Department of Physics and Astronomy
University of Heidelberg

Master Thesis in Physics
submitted by

Felix Schlepper

born in Wiesbaden (Germany)

2023

Electron identification with the ALICE TRD in LHC Run 3 and 4

This Master Thesis has been carried out by Felix Schlepper at the
Physikalisches Institut in Heidelberg
under the supervision of
Priv. Doz. Dr. Yvonne Pachmayer

Abstract

This thesis presents the implementation and first performance evaluation of electron identification with the ALICE experiment's Transition Radiation Detector (TRD) at the Large Hadron Collider for Run 3 and 4. The TRD is a crucial component for electron/pion separation, contributes to the track reconstruction in the experiment, and is used for space-charge corrections in the Time Projection Chamber. In LHC Run 3 and 4, the readout of the experiment was changed entirely to take data in Pb−Pb collisions at $\sqrt{s_{\mathrm{NN}}} = 13.6\,\mathrm{TeV}$ with an interaction rate of 50 kHz.

The study provides the first results of the TRD electron identification performance, including developing and implementing the algorithms in the O^2 framework, which is the new data acquisition and processing framework for ALICE in LHC Run 3. The evaluation includes the pion efficiency as a function of electron efficiency and momentum. In the first part of this thesis, two classical multidimensional likelihood methods were investigated to achieve electron/pion separation. In addition, a popular machine learning algorithm, XGBoost, was used to improve the separation over the classical methods. Lastly, details on the implementation in the O^2 framework of the electron identification and extensions to the TRD track parameters on the physics analysis level are given.

Kurzfassung

In dieser Arbeit wird die Implementierung und erste Leistungsbewertung der Elektronidentifikation mit dem Transition Radiation Detector (TRD) des ALICE-Experiments am Large Hadron Collider für Run 3 und 4 vorgestellt. Der TRD ist eine wichtige Komponente für die Elektron/Pion-Trennung, trägt zur Spurrekonstruktion im Experiment bei und wird für Raumladungskorrekturen in der Time Projection Chamber verwendet. Im LHC Run 3 und 4 wurde das Ausleseverfahren des Experiments vollständig geändert, um Daten bei Pb−Pb Kollisionen bei $\sqrt{s_{\mathrm{NN}}} = 13.6\,\mathrm{TeV}$ mit einer Interaktionsrate von 50 kHz aufzunehmen.

Die Studie liefert die ersten Ergebnisse zur Leistung der TRD-Elektronidentifikation, einschliesslich der Entwicklung und Implementierung von Algorithmen im O^2-Framework, welches das neue Datenerfassungs- und Verarbeitungs-Framework für ALICE im LHC Run 3 ist. Die Analyse umfasst die Pion-Effizienz als Funktion der Elektron-Effizienz und des Impulses. Im ersten Teil dieser Arbeit wurden zwei klassische mehrdimensionale Likelihood-Methoden untersucht, um die Elektron/Pion-Trennung zu erreichen. Darüber hinaus wurde ein beliebter Machine-Learning-Algorithmus, XGBoost, verwendet, um die Trennung gegenüber den klassischen Methoden zu verbessern. Zusätzlich werden in dieser Arbeit Details zur Implementierung der Elektronidentifikation im neuen O^2-Framework und Erweiterungen der TRD-Spurparameter auf der Physikanalyse-Ebene gegeben. Die Ergebnisse dieser Studie sind von Bedeutung für die zukünftigen Datennahmen von ALICE am LHC und können zu einem besseren Verständnis der Physik von schweren Ionenkollisionen beitragen.

Table of Contents

1 Introduction

This introductory chapter briefly summarizes the history of particle physics in Section 1.1, including the discovery of fundamental particles and the development of theories to explain their behavior. In Section 1.2, an overview of the Standard Model is provided. Section 1.3 describes the Quark-Gluon Plasma, which leads to the motivation for this thesis. Finally, in Section 1.4, an overview of the thesis is provided.

1.1 History of Particle Physics

The field of particle physics, also known as high-energy physics, deals with understanding the fundamental building blocks of matter and the forces governing their interactions. It seeks answers to some of the most fundamental and profound questions about the nature of the universe, such as what the universe is made of and how it evolved.

The study of subatomic particles can be traced back to the end of the 19th century with the discovery of the electron by J. J. Thomson in 1987 [1]. In 1919, E. Rutherford discovered the proton, the positively charged particle found in the nucleus of atoms [2], and later also suggested the existence of the neutron. The neutron was found in 1932 by J. Chadwick, a neutral particle also found in the nucleus of atoms [3]. Around 1928, P. Dirac proposed that electrons can have both a positive and negative charge [4], the former known as the positron, the antiparticle of the electron. Antiparticles are particles with the same mass as their corresponding particles but opposite charges. Experimentally, C. Anderson discovered the positron in 1932, confirming the existence of antimatter [5], and in 1937 S. Neddermeyer et al. detected the muon [6]. Furthermore, W. Pauli postulated in 1933 the neutrino to explain the continuous energy spectrum of the β^- decay. The neutrino was first confirmed experimentally in 1956 by F. Reines and C. L. Cowan [7].

The discovery of the neutron first sparked interest in developing particle accelerators to discover more of the fundamentals of the known matter in the 1930s. Particle accelerators are machines that accelerate particles to speeds close to the speed of light and then collide with other particles or a target. These collisions create showers of particles that can be studied and analyzed to reveal their properties. The invention of the cyclotron in the 1940s and the synchrotron in the 1950s led to the discovery of many new particles. In 1947, the first strange particle, the kaon, was discovered by G. Rochester and C. Butler [8]. A strange particle is a particle that contains at least one strange quark. In 1955, C.N. Yang and R. Mills proposed a non-abelian gauge theory, which forms the basis of the electroweak and strong nuclear forces [9]. The quark model was independently proposed by M. Gell-Mann and G. Zweig, which explained the observed mesons and baryons as composed of different combinations of three generations of quarks [10, 11]. Mesons are particles with one quark and one antiquark, while baryons have three quarks. This model also predicted the existence of quarks, such as the charm quark, which was discovered in 1974 with the independent discovery of the J/Ψ at the Stanford Linear Accelerator Center (SLAC) and Brookhaven National Laboratory (BNL). Quarks interact strongly by exchanging gluons.

Gluons were indirectly observed in 1979 [12]. Additionally, it was observed empirically that quarks only exist in color-neutral bound states, a property known as color confinement. Next, in 1983, the W and Z bosons were observed at Conseil Européen pour la Recherche Nucléaire (CERN) [13]. The Higgs boson, predicted in 1964 by P. Higgs [14] and F. Englert and R. Brout [15], was discovered in 2012 independently at the Large Hadron Collider (LHC) by A Torodial LHC Apperatus (ATLAS) [16] and the Compact Muon Solenoid (CMS) [17].

1.2 Overview of the Standard Model

All the significant accomplishments listed in Section 1.1 culminated in formulating and extending the Standard Model (SM) of physics. It is the most well-tested and successful theory of particle physics that explains the structure of matter and the interactions between elementary particles. The model describes three closely related Yang-Mills systems (strong, weak and electromagnetic), one Higgs field and several Dirac matter fields.

The SM describes two types of particles: fermions and bosons. Fermions are the elementary building blocks of matter. There are six types of quarks and six types of leptons, which are organized into three generations, and each has an antiparticle with the opposite charge. The first generation includes the up and down quarks, the valence quarks of the proton and neutron, and the electron and its neutrino. Valence quarks are the constituent quarks of hadrons responsible for their quantum numbers. In contrast, sea quarks are virtual particles that arise due to the strong interaction and contribute to the structure of hadrons. The second and third generations include heavier versions of these particles.

The interactions between particles in the Standard Model are mediated by force-carrying particles called bosons. Photons mediate the electromagnetic force, while the weak force is mediated by W and Z bosons, and gluons mediate the strong force. The particles interact with the Higgs field, which gives them their masses.

The electromagnetic force, i.e., Quantum Electrodynamics (QED), describes the interaction between charged particles like electrons and protons. The weak force is best known for its role in radioactive decay and is fundamental to flavor physics. In contrast, the strong force, i.e., Quantum Chromodynamics (QCD), describes the interaction between quarks and gluons and the formation of hadrons. The leptons interact electromagnetically (except neutrinos) and weakly because they do not carry color charge and thus do not interact with the gluon. The confinement of quarks to colorless states has been a fundamental challenge in understanding the strong force, as direct observations of individual quarks are impossible. This phenomenon is attributed to the self-interaction of gluons, which causes the strong force to increase as the distance between two quarks increases. As a result, quarks cannot exist in isolation and are only observed in bound states, providing insight into the nature of the strong force and the fundamental constituents of matter. Figure 1.1 shows a graphical representation of the Standard Model.

While the SM is a very successful theory, it is not complete. The incompleteness can already be seen because the gravitational force, although negligible at the subatomic scale, is not

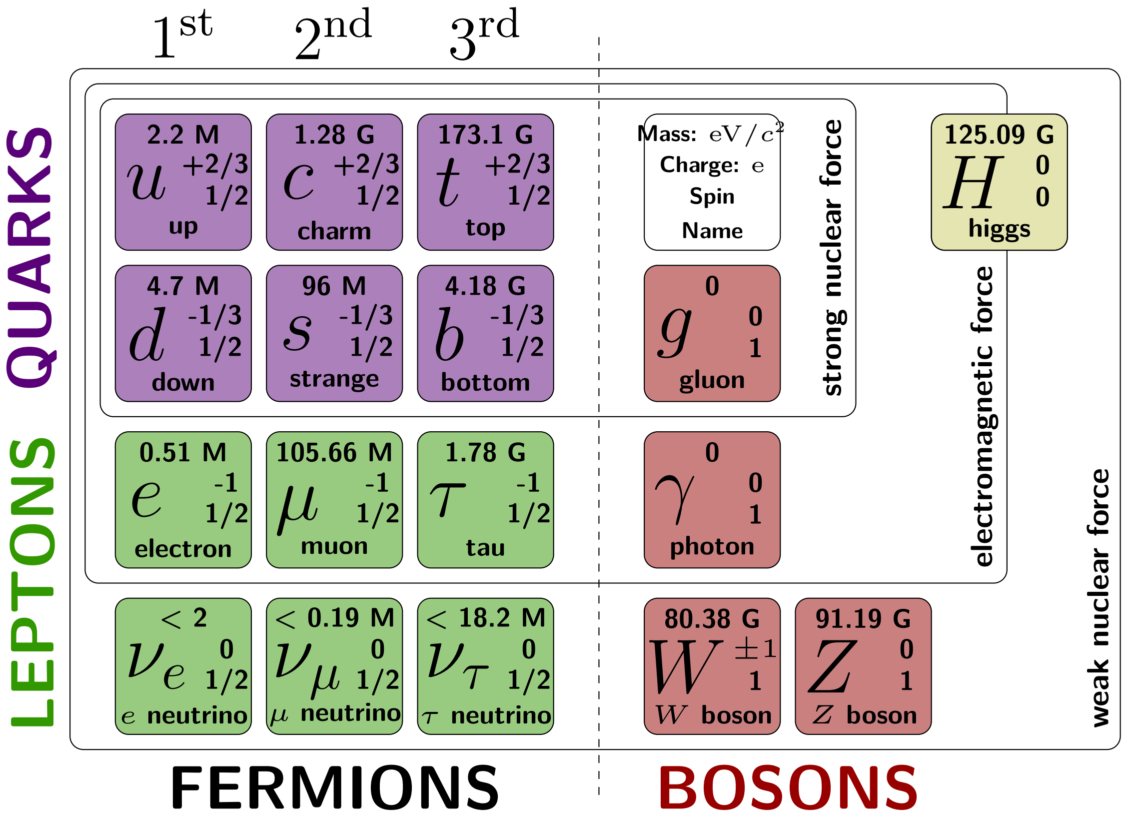


Figure 1.1: Standard Model of Physics [20].

accounted for. Additionally, the Standard Model fails to explain the existence of dark matter [18] and the matter-antimatter asymmetry in the universe [19]. To address these gaps in knowledge, physicists must continue to probe the SM through experiments.

1.3 Quark-Gluon Plasma

The QCD phase diagram as a function of temperature T and baryon chemical potential μ_B, as shown in Figure 1.2, is only known schematically, and the precise boundaries are not known experimentally. In addition to the region of hadrons at low T and μ_B, there is also a region characterized by color superconductivity at high μ_B and low T. The third phase at high T and μ_B is called Quark-Gluon Plasma (QGP). A first-order phase transition as a function of temperature at finite μ_B is expected to end at a critical point that is not yet established experimentally. One tool to map out this diagram is lattice QCD, which has technical difficulties at non-vanishing μ_B or n_B, where n_B denotes the net baryon number density. Lattice QCD predicts a cross-over transition at $T_c = (156.5 \pm 1.5)\,\mathrm{MeV}$ and zero net-baryon density [22].

The QGP phase is a '(locally) thermally equilibrated state of matter in which quarks and gluons are deconfined from hadrons propagating over nuclear, rather than nucleonic, volumes' [21]. In this phase, quarks and gluons are asymptotically free, i.e., breaking free

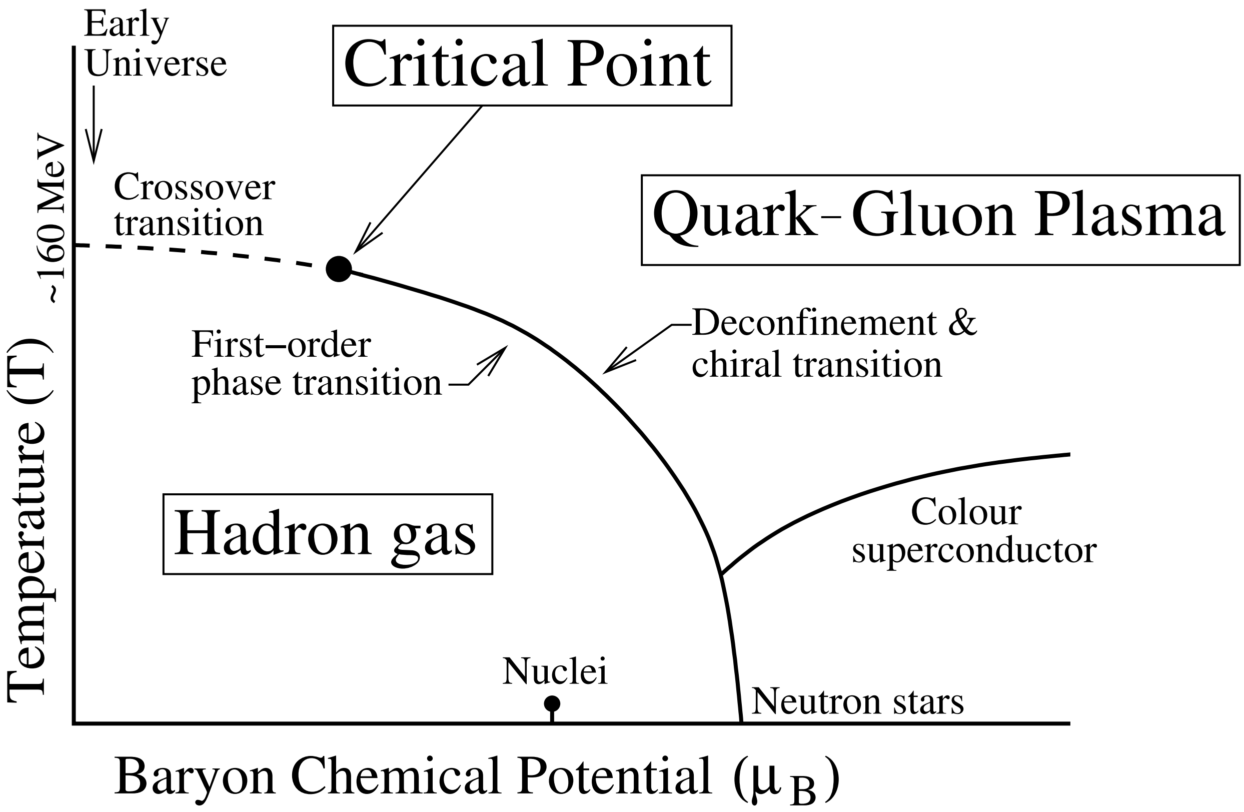


Figure 1.2: QCD Phase diagram. Adapted from [21].

from color confinement, and chiral symmetry is restored [23]. It is believed that the QGP existed up to a few microseconds after the Big Bang before the universe cooled down and reached T_c, at which point quarks hadronized, forming the matter known today. While the QCD phase diagram is not yet fully understood experimentally, continued research and advancements in lattice QCD and other techniques can help better understand the early universe and the fundamental forces at work.

1.3.1 Experiment and Evolution

Experimental evidence for the QGP has been obtained through heavy-ion collisions at facilities such as CERN Super Proton Synchrotron (SPS), Relativistic Heavy Ion Collider (RHIC) at BNL and LHC at CERN. The highly Lorentz contracted nuclei form a coherent cloud of partons before the collision. During collision, a significant fraction of the incoming kinetic energy is deposited in the overlap region leading to a high-energy-density fireball. A locally thermalized state of matter (QGP) is created in less than 1 fm/c. As the system evolves, it eventually cools down, leading to hadronization at T_C, where the quarks and gluons recombine to form hadrons. Chemical freeze-out occurs when inelastic collisions among these hadrons stop, followed by kinematic freeze-out, where the elastic collisions stop. Afterwards, the hadrons can be measured with the surrounding detectors. A depiction of the evolution of a single heavy-ion collision at LHC energies, including the relevant time scales, is shown in Figure 1.3.

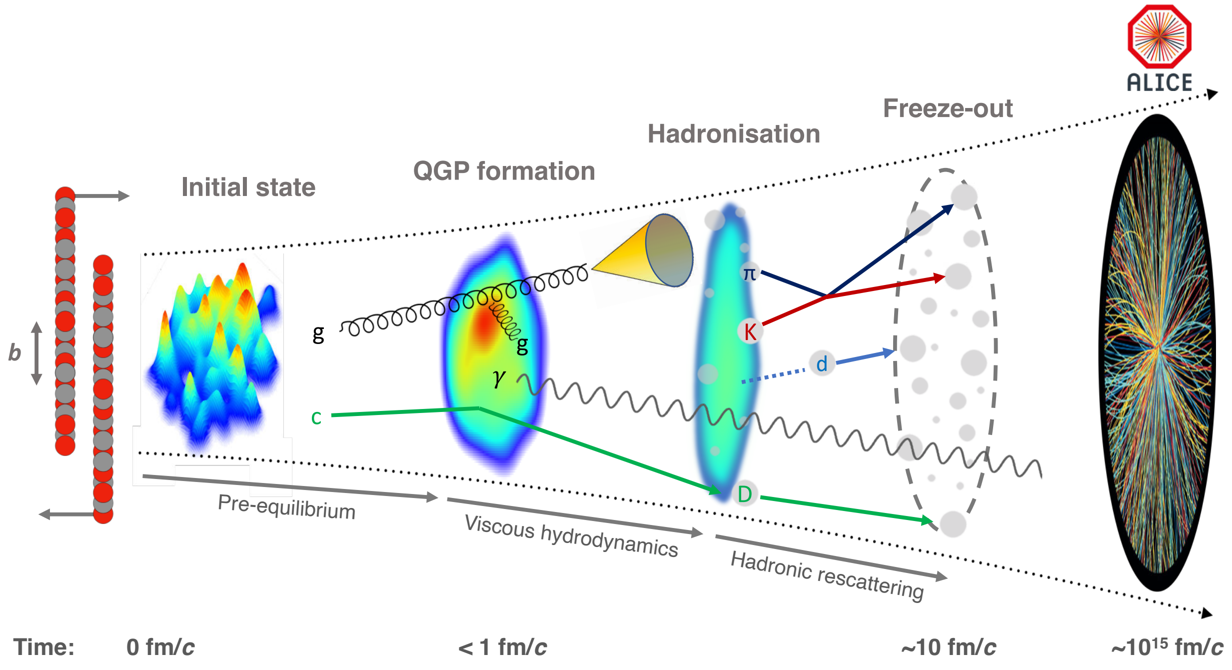


Figure 1.3: Evolution of a heavy-ion collision at LHC energies. Taken from [24].

1.3.2 Properties and Probing

Studying the properties of the QGP is at the core of the ALICE physics program. The QGP can only be observed indirectly using observables assuming its formation and evolution. Many suitable probes exist to determine the properties of this matter state, e.g., event-by-event fluctuations, jet quenching and strangeness enhancement [21]. Several essential probes of the QGP decay and can be detected via electrons.

- Quarkonia: Bound states of heavy quarks (charm or bottom) and their antiquarks called quarkonia. In a hot and dense medium such as the QGP, the attraction between quarks is screened, suppressing the formation of these states. The sequential melting of quarkonia as the temperature of the medium rises, ordered by their increasing binding energies, serves as a ‘thermometer’ for the medium. A systematic study of suppression patterns of J/Ψ and Υ families, together with baseline measurements in pp and pA collisions, over a broad energy range, would help to disentangle hot and cold nuclear matter effects [21]. Additionally, a regeneration effect is observed at lhc energies at low momenta due to the large $c\bar{c}$ cross-section. Quarkonia can decay via a dielectron channel (i.e., $J/\Psi \rightarrow e^+e^-$).

- Direct Photons: Throughout the whole evolution, different processes emit photons. Since photons only interact electromagnetically, they remain undisturbed on their way through the medium and retain the complete information about the system providing a measure of the system’s temperature. Through pair-production ($\gamma \rightarrow e^+e^-$), photons convert to electrons while interacting with the detector.

- Dileptons: In the QGP phase, e.g., a quark can interact with an antiquark to form a virtual photon that decays into two leptons. Similar to photons, their electromagnetic nature makes them powerful probes for the space-time evolution of the expanding fireball.

These observables require clean electron/positron identification, and selecting events with rare probes is crucial to explore and measuring the properties of the QGP. To effectively identify electrons and select events with these rare probes, the Transition Radiation Detector is integrated into the ALICE detector setup, providing both capabilities.

1.4 Thesis Overview

This thesis deals with evaluating and integrating electron identification in the new O^2 framework for the ALICE Transition Radiation Detector. As a first step, the thesis will introduce CERN, the LHC and ALICE in Chapter 2. In Chapter 3, the new computing system O^2 is described. Chapter 4 presents the working principle of the Transition Radiation Detector and describes its operation and performance in the current LHC Run 3. Following this, Chapter 5 presents the thesis analysis and defines the algorithms for electron identification. Subsequently, the obtained results are presented in Chapter 6. Afterwards, the implementation of the electron identification algorithms in the O^2 is described in Chapter 7. Finally, Chapter 8 provides the conclusion, including further steps.

2 CERN, LHC and ALICE

This chapter starts with Section 2.1, in which an overview of the fascinating history of CERN, one of the world's most renowned centers for particle physics research, is given. Afterwards, Section 2.2 introduces the LHC and its injector chain, representing the most advanced and powerful tools for particle physics research today. Furthermore, Section 2.3 describes ALICE, one of the four major experiments at the LHC, which focuses on studying the quark-gluon plasma, a unique state of matter believed to have existed just moments after the Big Bang. In the end, an overview of the physics program of ALICE in LHC Run 3 in Section 2.4 is given. A comprehensive understanding of the vital role played by CERN in advancing knowledge of the universe is aimed to be provided by exploring CERN's history, the impressive capabilities of the LHC, and the cutting-edge research conducted by ALICE throughout this chapter.

2.1 Conseil Européen pour la Recherche Nucléaire

At the end of World War II, European science was severely drained. However, a small group of established scientists recognized the need for a joint post-war European research facility to regain the ability to perform world-class science and promote unity among nations. In 1952, the first meeting of the Conseil Européen pour la Recherche Nucléaire (CERN) Council was held, and in 1955 Felix Bloch, the Director-General, laid the first foundation stone of the organization [25].

Since then, CERN has become the world's leading research facility, particularly in particle physics, and has made numerous world-changing contributions, including developing World Wide Web and changing our fundamental understanding of the universe. CERN is home to the largest particle accelerator, the Large Hadron Collider (LHC), and several smaller accelerator complexes. The collisions produced in these accelerators are detected by several experiments, including the major science experiments situated at the LHC ring: A Torodial LHC Apperatus (ATLAS) [26], Compact Muon Solenoid (CMS) [27], LHC-beauty (LHCb) [28] and A Large Ion Collider Experiment (ALICE) [28].

From its humble beginnings in the aftermath of World War II, CERN has grown into a hub of international collaboration and scientific discovery, with an impressive legacy of groundbreaking achievements.

2.2 Large Hadron Collider

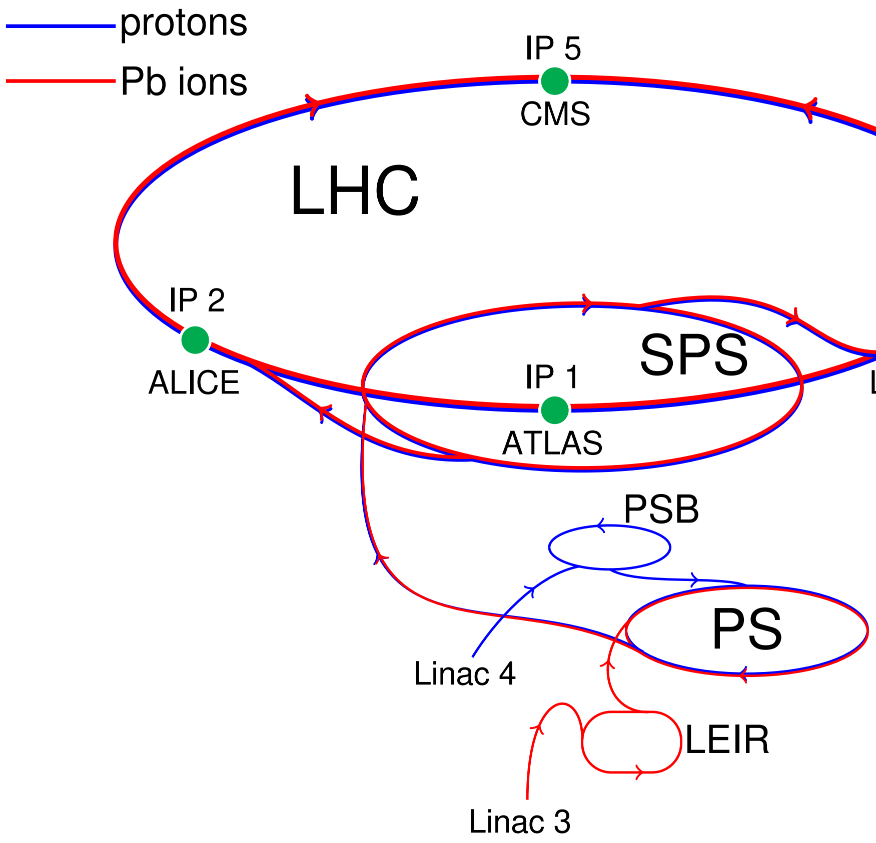


Figure 2.1: LHC injection ring. Adapted from [29].

The Large Hadron Collider (LHC) is a particle accelerator composed of a ring of superconducting magnets spanning 27 km in circumference and located between Lake Geneva and the Jura Mountains at CERN. Inside the accelerator, two high-energy particle beams collide at four intersection points. The LHC is designed to collide proton beams with a center-of-mass energy of up to $\sqrt{s} = 14\,\text{TeV}$ and a bunch spacing of 25 ns. Like its predecessor, the Large Electron-Positron Collider (LEP), LHC uses the preexisting CERN accelerator complex as an injector. The schematic layout of the LHC injection chain is shown in Figure 2.1.

Protons (or Ions) are first accelerated in the linear accelerator LINAC 4 (LINAC 3) and then injected into the Proton Synchrotron Booster (PBS) (Low Energy Ion Ring (LEIR)). Next, they are injected into the Proton Synchrotron (PS) and accelerated to 25 GeV/c. From the PS, the beams are directed to the Super Proton Synchrotron (SPS), accelerating them to 450 GeV/c, and finally into the LHC [30].

2.3 A Large Ion Collider Experiment

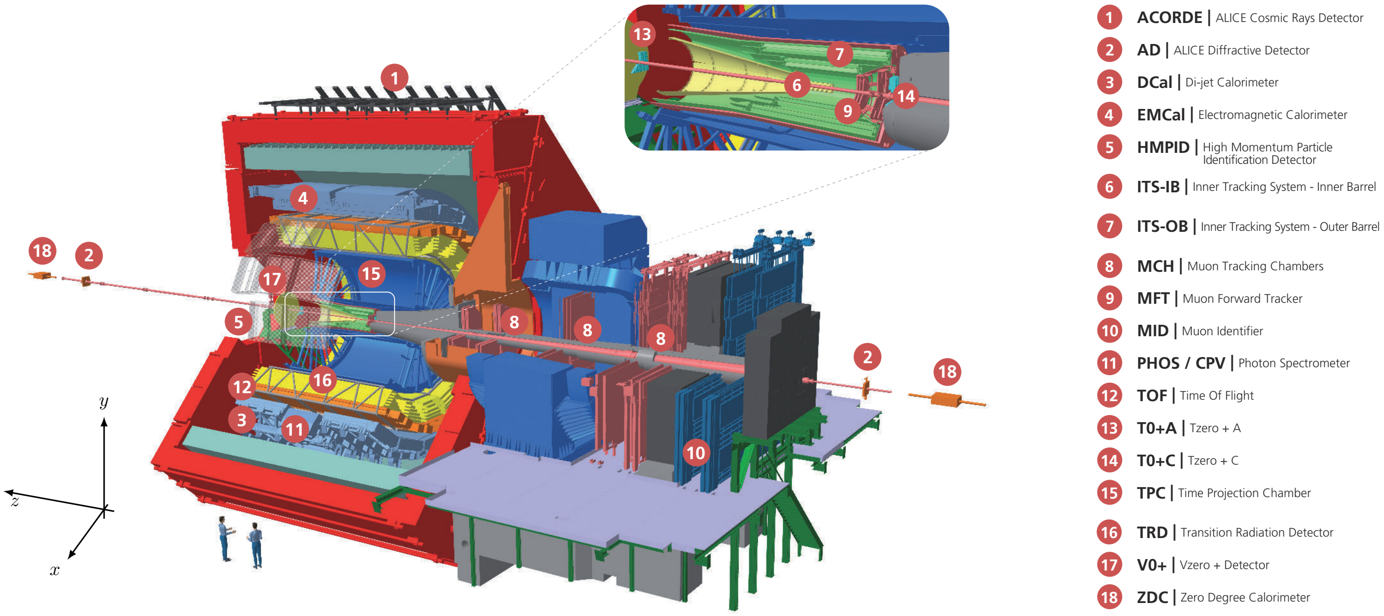


Figure 2.2: ALICE detector in Run 3. The central barrel detectors are housed in the large solenoid magnet at mid-rapidity. The muon arm is installed at forward rapidity. The global coordinate system is indicated on the left. Adapted from [31].

A Large Ion Collider Experiment (ALICE) was proposed in 1993 [32] as a general-purpose heavy-ion detector for the LHC at CERN, focusing on QCD [33]. Its main objective is to study matter under extreme conditions.

Constructed by a collaboration of over 1000 physicists and located at point IP 2 of the LHC, A Large Ion Collider Experiment (ALICE) has successfully been operated during the LHC Run 1 and 2, as discussed in [34]. To achieve its physics program, ALICE must be able to access a wide range of transverse momenta. It provides robust Particle Identification (PID) for momenta as low as $p_T \approx 150\,\mathrm{MeV/c}$ and up to $p_T \approx 20\,\mathrm{GeV/c}$ [24]. Additionally, high-energy heavy-ion collisions present a unique challenge for particle track reconstruction due to the typically high density of charged particles.

To meet these requirements set by the physics program, ALICE is based on a so-called central barrel that covers the whole azimuth angle and the pseudorapidity region $|\eta| \leq 0.9$ [24]. Furthermore, ALICE houses several other detectors, as shown in Figure 2.2, which further extend its physics capabilities. Additionally, it includes a dedicated muon arm at forward rapidity.

The following sections introduce the global coordinate system in Section 2.3.1, describe the central barrel detectors in more detail in Section 2.3.2, and present the planned physics program for ALICE in LHC Run 3 and Run 4 in Section 2.4.

2.3.1 Global Coordinate System

The global coordinate system is a right-handed Cartesian system with the origin placed at the nominal interaction point [35]. The z-axis is aligned with the beam and pointing away from the muon arm, the x-axis pointing inward to the center of the LHC and the y-axis upwards, as shown in Figure 2.2.

2.3.2 Central Barrel Detectors

The central barrel detectors of the ALICE experiment play a crucial role in the track reconstruction and PID capabilities. The relevant detectors for this thesis, arranged from inward to outward, are the Inner Tracking System (ITS), the Time Projection Chamber (TPC), the Transition Radiation Detector (TRD), and the Time-of-Flight Detector (TOF) [33]. A brief description of each detector's main principle and purpose follows.

Inner Tracking System

The Inner Tracking System (ITS) is a silicon-based detector that provides essential information for vertexing and improves the TPC tracking resolution. For Run 3, the ITS has been entirely replaced by a new 7-layer silicon detector design using the Monolithic Active Pixel Sensor (MAPS) technology [36]. The design is split into the Inner Barrel and Outer Barrel. The Inner Barrel has three layers at radii of 2.3 cm, 3.0 cm, and 3.7 cm from the nominal interaction point, and the Outer Barrel has four layers ranging from 19.6 cm to 39.2 cm. The main working principle of the ITS is the energy deposition of particles in depleted silicon regions. Compared to the previous ITS used in LHC Run 1 and 2, the new design has a significantly reduced material budget (radiation length of 0.3% X_0 per layer for the Inner Barrel) and improved the intrinsic spatial resolution ($4\,\mu$m), which is particularly important in separating primary and secondary vertices [36]. Additionally, the new ITS is designed for continuous readout and an interaction rate of 1 MHz (50 kHz) for pp (Pb−Pb) collisions. The ITS is shown in Figure 2.2 as numbers 6 and 7 for the Inner and Outer Barrel, respectively.

Time Projection Chamber

The Time Projection Chamber (TPC) is the primary tracking detector in the central barrel of the ALICE experiment, providing good track separation, charged-particle momentum measurements, PID (fully calibrated energy resolution of 5%), and vertex determination [31]. The TPC is a gas-filled drift chamber of about $90\,\mathrm{m}^3$ located between radii of 85 cm and 247 cm from the nominal interaction point [33], as shown in Figure 2.2 as number 15.

