UNIVERSITÀ DEGLI STUDI DI PADOVA

Dipartimento di Fisica e Astronomia "Galileo Galilei"

Master Degree in Physics

Final Dissertation

Study of charge asymmetries in pion and kaon identification with the Belle II detector

Thesis supervisor

Prof. Alessandro Gaz

Candidate

Cecilia Antonioli

Academic Year 2022/2023

Tutto, tutti, sempre.

A pupi lupi, il mio cuore.

Abstract

In this work, we study the performance of the charged kaon and pion identification at the Belle II experiment. The analysed data set corresponds to an integrated luminosity of 404 fb^{-1}, the data were collected from 2019 to 2022 and are compared with Monte Carlo simulation. Efficiencies and mis-identification rates of kaons and pions are measured using a sample of $D_s^+ \to \phi(K^+K^-)\pi^+$, they are defined as the ratio of signal yields after and before applying the identification selection to particle tracks. The signal yields are determined from fits to the invariant mass distributions of the resonances involved. Several preselection criteria are applied to reduce the background. The identification observables are studied as a function of the main kinematic variables: momentum p and $\cos\theta$. Furthermore, we measure the corresponding charge asymmetries and make a comparison of the results obtained from both real data and simulation. The goal is to determine whether the experimental data show significant charge asymmetries compared to those observed with the simulation. This study describes the performance of global particle identification at Belle II. In addition, we discuss the inclusion of calibration weights and evaluate the performance of individual sub-detectors.

In questo lavoro, si studiano le prestazioni di identificazione di kaoni e pioni carichi nell'esperimento Belle II. L'insieme dei dati analizzati corrisponde ad una luminosità integrata di 404 fb^{-1}, i dati sono stati raccolti dal 2019 al 2022 e sono confrontati con la simulazione Monte Carlo. Le efficienze e i tassi di errata identificazione di kaoni e pioni sono misurati utilizzando un campione di $D_s^+ \to \phi(K^+K^-)\pi^+$, e sono definiti come il rapporto tra gli eventi di segnale dopo e prima l'applicazione della selezione di identificazione alle tracce di particelle. Gli eventi di segnale sono determinati da fit delle distribuzioni di massa invariante delle risonanze coinvolte. Vengono applicati alcuni criteri di preselezione per ridurre il fondo. Le osservabili di identificazione sono studiate in funzione delle principali variabili cinematiche: quantità di moto p e $\cos\theta$. Inoltre, si misurano le corrispondenti asimmetrie di carica e si confrontano i risultati ottenuti dai dati reali e dalla simulazione. L'obiettivo è determinare se i dati sperimentali mostrano asimmetrie di carica significative rispetto a quelle osservate con la simulazione. Questo studio descrive le prestazioni globali di identificazione delle particelle a Belle II. Inoltre, si discute l'inclusione dei pesi di calibrazione e si valutano le prestazioni dei singoli rilevatori.

Contents

Chapter 1

Motivation and theoretical framework

1.1 Introduction

The Standard Model (SM) provides a precise description of elementary particles and their interactions. Despite the high precision of the experimental results that prove the validity of SM, there are many fundamental questions still unsolved. Among them there is the matter/anti-matter asymmetry observed in the universe [1]. CP violation, discussed in section 1.2.2, is a necessary condition for the evolution to a matter-dominated universe, however the CP violation predicted by the SM is several orders of magnitude too small to explain the observed asymmetry. Furthermore, the flavour mixing Cabibbo-Kobayashi-Maskawa matrix is roughly diagonal and the origin of this hierarchy is unknown, since its elements are free parameters of the theory. This may suggest the presence of some new flavour symmetry that exists unbroken at higher energy scales. Many new physics theories, including new particles and processes, have been proposed to explain the effects not described by the SM.

Experiments in high-energy physics are designed to address the unsolved questions through searches of New Physics (NP) using complementary approaches. At the energy frontier, the Large-Hadron-Collider (LHC) experiments [2], [3], [4] allow to directly search for NP effects in proton-proton collisions at the TeV scale. At the intensity frontier, signatures of NP can be observed through measurements of suppressed flavour physics reactions or from deviations from SM predictions. This is the approach of the Belle II experiment, introduced in chapter 2, which measures many processes with unprecedented precision and allows to appreciate any potential discrepancies due to new particles and processes occurring at mass scales greater than those achieved by the LHC.

The Belle II experiment is currently taking data at SuperKEKB e^+e^- collider. One of its main goals is the study of matter/anti-matter asymmetries in the decays of B and D mesons. The studies of CP violation phenomena show the different behaviour for matter and anti-matter, and time-dependent analyses allow to precisely measure fundamental parameters of the SM and to search for NP.
For many CP violation studies we compare probability amplitudes of particles and anti-particles decay processes, therefore it is necessary to determine the flavour of the mesons (whether they are $B\,(D)$ or $\overline{B}\,(\overline{D})$), even if their decay chain can not be fully determined. Particle tracking and particle identification (PID) are crucial information for reconstructing flavours and conducting CP violation studies, in particular excellent pion and kaon separation capabilities are required. Particles and anti-particles behave differently in their detection and we expect different performance results in the determination of PID efficiencies, for example depending on the charge. These spurious asymmetries arising in the interaction of particles with the detector material, must be under very firm control because they impact the sensitivity of the experiment to CP violation asymmetries and their related systematic uncertainties must be considered.
Identification performance of charged kaons and pions, measuring identification efficiencies and mis-identification rates, are usually studied using $D^{*+} \to D^0(K^-\pi^+)\,\pi^+$ decays [5], which is a relatively clean sample and charged pions and kaons can be obtained with minimal selection criteria. An alternative sample is $D_s^+ \to \phi(K^+K^-)\,\pi^+$ and its conjugated mode. This topology allows to study the

identification efficiencies of the positive and negative kaons from the same sample, and also to study the identification efficiencies of the accompanying pion, regardless of the charge asymmetries arising from the rest of the event. The goal of this work is to evaluate the PID performance at Belle II using $D_s^+ \to \phi(K^+K^-)\pi^+$ as a control sample, and to provide a measurement of the charge asymmetries in the identification of charged pions and kaons. Specifically, the PID efficiencies are studied as a function of the main kinematic variables, for both real data and Monte Carlo simulation. Our study aims to compare the performance of global identification with that of the Time Of Propagation detector alone, the main PID device that covers the Belle II barrel region. We also evaluate the performance of the Central Drift Chamber detector alone, the main tracking device. Possible discrepancies between experimental data and simulation results are investigated to determine whether they arise from a detection effects of the PID system or whether they originate from the tracking of charged particles.

