  // Some constants to be set before
2  int nPadsCRU; // Number of pads in the current CRU
3  int nPadsRandom; // Number of pads (around 10) at random distance to the current to
       pad. To be compared for additional check if the current pad is an empty pad
4  int nPadsMin; // Minimum number of pads (around nPadsRandom/2) required to have Q
       very close to current pad
5  float Q_thr1; // Optimized threshold 1 (comparable to 2*noise)
6  float Q_thr2; // Optimized threshold 2 (comparable to noise)
7  //
8  // Main CM correction algorithm
9  for each time bin{
10     // Calculate mean baseline of empty pads
11     vector<float> Q_pad_array; // Array to hold Q of empty pads of a CRU
12     for each pad{
13         float k_pulser = GetNormQPulser(padID); // from 2D pad-by-pad map
14         // Simple check if the current pad is empty
15         if (Q_pad <= Q_thr1){
16             // Additional check if the current pad is empty pad
17             int nPadsOK = 0;
18             for (int i = 0; i < nPadsRandom; i++){
```

```
19            int padRandomID = GetRandomPad(); // Randomly select a pad
20            float k_pulser_rndm = GetNormQPulser(padRandomID);
21            if (abs(Q_pad/k_pulser - Q_padRndm/k_pulser_rndm) < Q_thr2)
22            nPadsOK++;
23        }
24        // If empty, add charge of current pad into array, scale accordingly
25        if (nPadsOK >= nPadsMin) Q_pad_array.push_back(Q_pad/k_pulser);
26    }
27    }
28    // Calculate mean baseline of the non-signal pads
29    Q_baseline = mean(Q_pad_array); // Alternatively use median
30    // Apply common-mode correction
31    for each pad{
32        float k_pulser = GetNormQPulser(padID); // From 2D pad-by-pad map
33        Q_pad = Q_pad - Q_baseline*k_pulser;
34    }
35 }
```

Note that `Q_pad` and `Q_padRndm` are the charge after the pedestal subtraction. The above algorithm can be made more robust by using a median estimator instead of the mean for the baseline estimation. Furthermore, since the mean method can incorrectly classify pads with low-signal amplitudes as empty, biasing the baseline estimate, the selection of empty pads can be improved by performing a second iteration after the baseline calculation and before the common mode correction (line 29). In this case, the simple check in line 15 becomes (`Q_pad - Q_baseline*k_pulser <= Q_thr1`), which allows a better classification of the empty pads. Figure 24 shows the fraction of pads identified as empty pads for the mean and the mean 2nd iteration common-mode correction methods as a function of occupancy.
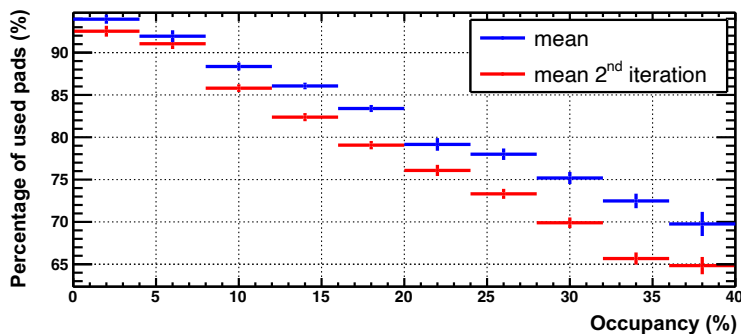
Figure 24. Fraction of pads used for the common-mode baseline estimation as a function of occupancy, for the mean (blue) and the mean 2nd iteration (red) methods. Only the common-mode effect is simulated.