The principle of operation is the ionization of gas by traversing charged particles. An applied electric field makes electrons drift toward the end caps, where the readout chambers are located. For Run 3, the Front-End Electronics (FEE) and the Multi-Wire Proportional Chambers (MWPC) have been replaced. The FEE have been replaced by Common Readout Units (CRUs), allowing for a significant data bandwidth increase. The MWPC were replaced

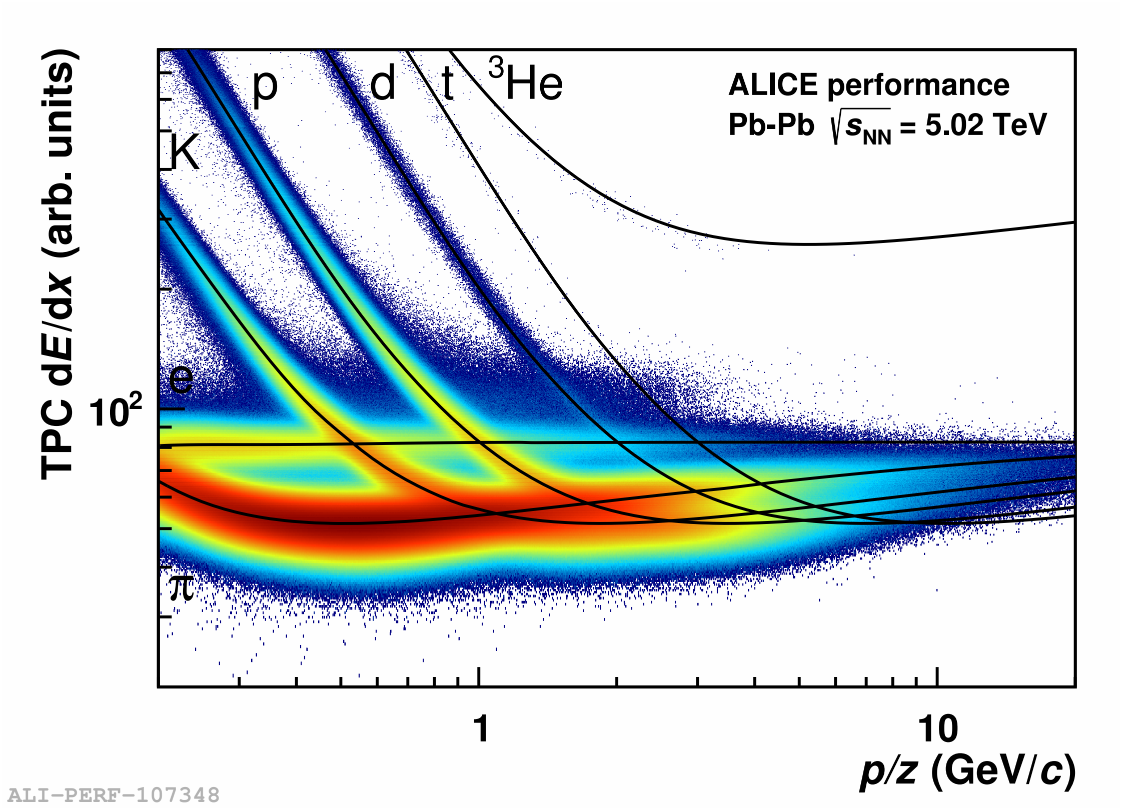


Figure 2.3: TPC dE/dx as a function of momentum with superimposed Bethe-Bloch lines for various particle species for Run 2 Pb−Pb low interaction rate runs at $\sqrt{s_{\mathrm{NN}}} = 5.02\,\mathrm{TeV}$.

by Gas Electron Multiplier (GEM) foils, allowing the removal of the gating grid [34]. Hence, the TPC is now operating in continuous readout mode.

In addition to tracking, the TPC also provides excellent PID capabilities. The mean energy loss (TPC dE/dx) as a function of particle velocity β and Lorentz factor γ is parameterized by a Bethe-Bloch-Aleph $f(\beta\gamma)$ [1] curve [37]:

$$f(\beta\gamma) = \frac{P_1}{\beta^{P_4}}\left[P_2 - \beta^{P_4} - ln\left(P_3 + \frac{1}{(\beta\gamma)^{P_5}}\right)\right], \tag{2.1}$$

where P_i are fit parameters to data.

The expected energy ($\langle dE/dx\rangle$) loss can then be calculated with:

$$\langle dE/dx\rangle(\beta\gamma) = \mathrm{MIP} * f(\beta\gamma), \tag{2.2}$$

where MIP denotes the minimum energy for ionizing particles. Figure 2.3 shows the TPC dE/dx measurement in Pb−Pb data at $\sqrt{s_{\mathrm{NN}}} = 5.02\,\mathrm{TeV}$. The critical factor for the PID

---

[1] Apparatus for LEP Physics (ALEPH) was a particle detector at LEP [37].

of a track is the measurement's energy resolution $\sigma_{dE/dx}$. With this, the deviation of the measurement ($\langle \frac{dE}{dx} \rangle_{\mathrm{meas.}}$) from the expected energy loss ($\langle \frac{dE}{dx} \rangle_i$) can be calculated as:

$$n^i\sigma_{\mathrm{TPC}} = \frac{\langle \frac{dE}{dx} \rangle_{\mathrm{meas.}} - \langle \frac{dE}{dx} \rangle_i}{\sigma_{dE/dx}}, \tag{2.3}$$

where $n^i\sigma_{\mathrm{TPC}}$ gives the distance in standard deviations from the expected energy deposition for a charged particle of species i, e.g., protons, pions and kaons.

Transition Radiation Detector

The primary purpose of the Transition Radiation Detector (TRD) is to provide electron identification by measuring the specific energy loss of particles. The electron's small mass gives them a large $\beta\gamma$ producing Transition Radiation (TR), which induces an additional signal in the detector, while heavier particles with the same momentum do not. Additionally, the TRD contributes to the tracking capabilities of the central barrel and the calibration of the TPC. Chapter 4 presents a detailed description of the design, the principle of operation and the readout scheme. It is depicted in Figure 2.2 as number 16.

Time-of-Flight Detector

The Time-of-Flight Detector (TOF) detector further extends the particle identification and tracking capabilities at intermediate momenta. Particle identification uses exact time-of-flight measurements and track length information. TOF uses a large array of Multigap Resistive Plate Chambers (MRPC) that achieve a timing resolution better than 80 ps [37]. No major upgrade was needed to meet the physics requirements of Run 3 and 4.

TOF measures the time-of-flight $\tau = t_{\mathrm{TOF}} - t_0$, where t_0 is the interaction time measured by the Fast Interaction Trigger (FIT) and t_{TOF} is the measured time at the detector. Furthermore, through calculation of the track length L, the particle velocity β can be determined as $\beta = \frac{L}{\tau}$. The particle velocity against the particle momentum distribution is shown in Figure 2.4.

Additionally, from these measurements, the squared invariant mass of the particle can be calculated via:

$$m^2 = p^2 \left[\left(\frac{c\tau}{L} \right)^2 - 1 \right]. \tag{2.4}$$

Similar to the TPC, the separation power $n^i\sigma_{\mathrm{TOF}}$ can be defined by taking the difference of measured time $\tau_{\mathrm{meas.}}$ and the expected time $\tau^i_{\mathrm{exp.}}$ for particle i:

$$n^i\sigma_{\mathrm{TOF}} = \frac{\tau_{\mathrm{meas.}} - \tau^i_{\mathrm{exp.}}}{\sigma_{\mathrm{TOF}}}. \tag{2.5}$$

The time difference of two particles ($i = 1, 2$) can be calculated accordingly using Equation 2.4:

$$\tau_1 - \tau_2 \approx \frac{L}{2p^2 c}\left(m_1^2 - m_2^2\right). \tag{2.6}$$

Therefore, the separation power decreases quadratically with increasing momentum for particles with similar masses. TOF allows particle separation of π/K below 2.5 GeV/c and p/K up to 4 GeV/c above 3σ [37]. TOF is sketched in Figure 2.2 as number 12.

Equations 2.3 and 2.5 can be combined to use the full information and enhance the PID capabilities. A single variable can be defined as:

$$n^i\sigma_{\text{comb.}} = \begin{cases} |n^i\sigma_{\text{TPC}}| & \text{if signal only in TPC} \\ \sqrt{\frac{1}{2}\left[\left(n^i\sigma_{\text{TOF}}\right)^2 + \left(n^i\sigma_{\text{TPC}}\right)^2\right]} & \text{if signal also in TOF} \end{cases}. \tag{2.7}$$

Equation 2.7 has the advantage of combining two detectors complementarily if the TOF signal is also available [38].

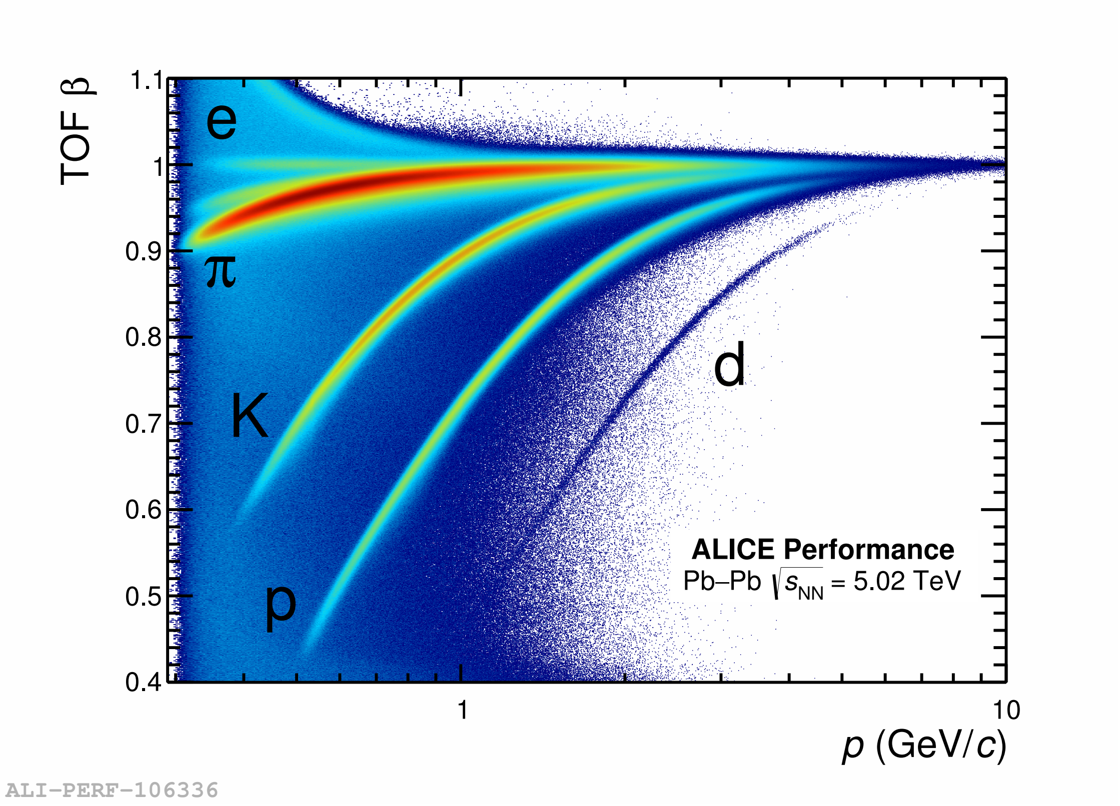


Figure 2.4: TOF β vs. momentum plot in Pb−Pb at $\sqrt{s_{\text{NN}}} = 5.02$ TeV.

2.4 ALICE physics program in LHC Run 3 and 4

After the Long Shutdown 2 (LS2) of the LHC, ALICE underwent an intense period of upgrades to prepare for its physics program. In 2021, a pilot beam for pp and Pb−Pb collisions marked the first successful data-taking for ALICE in LHC Run 3. In early 2022, ALICE recorded its first collisions in stable pp beams at a center-of-mass energy of $\sqrt{s} = 13.6\,\mathrm{TeV}$. For pp collisions, the goal is to go up to a 1 MHz interaction rate. The Pb−Pb run in 2022 was canceled due to an energy crisis. However, in 2023 Pb−Pb collisions will proceed as planed with an interaction rate of 50 kHz. The primary focus of the targeted physics program is the study of rare probes and their coupling with the medium and hadronization processes. These rare probes include heavy-flavor particles, quarkonium states, real and virtual photons, jets, and their correlations with other probes [39]. Due to their rare abundance and very low signal-to-background ratios, studying these probes requires extensive statistics. Therefore, the program aims to integrate a luminosity of $13\,\mathrm{nb}^{-1}$ in Pb−Pb collisions, $200\,\mathrm{pb}^{-1}$ for pp and $1\,\mathrm{pb}^{-1}$ for p−Pb [40]. Overall, the CERN ALICE Run 3 and 4 physics program is a crucial effort to deepen the understanding of the early universe and the fundamental laws of nature that govern it. By studying rare probes and their interactions in the QGP, ALICE will shed light on the properties of this unique state of matter and the strong force that binds it together.

3 Online and Offline Computing System

This chapter introduces in Section 3.1 the newly developed data processing O^2 framework for ALICE in LHC Run 3 and 4. Then in Section 3.2, the basis of the O^2 framework is described. Afterwards, Section 3.3 explains the whole flow of data from the detector to the computing facilities, including calibration steps.

3.1 Introduction

The major detector upgrades and new physics program in LHC Run 3 and 4, described in Chapter 2 require a fundamental shift in the data processing model. The TPC will operate in a continuous readout mode to keep up with the 50 kHz interaction rate in Pb−Pb collisions at $\sqrt{s_{\mathrm{NN}}} = 13.6\,\mathrm{TeV}$. The resulting throughput is estimated to be greater than 3.5 TB/s for Pb−Pb events, roughly two orders larger than in Run 1 [39]. The approach to data processing developed in Run 1 and 2 cannot meet the new throughput requirements, necessitating a new data processing model. The new computing model in LHC Run 3 and 4 is designed to achieve the maximal data volume reduction and is split into two stages. The first stage reduces the data volume online (synchronously) through track reconstruction and early calibration based on average running conditions. The second reconstruction stage will be performed offline (asynchronously), using the final calibration to achieve the required data quality [39].

In order to meet the requirements set by the physics program, a new software framework, combining the previously separated online and offline frameworks, was developed, named Online and Offline Computing System (O^2). O^2 is developed using general public libraries and tools such as ROOT [41], CMake [42] and Boost [43] in C++17 [44]. Additionally, O^2 utilizes two other frameworks called ALFA and FairRoot. ALFA is a framework jointly developed by ALICE and the Facility for Antiproton and Ion Research (FAIR), providing message-queue-based data transport, process load-balancing, and logging facilities [45]. FairRoot is a detector-oriented simulation, reconstruction, and data analysis framework developed by FAIR [46]. An overview of the dependencies of O^2 is depicted in Figure 3.1.

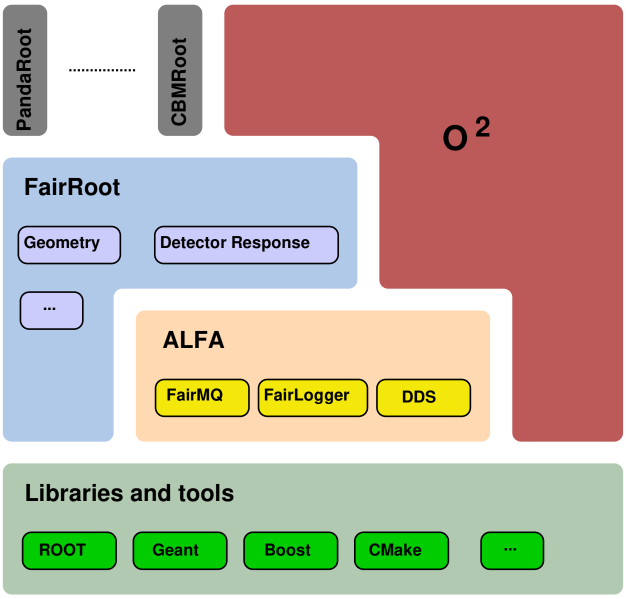


Figure 3.1: Overview of the O^2 dependencies. Taken from [47].

3.2 Data Processing Layer

At the heart of O^2 is the Data Processing Layer (DPL). This layer describes computation as a set of data processors implicitly organized in a logical data flow describing how data is transformed [48]. DPL devices can be configured within the framework to require particular inputs and specify their output, allowing the framework to automatically create workflows for a group of DPL devices by matching their input and output. A DPL device, for example, such as the TRD tracker, requires (ITS-)TPC tracks and TRD tracklets (local track segments) as input and provides (ITS-)TPC-TRD matched tracks as output. Additionally, the framework reorders the workflow topology to allow full utilization of heterogeneous computing systems [48], including hardware accelerators such as Graphics Processing Units (GPUs). The transport layer is implemented using the FairMQ message-passing framework developed at the Gesellschaft für Schwerionenforschung (GSI), which allows distributed computing over networks and shared memory.

3.3 Data flow

Overall, the data flow in the O^2 project is complex and involves multiple processing stages to reduce the raw data rate and ensure data quality. By providing an overview of the key components and processes involved, this section sets the stage for a more detailed discussion of the data analysis and results that follow. It is worth noting that these sections only give a rudimentary overview of the facility, and a more detailed description is provided in [39].

3.3.1 Timeframes

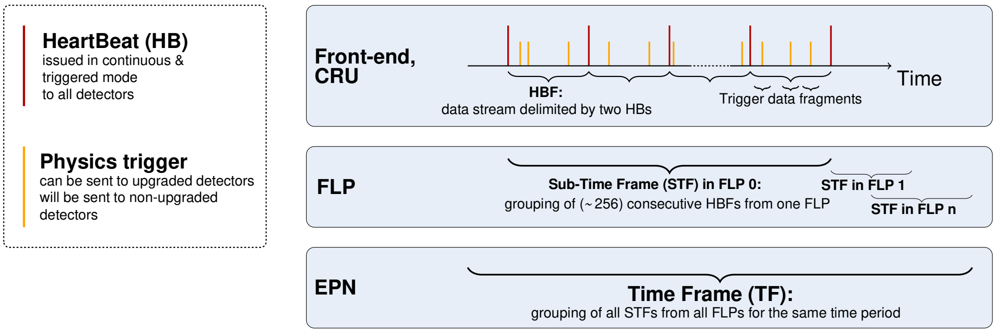


Figure 3.2: Building process of a Timeframe. Taken from [47].

The raw detector data from all detectors is initially produced at a rate exceeding 3.5 TB/s for minimum bias Pb—Pb at 50 kHz by the FEE. The main contributors are: TPC (92.5%), ITS (3.6%) and TRD (1.8%) [39]. To handle this high data rate, a computing farm consisting of First Level Processors (FLPs) is employed to reduce the data rate to 500 GB/s. The compression is achieved through techniques such as zero suppression and encoding. The reduced data is then split into chunks known as Heartbeat Frames (HBFs) and delimited by Heartbeats (HBs) issued by the Central Trigger Processor (CTP) to all detectors, regardless of their readout mode. The HBFs are grouped into Sub-Timeframes (STFs) on the FLPs and sent to the Event Processing Nodes (EPNs). On the EPNs the STFs from all detectors are aggregated into Timeframess (TFs). The TF length is configurable, for the year 2022/23 32 HBFs, equaling 2.9 ms of detector data [48], were used. The data layout is depicted in Figure 3.2. Finally, synchronous reconstruction is performed on the EPNs to further reduce the data into Compressed Timeframess (CTFs), resulting in a final data rate of 100 GB/s. This step involves, for example, the TPC clustering the detector data, performing local tracking and encoding the data.

Furthermore, in the case of detectors with a triggered readout mode, such as the TRD, the CTP provides physics triggers. These triggers also assign TRD data to TPC clusters in the same interaction [49]. This data processing and reduction pipeline provides an efficient and effective means of managing and analyzing the vast amounts of data produced by the O^2 system.

3.3.2 O^2 Facility

The flow of data in the experiment is depicted in Figure 3.3. Raw detector data, generated at a rate of 3.5 TB/s, are transmitted via optical fibers from Front-End Electronics to the Common Readout Unit (CRU) housed in the 200 First Level Processors (FLPs) in Counting Room 1. Most FLPs are equipped with up to 3 CRUs. A few are equipped with legacy CRORC cards for detectors where the readout electrons were not upgraded, e.g., the Electromagnetic Calorimeter (EMCal). The resulting data, now in STF format, are sent to the EPNs, located on the surface of Point 2 in Counting Room 0. The EPNs, which comprise a farm of 280 servers, are each equipped with up to 8 GPUs and can perform both synchronous and asynchronous reconstruction.

During pp collisions data taking, the EPNs farm is sufficient to reconstruct the data. However, during Pb−Pb data taking, it is necessary to delay or offload some of the asynchronous processing to other computing facilities, such as the CERN Computing Centre or other well-connected Grid sites [39]. At the end of the reconstruction chain, the output data are the Analysis Object Data (AODs) files, which are then committed to permanent storage and distributed globally for physics analyses.

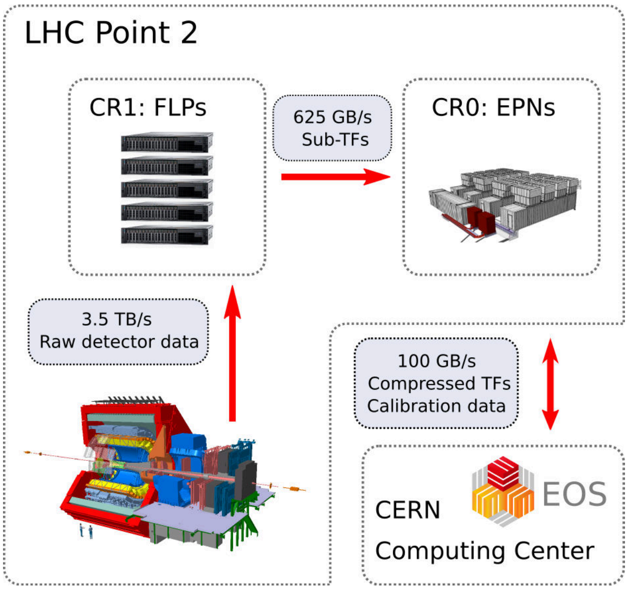


Figure 3.3: Overview of the O^2 data flow. Taken from [47].

3.3.3 Quality Control and Calibration

Quality control is essential at every step of the reconstruction process to ensure that the ALICE detector functions appropriately and produces high-quality data. To this end, O^2 provides valuable feedback for quality control throughout the reconstruction process. Calibration is integrated into O^2, allowing it to be performed at different processing stages. This early availability and usage of calibration data in the processing chain are critical for maintaining data quality and achieving high compression. The calibration data is uploaded to the Condition and Calibration Database (CCDB) for future reference and monitoring.

As an example of an online calibration, a DPL device updates correction values for each chamber every 15 mins to correct for magnetic field deflection effects ($E \times B$) in the TRD, as described in more detail in Section 4.2. Integrating calibration and quality control processes into O^2 is crucial for maintaining the high performance of the ALICE detector and ensuring the validity and good quality of the collected data.

4 Transition Radiation Detector

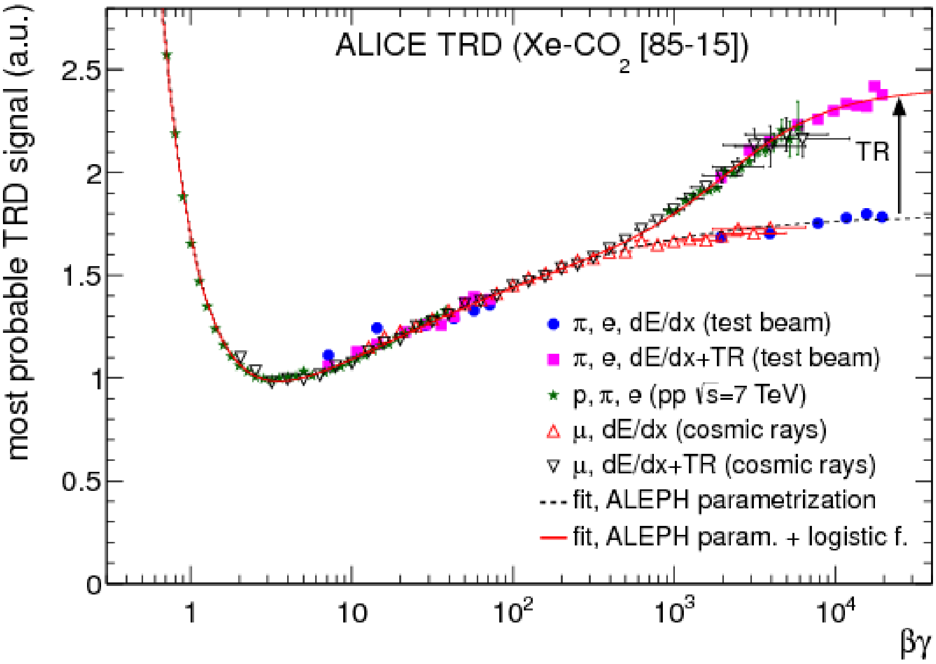


Figure 4.1: The most probable TRD signal as a function of $\beta\gamma$. Taken from [37].

In the context of heavy-ion collisions, the identification of electrons plays a crucial role in the measurement of various probes of the QGP, as already mentioned in Section 1.3.2. The Transition Radiation Detector (TRD) was initially added to the ALICE experiment to provide good electron/pion separation at high-p_T and triggers for various physics signatures, e.g., single high-p_T electrons and jets [50]. Transition Radiation (TR) was first predicted in 1945 by Ginzburg and Frank [51]. It occurs when a highly-relativistic particle (with $\beta\gamma \geq 800$) crosses the boundary between two media with different dielectric constants. The total yield is proportional to the fine-structure constant ($\alpha = 1/137$), and therefore, many boundaries are needed to obtain a satisfactory total radiation yield. Due to their low mass, electrons produce TR at momenta around 1 GeV/c, while pions start to produce TR around 140 GeV/c. Hence, TR is well-suited to differentiate electrons and pions. Figure 4.1 shows the most probable signal in the TRD as a function of $\beta\gamma$. One can see the Bethe-Bloch curve with and without TR. A clear separation of these curves is visible for higher $\beta\gamma$ due to the additional contribution to the signal of the TR.

In the following sections, an overview of the TRD detector will be given in Section 4.1. Subsequently, the calibration procedures will be described in Section 4.2. Section 4.3 describes how the TRD operates in LHC Run 3.

4.1 Detector Overview

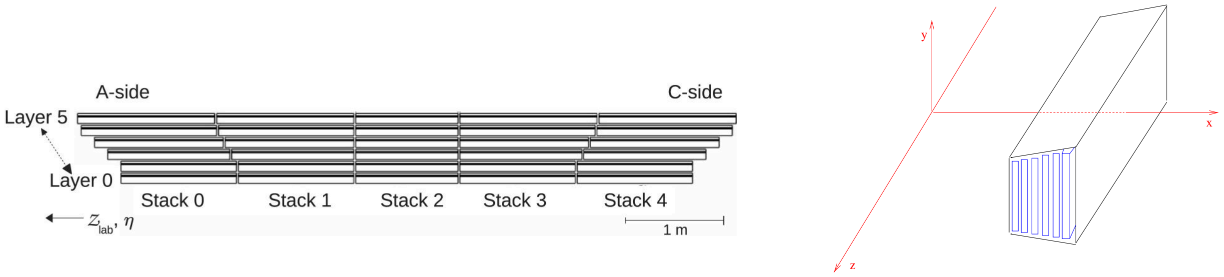
The TRD covers the whole azimuth angle at a radial distance from the nominal interaction point from 2.9 to 3.7 m, as illustrated in Figure 2.2. The active detector region covers approximately 85% of the azimuth angle due to the mechanical frame and segmented design [50].

4.1.1 Design

The TRD is divided into 18 segments (Supermodules (SMs)) in the ϕ-direction. Each supermodule comprises 30 chambers with five stacks arranged in z-direction along the beam direction and six in each stack. A cross-section of a supermodule is illustrated in Figure 4.2 (a), and a total of 522 chambers are installed. In SM 13, 14 and 15, stack 2 is removed to minimize the material in front of the photon detector Photon Spectrometer (PHOS).

4.1.2 Local Coordinate System

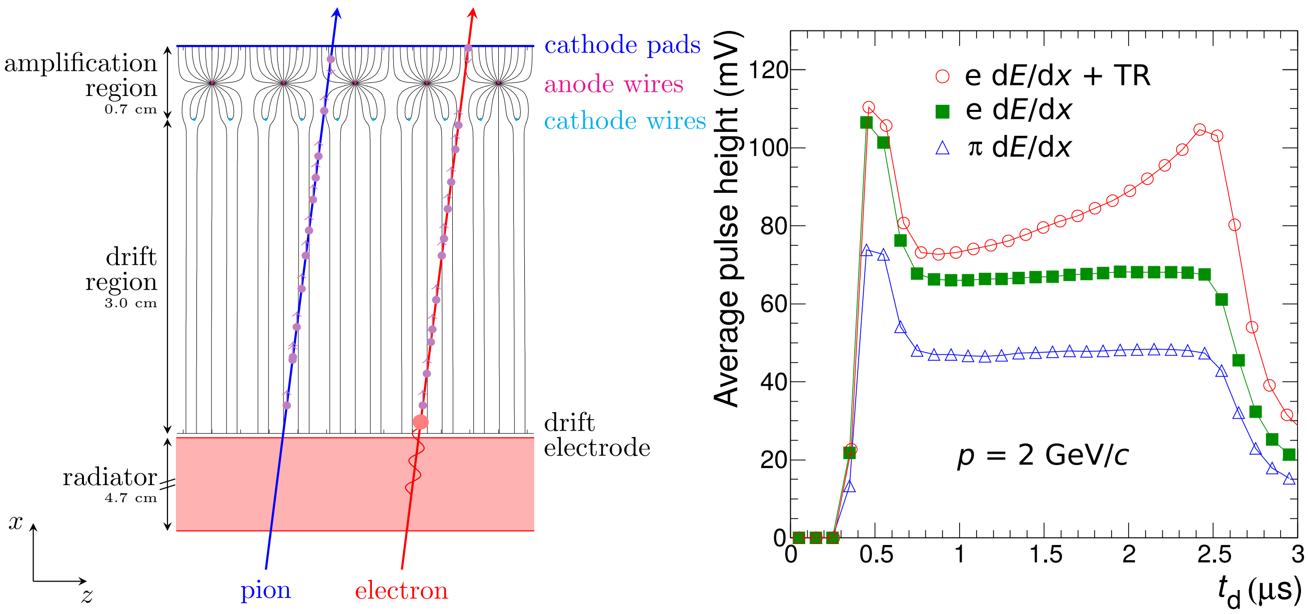
In order to simplify the description of the readout chamber, it is necessary to introduce a local coordinate system for the TRD. The local coordinate system is a right-handed orthogonal Cartesian system, as shown in Figure 4.2 (b). The origin lies at the nominal interaction point, the same as the global coordinate system in Section 2.3.1. The z-axis is also aligned with the global coordinate system. However, the x-axis is defined as perpendicular to the beam pipe and perpendicular to the chamber. The local y-axis specifies the distance relative to the center of a chamber.



(a) (b)

Figure 4.2: (a) Cross-section of a supermodule. Taken from [50]. (b) Local coordinate system of the TRD, defined by x-, y- and z-axis. Adapted from [52].

4.1.3 Readout chamber



(a) (b)

Figure 4.3: (a) Sketch of a TRD chamber in rz-direction, including a passing pion and electron. Only the highly-relativistic electron generates TR in the radiator. (b) Average pulse height as a function of drift time for pions, electrons without radiator and electrons with radiator for 2 GeV/c. Taken from [53].

Figure 4.3 (a) shows the basic buildup of a readout chamber. A radiator is mounted where transition radiation is produced before the 3 cm drift region. Behind the drift region is an amplification region followed by the cathode readout pads.

Radiator

As stated above, many layers are needed to achieve a good TR yield. The radiator is a sandwich design with two Rohacell foam sheets, each 0.8 cm thick, holding layers of polypropylene fiber mats with a thickness of 3.2 cm. Many boundaries with different dielectric constants are realized through these many-layered foam sheets. Highly-relativistic particles produce, on average, 1.45 TR photons [53].

Drift region

Charged particles and associated photons deposit energy in the detector by ionizing the $Xe - CO_2(85/15)$ gas mixture. CO_2 is used as a quencher and protects the detector from discharges. Xenon is preferred due to its high X-ray photon-absorption probability in a short absorption length. The absorption length in Xenon is about 1 cm for a TR photon energy of 10 keV [53], which implies that photons are primarily absorbed at the beginning of the drift region. Furthermore, charged particles uniformly ionize the gas, which results in the formation of an ionization trail. The electrons then drift toward the amplification region.

Amplification region

The cathode wire plane separates the drift and amplification regions. Electrons produced in the drift volume initiate an avalanche near the anode wires after passing through the cathode grid. The anode wire plane is situated between the pad plane and the cathode wire plane. The cathode wire plane enables independent drift velocity and gas gain adjustment by decoupling the drift field from the amplification region.

Readout pads

The segmented cathode pad plane is fabricated from thin Printed Circuit Boards (PCBs) glued onto a lightweight honeycomb and carbon fiber sandwich to ensure planarity and mechanical stiffness. Each PCB is segmented into 16 (12, for stack 2) pads along the z-direction (rows) and 144 pads in the direction of $r\phi$ along the anode wires (columns). Additionally, the pads are tilted, alternating between layers, improving the z-resolution during tracking.

Front-End Electronics

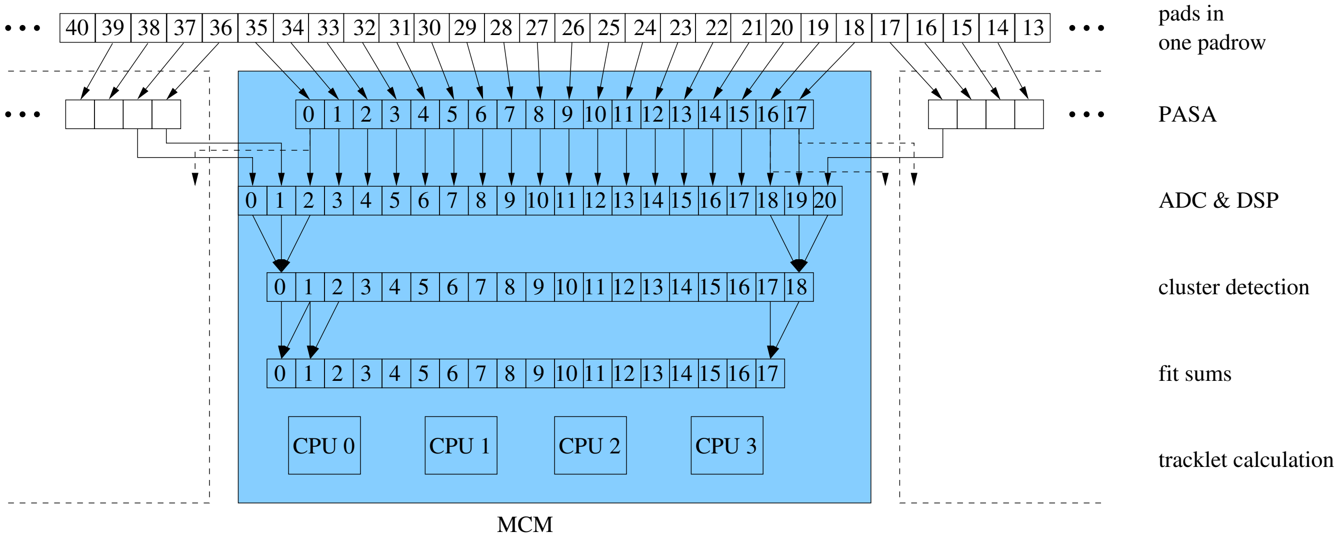
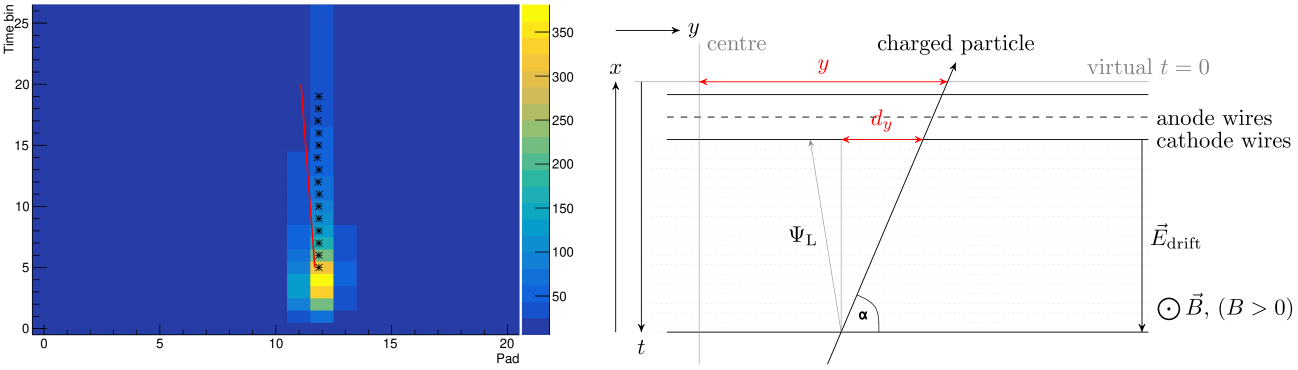


Figure 4.4: Pad connections to one MCM. Taken from [50].