1.2 The Standard Model

The Standard Model is the theory describing the interaction of particles through the strong, weak and electromagnetic fundamental forces (gravitational force is neglected). The particle content of the theory is summarized in Fig. 1.1. There are twelve fermions, six leptons and six quarks, categorized in three generations. For each particle of matter (the fermions) there is the corresponding anti-matter particle. The force carriers are the spin-1 gauge bosons, namely the photon, the gluons and the $W^{\pm}$ and Z^0 bosons. Finally, there is the Higgs boson H, the particle associated to the Higgs field.
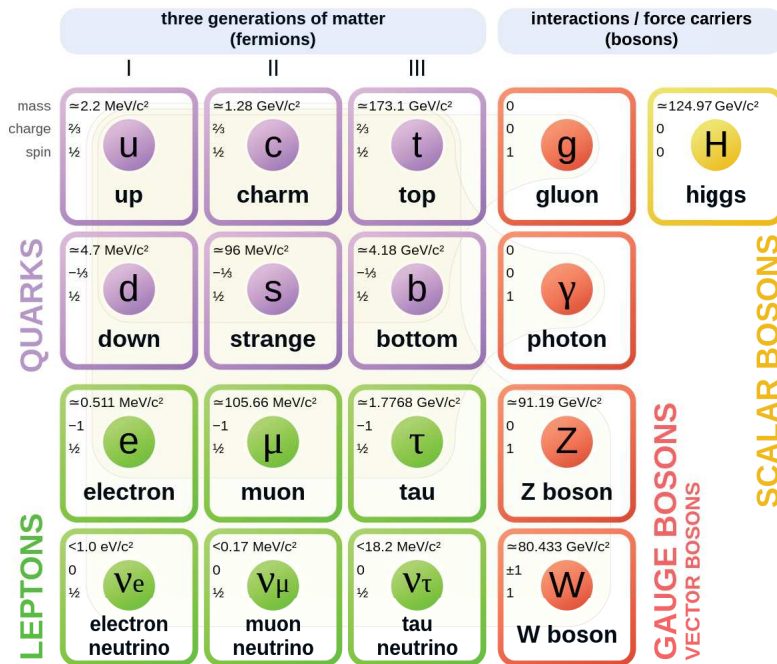
Figure 1.1: Elementary particle content of the Standard Model.

The SM Lagrangian can be written as the sum of two components

$$\mathcal{L}_{\text{SM}} = \mathcal{L}_{\text{QCD}} + \mathcal{L}_{\text{EW}} \tag{1.1}$$

where $\mathcal{L}_{\text{QCD}}$ is the Quantum-Chromo-Dynamics Lagrangian, describing strong interactions and invariant under local $SU(3)_C$ transformations, and $\mathcal{L}_{\text{EW}}$ is the Electro-Weak Lagrangian which includes electromagnetic and weak interactions (both neutral and charged current) with symmetry group $SU(2)_L \otimes U(1)_Y$. An important ingredient of the theory is the possibility of spontaneous symmetry breaking of the gauge group into a smaller group. This spontaneous symmetry breaking plays a key role in the Higgs mechanism: gauge bosons and fermions interact with the Higgs field acquiring their mass. The choice of the ground state breaks the SM symmetry group

$$G_{\text{SM}} = SU(3)_C \otimes SU(2)_L \otimes U(1)_Y$$

into

$$G_{\text{SM}}^{\text{SSB}} = SU(3)_C \otimes U(1)_{em}$$

resulting in three massive bosons $W^{\pm}$ and Z^0, eight massless gluons and a massless photon.

The interaction of fermions with the Higgs field is described by the Yukawa term of the SM Lagrangian, that, considering only the quark terms, reads

$$\mathcal{L}_{\text{Y}} = -\overline{Q}_L^i\, Y_{ij}^d\, \Phi\, d_R^j\, -\, \overline{Q}_L^i\, Y_{ij}^u\, \tilde{\Phi}\, u_R^j\, +\, \text{h.c.} \tag{1.2}$$

where $Y^{u,d}$ are the complex 3×3 Yukawa matrices, Φ is the Higgs field and $\tilde{\Phi} = i\sigma_2\Phi$ where σ_2 is the Pauli matrix, i and j are flavour indices. $\overline{Q}_L^i$ are the left-handed $SU(2)_L$ doublets, u_R^i and d_R^i are the right-handed $SU(2)_L$ singlets in the weak-eigenstate basis. After symmetry breaking, when the Higgs field acquires a vacuum expectation value $\langle\Phi\rangle = (0\,,v/\sqrt{2})$ ($v = 246\,\text{GeV}$), it is possible to diagonalize the Y matrices by means of four unitary matrices $V_{L,R}^{u,d}$. We pass to the mass basis, with fermion masses

$$M_{\text{diag}}^f = \frac{v}{\sqrt{2}}\, V_L^f\, Y^f\, V_R^{f\,\dagger}$$

where $f = u, d$. This change of basis does not affect the other fermionic sectors, with the exception of the weak charged current interaction sector: here the interaction is among u-type and d-type fermions and the unitarity property of the transformation can not be exploited. The charged current interaction Lagrangian in the mass basis reads

$$\mathcal{L}_{\text{CC}}^{\text{mass}} = -\frac{g}{\sqrt{2}} \left[\overline{u}_L^i\, V_{\text{CKM}}^{ij}\, W^+\!\!\!\!/\;\, d_R^j\, +\, \text{h.c.} \right] \tag{1.3}$$

where the flavour mixing matrix $V_{\text{CKM}} = V_L^u V_R^{d\,\dagger}$ has been introduced.