B Ion-tail correction algorithm

The following pseudo-code for the ion-tail correction is tested in section 5.1 using a toy MC simulation and implemented in the CRU firmware:

```
1 // Constant to be set before
2 float k0; // Multiplicative correction factor, same for all pads (around 0.9)
3 for each padID{
4     float k1 = GetFractionIT(padID); //k_fraction, from a static map
5     k1*=k0; // Scale IT fraction
```

```
6    float k2 = GetExpSlopeIT(padID); // = exp(-k_slope), from a static map
7    float Q_correction = 0;
8    for each time bin{
9        Q_out = Q_in - k1*(1-k2)*Q_correction;
10       Q_correction+=Q_in;
11       Q_correction*=k2;
12   }
13 }
```

The amount of correction for a given time bin is `k1*(1-k2)*Q_correction`, where `Q_correction` is the buffered charge value of the previous time bin and `k1` (fraction of the ion-tail integral with respect to the total charge of the pad signal) and `k2` (slope of the ion-tail assuming an exponential form) are the ion-tail parameters. Since these are pad-dependent parameters, they are provided as input to the CRUs in the form of two-dimensional, pad-by-pad maps. The pad-by-pad calibration of these parameters was obtained from data recorded with radioactive krypton isotopes in the gas volume [1]. Alternatively, the median value of the distributions shown in figure 15 can be used. The impact of this simplification is demonstrated in section 5.1. An additional scaling parameter, `k0`, whose value is the same for all pads, is introduced to account for the difference between the integral of the input charge with a continuous Gaussian form and the digitized signal, which has a discrete structure. The digitized signal is equal to or greater than the input, so the parameter `k0` is defined as $k0 \leq 1$. Note that the above filter can be applied continuously, even in the absence of signals. In this case, the filter has no effect on the measured charge.

In figure 25, the exponential filter is applied on the averaged laser signal of a pad. The ion-tail fraction and ion-tail slope parameters, as obtained from the fit, were used for the correction. It can be seen that a 100% correction (`k0=100%`) slightly overcorrects the data, which is due to the bias of the sampled charge. Figure 26 (top) shows the normalized data and the 100% correction for randomly selected pads of the nominal induction field setting, while the bottom panel shows the averaged data and the corresponding corrections.
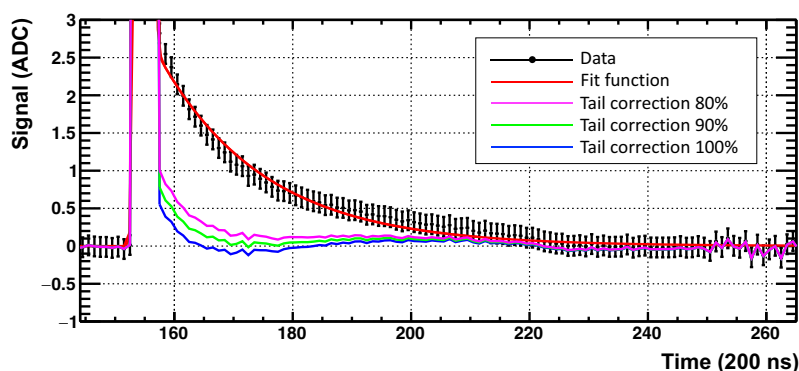
Figure 25. Demonstration of the ion-tail filter applied to laser data averaged over many signals on a given pad. The laser data are shown in black, the fit function in red, and the corrected data in blue, green, and magenta, for different values of the parameter `k0`. The convolution of a Gaussian and an exponential function was used to fit the data. The ion-tail parameters obtained from the fit were used for the correction.