The movement of electrons induces a positive signal on the cathode pads, amplified by a charge-sensitive Pre-Amplifier Shaper (PASA). The signals from 18 pads are connected to the PASA on one Multi Chip Module (MCM). The differential PASA outputs are fed into the Analog to Digital Converters (ADCs) of the Tracklet Processor (TRAP). The TRAP is a custom-designed digital chip that comprises cycling 10 bit ADCs for 21 channels and four two-stage pipelined Computer Processing Unit (CPU) with Hamming-protected instruction memories. Three excess ADCs channels receive signals from the two adjacent MCMs to prevent tracking inefficiencies at the MCM boundaries. An overview of the connections is shown in Figure 4.4. The MCMs are mounted on Readout Boards (ROBs) where 8 (6, for stack 2) ROBs cover a full detector chamber.

The signal is sampled in time bins of 100 ns inside the TRAP. The average time evolution of the signal is shown in Figure 4.3 (b), demonstrating a clear separation of the pion signal and electron signal. The first peak is the induced signal of the primary highly-relativistic particle at 0.5 μs for which ionization from both sides of the anode wires is measured. Afterwards, the signal of the drift electrons is a plateau since these travel along the electric field lines from the ionization tail. If TR photon was absorbed, the additional signal results in a second peak at 2.5 μs.

The local online tracking is carried out in parallel with the Front-End Electronics (FEE) of 65000 MCMs, each with data from 21 pads. As a first step, the signal from these pads is clusterized. A cluster is detected when the charge on three adjacent pads exceeds a configurable threshold and the center channel has the largest charge [54]. Next, a straight line fit is computed from these clusters via linear regression, see Figure 4.5 (a). The fit provides the local transverse position y, the deflection in the bending plane d_y, the longitudinal position z and the accumulated charge, see Figure 4.5 (b). The reconstructed values for y and d_y are corrected for systematic shifts caused by Lorentz drift (explained in Section 4.2.1) and the pad tilt. Finally, the information is packed into up to four 32 bit words for readout, which define the MCM header and word. A comprehensive overview of the data format is provided later in Figure 4.13 (a). Moreover, on the MCM for each tracklet, the charge of found clusters is summed within configurable time windows during local tracking.

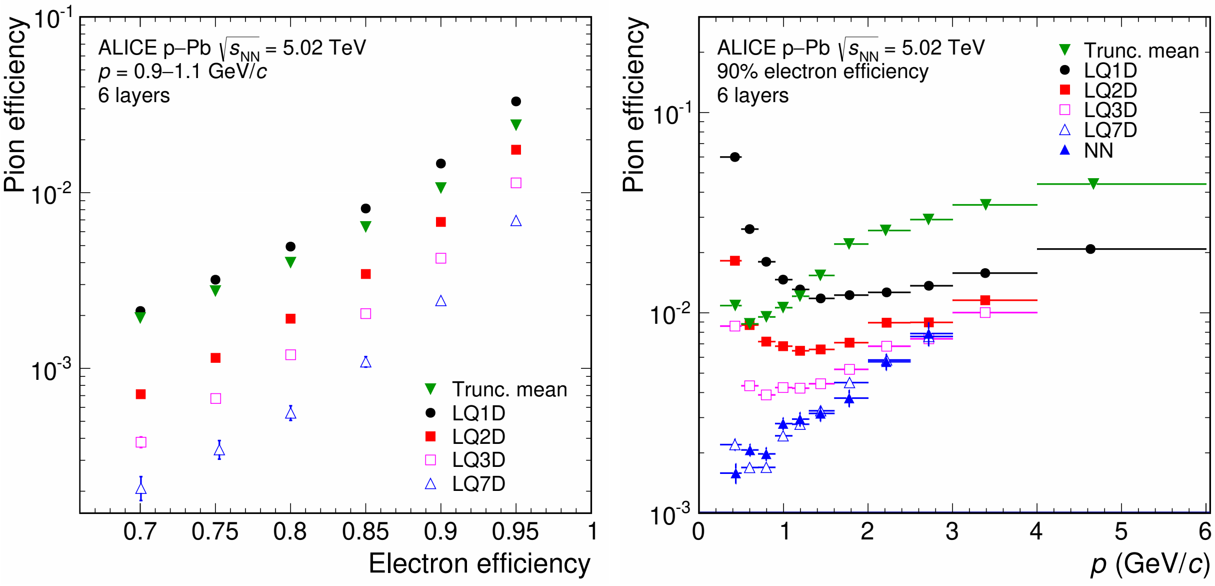


(a)

(b)

Figure 4.5: (a) Example tracklet in one MCM. The ADC data for 26 bins (100 ns each) from 21 channels are shown. The found clusters are marked as asterisks and the final tracklet, calculated as a straight line fit through the cluster, with a Lorentz correction as a red line. (b) The tracklet comprises the information on y, d_y, z and charge. The magnetic and electric field and the effect of the Lorentz angle Ψ_L are also indicated. Adapted from [50].

4.1.4 Particle Identification



(a) (b)

Figure 4.6: (a) The pion efficiency vs. the electron efficiency. (b) Momentum dependence of the pion efficiency for the truncated mean, LQ1D, LQ2D, and NN methods. The results are for 90% electron efficiency and tracks with signals in six layers in LHC Run 1 and 2. Taken from [37].

In order to perform precise electron identification, the temporal evolution of the signal is divided into slices. Slices are numbered from the readout-end farthest from the detector. These different slices cover separated parts of the pulse-height spectrum. For performance evaluation, two quantities are essential: the electron efficiency (ϵ_e) and pion efficiency (ϵ_π). Usually, the pion rejection factor is defined as the reciprocal of the pion efficiency. The pion rejection factor is a crucial figure of merit for the TRD, as it describes the pion suppression relative to the electron efficiency. Figure 4.6 (a) shows the pion efficiency as a function of the electron efficiency, displaying an almost logarithmic increase with increasing electron efficiency. Finally, Figure 4.6 (b) shows the pion rejection factor of different methods for different momenta. A thorough explanation of how this is calculated will be given in Section 5.3.1.

4.2 Calibration

The basic four calibration parameters for the TRD relevant for this thesis are time offset, drift velocity, Lorentz angle and gain calibration. Around $0.5\,\mu$s a peak, caused by the position of the anode wires and a falling edge at $2.8\,\mu$s due to the entrance window, is visible, see Figure 4.3 (b). The position of the peak provides the time offset. The distance between peak and edge is inversely proportional to the drift velocity. The plateau height is proportional to the gain. Although ionization electrons are drawn towards the anode wires by an electric field E, a magnetic field B, perpendicular to it, leads to a Lorentz force acting on the electron's drift direction. Knowledge of this deflection is necessary for the precise reconstruction of the tracklets [50].

4.2.1 v_{drift} and $E \times B$

The number of pads onto which a track is projected in the pad plane spreads out for lower transverse momenta. Additionally, the track incident angle a (in Figure 4.5 (b)) plays a significant role since smaller incident angles lead to more pads measuring a signal. Another contributor to this spread is the Lorentz angle. Negatively (positively) charge particles drift along (opposite to) the track inclination independent of the magnetic field [50]. One can see how this spread, due to the particle's charge, is corrected in Figure 4.7. The gain and the v_{drift} calibrations are based on new developments. The residual angle of the track and the attached tracklet $\delta\alpha$ is measured as a function of the track impact angle for each chamber. A straight line fit gives the Lorentz angle and an effective v_{drift}^{eff} [31].

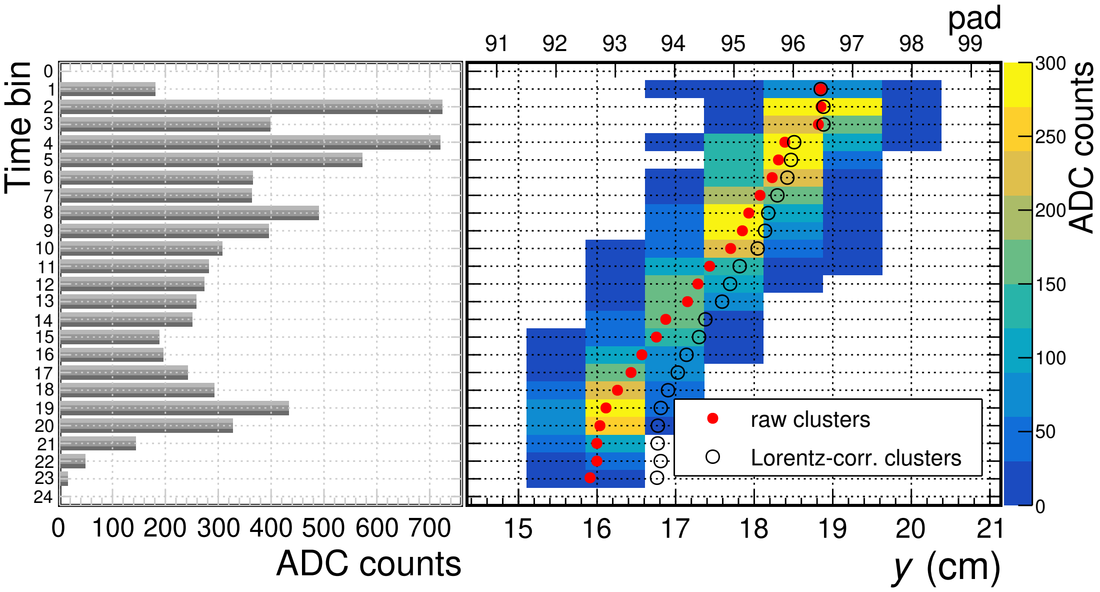


Figure 4.7: Signal produced by a positively charged particle ($p_T = 0.5\,\mathrm{GeV/c}$). Left: Total charge per time bin used for electron identification. Right: Ionization signal vs. pad number and time bin. The cluster positions are shown as reconstructed from the charge distribution (raw clusters) and after correction for the $E \times B$ effect (Lorentz-corr. clusters). Taken from [50].

4.2.2 Gain

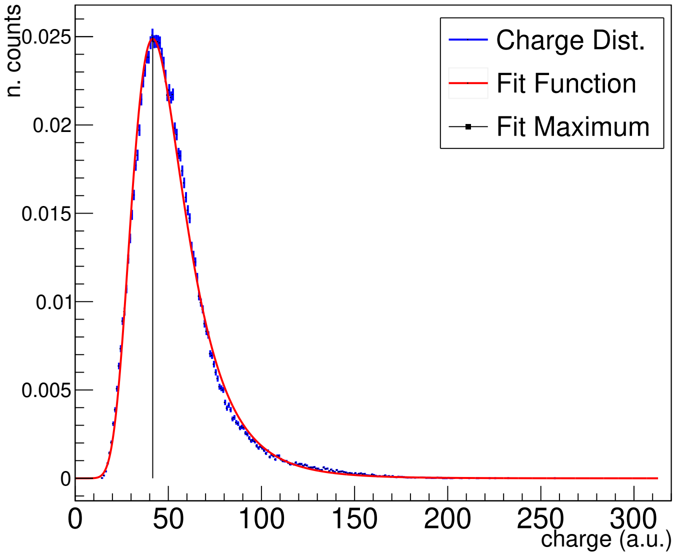
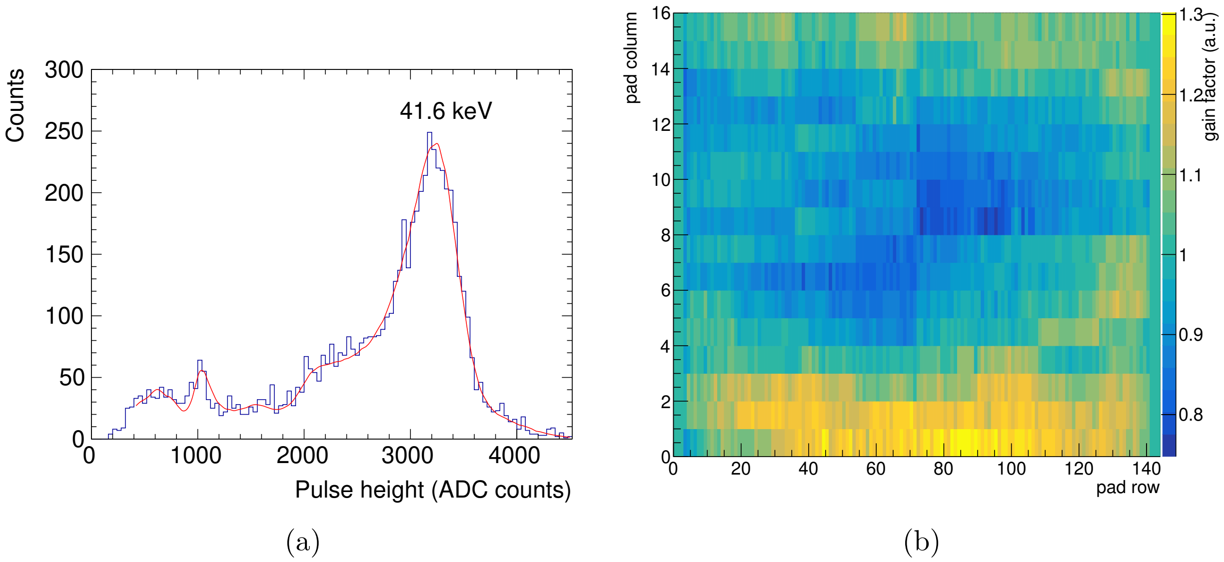


Figure 4.8: Total integrated charge divided by path length of one chamber including the fit [55] in Run 3 in pp collisions at $\sqrt{s_{\mathrm{NN}}} = 13.6\,\mathrm{TeV}$.

The gain calibration factor for each chamber is determined by histogramming, the total integrated charge divided by the path length and by extracting the mean of this distribution. A fit is performed as suggested in [56] by the convolution of a Gaussian and a Landau with an exponential distribution to determine the mean.

4.2.3 Krypton

A gain uniformity better than 5% is essential for suitable electron identification [58]. Hence, pad-by-pad calibration is performed with a ^{83}Rb source, which decays into gaseous Kr occupying the isometric state $^{83\mathrm{m}}$Kr with an excitation energy of 41.6 keV [50]. The Krypton is injected into the gas system of each chamber once or twice per year, where a pulse-height spectrum for each pad is recorded. Each spectrum is fitted with a reference spectrum, as shown in Figure 4.9 (a). Afterwards, the relative position of the peak is converted and normalized to a gain factor within each chamber. Figure 4.9 (b) shows a complete map of these gain factors for an individual chamber. The gain factors fluctuate 30% around the chamber mean. The data is from the dedicated Krypton runs in September 2021 and May 2022. The integrated charges have to be corrected with these gain factors.



(a) (b)

Figure 4.9: (a) Pulse height spectrum accumulated for one pad during the Kr-calibration run. The smooth solid line represents the fit with a reference distribution from which the gain is extracted [50]. (b) Relative pad gains for one chamber [57].

4.2.4 Time-Offset

Individual time offset corrections have to be applied to correct for differences in the timing of the MCMs and the resulting position resolution during tracking. Currently, the assumption is that these correction factors are absorbed in the global alignment of the experiment. However, this will still need to be confirmed and monitored. Monitoring is achieved by fitting the amplification peak of the pulse-height spectrum with a Landau curve. Then, the maximum value can be extracted and a nominal correction factor is calculated [50]. The pulse-height spectrum and the fit for one chamber is shown in Figure 4.10.

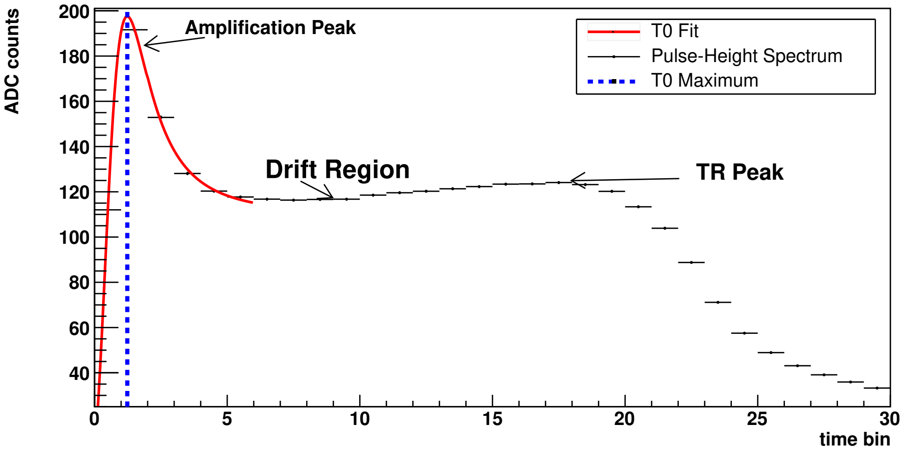


Figure 4.10: t_0 calibration for one individual chamber [59] in Run 3 in pp collisions at $\sqrt{s_{\mathrm{NN}}} =$ 13.6 TeV.

4.3 Operation in LHC Run 3 and 4

4.3.1 Tracking

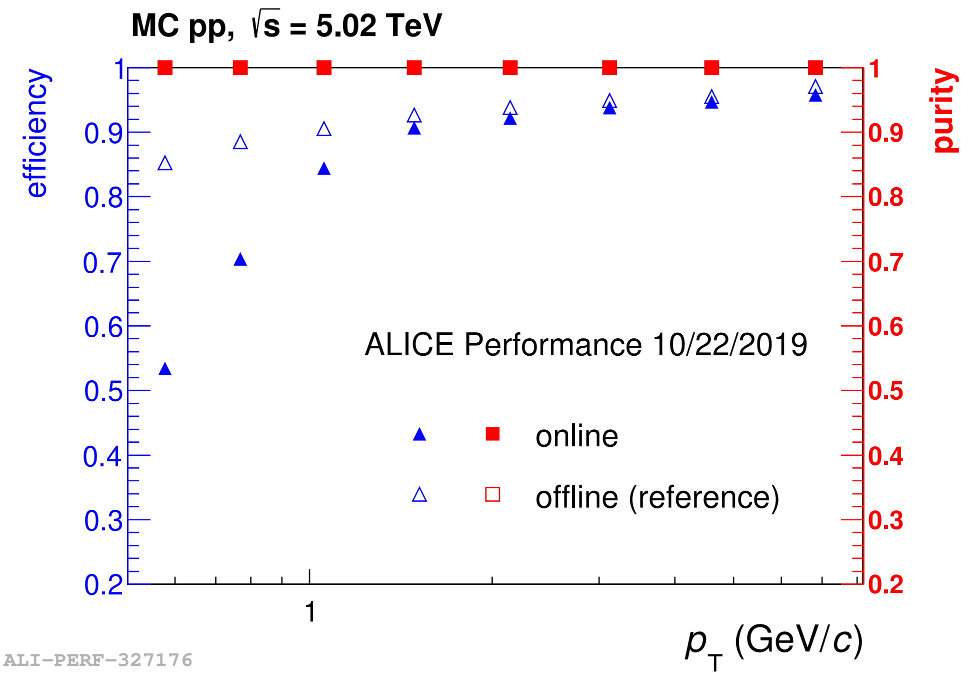


Figure 4.11: TRD acceptance corrected efficiency for at least two tracklets attached to the track as a function of track p_T for the offline reconstruction based on offline tracklets calculated from clusters and for the online tracking based on online tracklets which are optimized for high-p_T tracks for triggering. The purity is the fraction of tracks without any fake tracklets attached.

The TRD plays a crucial role in tracking through various mechanisms:

1. It increases the lever arm by approximately 70 cm, significantly improving the momentum resolution for high-p_T tracks.
2. It enhances the accuracy and effectiveness of attaching clusters from larger radius detectors to propagated tracks, especially for TOF the purity of attaching clusters increases.
3. It serves as a reference to obtain correction maps for TPC distortions that occur due to the accumulation of space charge at high interaction rates, so-called space-charge distortions [50].

The fraction of global tracks to matched TRD is shown in Figure 4.11. One can see that the online tracking performance is comparable to the offline tracking efficiency without any fake tracklets for $p_T \geq 1.5\,\mathrm{GeV/c}$. As expected, the efficiency decreases for smaller momenta.

4.3.2 Calibration

As of writing this thesis, only the Lorentz correction and the Krypton calibration are fully implemented in O^2. For the Lorentz correction, data is aggregated online for 15 min and new correction factors are applied. The time offset calibration, for now, is only available for monitoring and the gain calibration was close to being finalized. These calibrations are expected to be fully implemented in the O^2 framework soon.

4.3.3 Readout

The CRUs receive data directly from the FEE via 1044 upgraded Direct Detector Links (DDLs). In total, 36 CRUs (two per TRD SM) are in use and housed in twelve FLPs. The CRU controls the readout process of the detector and receives, buffers and formats the data for O^2. All CRUs are connected to the Local Trigger Unit (LTU) to receive trigger information and to signal a detector busy status to the CTP. Each CRU determines an individual busy status contribution depending on the status of the readout of the connected FEE links. The CTP combines the busy status contributions from all CRUs to determine the global busy status of the TRD detector.

The following points describe the process of acquiring an event sequentially, summarizing Figure 4.12:

- The CTP sends a trigger to the FEE and the CRUs in parallel via legacy and new trigger distribution networks, respectively.
- Upon the arrival of the trigger, the FEE starts recording data while the CRUs open a time window to wait for the input links to send all acquired data. The CRUs store all information from the trigger message and send their busy status contribution to the CTP.
- When the FEE has acquired and processed the data, it appends end markers and starts shipping it via optical links. The CRUs record the received data and release their busy status confirming the end marker's reception.
- The CRUs pack the data into packets and equip them with headers before shipping them to the readout system on the FLPs. The headers contain information needed for data reconstruction.

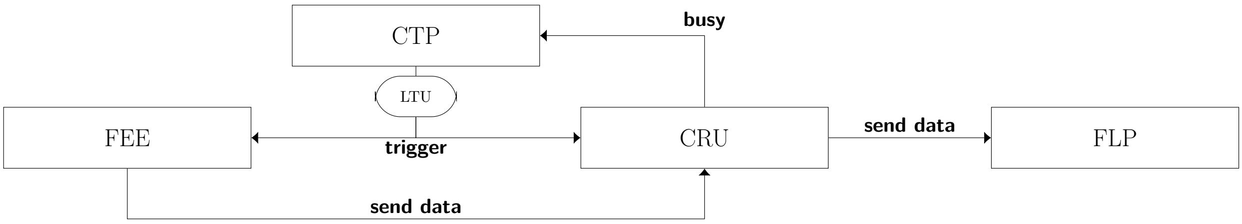


Figure 4.12: Sketch of the readout scheme for the TRD in Run 3. Derived from [31].

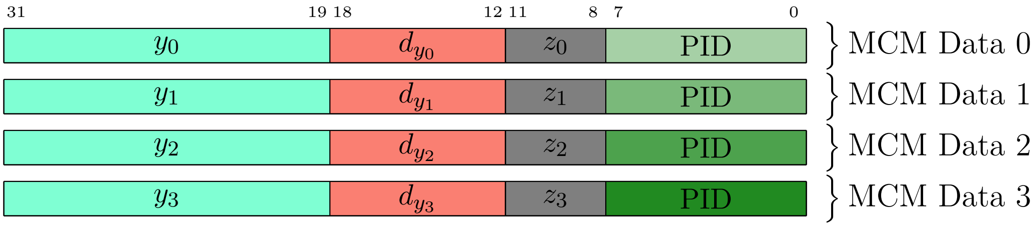
There are two types of triggers: physics data (tracklet-mode) and calibration data (digit-mode). In tracklet-mode physics data is recorded, while in digit-mode the complete raw data is read out from the FEE. Tracklet-mode was developed especially for Run 3 and 4 since the read out of a digit takes too long for a trigger. A calibration trigger is sent every 1000 events, allowing for accurate monitoring and quality assurance.

4.3.4 Data Format

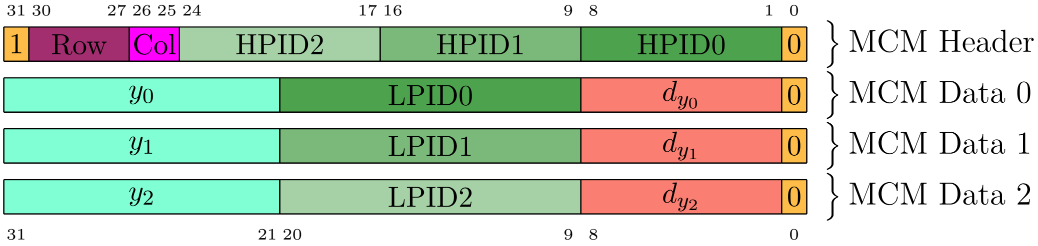
Only tracklets are read out (tracklet-mode) to achieve a high event-readout-rate in Run 3. The maximum data volume per MCM is four words of 32 bit each. In Run 1 and 2, each MCM processed and transmitted up to four tracklets, as seen in Figure 4.13 (a). The tracklet-mode was only used to derive a trigger on high-p_T particles and electrons, while the analysis was based on the reconstructed raw data. The PID field was a truncated float number containing the electron likelihood translated from the measured charge via a Lookup Table (LUT). Due to the high data rates in LHC Run 3 and 4, saving the raw data is no longer feasible, as explained in Section 3.1. For the TRD, the problem was that reading out the raw data would take too long for every trigger. That is why the accumulated charge information is now written into the tracklet word and PID is deferred to the software. Theoretically, the same format as in Run 1 and 2 could have been kept, but this would allow only for a single charge slice, which is not precise enough for PID. However, even in most central Pb−Pb collisions, a track density of four tracklets is rare (below 1%) [31]. Hence, in Run 3, only three tracklets are allowed and the freed-up word is used as a header, as seen in Figure 4.13 (b). This header allows each tracklet to have three charge slices. The tracklet positions y_i and slope d_i is now stored with higher precision. For the PID information, 20 bit per tracklet are available, which will be used to store integer-precision information from three time slices, see Figure 4.13 (c). The MCM header combined with the corresponding MCM data word (i.e., data coming from the same CPU) to retrieve the accumulated charge information. Q_0 (7 bit) covers the amplification peak, Q_1 (7 bit) the TR peak and Q_2 (6 bit) the plateau of the pulse-height spectrum in Figure 4.3 (b). On the MCMs, the accumulated charge of each slice is divided by a scaling factor to shift the information into the respective value range. Values exceeding this range get truncated to the respective maximal value. In practice, $Q_{0,1}$ are calculated directly on the hardware, while Q_2 is computed in the software. The relevant charge slice configuration of the pulse-height spectrum for this thesis is given in Table 4.1.

Table 4.1: TRD charge slices configuration. Each bin is 100 ns.

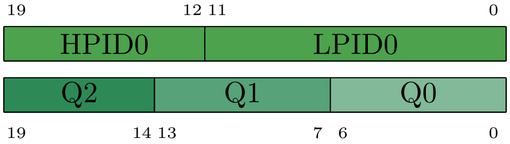
Slice	Start Bin	End Bin	Scale Factor	Description	Precision (bit)
Q_0	1	6	16	Amplification peak	7
Q_1	14	20	16	TR peak	7
Q_2	7	13	16	Plateau region	6



(a) LHC Run 1 and 2.



(b) LHC Run 3 and 4.



(c) Charge information conversion in LHC Run 3 and 4.

Figure 4.13: Comparison of the data formats with up to four tracklet words per MCM (a) and the new data format limited to three tracklets per MCM with 20 bits available for each tracklet (b). Additionally, in (c) an exemplary conversion of the charge information from the MCM header and word is shown. Adapted from [31].

4.3.5 Chamber Status

After Run 1, it was revealed during the disassembly of a supermodule that a couple of broken filter capacitors (4.7 nF, or a few 2.2 nF) were the cause of strange current behaviors (trending vs. time). Hence, in the last five SMs these capacitors were left out. Nine TRD supermodules were extracted during the LS2. In those supermodules, the capacitors in the high-voltage distribution were removed via non-invasive methods. As the capacitors were meant to buffer high-charge deposits in the chambers, their removal results in larger induced common-mode signals on readout pads in the same voltage segment as a function of the particle multiplicity. Based on the measured local charge deposits, this effect will be corrected at the software level.

Thanks to the repair work, the acceptance of the detector could be vastly increased compared to Run 1, see Figure 4.14. The figure shows the average number of tracklets attached to a track. The hole comes from the missing stack twos to minimize material before the PHOS detector.

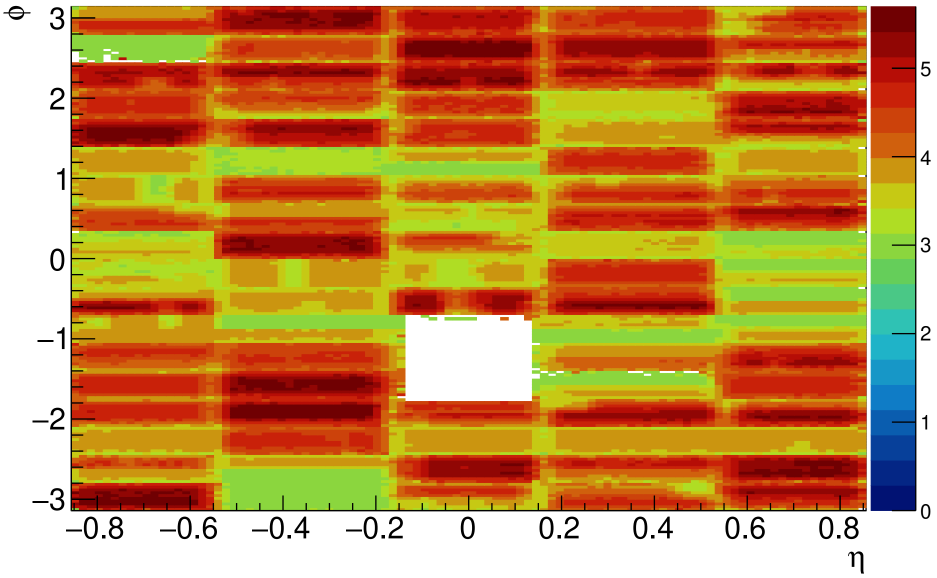


Figure 4.14: Average number of tracklets per track in each stack [60] in Run 3 in pp collisions at $\sqrt{s_{\mathrm{NN}}} = 13.6\,\mathrm{TeV}$.

5 Analysis

While the previous chapters lay the groundwork for electron identification with the TRD, this chapter describes the electron identification methods and the performance of this tracklet-based PID approach. First, the data used in the analysis are examined in Section 5.1. Then the performance of the electron identification is evaluated in Section 5.3 using a multidimensional likelihood approach and machine learning algorithms. The obtained results are presented in the subsequent Chapter 6.

5.1 Data Skimming

The data used in this thesis is from LHC Run 3, i.e., pp collisions at $\sqrt{s_{\mathrm{NN}}} = 13.6\,\mathrm{TeV}$. The exact specifications for this dataset are shown in Appendix A. Selection criteria are used to identify clean electron and pion samples. The Bethe-Bloch-Aleph parametrization for the energy loss of different particles in the TPC was taken from the CCDB, given in Appendix C. The TPC resolution was estimated to be 9% since not the full calibration was yet performed during reconstruction. The TOF timing resolution is roughly 200 ps due to the collision time uncertainty since the data was not yet matched to the FIT detector. The exact selection criteria used are given in Appendix Table 8.1. A χ^2 criterium for the global track fit is used to ensure a minimal amount of incorrectly matched tracklets to tracks. Additionally, at least four tracklets from not noisy MCMs must be attached to a track, and these tracklets cannot cross a pad-row and have neighboring tracklets. After these selection criteria, most particles ($\geq 99\%$) are pions. In total, roughly 4.2×10^7 pions and 3.0×10^5 electrons over the whole momentum range were identified. (ITS-)TPC-TRD(-TOF) tracks were analyzed to extract the necessary data for the following analysis.

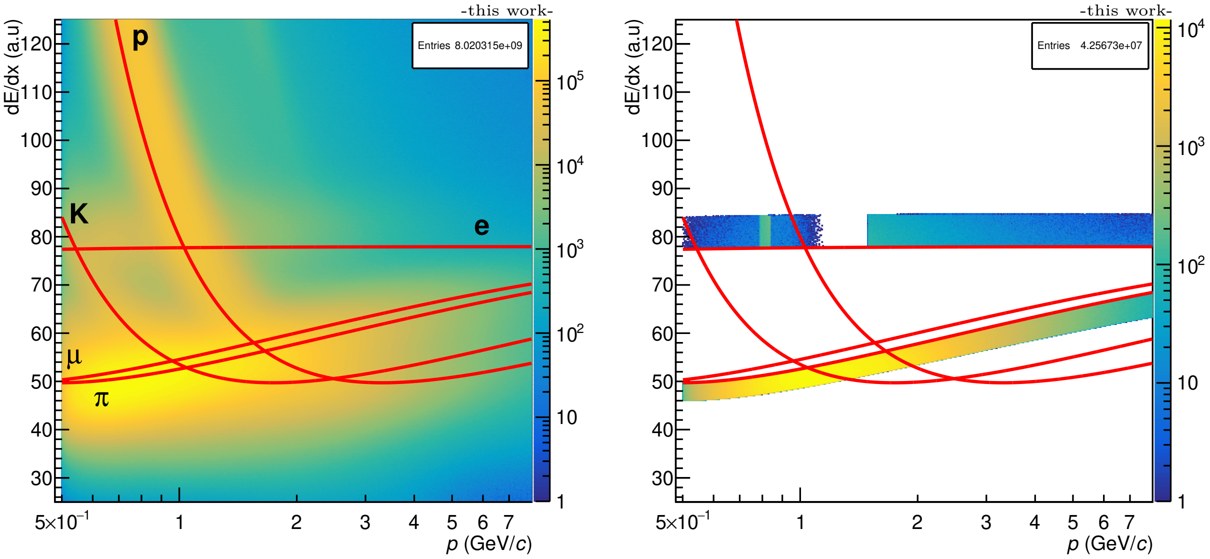


Figure 5.1: TPC dE/dx before (left) and after (right) selection criteria.