1.2.1 The CKM matrix

The Cabibbo-Kobayashi-Maskawa (CKM) matrix [6] is a 3×3 unitary matrix with flavour mixing properties that gives the relation between the weak eigenstates and the mass eigenstates

$$V_{\text{CKM}} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix}$$

A general $n \times n$ complex matrix is specified by $2n^2$ real parameters, then the request of unitarity ($VV^{\dagger} = V^{\dagger}V = \mathbb{1}$) removes n^2 free parameters. With n quark generations we are free to redefine $2n - 1$ phases of the quark mass eigenstates. Therefore, the physically meaningful parameters are $2n^2 - n^2 - (2n - 1) = (n - 1)^2$. For $n = 3$ quark generations, the physics of the CKM matrix is determined by four real parameters: three mixing angles and one irreducible phase. We are left with one complex phase that can not be removed by a redefinition of the quark fields, this irreducible phase is responsible for CP violation in the SM.

The standard parameterization of the CKM matrix is obtained as the product of three rotation matrices, one of which contains the complex phase δ

$$\begin{aligned} V_{\text{CKM}} &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & c_{23} & s_{23} \\ 0 & -s_{23} & c_{23} \end{pmatrix} \begin{pmatrix} c_{13} & 0 & s_{13}e^{-i\delta} \\ 0 & 1 & 0 \\ -s_{13}e^{i\delta} & 0 & c_{13} \end{pmatrix} \begin{pmatrix} c_{12} & s_{12} & 0 \\ -s_{12} & c_{12} & 0 \\ 0 & 0 & 1 \end{pmatrix} = \\ &= \begin{pmatrix} c_{13}c_{12} & c_{13}s_{12} & s_{13}e^{-i\delta} \\ -c_{23}s_{12} - s_{23}s_{13}e^{i\delta}c_{12} & c_{23}c_{12} - s_{23}s_{13}e^{i\delta}s_{12} & s_{23}c_{13} \\ s_{23}s_{12} - c_{23}s_{13}e^{i\delta}c_{12} & -s_{23}c_{12} - c_{23}s_{13}e^{i\delta}s_{12} & c_{23}c_{13} \end{pmatrix} \end{aligned} \tag{1.4}$$

where θ_{ij} are the mixing angles between the i-th and j-th generation, and $c_{ij} = \cos\theta_{ij}$, $s_{ij} = \sin\theta_{ij}$. The rotation angle between the first two generations, θ_{12}, is known as the Cabibbo angle.

From experimental observations $|V_{ub}|^2 \ll |V_{cb}|^2 \ll |V_{us}|^2$ (alternatively $s_{13} \ll s_{23} \ll s_{12} \ll 1$), and it is convenient to exhibit this hierarchy introducing an expansion. By defining

$$s_{12} \equiv \lambda \qquad s_{23} \equiv A\lambda^2 \qquad s_{13}e^{-i\delta} \equiv A\lambda^3(\rho - i\eta)$$

the Wolfenstein parameterization [7] of the CKM matrix is

$$V_{\text{CKM}} = \begin{pmatrix} 1 - \lambda^2/2 & \lambda & A\lambda^3(\rho - i\eta) \\ -\lambda & 1 - \lambda^2/2 & A\lambda^2 \\ A\lambda^3(1 - \rho - i\eta) & -A\lambda^2 & 1 \end{pmatrix} + \mathcal{O}(\lambda^4) \tag{1.5}$$

The unitarity of the CKM matrix

$$\sum_{j=u,c,t} V_{ji}V_{jk}^* = \delta_{ik} \tag{1.6}$$

leads to nine distinct complex relations among the matrix elements. The six relations with $i \neq k$ can be represented geometrically as six unitarity triangles (three independent) in the complex plane. The length of the sides of these triangles is related to the magnitude of the different CKM matrix elements and the angles are related to the CP violation in the decay of quarks and anti-quarks. Therefore, measurements of the different sides and angles allows to constrain the CKM matrix.

The system of B mesons is particularly interesting for CP violation measurements because a sizable CP violation is expected in the b sector. The unitarity triangle particularly relevant for the study of B physics is

$$V_{ud}V_{ub}^* + V_{cd}V_{cb}^* + V_{td}V_{tb}^* = 0 \tag{1.7}$$

Since the terms are of the same order in λ (i.e. λ^3), this equation defines a non-degenerate triangle shown in Fig. 1.2. Usually, the sides are rescaled and rotated so that one of them lies on the real axis and two of its vertices have coordinates (0,0) and (1,0). The amount of CP violation generated by the CKM matrix can be determined by the coordinates $(\bar\rho, \bar\eta)$ of the upper vertex, related to (ρ, η) by

$$\bar\rho = \rho\left(1 - \frac{\lambda^2}{2}\right) \qquad \bar\eta = \eta\left(1 - \frac{\lambda^2}{2}\right)$$

The angles of the unitarity triangle are related to the elements of the CKM matrix through the relations

$$\phi_1 = \beta = \arg\left(\frac{V_{cd}V_{cb}^*}{V_{td}V_{tb}^*}\right) \qquad \phi_2 = \alpha = \arg\left(\frac{V_{td}V_{tb}^*}{V_{ud}V_{ub}^*}\right) \qquad \phi_3 = \gamma = \arg\left(\frac{V_{ud}V_{ub}^*}{V_{cd}V_{cb}^*}\right)$$
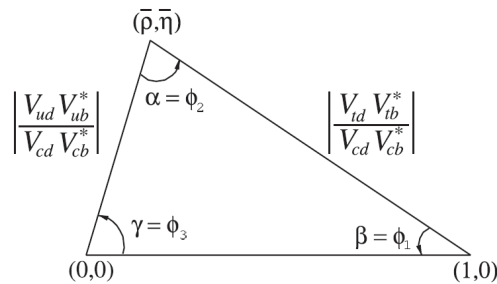
Figure 1.2: The unitarity triangle.