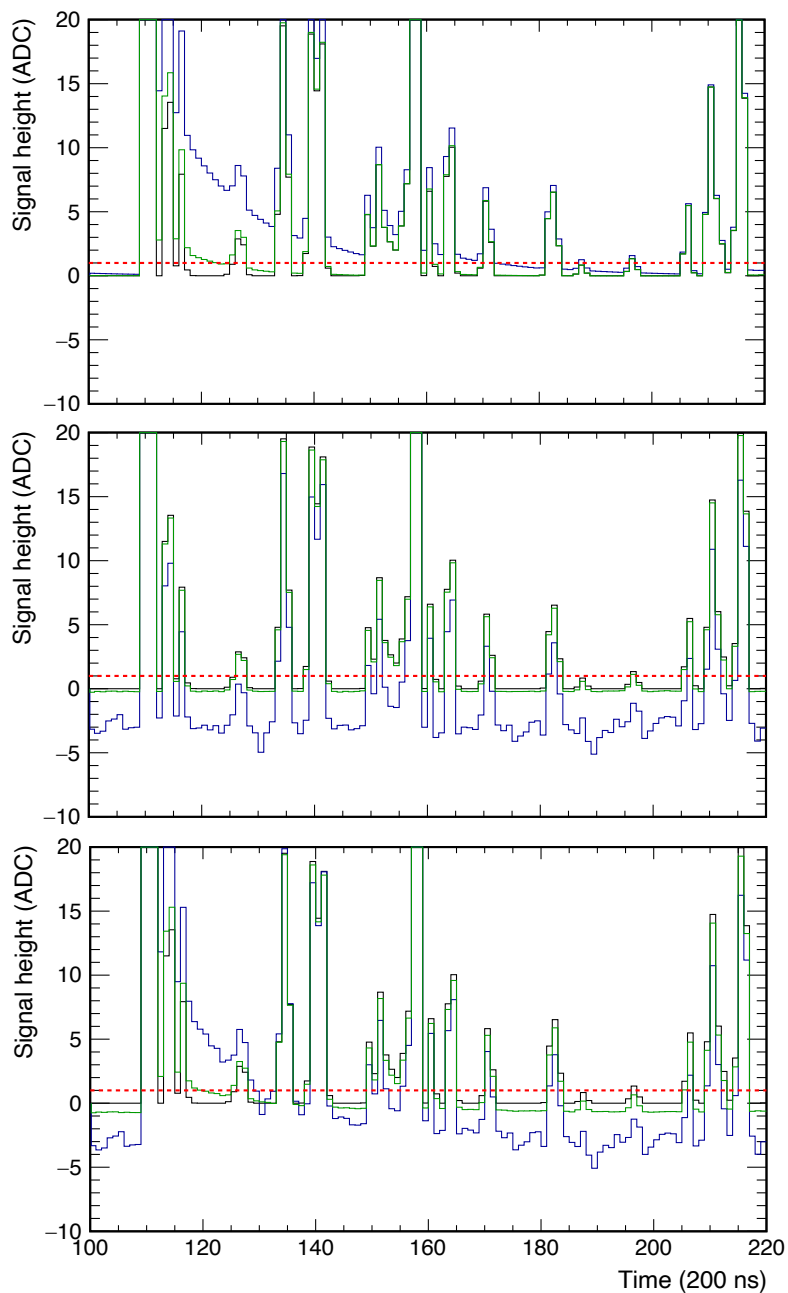
In figure 27 the simulation and correction of both effects are demonstrated. To show the effects more clearly, noise is not included. For the ion-tail correction, the ion-tail parameters of the
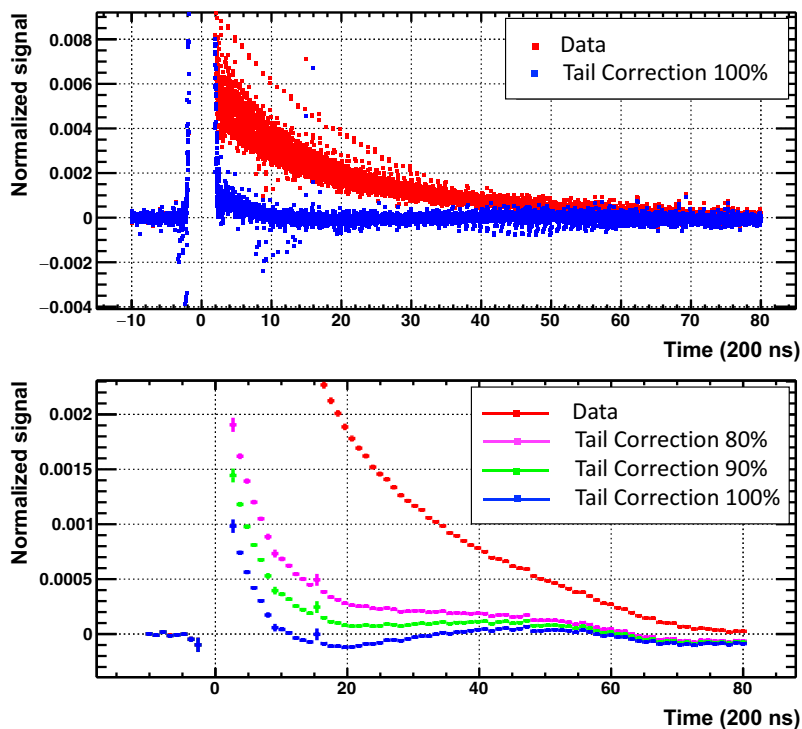
Figure 26. Top: normalized ion tail before (red) and after (blue) correction for randomly selected pads. For each pad, the fitted tail parameters were used for the correction with `k0=100%`. Bottom: averaged data with the corresponding corrections of 80%, 90% and 100%. The error bars correspond to the RMS of entries for each time bin.