Figure 5.1 shows the resulting TPC dE/dx distributions as a function of p before and after the selection criteria. The expected energy losses, parameterized by the Bethe-Bloch-Aleph curves, are also drawn for different particles. At low momenta, there are few entries due to the low tracking efficiency and a minimum of p due to the magnetic field. Where the kaon and proton lines intersect with the electron line, the separation power of the TPC is too low. Hence, only tracks also matched to TOF could be used. At $p \approx 0.8\,\mathrm{GeV/c}$, the TPC separation is high enough. In the range $p \approx 1.0\,\mathrm{GeV/c}$ to $1.3\,\mathrm{GeV/c}$, the TOF separation power was insufficient to separate protons from electrons. Beyond that, the TPC separation power and a graphical cut are used. The proton and kaon curves intersect the pion curve at $p \approx 1.0\,\mathrm{GeV/c}$ to $2\,\mathrm{GeV/c}$ since only the Bethe-Bloch-Aleph parametrizations are used. Otherwise, there would be no data in this interval. This leads to substantial proton/kaon contamination in this interval. The muon curve follows the pion curve closely. Thus no cut could be made, leading to proportional muon contamination.

5.2 Exploratory Data Analysis

5.2.1 Charge correction

The correction of the charges includes the gain factor from the krypton calibration from Section 4.2.3 and gain calibration from Section 4.2.2. Additionally, the charge was normalized to the track length by applying the following:

$$Q' = Q \times \frac{1}{\sqrt{\left(\frac{dx}{dx}\right)^2 + \left(\frac{dy}{dx}\right)^2 + \left(\frac{dz}{dx}\right)^2}} = Q \times \frac{1}{\sqrt{1 + \tan(\phi)^2 + \tan(\lambda)^2}}, \tag{5.1}$$

where $\tan(\phi)$ is the local inclination in the xy-plane and $\tan(\lambda)$ is the local inclination in the xz-plane.

Figure 5.2 shows the average total accumulated (un-)corrected charge as a function of the track's η and ϕ measured at the entrance of the drift chamber. An apparent increase for higher $|\eta|$ is visible on the left side. This is expected since tracks at higher $|\eta|$ have, on average, a larger slope in the drift chamber of the TRD and thus produce more clusters leading to a higher average charge. On the right-hand side, the corrected charges show no correlation by using the correction in Equation 5.1, which normalizes the charges across the whole $|\eta|$ range. This indicates that the correction with Equation 5.1 works well. In addition, the segmented structure of the trd is visible since tracklets have, on average, a smaller charge at the edges of the readout chambers. One can see the 18 supermodules in the ϕ-direction and the five stacks per supermodule in the η-direction. The PHOS hole from Section 4.3.5 is not visible in this plot since, very rarely, tracklets from the adjacent readout chambers can be attached to a track going through the hole.

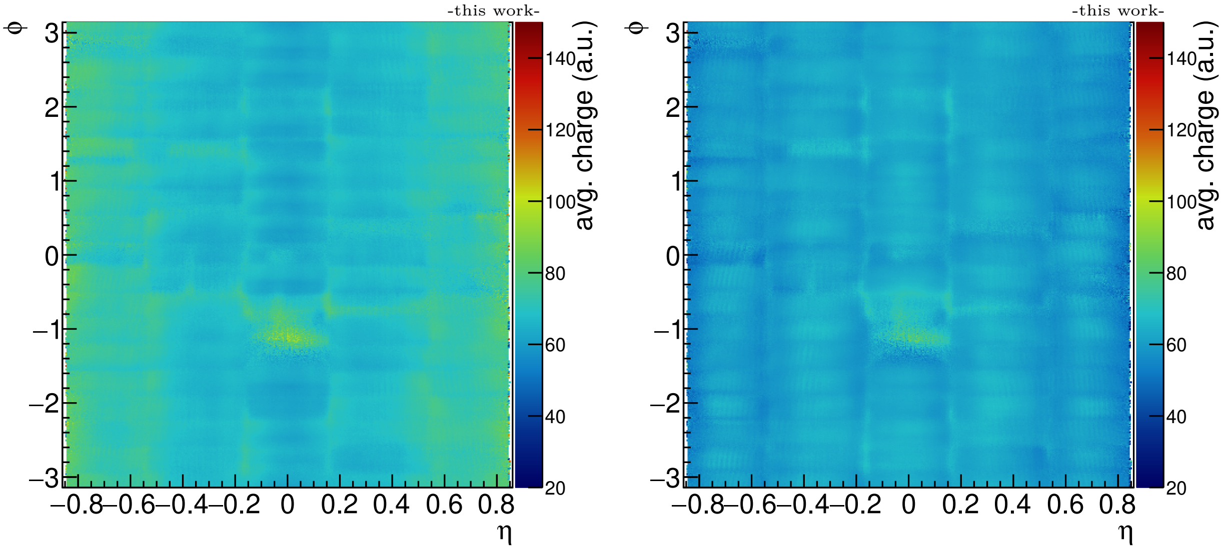
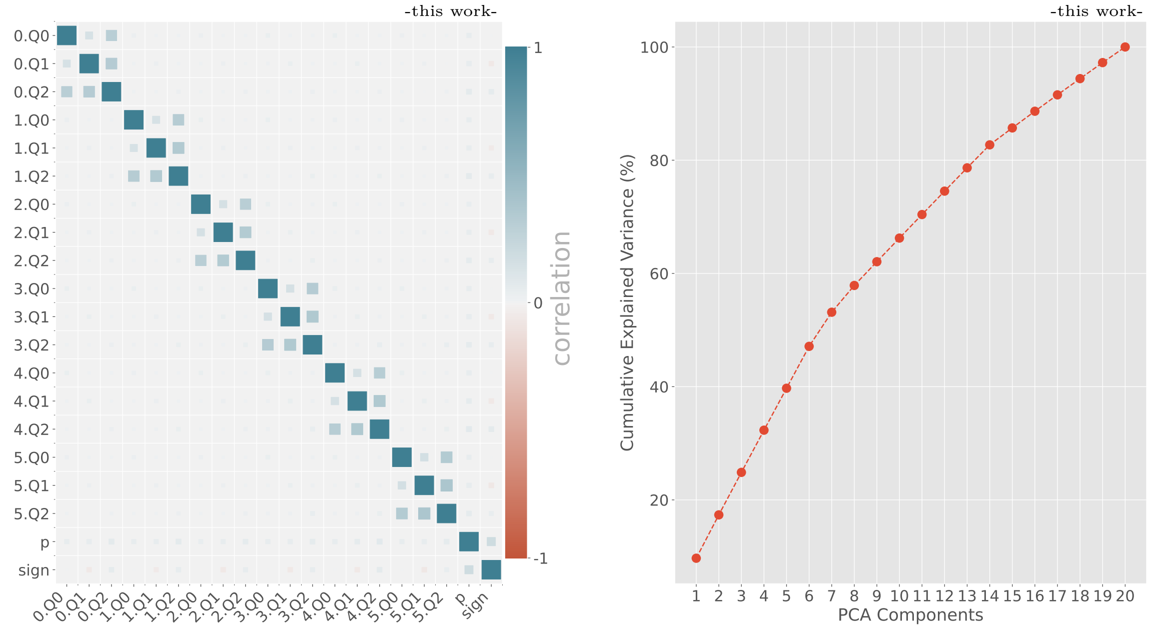


Figure 5.2: Average total accumulated uncorrected (left) and corrected (right) charge as a function of η and ϕ for all momenta.

5.2.2 Correlation and principal components

The extracted features from the datasets are the sign of the particle charge, the average momentum p of the track in the TRD and the corrected charge $i.Qj$ from layer i and slice j. The correlation of the features is shown in Figure 5.3 (a). The size of the marker and the color indicate the degree of correlation. As expected, there is a high correlation of the charges within each layer but no correlation between layers. Of course, the reason for the high correlation of charges within a layer is that the measured charge comes from the same pulse-height spectrum. In addition, the correlation of Q_0 and Q_1 for all layers is lower since Q_1 covers the TR peak, an independent physical effect.

Principal Component Analysis (PCA) is used to check if the amount of information the features offer is beneficial. PCA is a statistical technique used to reduce the dimensionality of a dataset while retaining most of its variability. It works by creating new variables, called principal components, which are linear combinations of the original variables in the dataset. The first principal component is a linear combination of the original variables that explains the largest variability in the dataset. Each subsequent principal component explains the remaining variability not explained by the previous components. The PCA for the data can be seen in Figure 5.3 (b). From this, it can be inferred that 11 features account for $\approx 70\%$ of the information in the dataset. No apparent kink in the plot indicates that each feature contributes valuable information and should not be discarded. However, PCA should not be used in isolation to discard features [61]. This is particularly important for machine learning models to avoid overfitting, which is explained in more detail in Section 5.3.3.



(a) (b)

Figure 5.3: (a) The correlation in the data of the features. ($i.Qj$ means layer i and slice j) (b) Principal components of the features. No feature names can be given to the components after the linear combination. Shown is the whole dataset, i.e., data from pions and electrons

5.2.3 Charge distributions

Figure 5.4 shows the normalized total accumulated charge distributions of the identified electrons and pions in one layer for positively and negatively charged particles in one momentum interval. Due to the $E \times B$ effect, the distributions differ slightly and the distribution for negatively charged particles is sharper than that for positively charged particles. For both charges, the distributions of the electrons are separated and shifted to higher values, as the electrons deposit, on average, more charge in the TRD than pions. This is due to the specific energy loss by ionization and the additional TR contribution in this momentum range. Additionally, as suggested in Section 4.2.2, a fit is performed to smooth out the distributions. The distributions for the other momentum intervals are shown in Appendix D.

Figure 5.5 shows the normalized charge distributions of the different slices of negatively charged electrons and pions in one layer for one momentum interval. The distributions for other momentum intervals and positively charged particles are shown in Appendix D. For $Q_{0,1}$, there are entries at lower charges which stem from a binning effect with the scaling factor in Section 4.3.4. These were excluded from the fits.

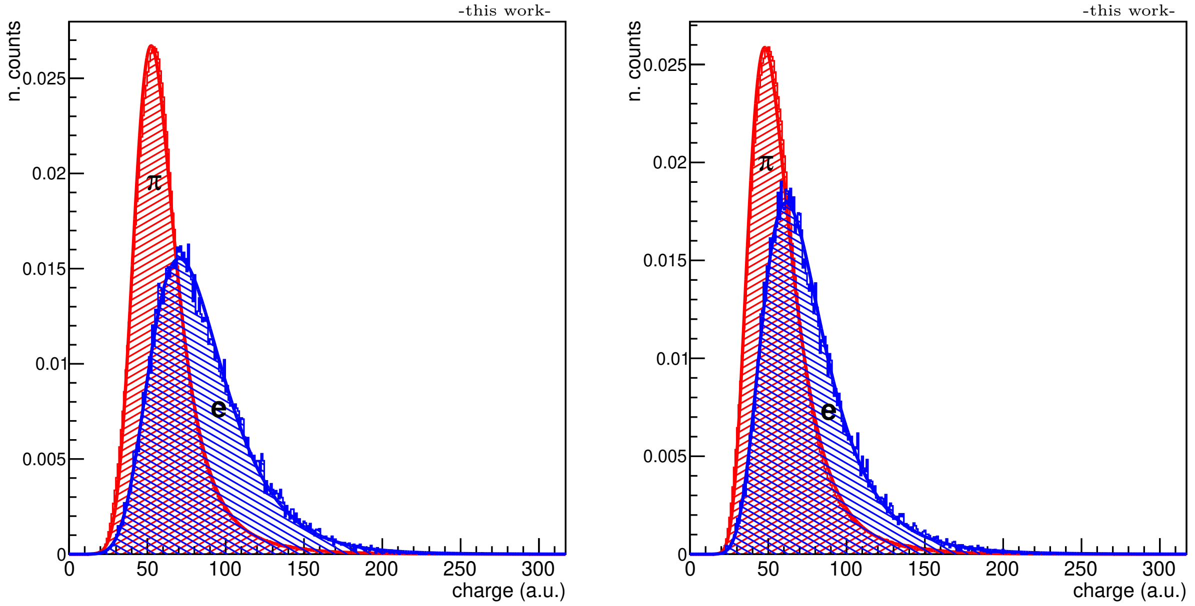


Figure 5.4: Total accumulated charge of electrons (blue) and pions (red) in layer 0, obtained from tracklets of tracks for (left) positively charged particles and (right) negatively charged particles for $p = 1.0\,\mathrm{GeV/c}$ to $2.0\,\mathrm{GeV/c}$, including the fits.

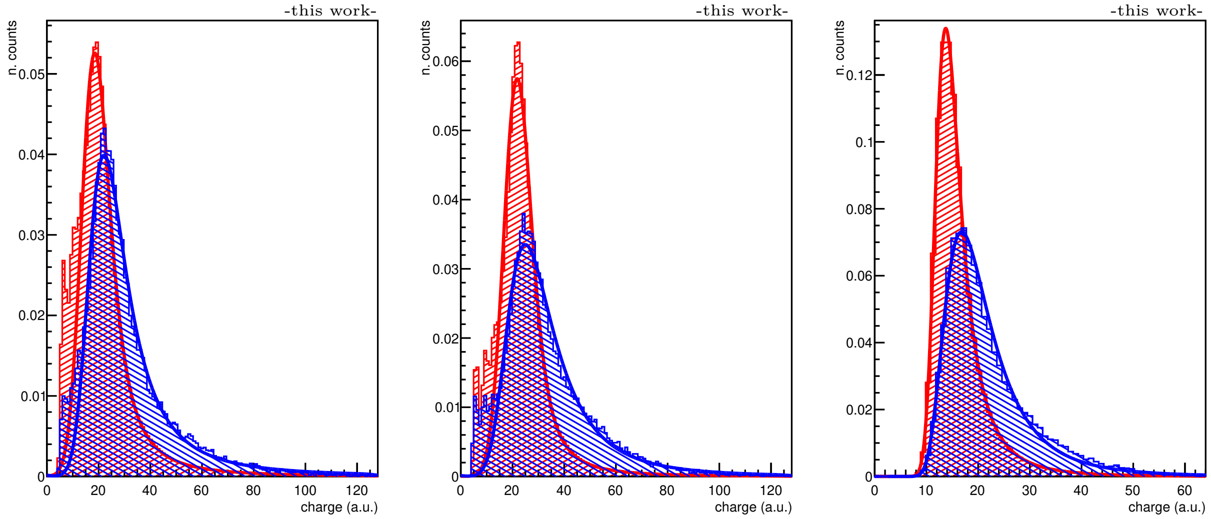
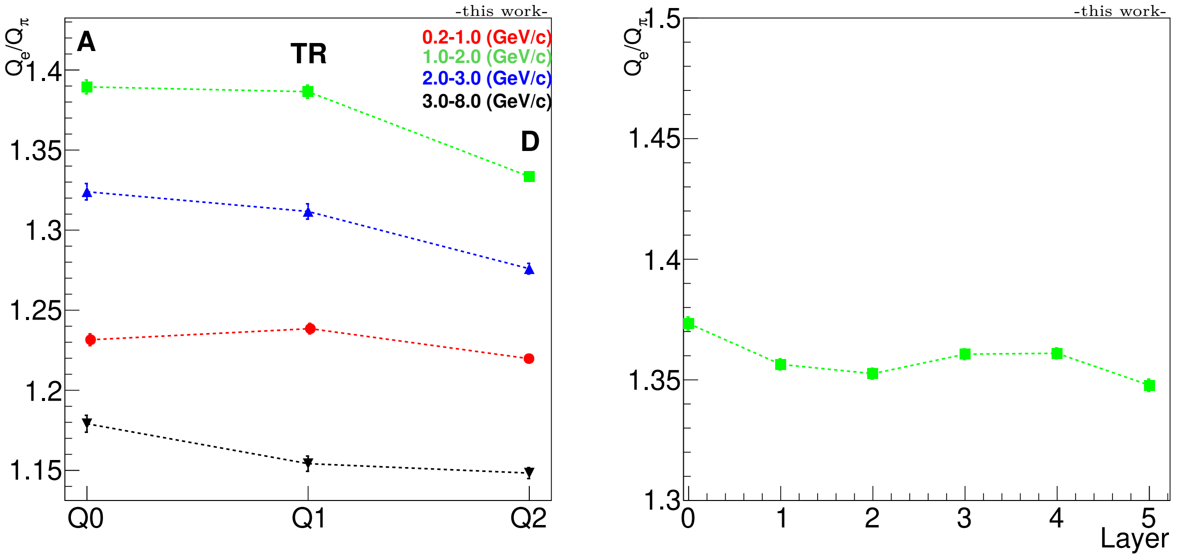


Figure 5.5: Charges ($Q_{0,1,2}$, left to right) of electrons (blue) and pions (red) in layer 0, obtained from tracks for negatively charged particles for $p = 1.0\,\mathrm{GeV/c}$ to $2.0\,\mathrm{GeV/c}$, including the fits.

The ratio of the different charge slices for different momenta is shown in Figure 5.6 (a). As expected, Q_0 and Q_1, as they cover the amplification (A) and TR peak (TR), have the most crucial information and Q_2 (D), which covers the plateau, has the least. Up to $p = 1\,\mathrm{GeV/c}$, there is no transition radiation, leading to a lower ratio. Since TR for electrons starts at around $p = 1\,\mathrm{GeV/c}$, the ratio is the highest. Afterwards, the ratios are lower due to data contamination due to the relativistic rise at higher p. The ratio of the mean total accumulated charge for pions to that of electrons, shown in Figure 5.6 (b), fluctuates little as a function of layers, indicating that each layer holds the same information for separation. Although layer 0 is the highest, possibly due to secondary particles leading to a high multiplicity which can cause clusters to be added incorrectly (i.e., causing a higher average charge).



(a) (b)

Figure 5.6: (a) The most-probable value of the accumulated charge ratio for different momenta as a function of the slice number, including the uncertainty. (b) The ratio of the most-probable total accumulated charge for pions and electrons versus layer number, including the uncertainty.

5.3 Particle Identification

5.3.1 1-dimensional Likelihood

After normalization, both distributions (i.e., the fits) in Figure 5.4 describe the conditional probability function $P(Q|k)$, which gives the probability of charge Q given particle k. In the 1-dimensional case, Q is defined as $Q = Q_0 + Q_1 + Q_2$, as explained in Section 4.3. Bayes' theorem relates $P(e|Q)$ and $P(Q|e)$ as follows:

$$P(e|Q) = \frac{P(Q|e)P(e)}{P(Q)} \approx P(Q|e)\frac{P(e)}{P(e) + P(\pi)}. \tag{5.2}$$

Since pions are the main source of background in hadron collisions, only they are considered. However, in pp and Pb−Pb collisions, the pions have a much higher prior probability compared to the one of the electrons, leading to $P(e|Q) \xrightarrow{P(e) \ll P(\pi)} 0$. Instead, the electron likelihood $L(e|Q)$, ignoring the priors, is defined as [62]:

$$L\left(e|Q\right) = \frac{P\left(Q|e\right)}{P\left(Q|e\right) + P\left(Q|\pi\right)}. \tag{5.3}$$

For each layer i with charge deposition Q_i, the likelihood L_i is:

$$L_i\left(e|Q_i\right) = \frac{P_i\left(Q_i|e\right)}{P_i\left(Q_i|e\right) + P_i\left(Q_i|\pi\right)}. \tag{5.4}$$

The combined electron likelihood is then defined as:

$$L\left(e|\vec{Q}\right) = \frac{\prod_i^6 P\left(Q_i|e\right)}{\prod_i^6 \sum_j P\left(Q_i|j\right)} = \frac{\prod_i^6 L_i(e|Q_i)}{\prod_i^6 L_i(e|Q_i) + \prod_i^6 (1 - L_i(e|Q_i))}, \tag{5.5}$$

where $\vec{Q} = Q_0, \ldots, Q_5$ are the total accumulated charges from each layer i and j denotes different particle species, assuming the likelihoods are not correlated. This is a reasonable assumption, as explained in Section 5.2.2. The electron likelihood is computed via Equation 5.4 as a lookup table of the electron and pion distribution for all layers, respectively. This method is also referred to as LQ1D. Figure 5.7 shows the resulting LUTs for all layers in one momentum interval.

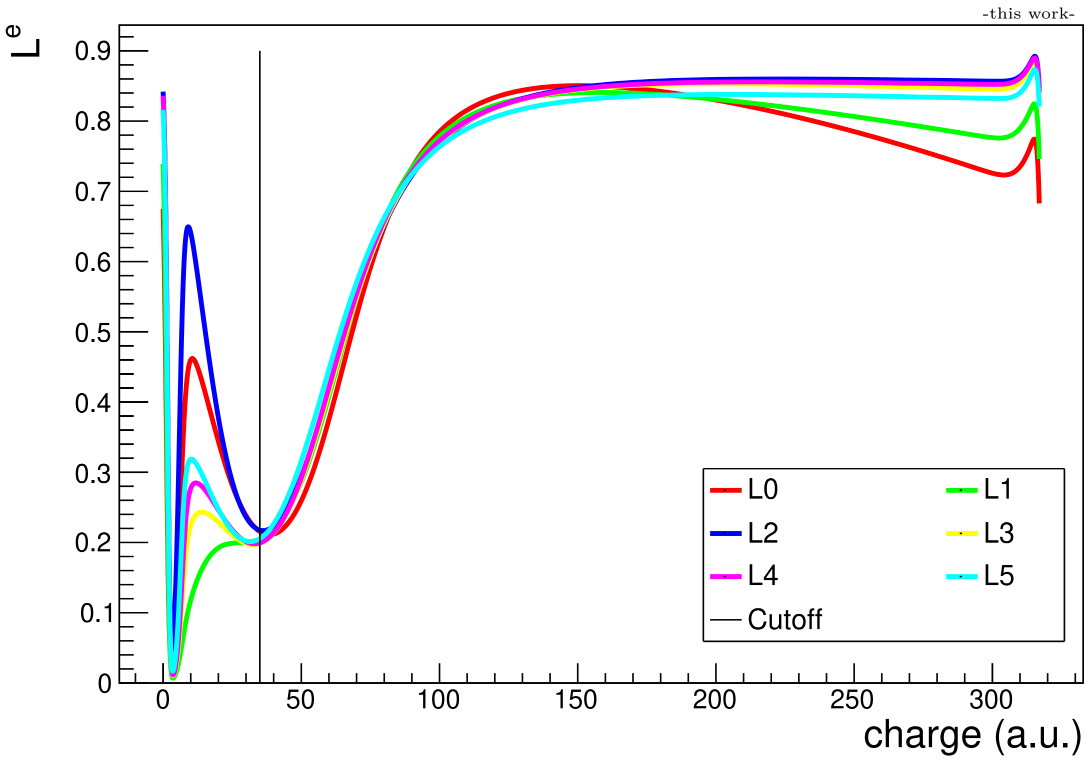
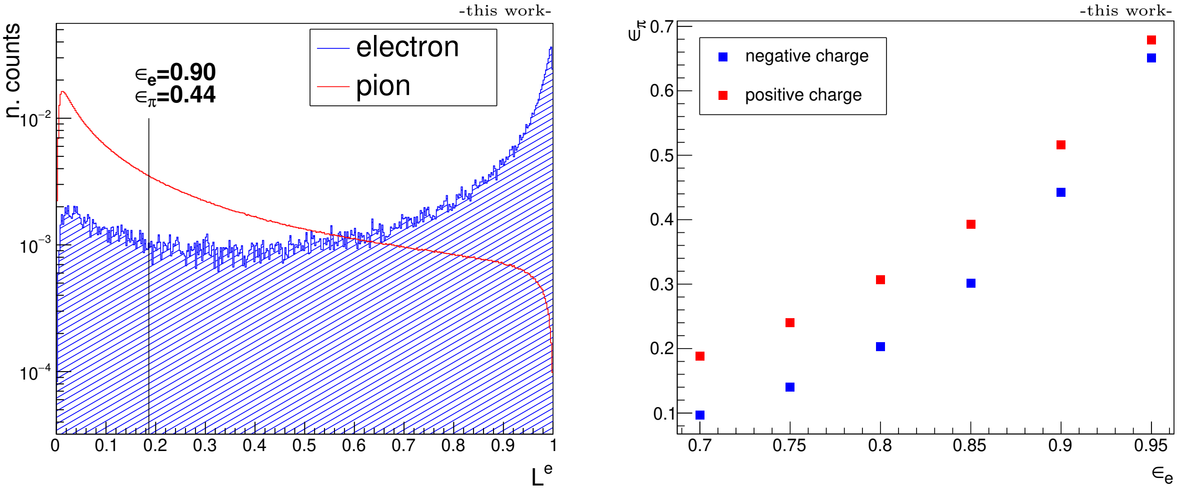


Figure 5.7: LUTs for the total accumulated charge of negatively charged particles with a line marking the cutoff for $p = 1.0\,\mathrm{GeV/c}$ to $2.0\,\mathrm{GeV/c}$.

In Figure 5.7 an apparent rise in electron likelihood is observable since electrons produce, on average, more charge in the TRD. A cutoff value is used to avoid strong fluctuations of the fits at low charges, setting the electron likelihood to zero. At higher charges, these fluctuations do not pose a problem due to the small prior probability of getting these high charges. It is clear that, particularly in the region of the steep rise, the LUTs are very similar. The difference in the different layer come from slight variations in the charge distribution, visible in Figure 5.6 (b). The distributions for the other momentum intervals are given in Appendix D.



(a) (b)

Figure 5.8: (a) Electron likelihood distribution for the total accumulated charge of negatively charged particles with a black line marking $\epsilon_e = 90\%$ for $p = 1.0\,\mathrm{GeV/c}$ to $2.0\,\mathrm{GeV/c}$. (b) Corresponding pion efficiency versus electron efficiency for positively and negatively charged particles.

Afterwards, the combined electron likelihood is calculated using Equation 5.5 and the electron likelihood distribution is analyzed to assess the pion rejection factor. Figure 5.8 (a) shows the electron likelihood distribution for negatively charged particles in one momentum interval for the identified electrons and pions. A clear peak at $L^e = 1$ for the electrons is visible, while the pions are mostly found around $L^e = 0$. This meets expectations from Run 2 data, coming from clean electron samples. A rise of the electron curve close to zero could indicate that the purity of the electron sample is not high enough. The quickly falling edge of the pion curve demonstrates good separability at high L^e. Based on this electron likelihood distribution, the performance at different ϵ_e can be assessed. In practice, the normalized electron likelihood is summed from 1 stepwise down until the interval, where approximately $x\%$ of the electron likelihood is integrated, which defines t_e^x. From this threshold, the pion efficiency is calculated by summing up the intervals to 1 for the pion distribution. Usually, the actual threshold is slightly undershot due to the binning, but this effect is minuscule if the binning is granular enough. Figure 5.8 (b) shows the pion efficiency at different electron efficiencies for positively and negatively charged particles. As expected, the pion efficiency increases with increasing electron efficiency. In addition, a clear difference in the pion efficiencies for negatively and positively charged particles due to $E \times B$ effect is visible. At $\epsilon_e = 0.9$, there is still a high pion content left, leading to a pion rejection factor of 2.28. The electron likelihood distributions for other momentum intervals are given in Appendix D.

5.3.2 3-dimensional Likelihood

In Run 3, as mentioned in Section 4.3, the charge is accumulated in three slices, Q_0, Q_1, and Q_2, to cover the average temporal evolution of the signal generated by an electron, making the method more sensitive to transition radiation. This separation of charge slices can potentially improve the rejection factor for pions. This method is also referred to as LQ3D. The first slice (Q_0) covers the amplification peak, the second (Q_1) the TR peak and the last (Q_2) part of the drift region. The exact time intervals are given in Table 4.1. The electron likelihood for charge slice j in layer i can be defined as $L_{i,j}(e|Q_j)$ ($:= L^e_{i,j}$), which allows for a generalization of Equation 5.4. This leads to the following expression for the electron likelihood in layer i:

$$L^e_i = \frac{\prod_j L^e_{i,j}}{\prod_j L^e_{i,j} + \prod_j L^\pi_{i,j}} = \frac{\prod_j L^e_{i,j}}{\prod_j L^e_{i,j} + \prod_j \left(1 - L^e_{i,j}\right)}. \tag{5.6}$$

This allows the use of more granular information and to differentiate electrons and pions better. The combined electron likelihood in Equation 5.5 can be extended by using Equation 5.6 to:

$$L^e = \frac{\prod_i L^e_i}{\prod_i L^e_i + \prod_i L^\pi_i} = \frac{\prod_{i,j} L^e_{i,j}}{\prod_{i,j} L^e_{i,j} + \prod_{i,j} \left(1 - L^e_{i,j}\right)}. \tag{5.7}$$

For each slice, a separate LUT is calculated with the distributions from Figure 5.5. As in the 1-dimensional case, cutoff values at low charges suppress the fluctuations from the fits. All curves show the same trend, with a steep rise at small charges and a plateau for higher charges. Only Q_2 shows an upwards trend at high charges. These lookup tables can then provide the electron likelihood per individual layer.

Figure 5.10 (a) shows the resulting electron likelihood distributions for one momentum interval for electrons and pions. The distributions are shifted more to the edges due to how the combined likelihood is calculated compared to the 1-dimensional case, making full use of the specific temporal evolution of the signal. However, this also results in a peak rise for pions at $L^e = 1$. Still, a clear improvement at 90% electron efficiency is visible, yielding a pion rejection factor of 3.36. Figure 5.10 (b) shows the pion efficiency at different electron efficiencies for positively and negatively charged particles. The plot shows the same features as in Figure 5.8 (b), although with a better pion rejection factor. The results for the other momentum intervals are in Appendix E.

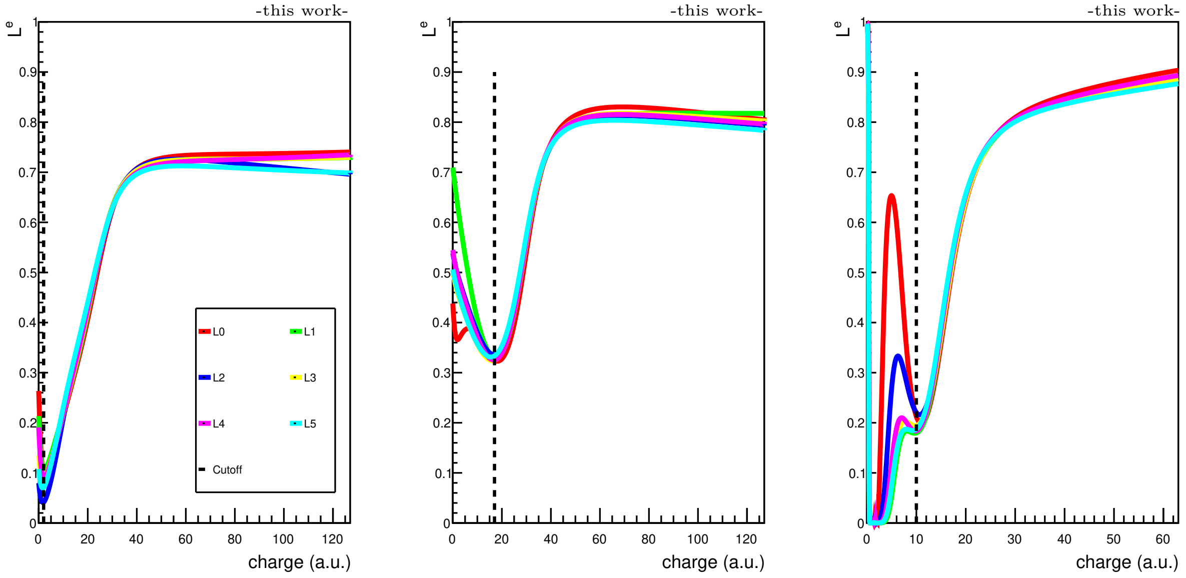
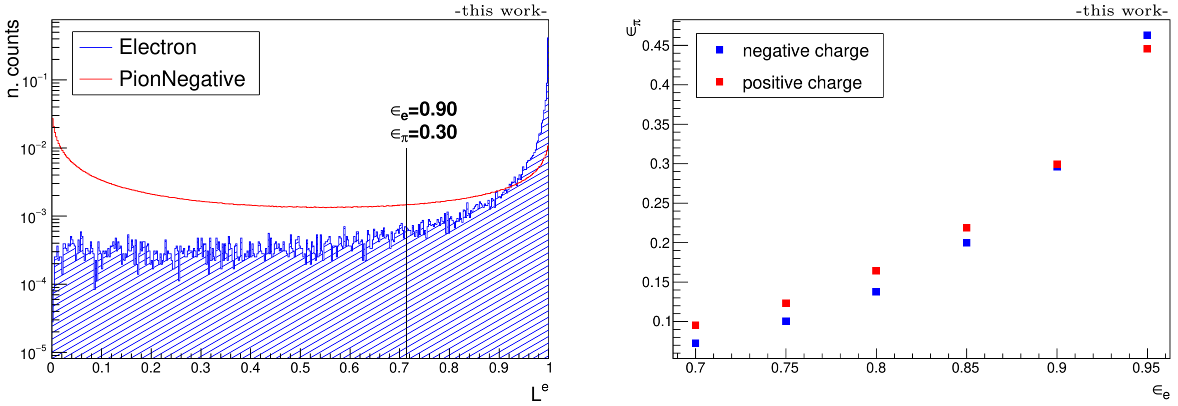


Figure 5.9: LUTs for $Q_{0,1,2}$ (left to right) of negatively charged particles with lines marking the cutoff for $p = 1.0\,\mathrm{GeV/c}$ to $2.0\,\mathrm{GeV/c}$.



(a)

(b)

Figure 5.10: (a) Electron likelihood distribution of negatively charged particles with a black line marking $\epsilon_e = 90\%$ for $p = 1.0\,\mathrm{GeV/c}$ to $2.0\,\mathrm{GeV/c}$. (b) Corresponding pion efficiency versus electron efficiency for positively and negatively charged particles.

5.3.3 Machine Learning

Machine Learning (ML) techniques have long been essential in high-energy physics, enabling researchers to solve otherwise intractable regression or classification problems [63]. In particular, supervised ML algorithms have been widely used to predict the class labels of particles in high-energy physics experiments. This section introduces the fundamental concepts of supervised machine learning and applies them to the binary classification problem of this thesis using Gradient Boosted Decision Trees (BDT).