In Fig. 1.3, the current results for the unitarity triangle parameters are shown. The Belle II experiment, described in chapter 2, aims to collect a $B\bar{B}$ sample corresponding to 50 ab^{-1}, allowing to precisely measure the unitarity triangle. In Fig. 1.4 the expected precision on the unitarity triangle parameters with 50 ab^{-1} is shown.
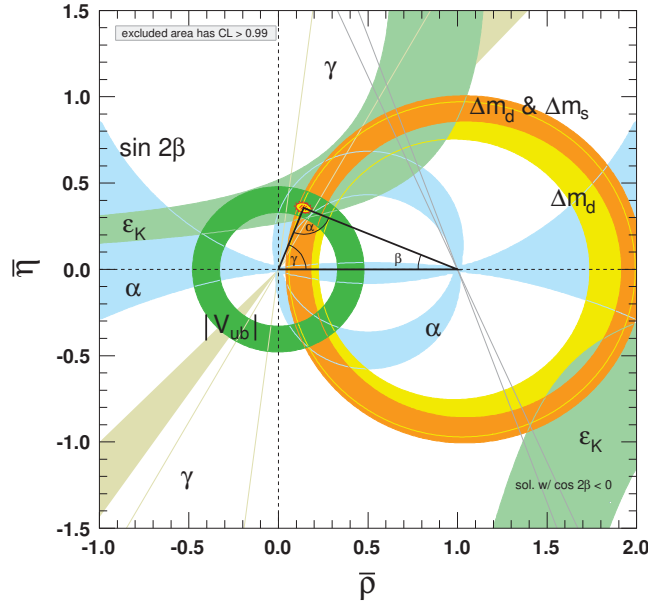
Figure 1.3: Current constraints on the $(\overline{\rho}, \overline{\eta})$ plane [8].
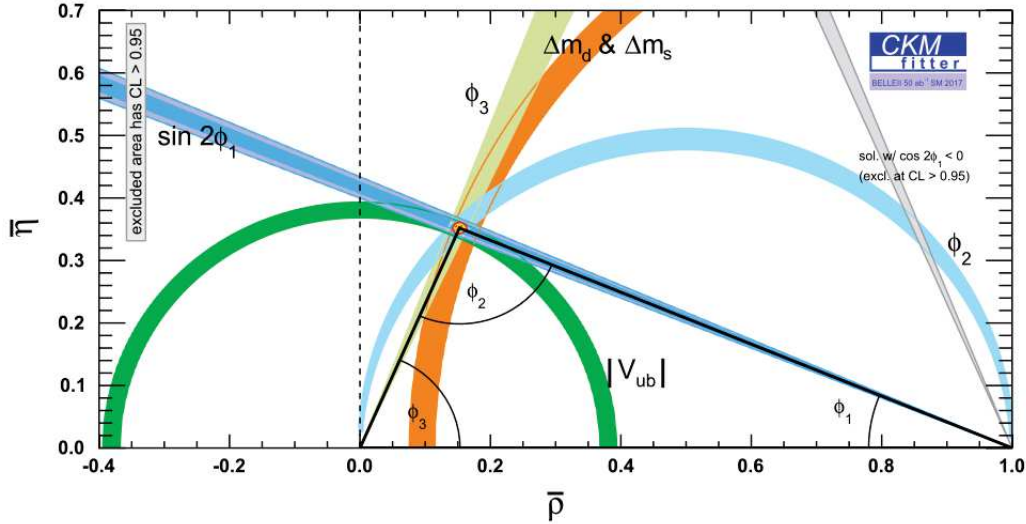
Figure 1.4: Unitarity triangle fit extrapolated to 50 ab^{-1} [9].

1.2.2 CP violation

Symmetries are a very important concept in physics since the invariance of physics equations under a certain transformation is connected to conservation laws by the Noether theorem.

In particle physics there are three important discrete symmetries: parity (P), charge conjugation (C) and time reversal (T). The parity operator P action generates an inversion of the spatial field coordinates $\vec{x} \rightarrow -\vec{x}$. The charge conjugation operation C changes the sign of all charges. The time reversal operation T generates an inversion of the time coordinate $t \rightarrow -t$.

P and C are good symmetries of the strong and electromagnetic interactions, while the weak interactions violate P, C and CP. The parity violation was firstly observed in 1957 in the β decay of ^{60}Co [10], the charge conjugation violation was determined from neutrino helicity measurement by studying the electron capture in ^{152}Eu [11]. CP violation was then observed in the neutral kaon system, with the detection of the unexpected CP violating decay $K_{\mathrm{L}} \rightarrow \pi^{+}\pi^{-}$ [12]. The SM accounts for CP violation in the weak sector through the presence of a complex phase in the three-generations flavour mixing CKM matrix, introduced in 1973 by Kobayashi and Maskawa [6]. In 2001, the observation of CP

violation in the B meson system by the Belle and BaBar experiments provided a clear proof of the Kobayashi and Maskawa scheme for CP violation in the SM [13], [14]. Furthermore, CP violation has been observed in B_s^0 decays [15] and, more recently, in charm decays [16]. Future measurements of CP violation in K, D, B, and B_s^0 meson decays will provide additional constraints on the flavour parameters of the SM, and might be sensitive to contribution from NP.

In the SM there are three types of CP violation effects that can occur in the quark sector:

- **Violation in decays** (or **direct CP violation**)
  Let M be a meson decaying into the final state f. If CP were conserved, the probability of $M \to f$ should be equal to the probability of $\overline{M} \to \bar{f}$, where $\overline{M}$ is the anti-meson and $\bar{f}$ the conjugate state of f. Instead, in the direct violation of the CP symmetry the two probabilities are not equal. If we call A_f and $\overline{A}_{\bar{f}}$ the decay amplitudes, we can define the CP asymmetry

  $$\mathcal{A}_f = \frac{|\overline{A}_{\bar{f}}|^2 - |A_f|^2}{|\overline{A}_{\bar{f}}|^2 + |A_f|^2} \tag{1.8}$$

  This type of CP violation can be observed in both neutral and charged B mesons, for example in $B^0 \to K^+ \pi^-$ decays [8].

- **Violation in mixing** (or **indirect CP violation**)
  It happens when the eigenstates of the free Hamiltonian are not CP eigenstates. The probability of a M^0 to oscillate into a $\overline{M}^0$ is different from the probability of the conjugate process, oscillation occurs because mass and flavour eigenstates are different. The indirect CP violation can be measured via the asymmetry of the "wrong-sign" semileptonic neutral meson decays $M^0, \overline{M}^0 \to \ell^{\pm} X^{\mp}$ induced by oscillations

  $$\mathcal{A}_{\mathrm{SL}} = \frac{\Gamma(\overline{M}^0_{\mathrm{phys}}(t) \to \ell^+ X^-) - \Gamma(M^0_{\mathrm{phys}}(t) \to \ell^- X^+)}{\Gamma(\overline{M}^0_{\mathrm{phys}}(t) \to \ell^+ X^-) + \Gamma(M^0_{\mathrm{phys}}(t) \to \ell^- X^+)} \tag{1.9}$$

  It is a small effect that has been observed in the neutral kaon system. Note that this asymmetry of time-dependent decay rates is actually time-independent.