respective pad were used, and a value of `k0=80%` was chosen. For the common-mode correction, a two-iteration median correction was applied with `nPadsRandom=6`, `nPadsMin=4`, `Q_thr1=2`, `Q_thr2=2`.

Figure 27. Illustration of ion tail and common-mode correction of a simulated signal. The input signal is shown in black, the simulated effect (top: ion tail, middle: common-mode, bottom: ion tail and common-mode) in blue, and the corrected signal in green. Noise is not included in the pad signal for better visibility.

References

[1] J. Alme et al., *The ALICE TPC, a large 3-dimensional tracking device with fast readout for ultra-high multiplicity events*, *Nucl. Instrum. Meth. A* **622** (2010) 316 [arXiv:1001.1950].

[2] ALICE collaboration, *ALICE time projection chamber: Technical Design Report*, CERN-OPEN-2000-183, CERN, Geneva (2000).

[3] S. Rossegger and W. Riegler, *Signal shapes in a TPC wire chamber*, *Nucl. Instrum. Meth. A* **623** (2010) 927.

[4] ALICE collaboration, *Upgrade of the ALICE Experiment: Letter Of Intent*, *J. Phys. G* **41** (2014) 087001.

[5] F. Sauli, *The gas electron multiplier (GEM): Operating principles and applications*, *Nucl. Instrum. Meth. A* **805** (2016) 2.

[6] ALICE collaboration, *Upgrade of the ALICE Time Projection Chamber*, CERN-LHCC-2013-020 (2013).

[7] ALICE TPC collaboration, *The upgrade of the ALICE TPC with GEMs and continuous readout*, 2021 *JINST* **16** P03022 [arXiv:2012.09518].

[8] M. Arslandok et al., *Track Reconstruction in a High-Density Environment with ALICE*, *Particles* **5** (2022) 84 [arXiv:2203.10325].

[9] ALICE collaboration, *Readout scheme of the upgraded ALICE TPC*, ALICE-PUBLIC-2016-006 (2016).

[10] H. Hernández et al., *A Monolithic 32-Channel Front End and DSP ASIC for Gaseous Detectors*, *IEEE Trans. Instrum. Measur.* **69** (2019) 2686.

[11] ALICE collaboration, *The Laser calibration system of the ALICE time projection chamber*, *Czech. J. Phys.* **55** (2005) 1671 [nucl-ex/0511014].

[12] B.S. Nielsen et al., *Design Note on the ALICE TPC laser calibration system*, Tech. Rep. ALICE-INT-2002-022 (2002).

[13] H.J. Hilke, *Detector Calibration With Lasers — a Review*, *Nucl. Instrum. Meth. A* **252** (1986) 169.

[14] T.K. Ho, *Random decision forests*, in the proceedings of *3rd international conference on document analysis and recognition*, vol. 1, pp. 278–282, IEEE, (1995).

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

[16] P. Buncic, M. Krzewicki and P. Vande Vyvre, *Technical Design Report for the Upgrade of the Online-Offline Computing System*, CERN-LHCC-2015-006 (2015).

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

[18] *Magboltz — transport of electrons in gas mixtures*, https://www.ansys.com.

[19] *Garfield — simulation of gaseous detectors*, https://garfieldpp.web.cern.ch/garfieldpp/.

[20] ALICE TPC collaboration, *Ion backflow studies for the ALICE TPC upgrade with GEMs*, 2014 *JINST* **9** C04025.

[21] ALICE collaboration, *Usage of GPUs in ALICE Online and Offline processing during LHC Run 3*, *EPJ Web Conf.* **251** (2021) 04026 [arXiv:2106.03636].