Supervised learning and Predictions

Supervised learning is a type of machine learning that involves learning a mapping function from input features to an output space based on a labeled dataset. In other words, given the dataset $D = \{(x_i, y_i)\}_{i=1}^{n}$ ($|D| = n, x_i \in \mathbb{R}^m, y_i \in C$) with n examples and m features, where x_i denotes the input features and y_i the class labels, the goal is to find an approximation $\tilde{f}(\mathbf{x}) = \tilde{\mathbf{y}}$ given D of an optimal mapping $\hat{f}(\mathbf{x}) = \mathbf{y}$.

In order to achieve this goal, supervised learning algorithms typically minimize a loss function $\mathbb{L}(\mathbf{y}, \tilde{\mathbf{y}})$, that measures the discrepancy between the predicted output $\tilde{\mathbf{y}}$ and the true output $\mathbf{y}$. The loss function choice depends on the problem being solved; for binary classification, common choices include cross-entropy loss and binary classification loss.

Once the mapping function has been learned from the training data, it can be used to predict the class labels of new data points. Specifically, given a new input x_i, the class label $\tilde{y}_i$ can be predicted as follows:

$$\tilde{y}_i = \tilde{f}(x_i) \tag{5.8}$$

In this thesis, a binary classification ($|C| = 2$) problem of electron and pion separation will be solved. Hence, $\tilde{y}_i = 1$ corresponds to a prediction of an electron, while $\tilde{y}_i = 0$ corresponds to a prediction of a pion.

Decision Trees

Decision trees are flowchart-like structures, where each node represents a split on a feature and each leaf node represents a label. A branch of the tree is followed to generate a classification. Figure 5.11 shows an example of a decision tree. The real decision trees are much deeper since the feature space is larger (e.g., charge slices for all layers). The root node contains ten examples of pions and electrons, respectively. The first decision is on Q_0. If Q_0 is less than five, 6 pions are correctly identified and 2 electrons are misidentified. Otherwise, another classification rule is applied. Each sample traverses this tree and reaches a leaf (i.e., the classification of the sample). While a single decision tree is shown in Figure 5.11, they are seldom used alone nowadays, as they tend to be unstable and are therefore referred to as weak learners [61]. Instead, multiple trees are used simultaneously in what is referred to as an ensemble or forest. The final prediction $\tilde{y}_i$ for a tree ensemble

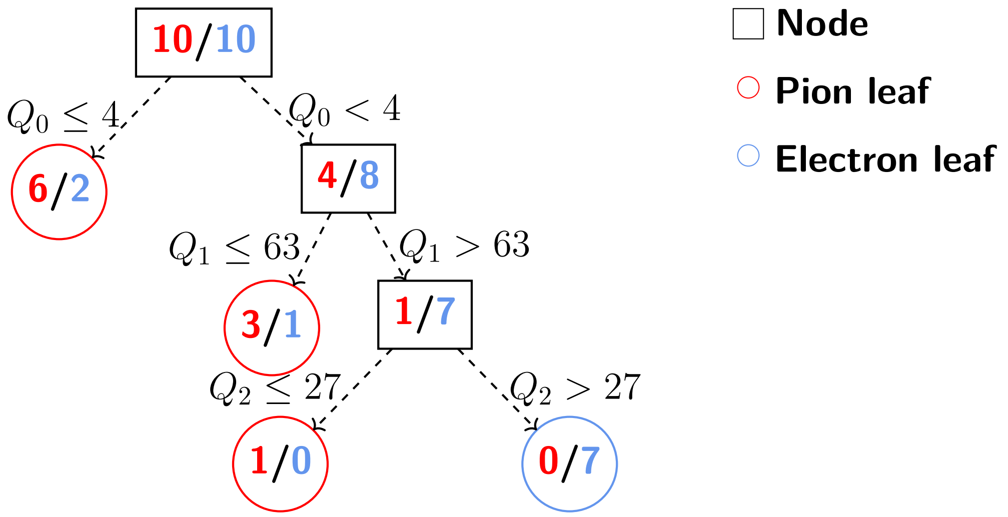


Figure 5.11: Schematic decision tree for a binary classification problem between electrons (blue) and pions (red). Nodes are drawn as rectangles, branches indicate the decision threshold and leafs are drawn as circles. Each node and leaf contain the remaining surviving data points.

is obtained by weighting the prediction of each tree:

$$\tilde{y}_i(x_i) = \tilde{f}(x_i) = \sum_{k=1}^{K} w_k \tilde{f}_k(x_i) = \sum_{k=1}^{K} h_k(x_i), \tag{5.9}$$

where K denotes the number of trees in the ensemble, x the input data, $\tilde{f}_k(x)$ the mapping and w_k the weight assigned to the contribution of the k-th tree. Finding the right split in a decision tree is crucial, so the branch yielding the highest separation or purity is chosen at each stage. Decision trees are also known for their interpretability, making them more accessible than other ML algorithms, such as neural networks.

Regularized learning objective

Overfitting is more frequent than underfitting, which is typically only a problem if the dataset is too small or the features are too few, which means that the algorithm learns the intricacies of the dataset but generalizes badly. A regularization term is introduced in the loss function to penalize the complexity of the model for addressing the problem of overfitting. Following this, the loss function $\mathbb{L}$ is modified:

$$\begin{gathered} \mathbb{L}\left(\mathbf{y}, \tilde{\mathbf{y}}\right) = \sum_i l\left(y_i, \tilde{y}_i\right) + \sum_k \Omega\left(h_k\right) \\ \text{where } \Omega\left(h_k\right) = \gamma T_k + \frac{1}{2}\lambda ||w_k||^2. \end{gathered} \tag{5.10}$$

In Equation 5.10, l represents a differentiable convex loss function and Ω the model complexity penalty, γ and λ are both regularization parameters and h_k corresponds to an independent tree k with T_k leaves and weight of w_k.

Gradient Tree Boosting

Gradient boosting is a technique in ML where the gradient boosting algorithm trains trees iteratively based on previous trees, adding them to the ensemble. In each iteration, a new weak learner is added, and the gradient of the loss function is slowly approximated using gradient descent of a loss function l. Formally, the loss function from Equation 5.10 at the i-th tree at the t-th iteration can be written as:

$$\mathbb{L}^{(t)} = \sum_{i}^{n} l(y_i, \tilde{y}_i^{(t-1)} + \eta_{\mathrm{LR}} h_t(x_i)) + \sum_{k} \Omega(h_k) \tag{5.11}$$

While this method can improve the model performance, it comes at the cost of interpretability and requires more training time. The parameter η_{LR} controls the degree learned from the previous iteration (i.e., the learning rate).

In this thesis, the weighted binary-logistic loss function is used for classification [61]:

$$l\left(y_i, \tilde{y}_i\right) = -\frac{1}{n}\left[y_i \log\left(\tilde{y}_i\right) + \left(1 - y_i\right)\log\left(1 - \tilde{y}_i\right)\right]. \tag{5.12}$$

XGBoost

Extreme Gradient Boosting (XGBoost) is a gradient tree boosting algorithm introduced by Chen and Guestrin in 2016 [64]. It has gained popularity due to its high accuracy and efficiency and provides parallel tree boosting. XGBoost is one of the leading machine learning libraries and offers a wide range of tunable hyperparameters. Hyperparameters are all model parameters that gradient descent cannot learn during training.

Training and (cross)-validation

In order to train a machine learning model, the dataset is partitioned into separate training, validation, and test datasets. The training dataset is used solely for training the model, while the validation dataset monitors the model performance during training and hyperparameter tuning. Crucially, one uses the test data to obtain an unbiased model performance evaluation.

A technique known as k-fold cross-validation is employed to obtain reliable estimates of the model generalization performance. This technique randomly splits the training dataset into k non-overlapping folds, where $k - 1$ is used for training and the remaining fold is used for testing. The process is repeated k times, using a different fold for testing, resulting in k models and performance estimates. With this approach, the sensitivity of the performance estimates to the particular sub-partitioning of the dataset is reduced [65]. It is a commonly used technique for estimating model performance.

This thesis uses a typical split for hyperparameter tuning and training, i.e., 70 : 15 : 15 (training:validation:test), where most of the data is used for training, and smaller portions are used for validation and testing. Hence, a total of 3.29×10^7 pions and 2.1×10^5 electrons for training, and 7.05×10^6 pions and 4.5×10^4 electrons for validation and testing. In particular, stratification was used to keep the same class imbalance in all sets. Additionally, for the final model, a $k = 10$-fold cross-validation with early-stopping was used to estimate the model performance. Early-stopping stops the training when the evaluation metric decreases for a configurable amount of boosting rounds. Afterwards, the best boosting round is used to evaluate the model performance.

The input features are the ones given in Section 5.2.2. Since the pions are the overwhelming majority, the training data was under-sampled, meaning only a fraction of the pion data was used. Under-sampling is often employed in machine learning on imbalanced data to avoid learning the underlying class distributions. Otherwise, the machine learning algorithm would learn the class distribution. This would result in training a classifier, which identifies everything as a pion. On average, training and validation on a RTX 3090 GPU took around one hour.

Hyperparameter tuning

In this thesis, a random search has been used to scan over the hyperparameters available in XGBoost. A random search algorithm initializes a starting hyperparameter configuration, which is subsequently extended at each iteration by randomly generated hyperparameters. The objective was to find hyperparameters for a model which provides the highest pion rejection factor. The hyperparameter search was done via the Optuna python library [66], which through its particular efficient implementation, allows for a fast search of the phase-space (spanned by the hyperparameters). The hyperparameters, scanned ranges and final values are shown in Table 5.1.

Performance evaluation

The final model was trained using 10-fold cross-validation with the hyperparameters in Table 5.1. Predictions are grouped into True Positive (TP) and True Negative (TN) for correctly classified labels, and False Negative (FN) and False Positive (FP), for miss-classified labels. FP is analogous to a type I error and FN to a type II error.

A Receiver Operating Characterisitc (ROC) curve can demonstrate the consistency of predictive model performance across different datasets. Specifically, suppose the training, validation, and test data curves are in close agreement. In that case, it indicates that the model is not overfitting to the training data and can generalize well to unseen data. The ROC is based on the model's False Positive Rate (FPR) and True Positive Rate (TPR). TPR and FPR are given by:

$$TPR = \frac{TP}{TP + FN} \text{ and } FPR = \frac{FP}{FP + TN}. \tag{5.13}$$

Table 5.1: Hyperparameters in XGBoost

Hyperparameter	Description	Range	Value
Evaluation metric	Training and monitoring the performance of the model	AUCPR, AUC, RMSE, ERROR	AUC
α parameter	L1 regularization term on weights for the loss function	$[0.001, 1.0]$	0.3453
λ parameter	L2 regularization term on weights for the loss function	$[0.001, 1.0]$	0.6624
γ parameter	Criterion to decide leaf node splitting used to control overfitting	$[0.01, 10.0]$	0.1730
Learning rate η_{LR}	Magnitude of change when updating trees	$[0.001, 1.0]$	0.3874
Max. depth	Maximal number of nodes along longest path from root to a leaf	$[5, 12]$	10
Early-stopping	Stops training early	-	5

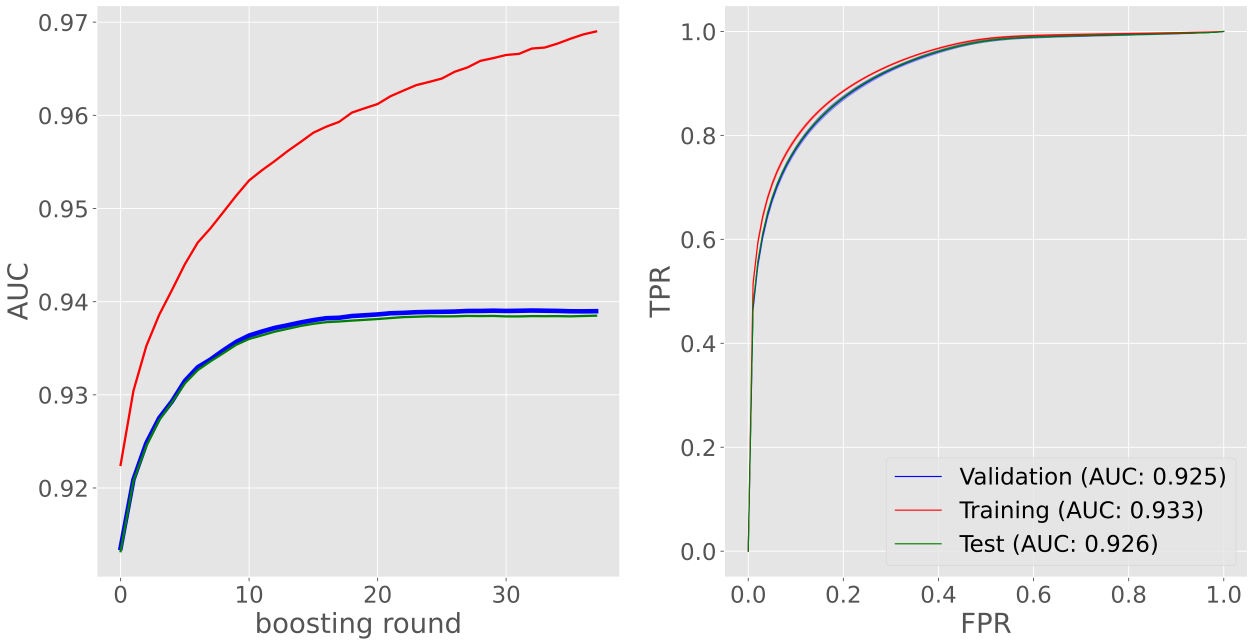


Figure 5.12: The left-hand side shows the evaluation metric against the boosting rounds for one exemplary fold. The right-hand side shows the ROC curve for the different datasets for the final model.

A standard measure used to compare models is the area under the curve, known as ROC area under the curve (AUC). AUC condenses the ROC into a single number, making it useful for model comparison.

Figure 5.12 shows on the left-hand side the evaluation metric against the boosting rounds for one exemplary fold. Validation and test datasets agree well with each other, which indicates that the model generalizes well. Training is, of course, higher and has an upward trend since the model starts to learn the underlying data. Boosting is stopped after around 35 iterations, where both curves no longer improve. The ROC curve of the final model is given on the right-hand side. The fact that the curves from the training, test and validation data set do not dramatically differ from the training curve is a good indicator that the model was not overfitted.

Figure 5.13 shows the electron likelihood distribution of negatively charged particles in one momentum interval for the final model. XGBoost nicely reproduces the distribution of the multidimensional methods, separating the pion and electron distributions. The additional bump for the pion distribution at $L^e = 0.2$ could indicate a problem with data quality due to contamination. The same trend for the pion efficiencies versus the electron efficiencies is observed. A pion rejection factor for negatively charged particles of 5.22 is achieved.

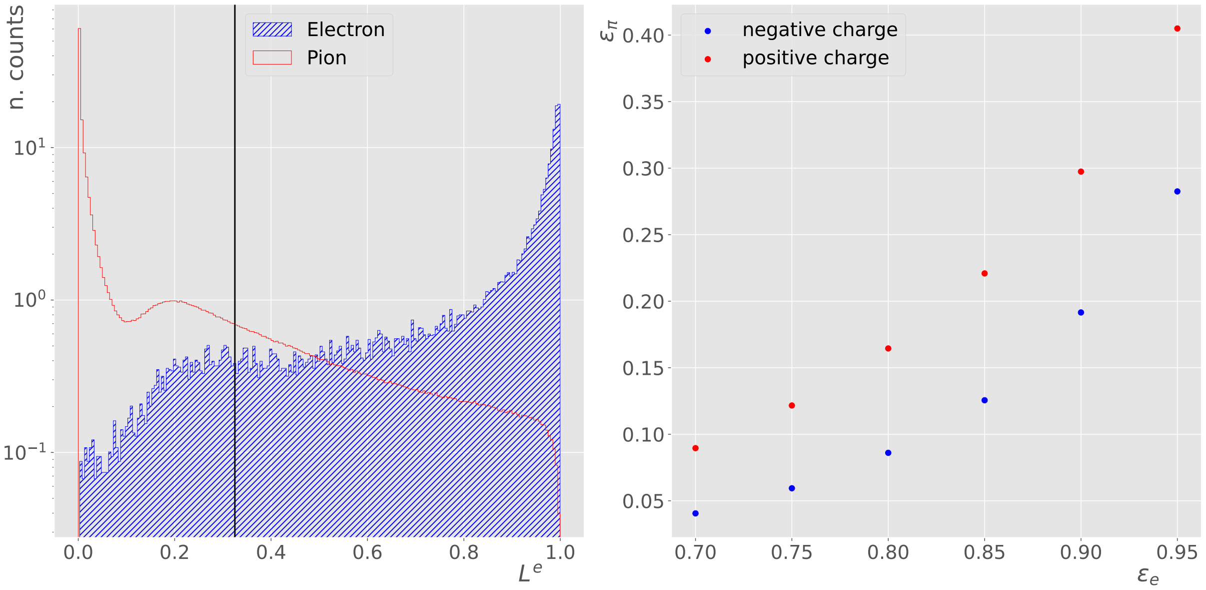


Figure 5.13: Left-hand side: Electron likelihood distribution for the machine learning algorithm of negatively charged particles with a black line marking $\epsilon_e = 90\%$ for $p = 1.0\,\mathrm{GeV/c}$ to $2.0\,\mathrm{GeV/c}$. Right-hand side: Corresponding pion efficiency versus electron efficiency for positively and negatively charged particles.

6 Results

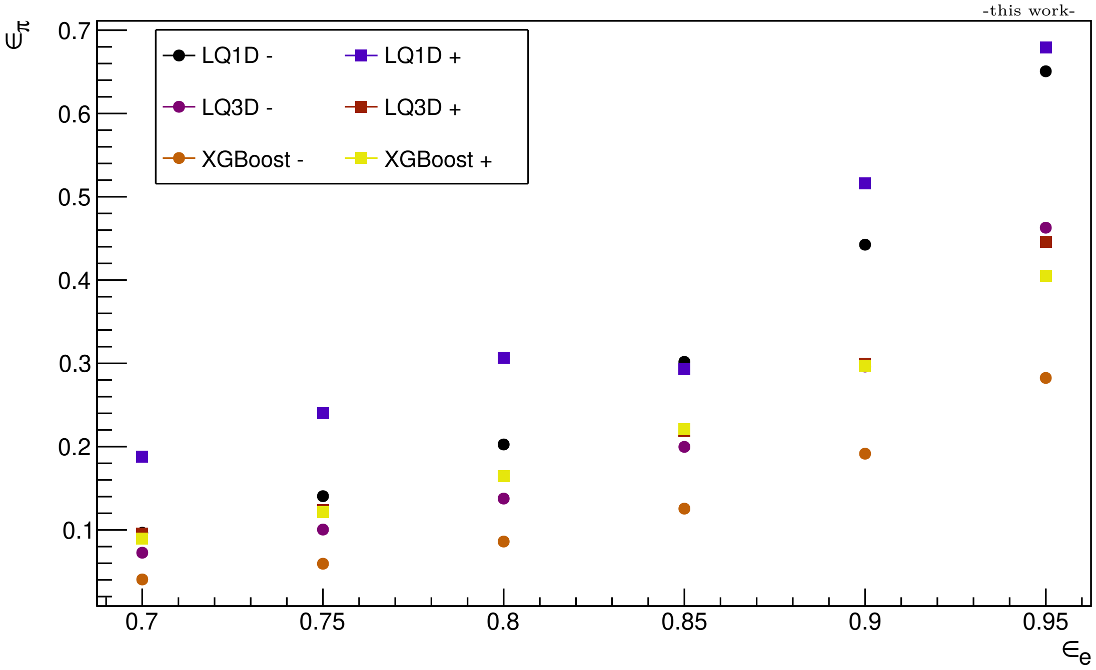


Figure 6.1: Pion efficiency versus electron efficiency for positively and negatively charged particles for all methods for $p = 1.0\,\text{GeV/c}$ to $2.0\,\text{GeV/c}$.

Figure 6.1 shows the pion efficiencies as a function of the electron efficiencies for all methods in the momentum interval $p = 1.0\,\text{GeV/c}$ to $2.0\,\text{GeV/c}$ for negatively and positively charged particles. In general, the pion efficiencies for negatively charged particles are better than the ones for positively charged particles. Moreover, the 3-dimensional likelihood achieves better results than the 1-dimensional case. Additionally, it is visible that in the respective categories, XGBoost equals, if not outperforms, the 3-dimensional likelihood method.

The momentum-dependent efficiency of pions at a 90% electron identification efficiency is illustrated in Figure 6.2 for different electron identification methods. All methods exhibit a decrease in pion efficiency initially, attributed to the onset of transition radiation production. Subsequently, pion efficiency gradually weakens as momentum increases due to an increase in pion energy loss in the relativistic rise of the Bethe-Bloch curve and saturation of transition radiation production, leading to a further decrease in pion efficiency. This behavior is well-captured by the data presented in Figure 6.2, and provides valuable insight into the performance of the electron identification methods used in the analysis. The 1-dimensional method performs the worst out of the three over the whole momentum range. As expected, the 3-dimensional method outperforms the 1-dimensional one since more granular information is used to separate electrons and pions. XGBoost performs slightly better than the 3-dimensional method, with the added cost and complexity, using a ML algorithm, which considers the data's correlation.

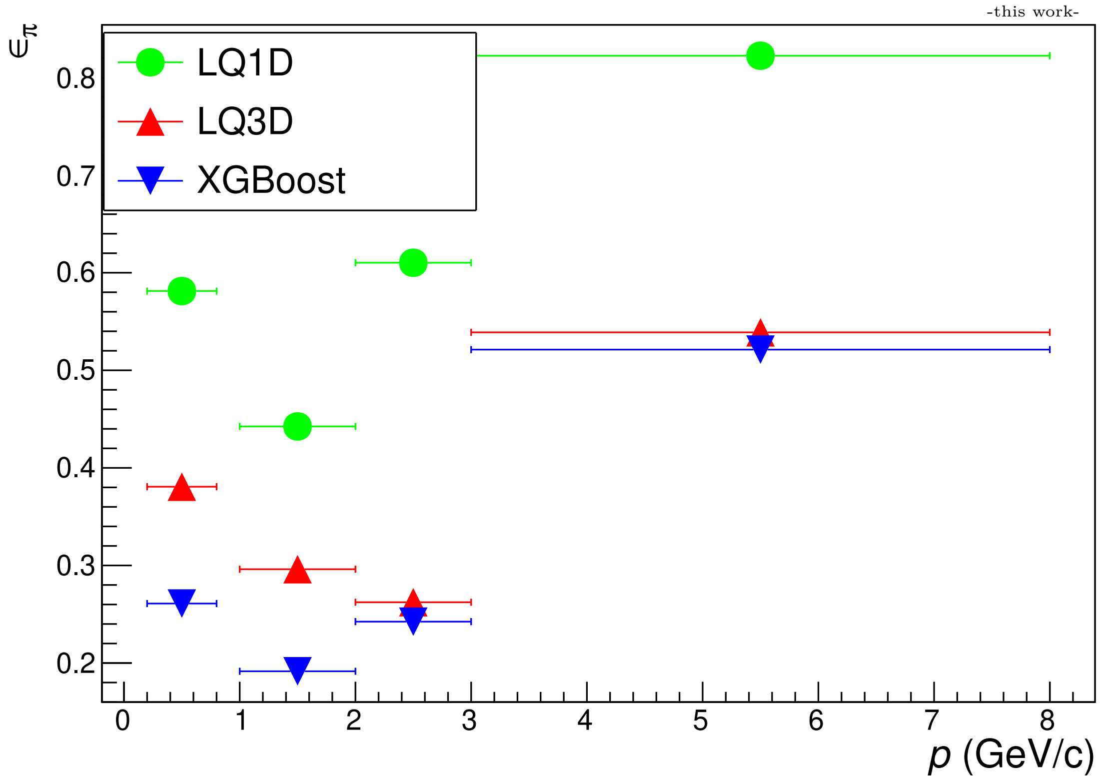


Figure 6.2: Momentum dependence of the pion efficiency at $\epsilon_e = 90\%$ for negatively charged particles for the three presented methods.

The efficiency of electron/pion separation is worse than the target result of a pion rejection factor of 100 for $p = 1.0\,\mathrm{GeV/c}$ to $2.0\,\mathrm{GeV/c}$ [50] due to various factors, such as lower precision of the input data and data contamination. Lower precision stems from adapting of the new TRD tracklet data format for Run 3, see Section 4.3.4. Previously, the pion rejection factor in Run 2 was calculated from clean V0 electron samples. These issues can significantly impact the performance of the analysis, leading to a decrease in the overall pion efficiency. Data contamination is the most significant factor that can adversely impact the efficiency of pion identification. The contamination is particularly problematic since it is easier to distinguish between electrons and background with a clean signal separation. Lower precision of the input data and data contamination significantly impact the efficiency of electron identification, and it is crucial to assess and mitigate these issues to achieve reliable results carefully. A thorough investigation of careful selection criteria to achieve satisfying electron/pion separation and filtering techniques can help to address these issues and improve the efficiency of electron identification.

7 Implementation in the ALICE O^2 framework

Within this thesis, the framework for a tracklet-based PID approach for the TRD has been implemented from scratch in the new O^2 framework, and usable first models were uploaded to the Condition and Calibration Database (CCDB), a database for general use by all ALICE members. For the LQND methods, Lookup Tables (LUTs) for different momenta were uploaded and the machine learning model was exported to the Open Neural Network Exchange (ONNX) format [67], which was subsequently used in O^2 via the ONNXRuntime [68]. Moreover, the electron likelihood is now written to the so-called Analysis Object Data (AODs) files, making it available for physics analysis. The following chapter gives an overview of the implementation.

7.1 Particle Identification

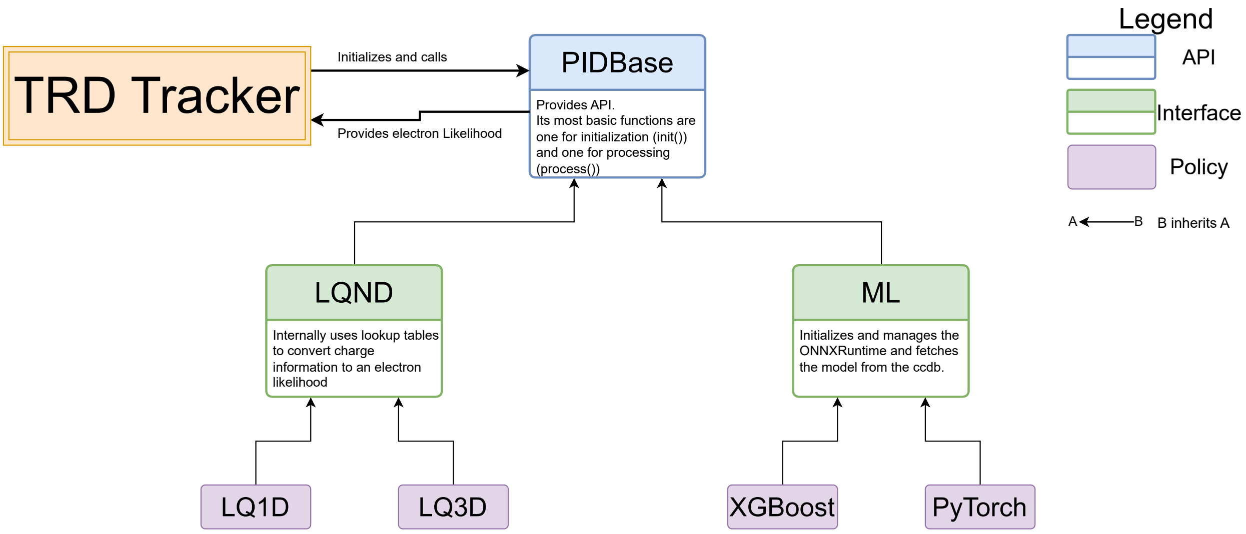


Figure 7.1: Overview of implementation in O^2.

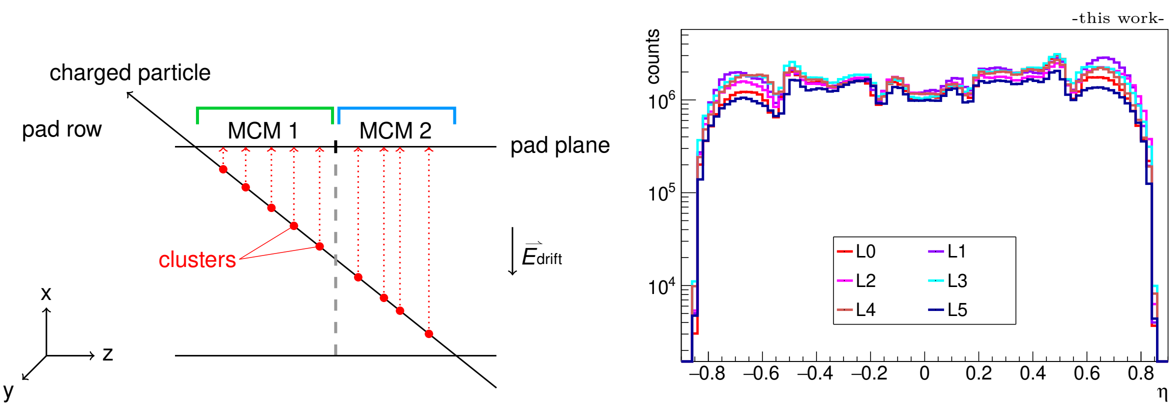
The full implementation is illustrated in Figure 7.1. The PID algorithm is implemented as a separate interface in O^2 in the ‘PIDBase’ class, which serves as a virtual base class. The tracking code manages this instance, and an electron likelihood is calculated for every (ITS-)TPC-TRD track. This electron likelihood is written to the ‘TRDTrack’ class, which propagates to the AOD producer, where the information is stored permanently. The method used to calculate an electron likelihood is specified by picking a ‘policy’ (e.g., one of the LQND or ML methods). LUTs in different momentum bins determine the N-dimensional likelihood per track for the LQND methods. For the ML method, the ONNXRuntime, an inference engine, is used to run various models exported to the ONNX format, an open-source file format designed to represent machine learning models, allowing inference and training across different frameworks and hardware platforms. Most popular machine learning libraries, such as XGBoost, provide ways to export their models to this format. The implementation was extensively tested and LQ3D is now turned on by default. Moreover, different policies can be selected easily by using the command line options:

```
o2-trd-global-tracking [--disable-pid|--policy=(default|LQ1D|LQ3D|XGB|PY)]
```

7.2 Tracklet z-row Merging

When a particle crosses the TRD under an angle such that its clusters are spread over multiple readout pads, more than one tracklet can be calculated per particle. This effectively 'splits' the tracklet in two. If the tracklets are split in the z-direction, they lie in different pad-rows. As one MCM covers only one pad-row, see Section 4.1.3, two separate non-communicating MCMs compute a tracklet, respectively. Consequently, the full charge information is unavailable during PID, and the information quality would deteriorate. However, after tracking is performed and a pad-row crossing is detected, the full charge information can be recovered by searching for close tracklets in space. Therefore, this so-called z-row merging of the tracklet charge information is implemented in the PID algorithm. Therefore, the PID algorithm implements z-row merging of tracklet charge information.

If a tracklet is split largely depends on the incident angle in the readout chamber and the track geometry. Around 11.4% of all tracklets cross pad-rows. The distribution of these split tracklets in the different layers as a function of η is shown in Figure 7.2 (b). As expected, this shows a symmetric distribution with fewer split tracklets at the readout chamber edges and a falling edge at high $|\eta|$ due to tracking inefficiencies.



(a)

(b)

Figure 7.2: (a) Sketch of split tracklet across two z-rows. Taken from [29]. (b) η-distribution of the split tracklets.

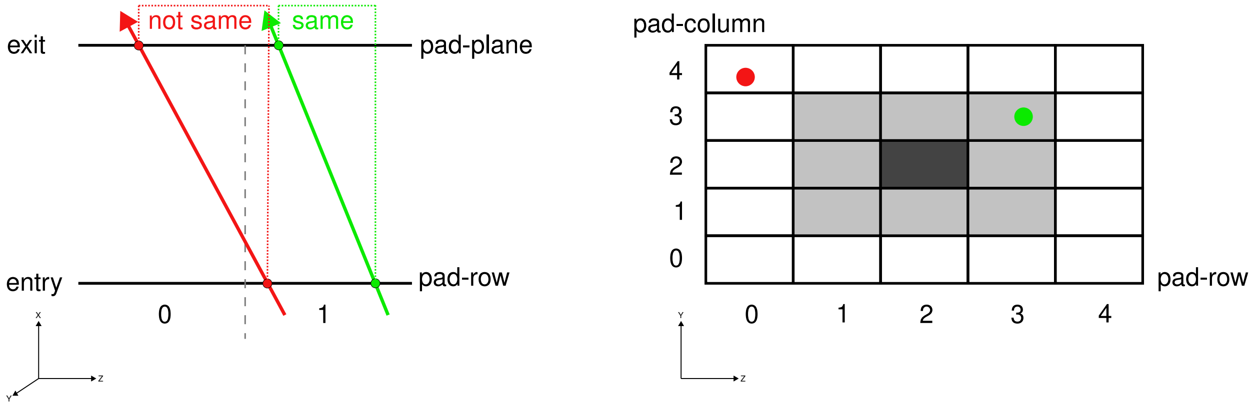


Figure 7.3: Calculation of pad-row crossings and search for neighboring tracklets.

The search for these split tracklets is implemented in the O^2 framework by propagating the track from the entry to the exit of the chamber. If the pad-row number at the entry is the same as that at the end, the track lies entirely in one pad-row. However, if the number is different, the track crosses a pad-row, as illustrated in Figure 7.3 on the left side. The red track has crossed pad-rows, and the green one has not.

Moreover, a search for neighboring tracklets was implemented. These neighboring tracklets can indicate two things: for one thing, the tracking code could have potentially other tracklet candidates, or a particle produced enough signal spread over close pads that in the MCMs two separate tracklets were calculated. In Figure 7.3 on the right side, the dark pad indicates the pad of the primary tracklet, and the gray area around it is the search radius. The green dot indicates a neighboring tracklet, while the red one represents one far enough away. For 8.8% of all tracklets, such a neighboring tracklet can be found. Since this largely depends on the track geometry and detector occupancy, the tracklets should only be discarded if one wants clean samples.

8 Summary and Outlook

For the LHC Run 3, a new software framework, called O^2, was developed to allow continuous data-taking in Pb−Pb collisions at 50 kHz. The data read out of the Transition Radiation Detector was changed entirely to cope with the requirements posed by these high interaction rates and operates now in a so-called tracklet-mode. Previously in Run 1 and 2, the tracklet-mode was only used to derive a trigger on high-p_T particles and electrons, while the analysis was performed on the complete raw data. However, the analysis is no longer based on the raw data but on reconstructed data from the Event-Processing-Nodes. For the TRD, the tracklet-mode is used for the reconstruction. Consequently, the data format was adapted and now includes the induced signal on the readout pads in a truncated format.