- **Violation in the interference between decays with and without oscillation**
  This type of CP violation is due to the interference of the decay without mixing $M^0 \to f$ and the decay with mixing $M^0 \to \overline{M}^0 \to f$. This effect can be observed in decays to final states common to both M^0 and $\overline{M}^0$. For final CP eigenstates f_{CP}, we can define the time-dependent asymmetry

  $$\mathcal{A}_{f_{CP}}(t) = \frac{\Gamma(\overline{M}^0_{\mathrm{phys}}(t) \to f_{CP}) - \Gamma(M^0_{\mathrm{phys}}(t) \to f_{CP})}{\Gamma(\overline{M}^0_{\mathrm{phys}}(t) \to f_{CP}) + \Gamma(M^0_{\mathrm{phys}}(t) \to f_{CP})} \tag{1.10}$$

  This CP asymmetry can be very large and can be cleanly related to CKM angles. This type of CP violation was observed for the first time in 2001 by the Belle and BaBar experiments in the decay of $B^0, \overline{B}^0$ to the CP eigenstate $J/\psi K^0$ [13], [14].

All of the observed CP asymmetries are consistent with the SM predictions of CP violation in the quark sector. Similar effects could also occur in decays of baryons, but have not yet been observed. Given that neutrino masses and lepton mixing have been established, it is expected that CP is violated also in the lepton sector. Discovering CP violation in the lepton sector is one of the main goals of current and near-future experiments. CP violation has not yet been observed in processes involving the t quark, nor in flavour-conserving processes. For these, any significant observation would be a clear indication of physics beyond the SM.

The B system provides the ideal environment for measurements of CP violation, indeed the effects observed are larger. Studies of charmless hadronic B decays give access to the angle α/ϕ_2, the least known angle of the CKM unitarity triangle, and probe contributions of NP in processes mediated by loop decay amplitudes. Charmless B decays are decays not mediated by $b \to c$ transitions, the

underlying quark level transitions are the tree level $b \to u$ and loop level $b \to d, s$ transitions. The most promising determination of α/ϕ_2 relies on the analysis of multiple isospin-related $B \to \pi\pi$ decays ($B^0 \to \pi^0\pi^0$, $B^+ \to \pi^+\pi^0$, and $B^0 \to \pi^+\pi^-$) [17]. In addition, isospin symmetry can be employed to build sum-rules, i.e. linear combinations of branching fractions and direct CP asymmetries, to test SM predictions. For the set of $B \to K\pi$ decays ($B^0 \to K^+\pi^-$, $B^+ \to K^0\pi^+$, $B^+ \to K^+\pi^0$, and $B^0 \to K^0\pi^0$) there is a reliable and sensitive test based on comparing the observed value of the sum-rule and the SM expectation [18]. Belle II is the only experiment capable of measuring jointly, and within a consistent experimental environment, all the isospin-related decays.

In these studies, particle identification of hadrons with opposite charges plays a very important role because the PID performance of pions and kaons affects directly the sensitivity of CP asymmetry measurements, and related corrections and uncertainties must be estimated. In particular, the measured asymmetries

$$\mathcal{A} = \mathcal{A}_{CP} + \mathcal{A}_{\mathrm{det}}$$

are the sum of the genuine CP asymmetry $\mathcal{A}_{CP}$ and the instrumental asymmetry $\mathcal{A}_{\mathrm{det}}$ due to differences in interaction and reconstruction probabilities between particles and anti-particles [19]. We estimate the instrumental asymmetry for charged pions and kaons using control modes, i.e. $D^+ \to K^0_{\mathrm{S}}\pi^+$ and $D^0 \to K^-\pi^+$. Therefore, it is crucial to have an excellent identification performance and a comprehensive understanding of charge dependent efficiencies, and their related asymmetries, through the study of control samples, such as $D^+_s \to \phi(K^+K^-)\pi^+$.

Chapter 2

SuperKEKB and Belle II

2.1 The SuperKEKB collider

SuperKEKB is an asymmetric-energy electron-positron collider located in Tsukuba, Japan, constructed by upgrading the KEKB B-Factory. The design luminosity of SuperKEKB is $6 \cdot 10^{35}$ cm^{-2} s^{-1} (30 times higher than that achieved by KEKB) and the goal is to accumulate an integrated luminosity of 50 ab^{-1} [20]. The high luminosity is achieved increasing the beam currents and significantly reducing the beam size at the Interaction Point (IP) with the use of the nano-beam scheme. During Phase III of SuperKEKB in early 2019 until mid-2022, the Belle II detector has recorded a total integrated luminosity of the order of 400 fb^{-1}, while the maximum instantaneous luminosity achieved is $4.65 \cdot 10^{34}$ cm^{-2} s^{-1} [21]. In Fig. 2.1 the recorded luminosity profile is shown.
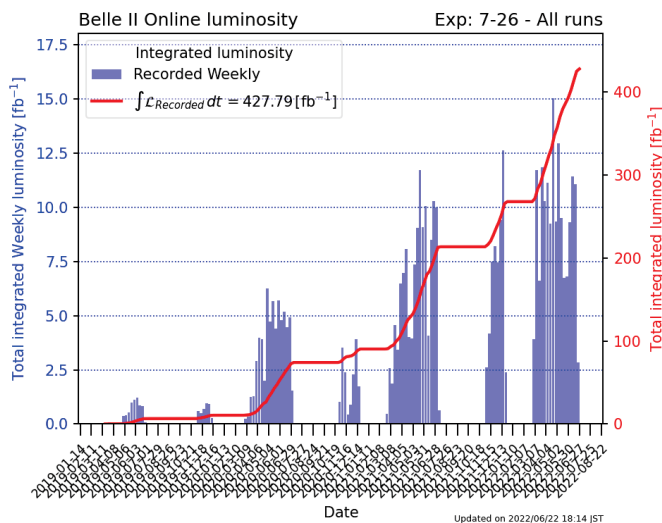
Figure 2.1: Total recorded integrated luminosity before Long Shutdown 1 [22].