Within this thesis, the implementation and very first performance evaluation of the electron identification with the TRD in the LHC Run 3 have been developed. Firstly, this thesis investigates the PID performance for a Run 3 dataset. Track selection criteria were used to reduce the number of incorrectly reconstructed tracks. In addition, the Time Projection Chamber and the Time-of-Flight detector were used to select electron and pion samples for the analysis. Afterwards, lookup tables were calculated for both the 1-& 3-dimensional electron likelihood methods in four momentum intervals separately for negatively and positively charged particles due to the $E \times B$ effect. In the momentum interval $p = 1.0\,\mathrm{GeV/c}$ to $2.0\,\mathrm{GeV/c}$ at 90% electron efficiency, a pion rejection factor of 2.28 and 3.36 for negatively charged particles for the 1-dimensional and 3-dimensional methods was achieved, respectively. Machine learning was investigated as a possible way to improve the electron/pion separation. Cross-validation and hyperparameter tuning was employed to find the most stable and best model. XGBoost, a popular machine learning library, bested the classical methods by achieving a higher pion rejection factor of 5.22. The model improved the separation power over the whole momentum range. Secondly, this thesis also describes the implementation of electron identification in the new O^2 framework. For the multidimensional likelihood methods, the calculated lookup tables were uploaded to the CCDB, a general-purpose database in O^2 for ALICE. The machine learning model was exported to the standardized ONNX file format for machine learning. The ONNXRuntime, an inference engine, allows using the model. Contamination of the electron and pion samples presented the biggest challenge in all results. Furthermore, the TRD track properties in AODs files, which are those used in physics analysis, have been extended for future physics analyses.

Finally, coming to an outlook. Further steps include finding a way to reduce the contamination of the of the electron and investigating the use of a non-linear transformation function in the Front-End-Electrons for the charge slices. The electron and pion samples were contaminated mainly due to incomplete TPC and TOF calibration. One possible way to reduce data contamination could be to propagate the charge information to the AOD level and use a V0 finder to select cleaner electron/pion samples. Clean samples, for example, can be obtained by selecting tracks originating from the decay of $\gamma \rightarrow e^+e^-$ and $K_s^0 \rightarrow \pi^+\pi^-$, respectively. However, this requires a custom patch of O^2 and a dedicated reconstruction run, which was not allowed when writing this thesis. Using a non-linear transformation function in the FEE could improve the resolution of the charge distributions in the region where the electron and pion charge distributions overlap.

Bibliography

[1] J. J. Thomson. “Cathode Rays”. In: *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 44.269 (1897), pp. 293–316. DOI: 10.1080/14786449708621070.

[2] Earnest Rutherford. “Collision of α particles with light atoms. An anomalous effect in nitrogen”. In: *Philosophical Magazine* 90 (1919), pp. 31–37. DOI: 10.1080/14786431003659230.

[3] James Chadwick. “Possible Existence of a Neutron”. In: *Nature* 129 (1932), p. 312. DOI: 10.1038/129312a0.

[4] Paul Dirac. “The quantum theory of the electron”. In: *Proceedings of The Royal Society A: Mathematical, Physical and Engineering Sciences* 117 (1928), pp. 610–624. DOI: 10.1098/rspa.1928.0023.

[5] Carl D. Anderson. “The Apparent Existence of Easily Deflectable Positives”. In: *Science* 76.1967 (1932), pp. 238–239. DOI: 10.1126/science.76.1967.238.

[6] Seth H. Neddermeyer and Carl D. Anderson. “Note on the Nature of Cosmic-Ray Particles”. In: *Phys. Rev.* 51 (10 1937), pp. 884–886. DOI: 10.1103/PhysRev.51.884.

[7] C. L. Cowan et al. “Detection of the Free Neutrino: a Confirmation”. In: *Science* 124.3212 (1956), pp. 103–104. DOI: 10.1126/science.124.3212.103.

[8] G. Rochester and C. Butler. “Evidence for the Existence of New Unstable Elementary Particles”. In: *Nature* 160 (1947), pp. 855–857. DOI: 10.1038/160855a0.

[9] C. N. Yang and R. L. Mills. “Conservation of Isotopic Spin and Isotopic Gauge Invariance”. In: *Phys. Rev.* 96 (1 1954), pp. 191–195. DOI: 10.1103/PhysRev.96.191.

[10] M. Gell-Mann. “A Schematic Model of Baryons and Mesons”. In: *Resonance* 24 (1964), pp. 923–925. DOI: 10.1007/s12045-019-0853-x.

[11] G. Zweig. “An SU(3) model for strong interaction symmetry and its breaking”. In: (1964). URL: https://inspirehep.net/literature/11881.

[12] D. P. Barber et al. “Discovery of Three-Jet Events and a Test of Quantum Chromodynamics at PETRA”. In: *Phys. Rev. Lett.* 43 (12 1979), pp. 830–833. DOI: 10.1103/PhysRevLett.43.830.

[13] Mary K. Gaillard, Paul D. Grannis, and Frank J. Sciulli. “The standard model of particle physics”. In: *Rev. Mod. Phys.* 71 (2 1999), pp. 96–111. DOI: 10.1103/revmodphys.71.s96.

[14] P.W. Higgs. “Broken symmetries, massless particles and gauge fields”. In: *Physics Letters* 12.2 (1964), pp. 132–133. DOI: 10.1016/0031-9163(64)91136-9.

[15] F. Englert and R. Brout. “Broken Symmetry and the Mass of Gauge Vector Mesons”. In: *Phys. Rev. Lett.* 13 (9 1964), pp. 321–323. DOI: `10.1103/PhysRevLett.13.321`.

[16] ATLAS Collaboration. “Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC”. In: *Physics Letters B* 716.1 (2012), pp. 1–29. DOI: `10.1016/j.physletb.2012.08.020`.

[17] CMS Collaboration. “Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC”. In: *Physics Letters B* 716.1 (2012), pp. 30–61. DOI: `10.1016/j.physletb.2012.08.021`.

[18] Gianfranco Bertone, Dan Hooper, and Joseph Silk. “Particle dark matter: evidence, candidates and constraints”. In: *Physics Reports* 405.5 (2005), pp. 279–390. DOI: `10.1016/j.physrep.2004.08.031`.

[19] Andrei D Sakharov. “Violation of CP invariance, C asymmetry, and baryon asymmetry of the universe”. In: *Soviet Physics Uspekhi* 34.5 (1991), p. 392. DOI: `10.1070/PU1991v034n05ABEH002497`.

[20] Matic Lubej. *Standard Model Graphical Representation.* `https://github.com/mlubej/standard-model`. [Accessed: 12.04.2023].

[21] Rajeev S. Bhalerao. *Relativistic heavy-ion collisions.* 2014. arXiv: `1404.3294`.

[22] A. Bazavov et al. “Chiral crossover in QCD at zero and non-zero chemical potentials”. In: *Physics Letters B* 795 (2019), pp. 15–21. DOI: `10.1016/j.physletb.2019.05.013`.

[23] Franz Gross et al. *50 Years of Quantum Chromodynamics.* 2022. arXiv: `2212.11107`.

[24] The ALICE Collaboration et al. *The ALICE experiment - A journey through QCD.* 2022. arXiv: `2211.04384`.

[25] *CERN History.* `https://www.home.cern/about/who-we-are/our-history`. [Accessed: 08.04.2023].

[26] *ATLAS Homepage.* `https://atlas.cern/`. [Accessed: 08.04.2023].

[27] *CMS Homepage.* `https://cms.cern/`. [Accessed: 08.04.2023].

[28] *ALICE Homepage.* `https://alice-collaboration.web.cern.ch/`. [Accessed: 08.04.2023].

[29] Hannah Klingenmeyer. *Tracklet-based particle identification with ALICE TRD for LHC Run 3.* 2017.

[30] Lyndon Evans and Philip Bryant. “LHC Machine”. In: *Journal of Instrumentation* 3.08 (2008), S08001. DOI: `10.1088/1748-0221/3/08/S08001`.

[31] The ALICE Collaboration et al. *ALICE upgrades during the LHC Long Shutdown 2.* 2023. arXiv: `2302.01238`.

[32] The ALICE Collaboration et al. *Letter of Intent for A Large Ion Collider Experiment [ALICE].* Tech. rep. CERN, 1993. URL: `https://cds.cern.ch/record/290825`.

[33] The ALICE Collaboration et al. “The ALICE experiment at the CERN LHC”. In: *JINST 3* (2008). DOI: `10.1088/1748-0221/3/08/S08002`.

[34] The ALICE Collaboration et al. *Upgrade of the ALICE Experiment: Letter of Intent.* Tech. rep. CERN, 2014. DOI: `10.1088/0954-3899/41/8/087001`.

[35] The ALICE Collaboration et al. “Definition of the ALICE Coordinate System and basic rules for Sub-Detector Components numbering”. 2003. URL: `https://edms.cern.ch/document/406391/2`.

[36] The ALICE Collaboration et al. *Technical Design Report for the Upgrade of the ALICE Inner Tracking System*. Tech. rep. CERN, 2014. DOI: `10.1088/0954-3899/41/8/087002`. URL: `https://cds.cern.ch/record/1625842`.

[37] The ALICE Collaboration et al. “Performance of the ALICE experiment at the CERN LHC”. In: *International Journal of Modern Physics A* (2014). DOI: `10.1142%2Fs0217751x14300440`.

[38] The ALICE Collaboration et al. “Particle identification in ALICE: a Bayesian approach”. In: *Eur. Phys. J. Plus* 131 (2016), p. 168. arXiv: `1602.01392`.

[39] The ALICE Collaboration et al. *Technical Design Report for the Upgrade of the Online-Offline Computing System*. Tech. rep. 2015. URL: `https://cds.cern.ch/record/2011297`.

[40] Z. Citron et al. *Future physics opportunities for high-density QCD at the LHC with heavy-ion and proton beams*. 2019. arXiv: `1812.06772`.

[41] R. Brun and F. Radmakers. *ROOT - An Object-Oriented Data Analysis Framework*. Tech. rep. 1997, pp. 81–86. DOI: `10.1016/S0168-9002(97)00048-X`. URL: `https://root.cern/`.

[42] Kitware. *CMake Software build system*. [Accessed: 07.04.2023]. URL: `https://cmake.org/`.

[43] Boost Organisation. *Boost C++ Libraries*. [Accessed: 07.04.2023]. URL: `https://boost.org/`.

[44] International Organization for Standardization. *C++17 Standard*. [Accessed: 07.04.2023]. URL: `https://en.cppreference.com/w/cpp/17`.

[45] The FAIR and ALICE Collaborations. “ALFA: The new ALICE-FAIR software framework”. In: *Journal of Physics: Conference Series* (2015).

[46] The FAIR Collaboration et al. “The FairRoot framework”. In: *Journal of Physics: Conference Series* (2012).

[47] Marten Ole Schmidt. “Space-point calibration of the ALICE TPC with track residuals”. PhD thesis. Physikalisches Institut der Universität Heidelberg, 2020. URL: `https://cds.cern.ch/record/2712602`.

[48] Giulio Eulisse et al. “Evolution of the ALICE software framework for Run 3”. In: *EPJ Web Conf.* (2019). DOI: `10.1051/epjconf/201921405010`.

[49] The ALICE Collaboration et al. *Upgrade of the ALICE Readout & Trigger System*. Tech. rep. 2013. URL: `https://cds.cern.ch/record/1603472`.

[50] The ALICE Collaboration et al. “The ALICE Transition Radiation Detector: Construction, operation, and performance”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 881 (2018), pp. 88–127. DOI: `10.1016/j.nima.2017.09.028`.

[51] Vitalii Lazarevich Ginzburg and I. M. Frank. “Radiation of a uniformly moving electron due to its transition from one medium into another”. In: 1945.

[52] Eva Sicking. "Alignment of ALICE TRD Modules Using Cosmic Ray Data". Institut für Kernphysik der Unversität Münster. Diploma Thesis. 2009.

[53] The ALICE Collaboration et al. *ALICE transition-radiation detector: Technical Design Report*. Tech. rep. 2001. URL: https://cds.cern.ch/record/519145.

[54] Jochen Klein. "Jet Physics with A Large Ion Collider Experiment at the Large Hadron Collider". PhD thesis. Physikalisches Institut der Universität Heidelberg, 2014.

[55] Gauthier Legras. *Gain calibraion*. private communications. Institut für Kernphysik der Unversität Münster.

[56] Markus Fasel. "Single-electron analysis and open charm cross section in proton-proton collisions at $\sqrt{s} = 7\,\text{TeV}$". Presented 31 Oct 2012. 2012. URL: https://cds.cern.ch/record/1551722.

[57] Jana Crkovska. *Krypton calibration*. private communications. Physikalisches Institut der Universität Heidelberg.

[58] Johannes Stiller. "Gain calibration of the ALICE TRD using the decay of $^{83\text{m}}$Kr by internalconversion1". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 706 (2013), pp. 20–22. DOI: 10.1016/j.nima.2012.05.019.

[59] Luisa Bergmann. *t_0 calibration*. private communications. Physikalisches Institut der Universität Heidelberg.

[60] Salman K. Malik. *Chamber status*. private communications. University of Jammu.

[61] Alan S. Cornell et al. "Boosted decision trees in the era of new physics: a smuon analysis case study". In: *Journal of High Energy Physics* (2022). DOI: 10.1007/jhep04(2022)015.

[62] Daniel Lohner. "Anisotropic flow of direct photons in Pb−Pb collisions at $\sqrt{s_{\text{NN}}} = 2.76\,\text{TeV}$". PhD thesis. Physikalisches Institut der Universität Heidelberg, 2013.

[63] Kim Albertsson et al. *Machine Learning in High Energy Physics Community White Paper*. 2018. arXiv: 1807.02876.

[64] Tianqi Chen and Carlos Guestrin. "XGBoost". In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2016. DOI: 10.1145/2939672.2939785.

[65] S. Raschka and V. Mirjalili. *Python Machine Learning*. 2017. ISBN: 9781787126022. URL: https://books.google.de/books?id=%5C_plGDwAAQBAJ.

[66] Takuya Akiba et al. *Optuna: A Next-generation Hyperparameter Optimization Framework*. 2019. arXiv: 1907.10902.

[67] Junjie Bai, Fang Lu, Ke Zhang, et al. *ONNX: Open Neural Network Exchange*. https://github.com/onnx/onnx. [Accessed: 07.04.2023]. 2019.

[68] ONNX Runtime developers. *ONNX Runtime*. https://onnxruntime.ai/. [Accessed: 07.04.2023]. 2021.

Acronyms

ADC Analog to Digital Converter

ALEPH Apparatus for LEP Physics

ALICE A Large Ion Collider Experiment

AOD Analysis Object Data

ATLAS A Torodial LHC Apperatus

AUC ROC area under the curve

BDT Gradient Boosted Decision Trees

BNL Brookhaven National Laboratory

CCDB Condition and Calibration Database

CERN Conseil Européen pour la Recherche Nucléaire

CMS Compact Muon Solenoid

CPU Computer Processing Unit

CRU Common Readout Unit

CTF Compressed Timeframes

CTP Central Trigger Processor

DDL Direct Detector Link

DPL Data Processing Layer

EMCal Electromagnetic Calorimeter

EPN Event Processing Node

FEE Front-End Electronics

FIT Fast Interaction Trigger

FLP First Level Processor

GEM Gas Electron Multiplier

GPU Graphics Processing Unit

GSI Gesellschaft für Schwerionenforschung

HB Heartbeat

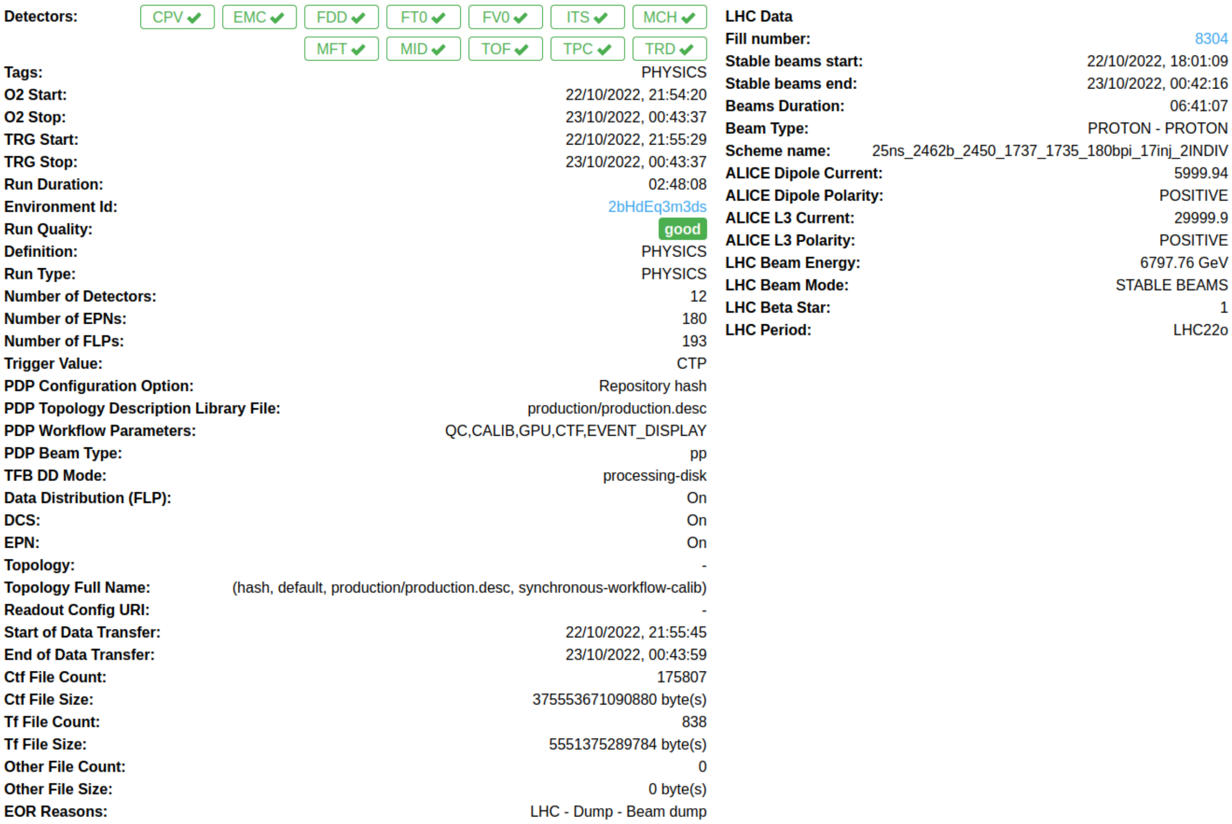
HBF Heartbeat Frame

ITS Inner Tracking System
LEIR Low Energy Ion Ring
LEP Large Electron-Positron Collider
LHC Large Hadron Collider
LHCb LHC-beauty
LS2 Long Shutdown 2
LTU Local Trigger Unit
LUT Lookup Table
MAPS Monolithic Active Pixel Sensor
MCM Multi Chip Module
ML Machine Learning
MRPC Multigap Resistive Plate Chambers
MWPC Multi-Wire Proportional Chambers
O^{2} Online and Offline Computing System
ONNX Open Neural Network Exchange
PASA Pre-Amplifier Shaper
PBS Proton Synchrotron Booster
PCA Principal Component Analysis
PCB Printed Circuit Board
PHOS Photon Spectrometer
PID Particle Identification
PS Proton Synchrotron
QCD Quantum Chromodynamics
QED Quantum Electrodynamics
QGP Quark-Gluon Plasma
RHIC Relativistic Heavy Ion Collider
ROB Readout Board
ROC Receiver Operating Characterisitc
SLAC Stanford Linear Accelerator Center
SM Standard Model
SM Supermodule
SPS Super Proton Synchrotron
STF Sub-Timeframe
TF Timeframes

TOF	Time-of-Flight Detector
TPC	Time Projection Chamber
TR	Transition Radiation
TRAP	Tracklet Processor
TRD	Transition Radiation Detector
XGBoost	Extreme Gradient Boosting

Appendix

A Specifics for run 527852



Detectors:	CPV ✓ EMC ✓ FDD ✓ FT0 ✓ FV0 ✓ ITS ✓ MCH ✓ MFT ✓ MID ✓ TOF ✓ TPC ✓ TRD ✓
Tags:	PHYSICS
O2 Start:	22/10/2022, 21:54:20
O2 Stop:	23/10/2022, 00:43:37
TRG Start:	22/10/2022, 21:55:29
TRG Stop:	23/10/2022, 00:43:37
Run Duration:	02:48:08
Environment Id:	2bHdEq3m3ds
Run Quality:	good
Definition:	PHYSICS
Run Type:	PHYSICS
Number of Detectors:	12
Number of EPNs:	180
Number of FLPs:	193
Trigger Value:	CTP
PDP Configuration Option:	Repository hash
PDP Topology Description Library File:	production/production.desc
PDP Workflow Parameters:	QC,CALIB,GPU,CTF,EVENT_DISPLAY
PDP Beam Type:	pp
TFB DD Mode:	processing-disk
Data Distribution (FLP):	On
DCS:	On
EPN:	On
Topology:	-
Topology Full Name:	(hash, default, production/production.desc, synchronous-workflow-calib)
Readout Config URI:	-
Start of Data Transfer:	22/10/2022, 21:55:45
End of Data Transfer:	23/10/2022, 00:43:59
Ctf File Count:	175807
Ctf File Size:	375553671090880 byte(s)
Tf File Count:	838
Tf File Size:	5551375289784 byte(s)
Other File Count:	0
Other File Size:	0 byte(s)
EOR Reasons:	LHC - Dump - Beam dump

LHC Data	
Fill number:	8304
Stable beams start:	22/10/2022, 18:01:09
Stable beams end:	23/10/2022, 00:42:16
Beams Duration:	06:41:07
Beam Type:	PROTON - PROTON
Scheme name:	25ns_2462b_2450_1737_1735_180bpi_17inj_2INDIV
ALICE Dipole Current:	5999.94
ALICE Dipole Polarity:	POSITIVE
ALICE L3 Current:	29999.9
ALICE L3 Polarity:	POSITIVE
LHC Beam Energy:	6797.76 GeV
LHC Beam Mode:	STABLE BEAMS
LHC Beta Star:	1
LHC Period:	LHC22o

Figure: ALICE Bookkeeping Entry. Accessed: 3.04.2023.

B Selection criteria

Table 8.1: Track selection criteria.

Selection	Criterion	Explanation
$\chi^2_{\text{red.}}$	≤ 6	Ensures that the track fits have converged and are reasonable.
Number of tracklets in track	≥ 4	Reduces the number of falsely attached tracklets to tracks.
$n^{e/K}_{\text{comb.}}$	≥ 2.5	Combined separation power calculated according to Equation 2.7.
$n^{e/p}_{\text{comb.}}$	≥ 2.5	Combined separation power calculated according to Equation 2.7.
$n^{\pi}\sigma_{\text{TPC}}$	$]-1, 0]$	TPC dE/dx pion separation, only taking the lower interval to reduce contamination calculated according to Equation 2.3.
Pad-row crossing tracklets	excluded	Tracklets which where split across z-rows as in Figure 7.2 (a).
Neighbor tracklets	excluded	Tracklets for which a neighboring tracklet is found.
MCMs with noisy Pads	excluded	Tracklets from these pads are mostly noise.
Manual dE/dx cut	-	For $p \geq 3\,\text{GeV/c}$ the separation power is insufficient to separate electrons and pions. Hence, a rough cut around the Bethe-Bloch curve for electrons was used.

C TPC dE/dx parametrization

For run 527852 the following parameters where used for Equation 2.1 and Equation 2.2.

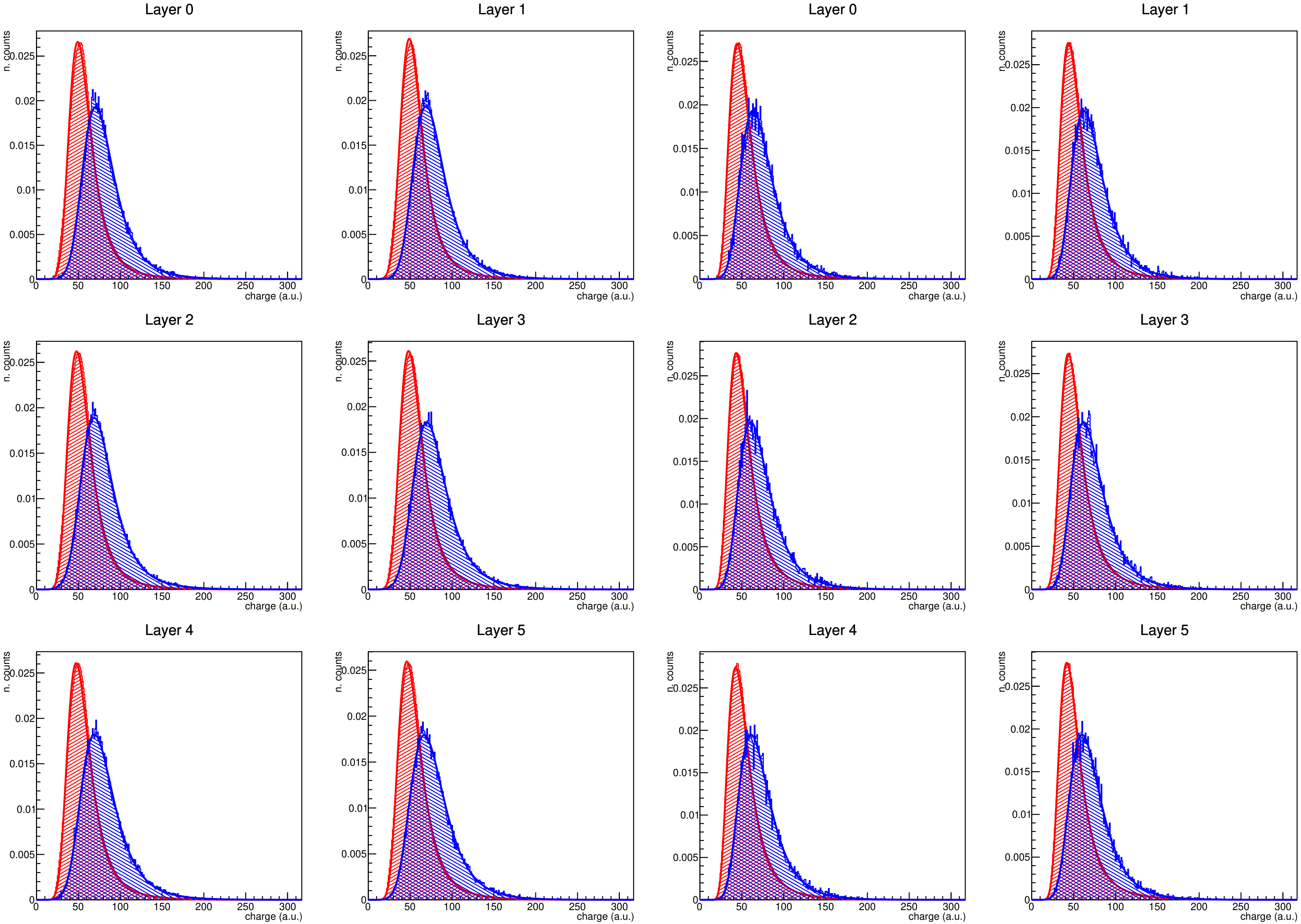
Parameter	Value
P_1	0.15429793
P_2	5.255017
P_3	0.0028807498
P_4	2.3966658
P_5	1.2244654
MIP	50.0

Table: TPC dE/dx parametrization

The parameters in the table above, where retrieved with:

```
o2-pidparam-tpc-response --mode pull --min-runnumber 527852
```

D. LQ1D

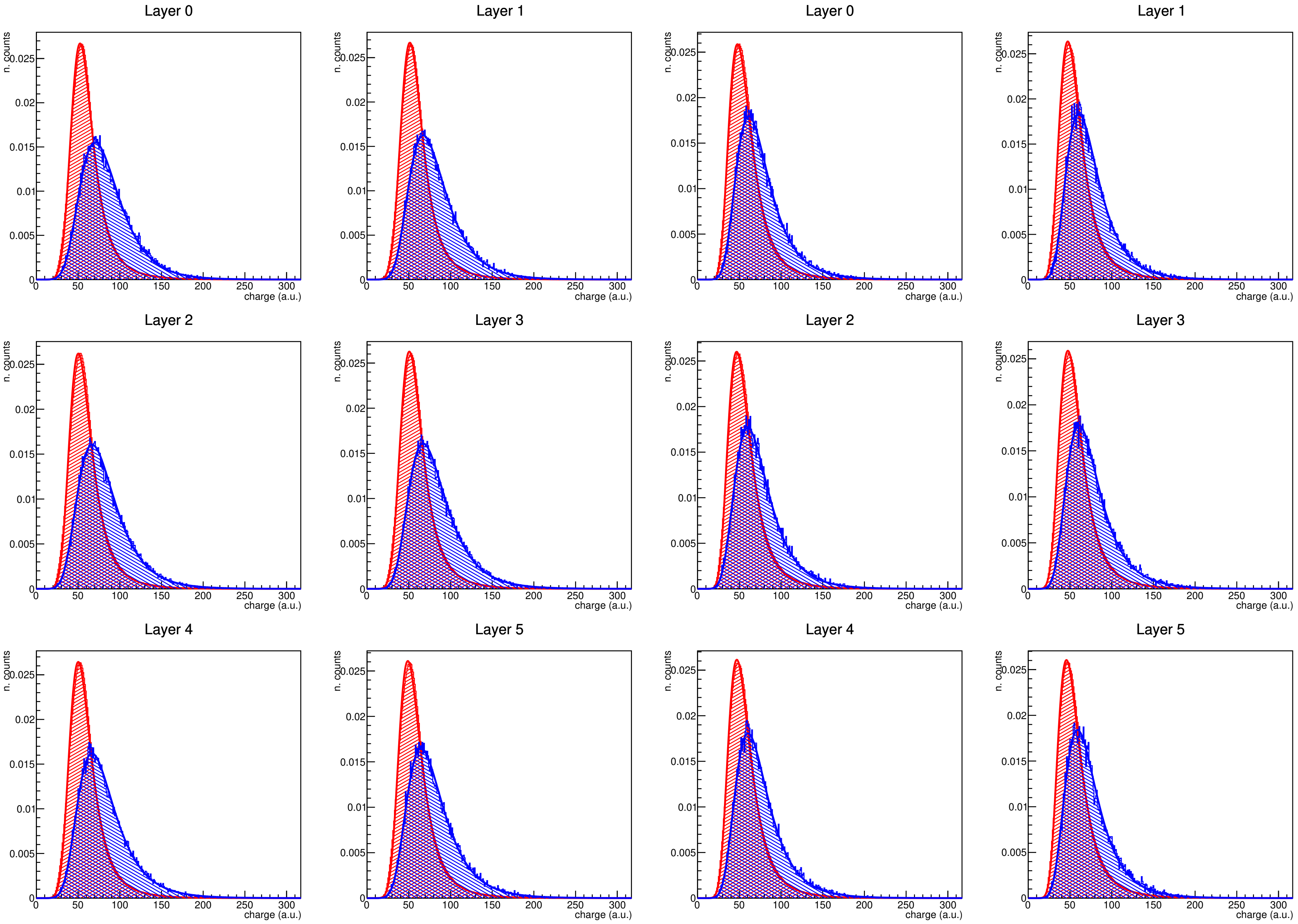


(a) negatively charged

(b) positively charged

Figure: Total accumulated charge distribution for $p = 0.2\,\mathrm{GeV/c}$ to $1.0\,\mathrm{GeV/c}$.

D. LQ1D

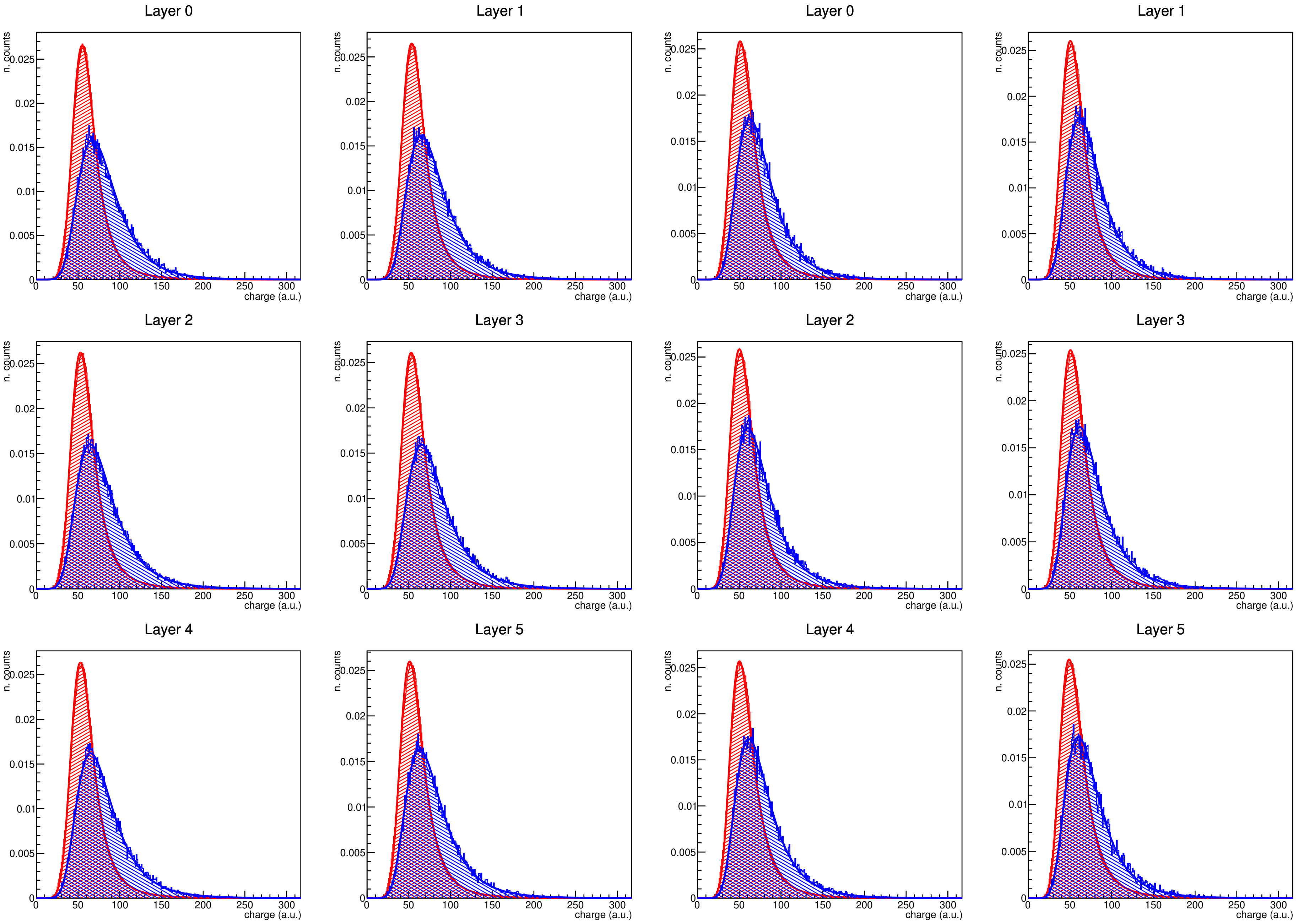


(a) negatively charged

(b) positively charged

Figure: Total accumulated charge distribution particles for $p = 1.0\,\mathrm{GeV/c}$ to $2.0\,\mathrm{GeV/c}$.

D. LQ1D

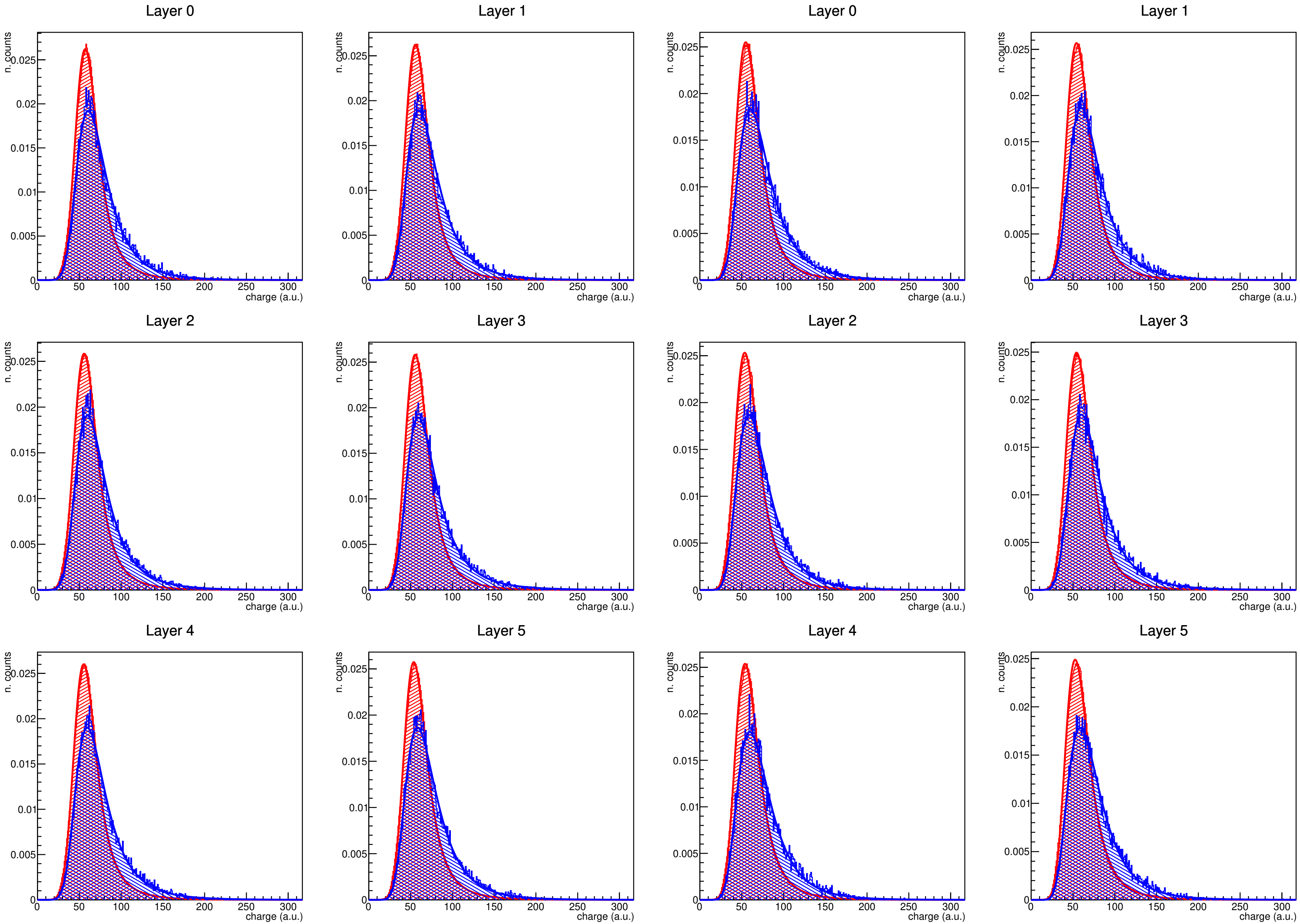


(a) negatively charged

(b) positively charged

Figure: Total accumulated charge distribution particles for $p = 2.0\,\mathrm{GeV/c}$ to $3.0\,\mathrm{GeV/c}$.

D. LQ1D



(a) negatively charged (b) positively charged

Figure: Total accumulated charge distribution particles for $p = 3.0\,\mathrm{GeV/c}$ to $8.0\,\mathrm{GeV/c}$.

D. LQ1D

(a) negatively charged (b) positively charged

Figure: LUT for $p = 0.2\,\mathrm{GeV/c}$ to $1.0\,\mathrm{GeV/c}$.

(a) negatively charged (b) positively charged

Figure: LUT for $p = 1.0\,\mathrm{GeV/c}$ to $2.0\,\mathrm{GeV/c}$.

D. LQ1D

(a) negatively charged

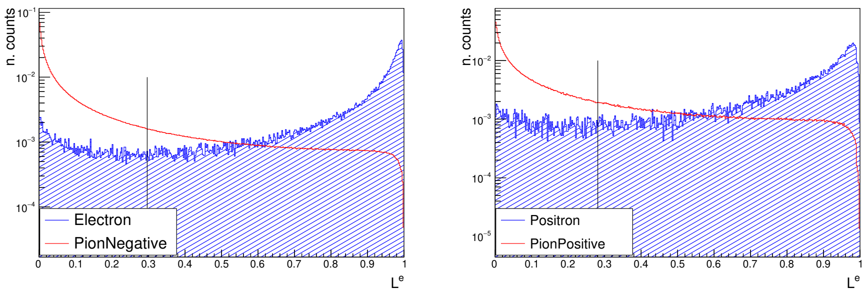
(b) positively charged

Figure: LUT for $p = 2.0\,\mathrm{GeV/c}$ to $3.0\,\mathrm{GeV/c}$.

(a) negatively charged

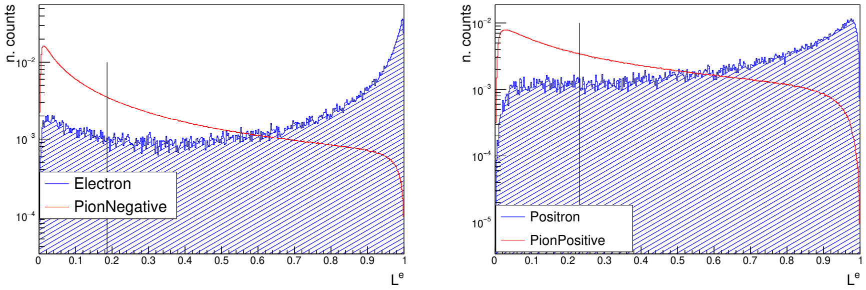
(b) positively charged

Figure: LUT for $p = 3.0\,\mathrm{GeV/c}$ to $8.0\,\mathrm{GeV/c}$.



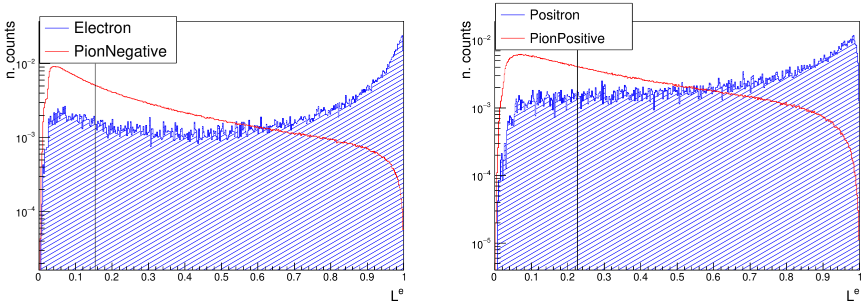
(a) negatively charged (b) positively charged

Figure: $p = 0.2\,\mathrm{GeV/c}$ to $1.0\,\mathrm{GeV/c}$.



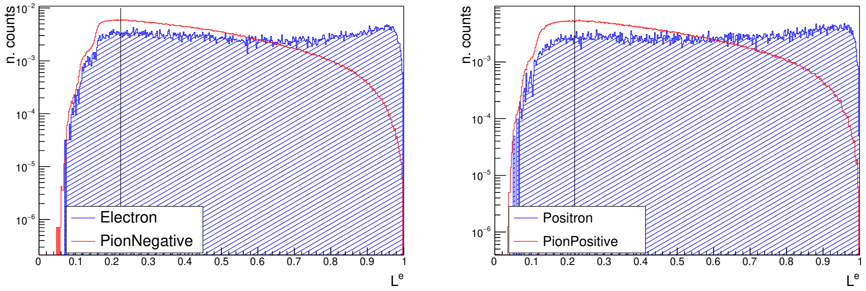
(a) negatively charged (b) positively charged

Figure: $p = 1.0\,\mathrm{GeV/c}$ to $2.0\,\mathrm{GeV/c}$.



(a) negatively charged (b) positively charged

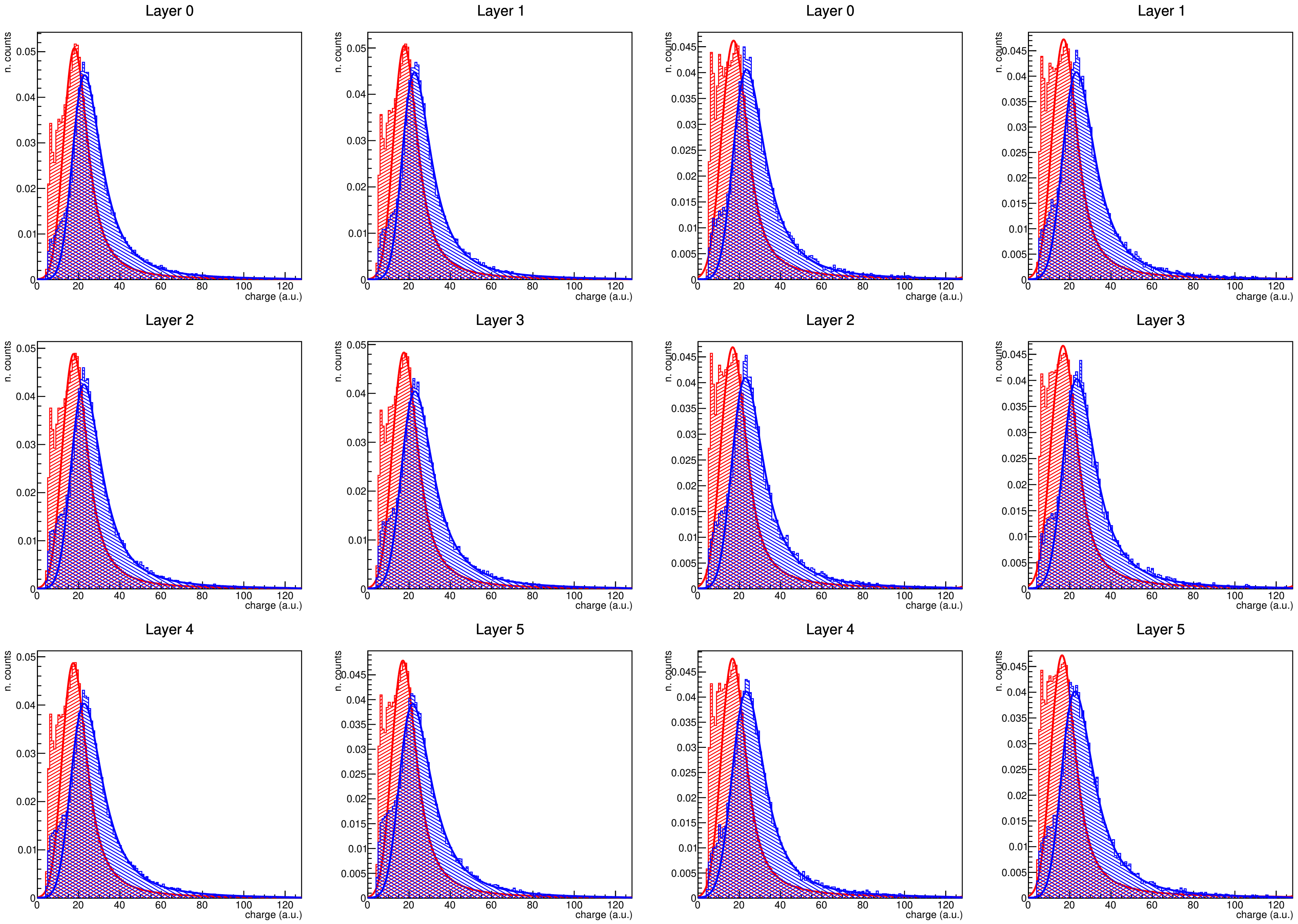
Figure: $p = 2.0\,\mathrm{GeV/c}$ to $3.0\,\mathrm{GeV/c}$.



(a) negatively charged (b) positively charged

Figure: $p = 3.0\,\mathrm{GeV/c}$ to $8.0\,\mathrm{GeV/c}$.

E. LQ3D

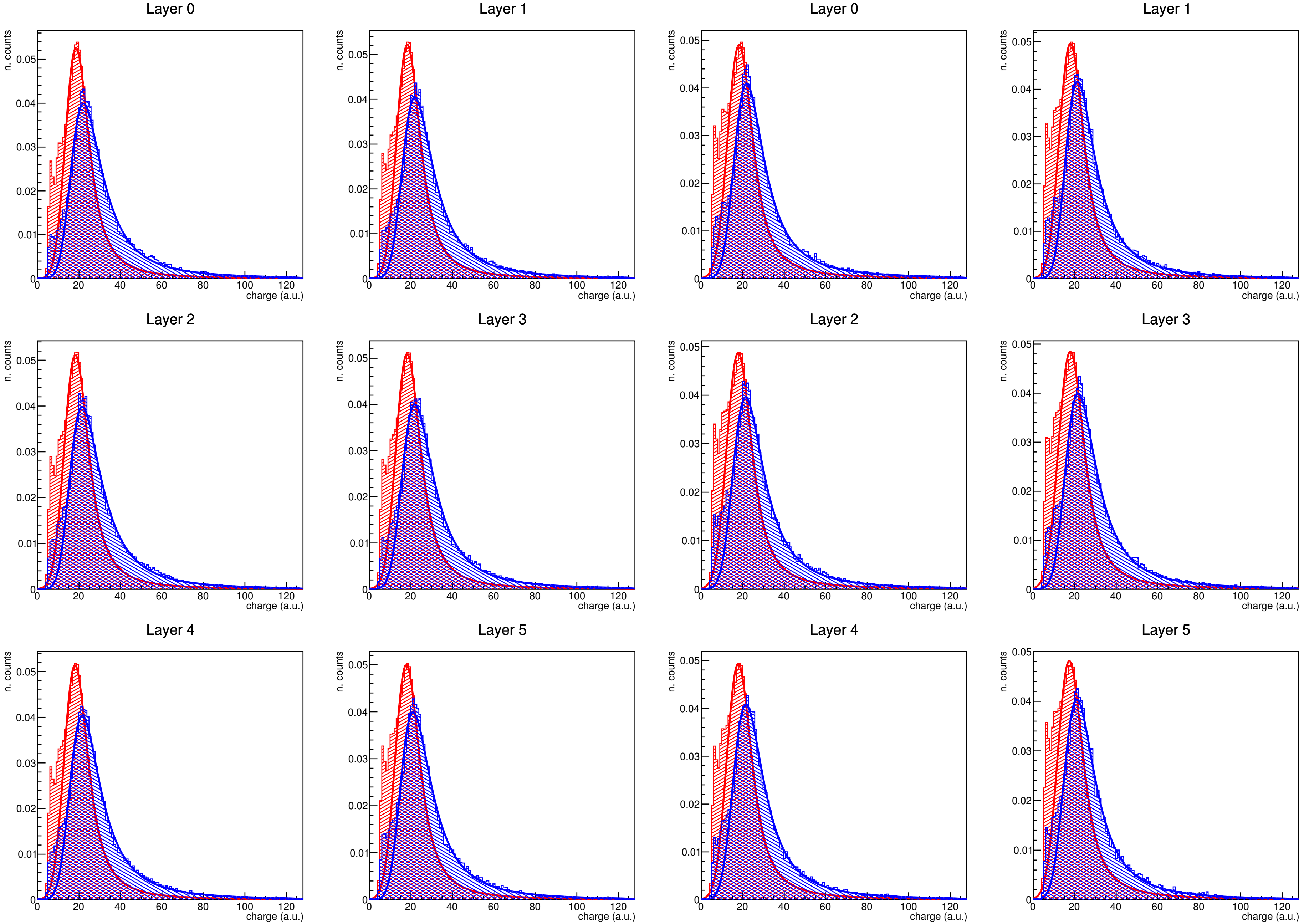


(a) negatively charged

(b) positively charged

Figure: Q_0 charge distribution for $p = 0.2\,\mathrm{GeV/c}$ to $1.0\,\mathrm{GeV/c}$.

E. LQ3D

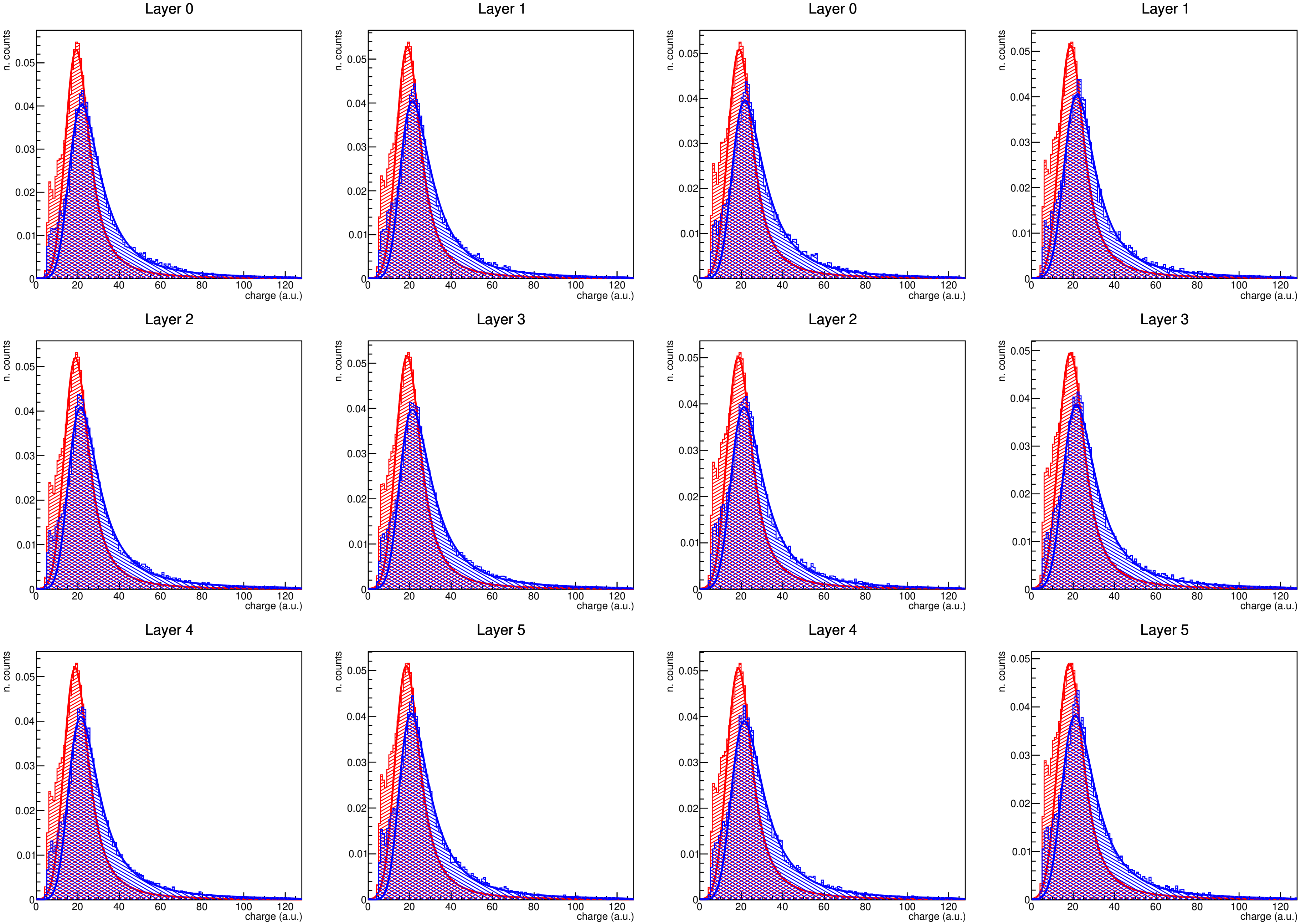


(a) negatively charged

(b) positively charged

Figure: Q_0 charge distribution for $p = 1.0\,\mathrm{GeV/c}$ to $2.0\,\mathrm{GeV/c}$.

E. LQ3D

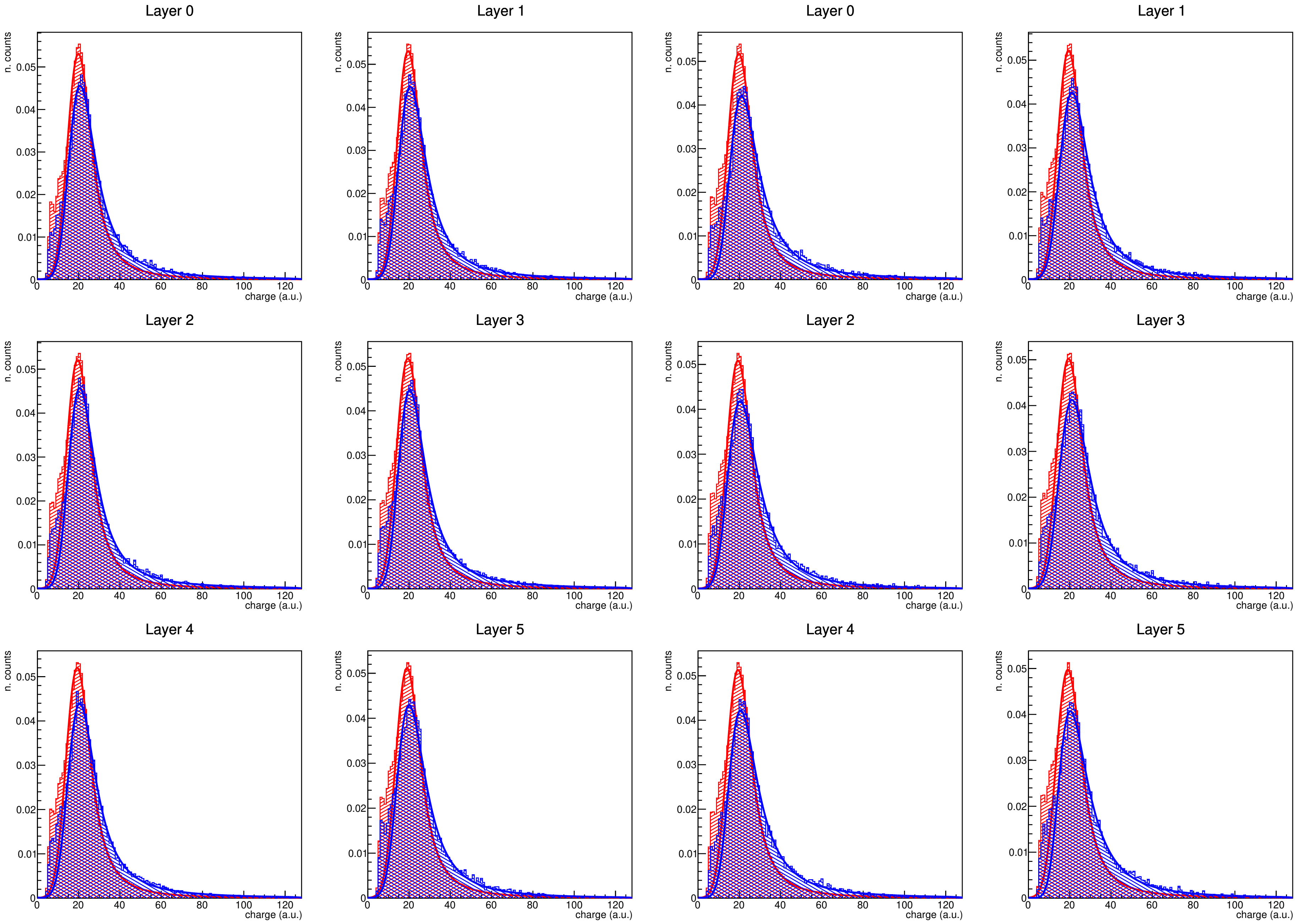


(a) negatively charged

(b) positively charged

Figure: Q_0 charge distribution for $p = 2.0\,\mathrm{GeV/c}$ to $3.0\,\mathrm{GeV/c}$.

E. LQ3D

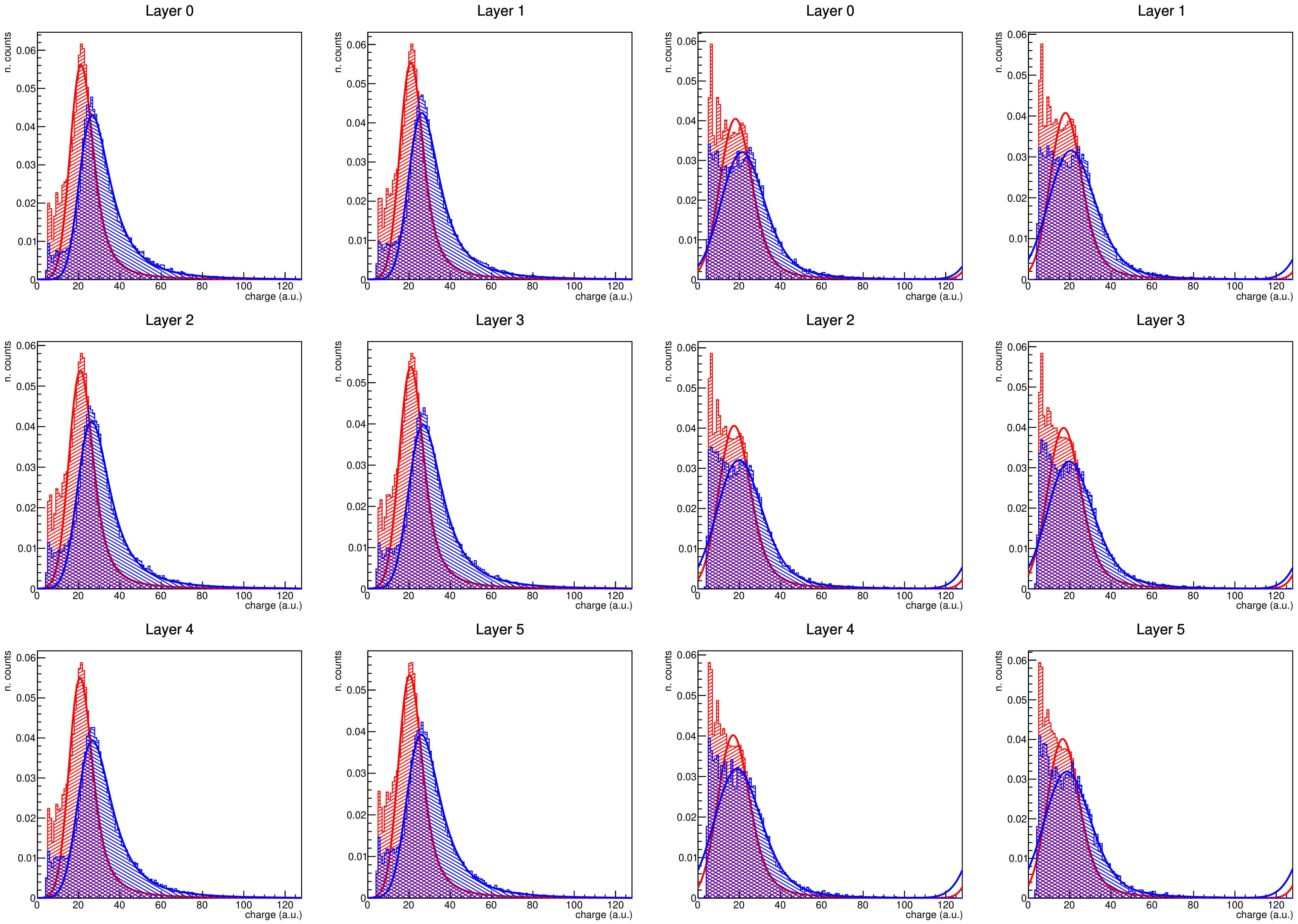


(a) negatively charged

(b) positively charged

Figure: Q_0 charge distribution for $p = 3.0\,\mathrm{GeV/c}$ to $8.0\,\mathrm{GeV/c}$.

E. LQ3D

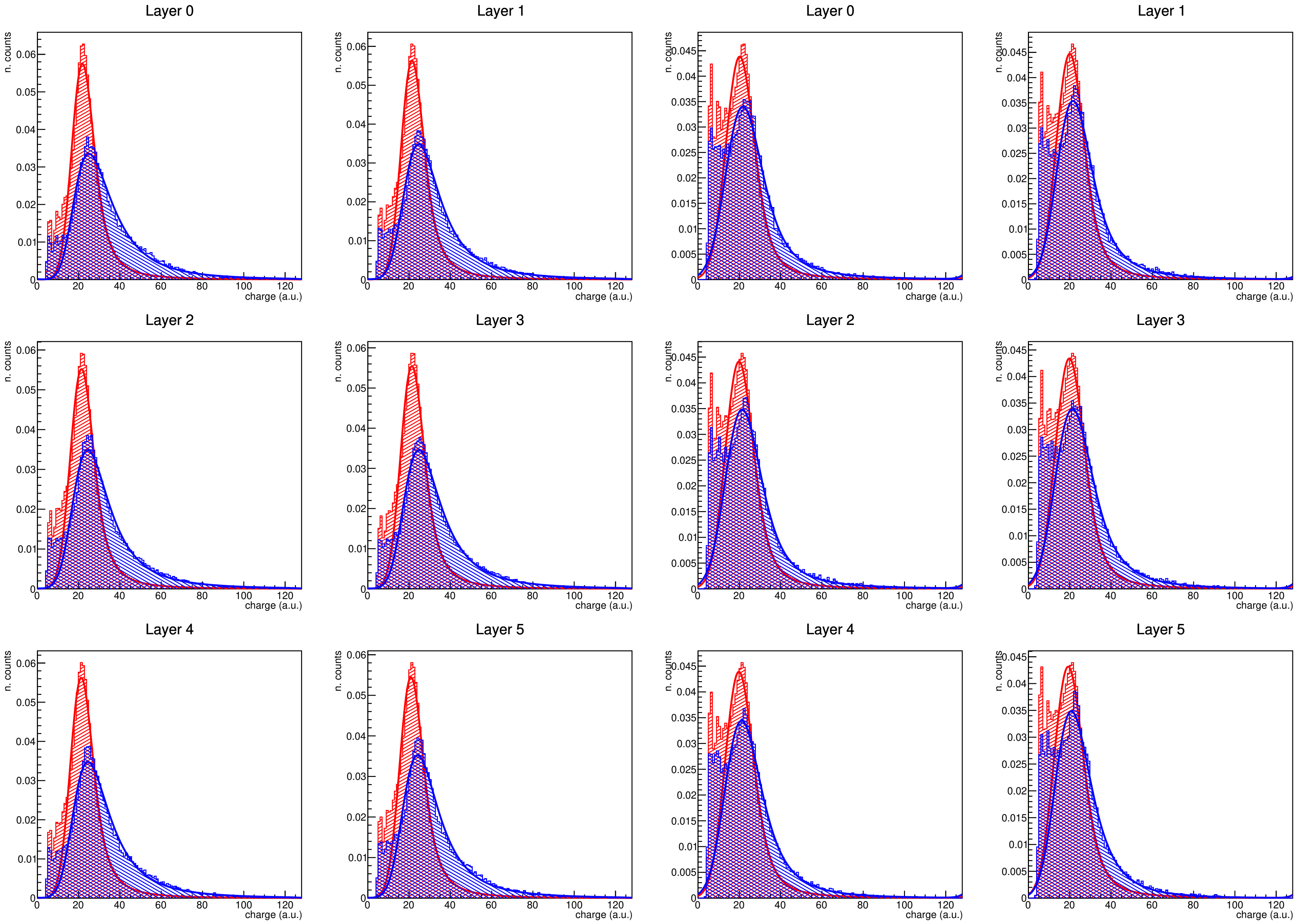


(a) negatively charged

(b) positively charged

Figure: Q_1 charge distribution for $p = 0.2\,\mathrm{GeV/c}$ to $1.0\,\mathrm{GeV/c}$.

E. LQ3D

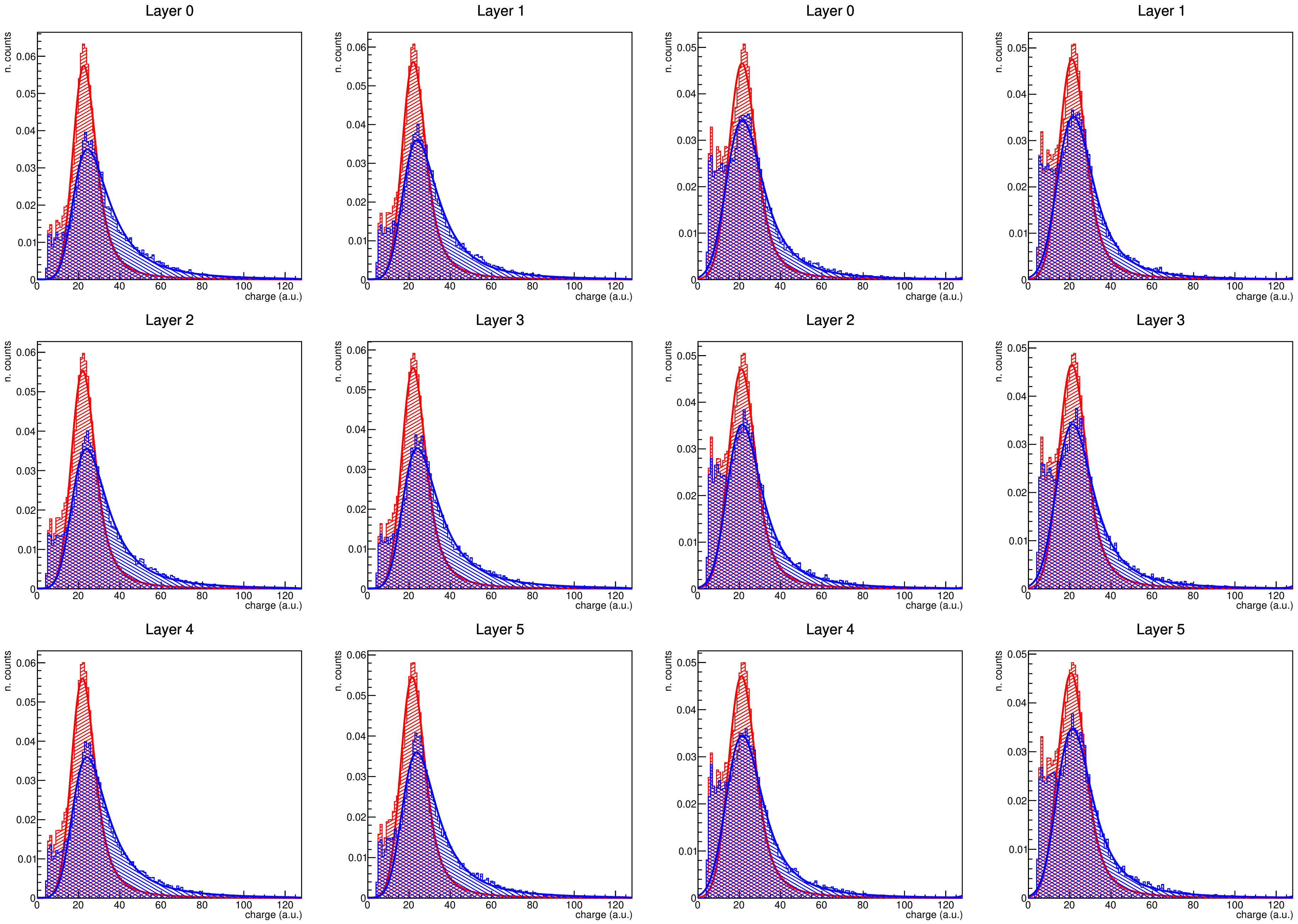


(a) negatively charged

(b) positively charged

Figure: Q_1 charge distribution for $p = 1.0\,\mathrm{GeV/c}$ to $2.0\,\mathrm{GeV/c}$.