The SuperKEKB configuration is shown in Fig. 2.2. The collider complex consists of a 7 GeV electron ring (HER, High-Energy Ring), a 4 GeV positron ring (LER, Low-Energy Ring), and an injector linear accelerator (LINAC) with a 1.1 GeV positron Damping Ring (DR). The SuperKEKB accelerator has the capacity to deliver e^+e^- collisions in the center-of-mass energy range from just below the $\Upsilon(1S)$ (9.46 GeV) to just above the $\Upsilon(6S)$ (11.24 GeV). The vast majority of data are taken at the center-of-mass energy $\sqrt{s} = 10.58$ GeV corresponding to the $\Upsilon(4S)$ resonance, the lightest that can decay strongly to a pair of B mesons. The asymmetric beam energies provide a Lorentz boost to the center-of-mass system with $\beta\gamma = 0.28$. As a consequence, the B or D mesons produced in the collision travel an appreciable distance before decaying, allowing to precisely measure lifetimes, mixing parameters and CP violation.
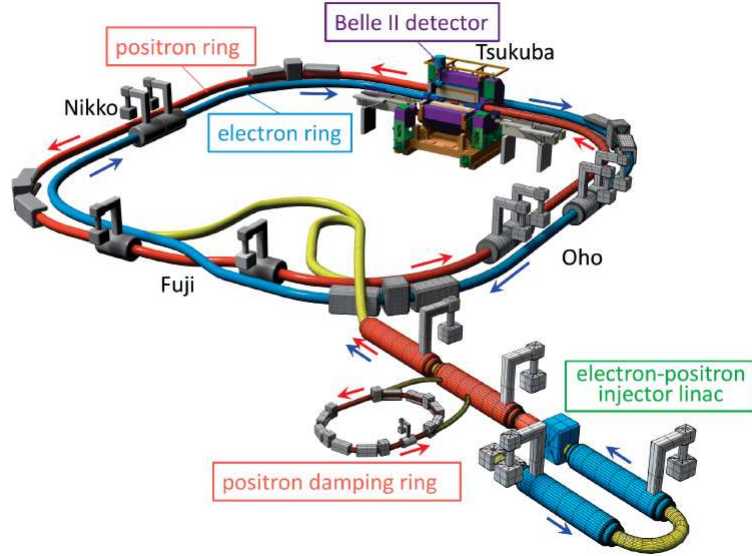
Figure 2.2: Schematic view of SuperKEKB [20].

2.2 Physics program of Belle II at SuperKEKB

The primary physics goals of Belle II are to search of NP in the flavour sector at the intensity frontier, measuring many processes with unprecedented precision and allowing to appreciate discrepancies with the SM predictions due to new particles and processes occurring at mass scales higher than the ones reached by the LHC.

As a next generation flavour factory, one of the questions addressed by the Belle II experiment is the explanation of observed matter/anti-matter asymmetries. New insights into the presence of new CP violating phases will come, for example, from measurements of time-dependent CP violation in penguin transitions of $b \to s$ and $b \to d$ quarks. Furthermore, the measurement of flavour transitions to τ leptons, such as $B \to \tau \nu$ and $B \to D^{(*)} \tau \nu$, will allow to search for charged Higgs bosons, that are predicted to exist by many extensions of the SM. Left-right symmetry and the presence of new flavour changing neutral currents from physics beyond the SM are investigated measuring processes such as $b \to s \nu \overline{\nu}$ transitions and precisely measuring $b \to d$, $b \to s$ and $c \to u$ transitions. The decays of τ leptons are studied to search for sources of lepton flavour violation beyond the SM. Moreover, besides the measurements in the flavour sector, Belle II is sensitive to dark matter looking at missing energy decays and it will contribute to the study of quarkonia. These analyses require a detector with good detection capabilities for all neutral and charged particles, and excellent particle identification (PID) capabilities. In the next section 2.3 a detailed description of the Belle II detector is given.

The main competitor of the Belle II experiment is LHCb [4] at LHC, where the B and D mesons are produced in proton-proton collisions at a centre-of-mass energy of up to 14 TeV, allowing the collection of larger statistics. However, the advantages of the Belle II experiment at SuperKEKB are the relatively clean environment of the e^+e^- collisions compared to the hadronic environment, the full solid angle detector coverage and the possibility of completely reconstructing the final states.

2.3 The Belle II detector

The Belle II detector is designed for the SuperKEKB collider and it is optimized to operate at a higher event rate and reduced center-of-mass boost with respect to its predecessor Belle. It is a system of multiple sub-detectors, each optimized to reconstruct some specific features of the collision final states. The detector has a cylindrical structure around the beam pipe (z-axis), while it has a significant forward-backward asymmetry to improve the solid angle acceptance in the boost (forward)