E. LQ3D

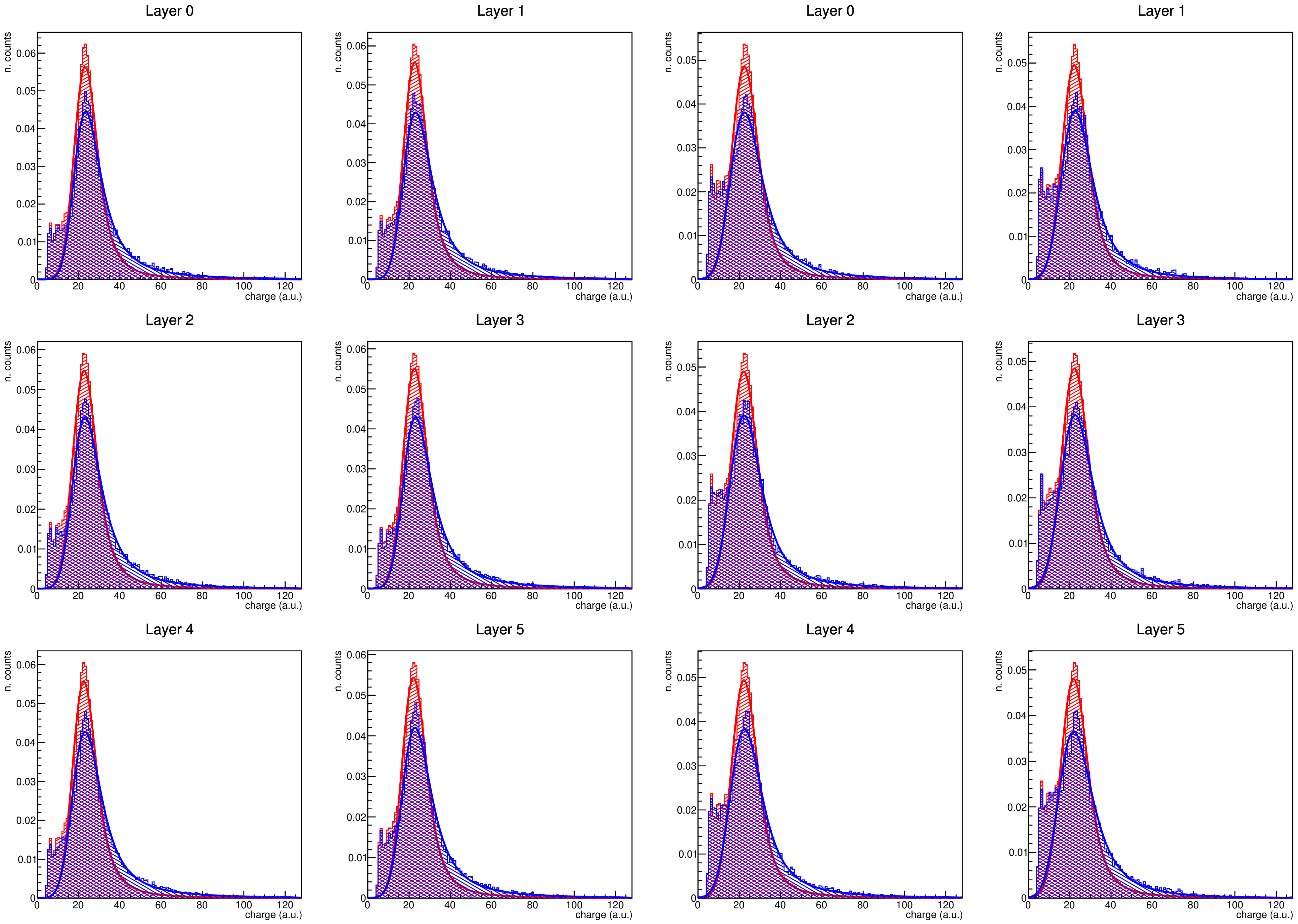


(a) negatively charged

(b) positively charged

Figure: Q_1 charge distribution for $p = 2.0\,\mathrm{GeV/c}$ to $3.0\,\mathrm{GeV/c}$.

E. LQ3D

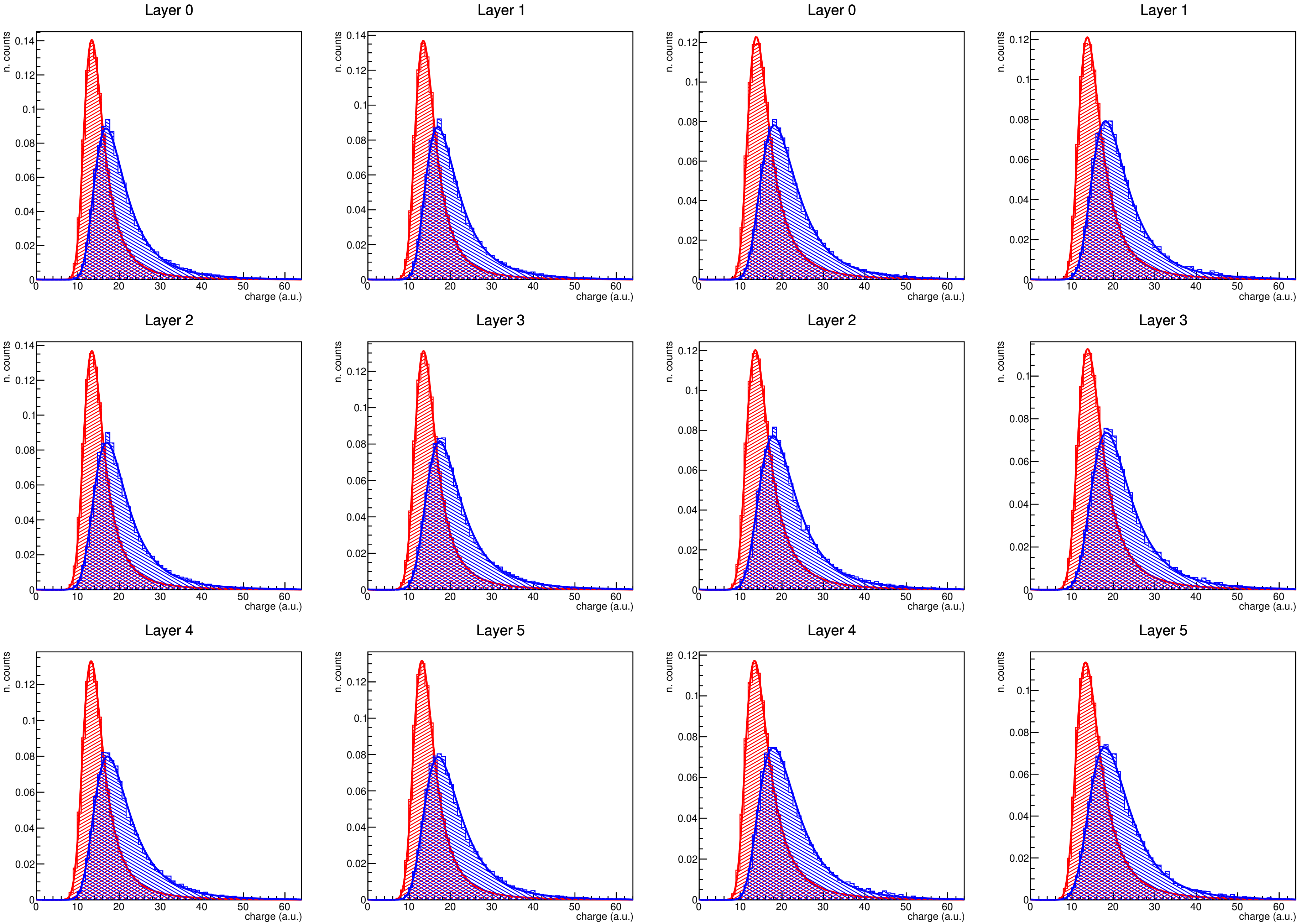


(a) negatively charged

(b) positively charged

Figure: Q_1 charge distribution for $p = 3.0\,\mathrm{GeV/c}$ to $8.0\,\mathrm{GeV/c}$.

E. LQ3D

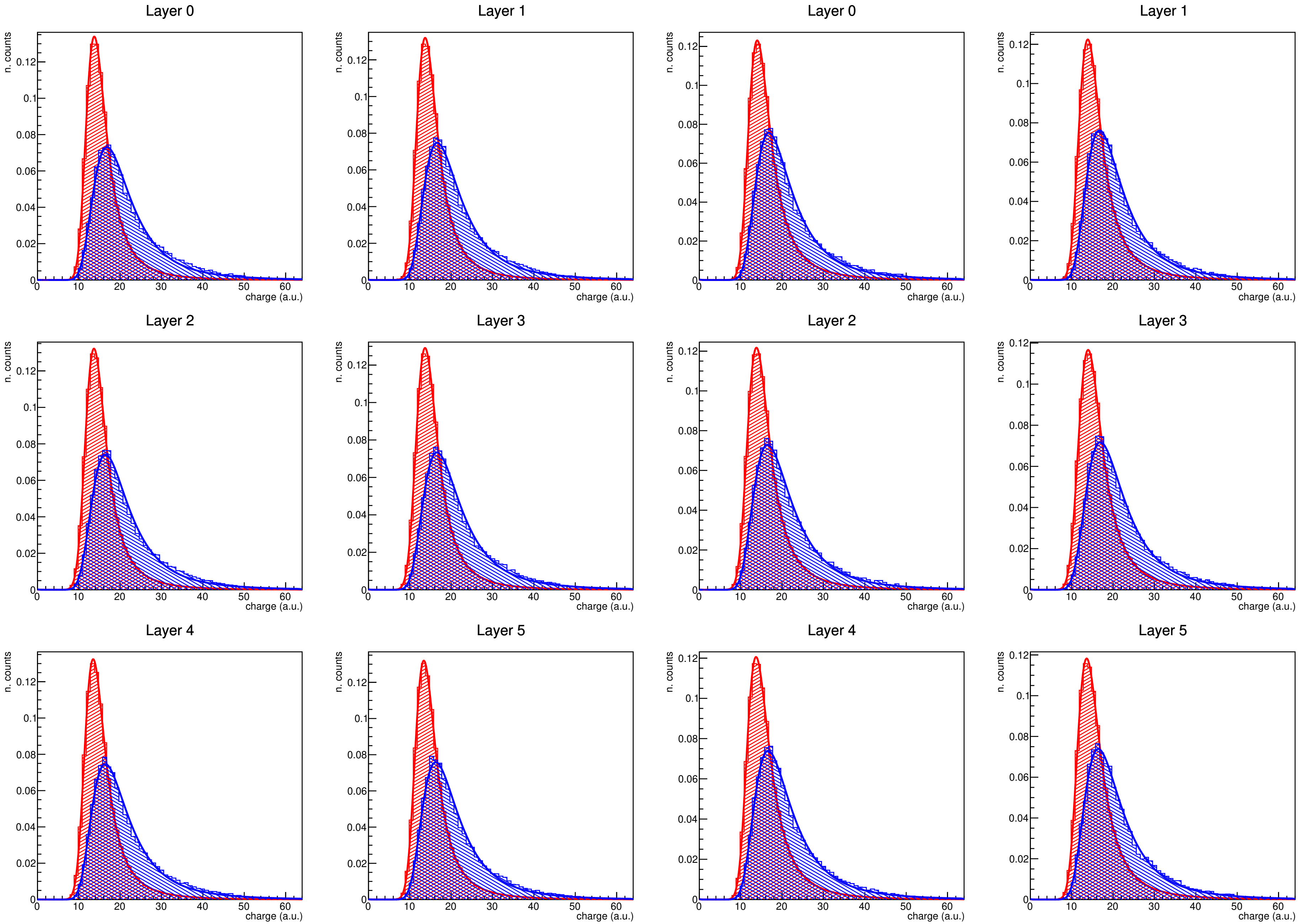


(a) negatively charged

(b) positively charged

Figure: Q_2 charge distribution for $p = 0.2\,\mathrm{GeV/c}$ to $1.0\,\mathrm{GeV/c}$.

E. LQ3D

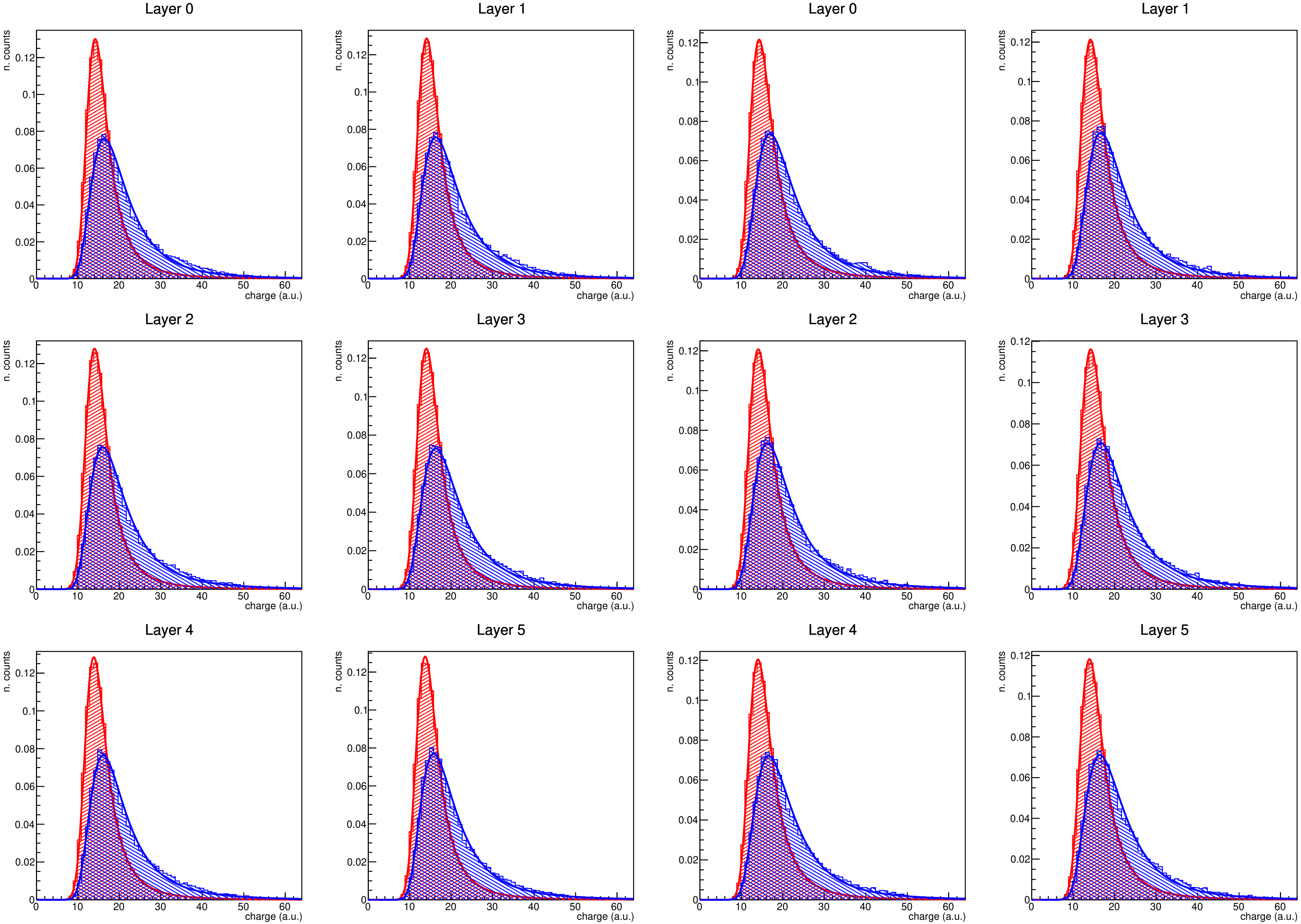


(a) negatively charged

(b) positively charged

Figure: Q_2 charge distribution for $p = 1.0\,\mathrm{GeV/c}$ to $2.0\,\mathrm{GeV/c}$.

E. LQ3D

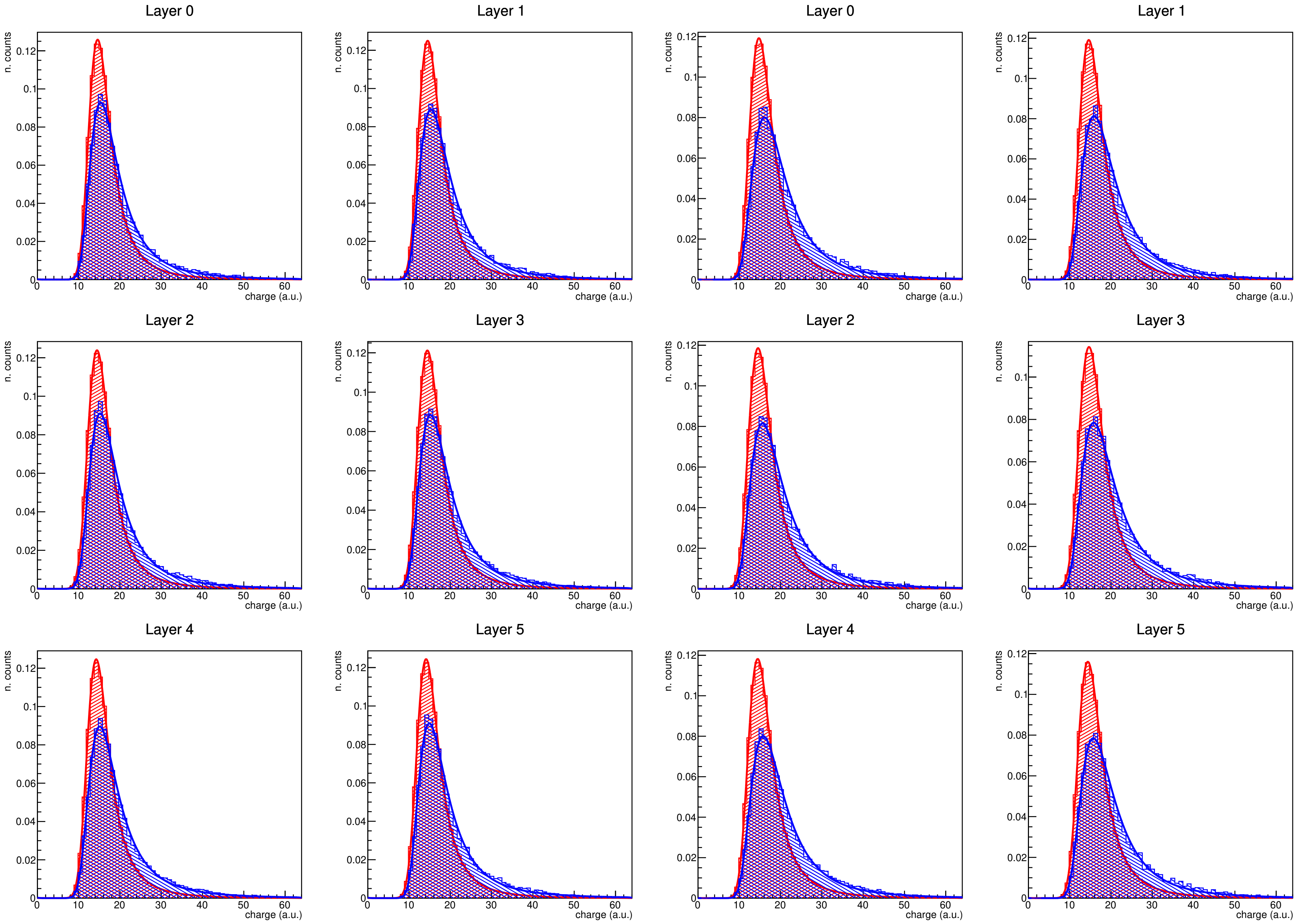


(a) negatively charged

(b) positively charged

Figure: Q_2 charge distribution for $p = 2.0\,\mathrm{GeV/c}$ to $3.0\,\mathrm{GeV/c}$.

E. LQ3D

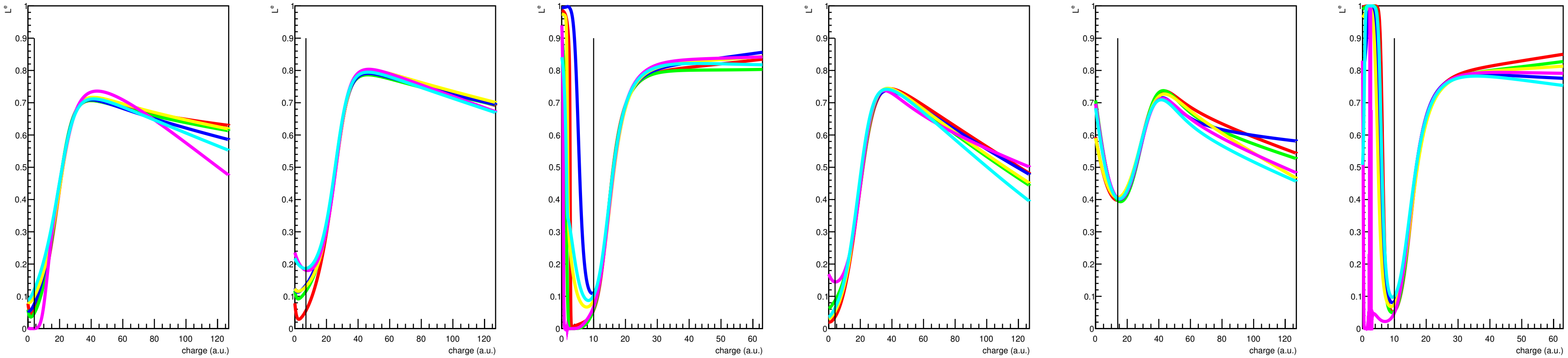


(a) negatively charged

(b) positively charged

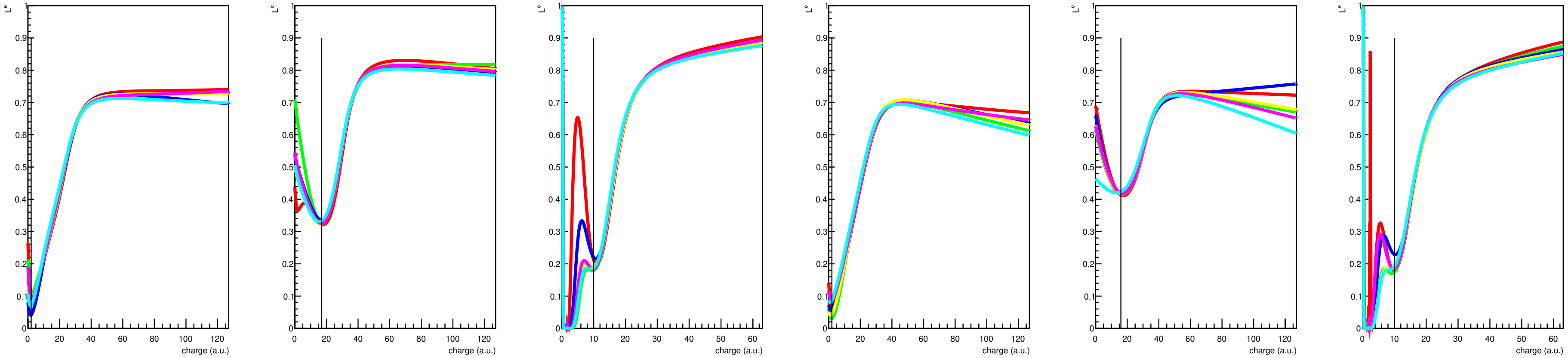
Figure: Q_2 charge distribution for $p = 3.0\,\mathrm{GeV/c}$ to $8.0\,\mathrm{GeV/c}$.

E. LQ3D



(a) negatively charged (b) positively charged

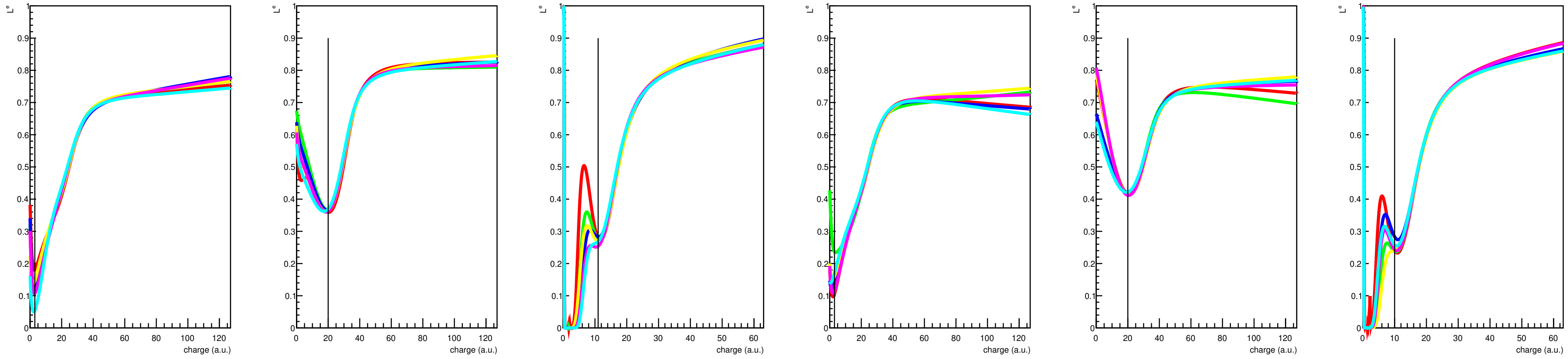
Figure: LUT for $Q_{0,1,2}$ (left to right) for $p = 1.0\,\text{GeV/c}$ to $1.0\,\text{GeV/c}$.



(a) negatively charged (b) positively charged

Figure: LUT for $Q_{0,1,2}$ (left to right) for $p = 1.0\,\text{GeV/c}$ to $2.0\,\text{GeV/c}$.

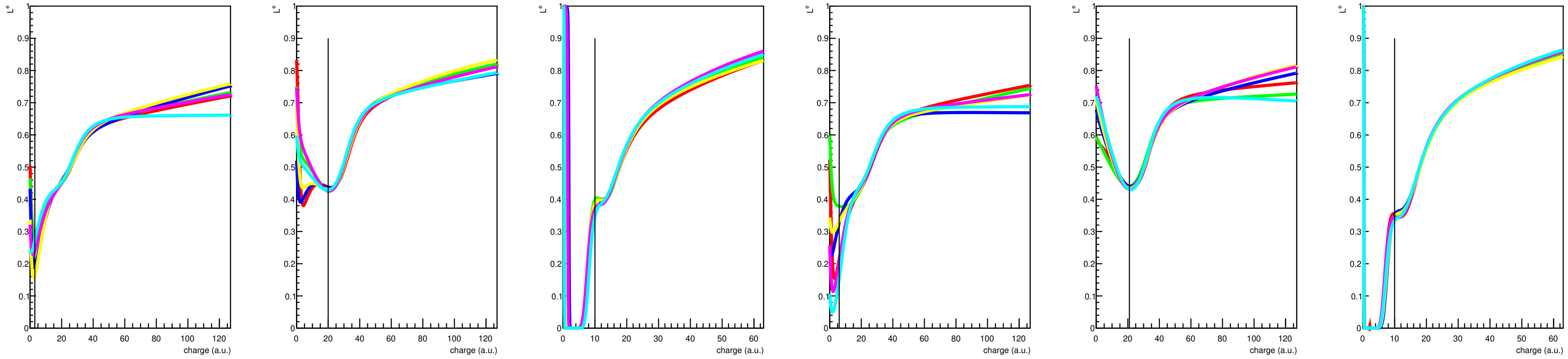
E. LQ3D



(a) negatively charged

(b) positively charged

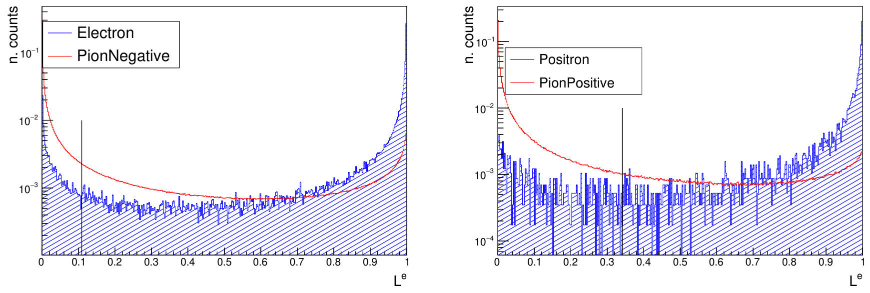
Figure: LUT for $Q_{0,1,2}$ (left to right) for $p = 2.0\,\mathrm{GeV/c}$ to $3.0\,\mathrm{GeV/c}$.



(a) negatively charged

(b) positively charged

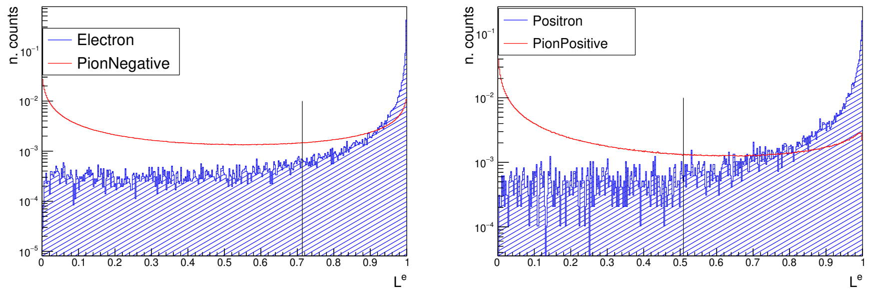
Figure: LUT for $Q_{0,1,2}$ (left to right) for $p = 3.0\,\mathrm{GeV/c}$ to $8.0\,\mathrm{GeV/c}$.



(a) negatively charged

(b) positively charged

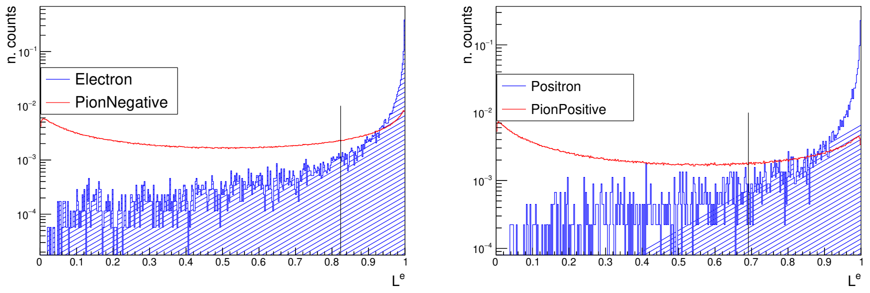
Figure: $p = 0.2$ GeV/c to 1.0 GeV/c.



(a) negatively charged

(b) positively charged

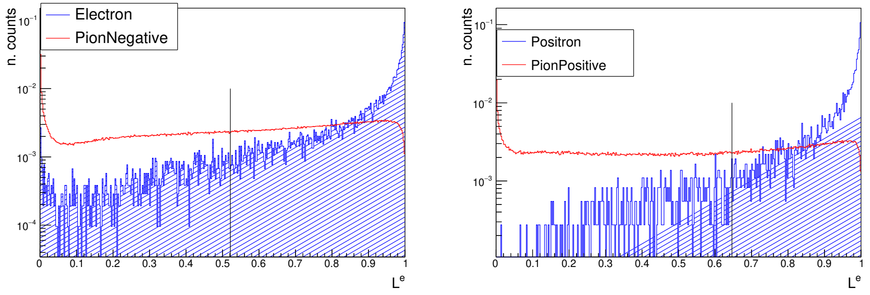
Figure: $p = 1.0$ GeV/c to 2.0 GeV/c.



(a) negatively charged

(b) positively charged

Figure: $p = 2.0$ GeV/c to 3.0 GeV/c.



(a) negatively charged

(b) positively charged

Figure: $p = 3.0$ GeV/c to 8.0 GeV/c.

Acknowledgments

> But there's no sense crying
> over every mistake.
> You just keep on trying
> 'til you run out of cake.
>
> *Portal - Still Alive*
> *(still my favorite video game)*

I would like to express my sincere gratitude to my supervisor Yvonne Pachmayer, whose guidance, support, and encouragement were instrumental in completing this Master's Thesis. Her deep knowledge, unwavering dedication, and critical feedback helped me to develop a comprehensive understanding of the research area. She gave me invaluable insights to overcome my challenges throughout this journey. Subsequently, I thank Dr. apl. Prof. Monica Dunford for being my second examiner.

I also extend my heartfelt appreciation to the members of my working group for their insightful discussions, technical expertise, and unwavering support. Their constructive feedback and diverse perspectives greatly enriched my research and enabled me to develop novel insights and ideas.

Additionally, I express my gratitude to Ole Schmidt, Martin Kroessen, Ilya Fokin and Andres Borquez for proofreading my thesis. Their attention to detail, critical feedback, and willingness to spend valuable time to help me improve my work have been invaluable. Also, their insightful comments and suggestions have helped me to refine my arguments, clarify my ideas, and improve the overall quality of my thesis. I genuinely appreciate the effort and care they put into reviewing my work.

Furthermore, I sincerely thank Sven Hoppner for being such a fantastic office mate during my thesis. His positive attitude has been a great source of motivation for me, and the minor distractions he provided were essential.

Finally, I would like to thank my family for their love, encouragement, and support. Their unwavering support and encouragement have been invaluable throughout my academic journey.

Thank you all for making this journey an enjoyable and unforgettable experience.

Cheers
Felix

Erklärung

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

Heidelberg, den 30.04.2023

Datum Unterschrift