direction. The detector is almost entirely located inside a superconducting solenoid, providing a 1.5 T magnetic field parallel to the beam direction in the internal region. The structure of the detector is shown in Fig. 2.3. The Belle II sub-detectors are presented in the following sections. A detailed description of the Belle II detector can be found in [9] and [23].
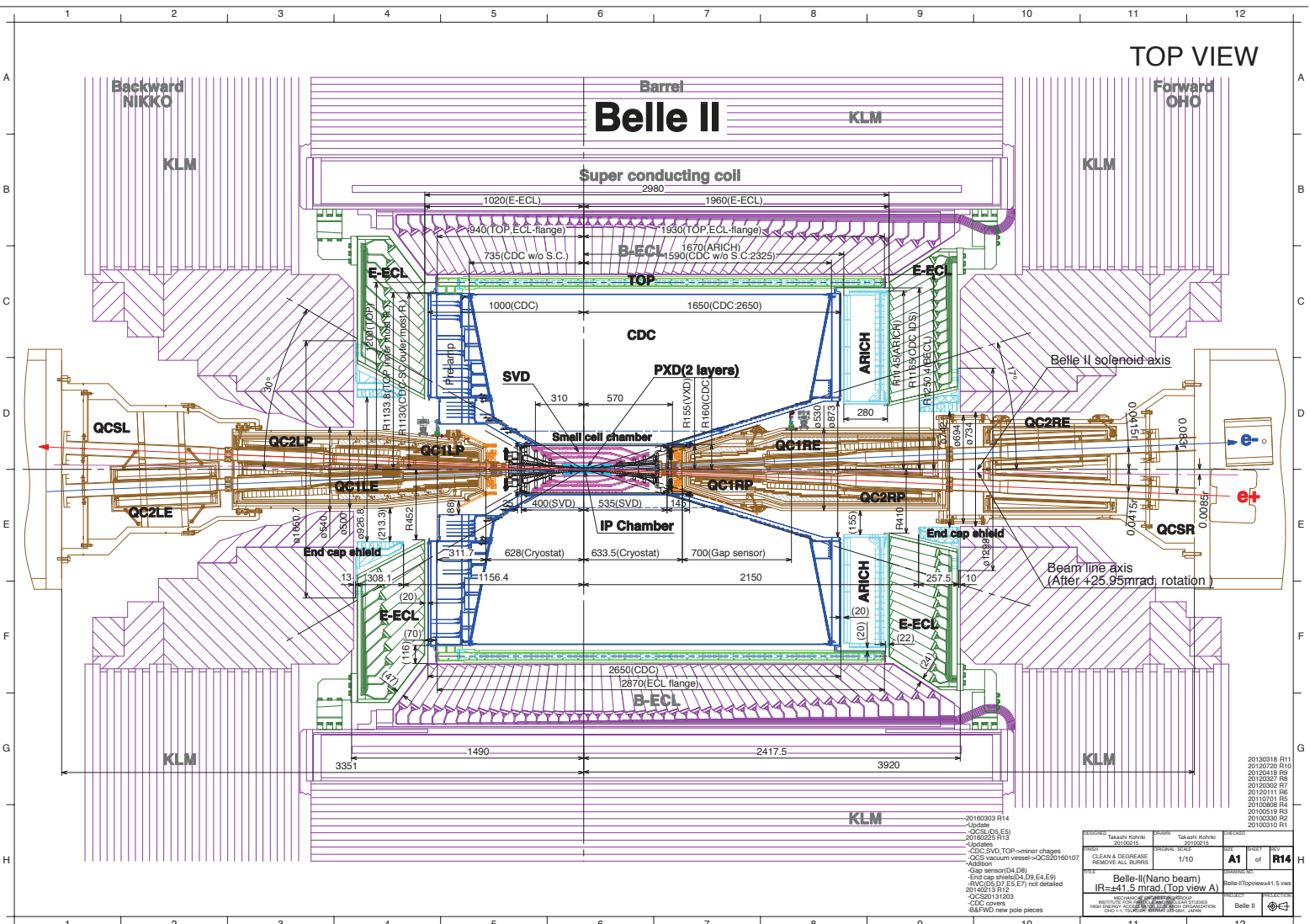
Figure 2.3: Top view of the Belle II detector [9].

2.3.1 Vertex Detector (VXD)

The innermost component of the Belle II detector is the Vertex Detector (VXD), and it comprises two devices: a silicon Pixel Detector (PXD) and a Silicon Vertex Detector (SVD), with altogether six layers around a 10 mm radius Be beam pipe. The first two layers are instrumented with pixel sensors, while the outermost four are strip detectors. The polar angle acceptance ranges from 17° to 150°. Efficient track reconstruction is of great importance since flavour physics final states are mostly composed of charged particles and analyses rely strongly on precise measurements of their momenta and the decay positions of their long-lived parent particles. The main purpose of the VXD sub-detector is to measure the B decays vertices for the measurement of CP asymmetries.

Pixel Detector (PXD)

The PXD is composed of two layers of pixelated sensors with DEPFET (DEPleted Field Effect Transistor) technology which allows for very thin (50 μm) sensors and fine granularity. The pixelated sensors have been chosen to sustain the higher hit rate due to the shorter distance from the IP and the higher luminosity with respect to Belle. The solution of using pixelated sensors rather than strips for the innermost layers allows to keep a much smaller detector occupancy, defined as the fraction of channels hit in each triggered event.

The layers are placed at 14 mm and 22 mm from the IP. The inner layer consists of eight planar modules ("ladder"), each with a width of 15 mm and a sensitive length of 90 mm. The outer layer consists of twelve modules with a width of 15 mm and a sensitive length of 123 mm. A schematic drawing of the sensor arrangement is shown in Fig. 2.4. The active area of each ladder is segmented into 250×768 pixels with sizes ranging from $50 \times 55 \ \mu m^2$ in the center of the inner layer to $50 \times 85 \ \mu m^2$ in the outer layer.

The DEPFET is a semiconductor detector concept that combines detection and amplification within one device. The DEPFET pixel consists of a depleted silicon bulk with a special potential structure and a FET on top. Electron-hole pairs are generated by traversing ionizing particles inside the sensor bulk. While holes drift to the backside, electrons are collected in the potential minimum of the so-called "internal gate".
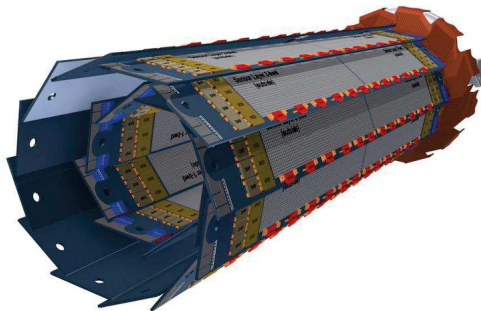
Figure 2.4: Schematic view of the geometrical arrangement of the sensors for the Belle II PXD. The light grey surfaces are the sensitive DEPFET pixels [24].

Silicon Vertex Detector (SVD)

Good characteristics for a vertex detector are low mass, high precision, immunity to background hits, radiation tolerance and long-term stability. The SVD of Belle II is designed with silicon strip sensors to avoid the huge channel count of pixels without compromising the vertex-detection capability of detector. The SVD also provides standalone tracking and particle identification via the measurement of ionisation energy loss.

The sub-detector is composed of four layers placed at 39 mm, 80 mm, 104 mm and 135 mm from the IP. These layers are equipped with DSSD (Double-sided Silicon Strip Detector) sensors, which perform

the vertex measurement and low-momentum tracking together with the PXD. The total of 172 DSSD sensors are distributed, from lowest to highest radius, in seven, ten, twelve, and sixteen ladders with two, three, four, and five sensors, respectively. The material budget average per layer is 0.7% of the radiation length X_0. Diamond sensors are installed on the beam pipe for radiation monitoring and to trigger fast beam aborts. A longitudinal schematic view of the SVD is shown in Fig. 2.5.
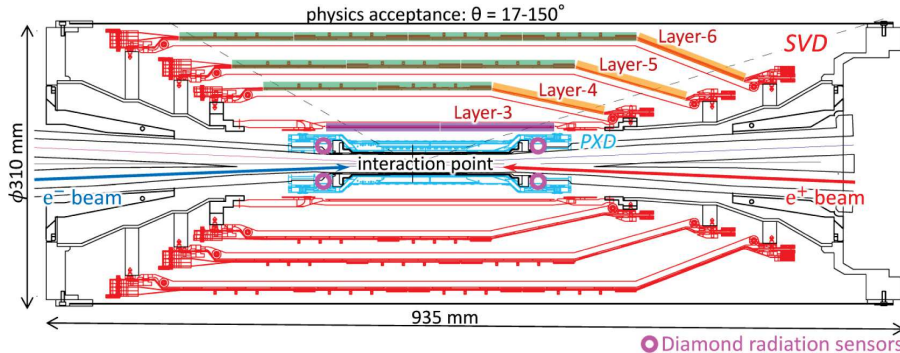
Figure 2.5: Schematic view of the Belle II Silicon Vertex Detector (SVD) [25].

Each DSSD sensor is based on an n-type bulk between 300-320 μm thick, equipped with implanted p- and n-doped sensitive strips on opposite sides. The metal strips for the readout are AC coupled on top of the implanted strips, separated by a dielectric SiO_2 layer, and alternated with strips which are not readout (floating strips). Along the sensors, strips are arranged in perpendicular directions on opposite sides in order to provide 2D spatial information: the p-side strips, orthogonal to the beam axis, measure the r-ϕ direction, instead the n-side strips provide information on the z coordinate along the beam line. Small rectangular sensors are used for the third layer, while the other layers are equipped with large rectangular sensors in the barrel region and slanted trapezoidal ones in the forward region. In total, the SVD covers an active area of 1.2 m^2 and has 224,000 readout strips.

2.3.2 Central Drift Chamber (CDC)

In the Belle II detector, the Central Drift Chamber (CDC) plays three important roles. First, it reconstructs charged tracks and measures their momenta precisely. Second, it provides particle identification information using measurements of energy loss within its gas volume. Low-momentum tracks, which do not reach the particle identification devices, can be identified using the CDC alone. Finally, it provides efficient and reliable trigger signals for charged particles.

When a charged particle traverses the CDC volume, it ionizes the gas inside the chamber, freeing electrons and positive ions from gas atoms. An applied electric field then moves these charges toward the sense wires, where high field gradients cause an abrupt acceleration with secondary ionization that induces an electric signal on the sensing wires. The particle trajectory is inferred from the time between the collision and the signal.

The CDC has a cylindrical shape with an inner radius of 16 cm, an outer radius of 113 cm, and a length of 232.5 cm at the outer radius. It has an asymmetric shape in the longitudinal direction to be able to treat asymmetric energy collisions. As for the SVD, the CDC covers a polar angle interval of $17° < \theta < 150°$, and has full coverage of the azimuthal angle. The chamber consists of 56 layers, arranged in nine super-layers. A smaller cell size design compared to Belle is adopted for the eight innermost layers to reduce the background occupancy. The axial (A) and stereo (U and V) super-layers are distributed according to the pattern AUAVAUAVA, as shown in Fig. 2.6. Combining information from the axial and stereo layers it is possible to reconstruct a full 3D helix track. Each cell consists of a sense wire surrounded by eight field wires, and the shape of the cell is nearly square. The innermost layer features the smallest cell size (0.6 cm), and the outermost layer has the largest (1.8 cm). The sense wires are made of gold-plated tungsten and the field wires are made of aluminium without plating, with diameters of 30 μm and 126 μm, respectively. There are in total 14336 sense wires and 42240 field wires. The chamber is filled with a mixture of 50% C_2H_6 and 50% He.
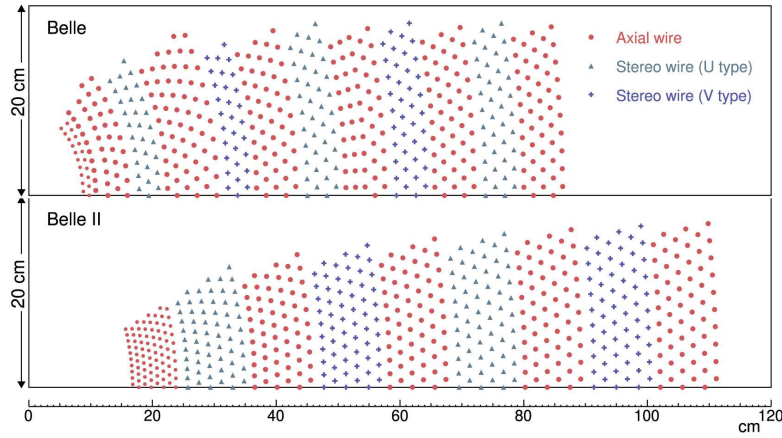
Figure 2.6: Layer configuration of the Belle and Belle II CDCs [26].

Specific ionisation (dE/dx) measurements are obtained for each reconstructed track in the CDC. Signal pulses on each sense wire of a cell, generated by the passage of a charged particle, are digitised and values over a nominal threshold are summed to yield the raw ADC (Analog-to-Digital Converter) readout. Geometrical corrections are applied for track length in each drift cell ("dx"): one correction for the projected r-z path length based on the polar angle of the track, and one for the projected r-ϕ path length based on the track geometry in the cell. Variations in wire-to-wire gain and run-to-run gain are calibrated with high-statistics samples of $e^+e^- \to e^+e^-\gamma$ radiative Bhabha events. These events are also used to map the gas-gain saturation vs. polar angle at the ionisation level of electrons, i.e. the "Fermi plateau" of the Bethe-Bloch formula. Empirical corrections for non-geometrical r-ϕ variations are also measured with radiative Bhabha samples.

Samples of e, μ, π, K, and p from dedicated control samples are used to determine gas-gain saturation effects for tracks with varying intrinsic ionisation. The same control samples are also used to parameterise the mean expected dE/dx as a function of $\beta\gamma = p/m$ and the dE/dx resolution as a function of $\beta\gamma$, polar angle, and the number of hits on the track. Fig. 2.7 shows the predicted dE/dx for six stable charged-particle hypotheses (e, μ, π, K, p, d) as a function of momentum.
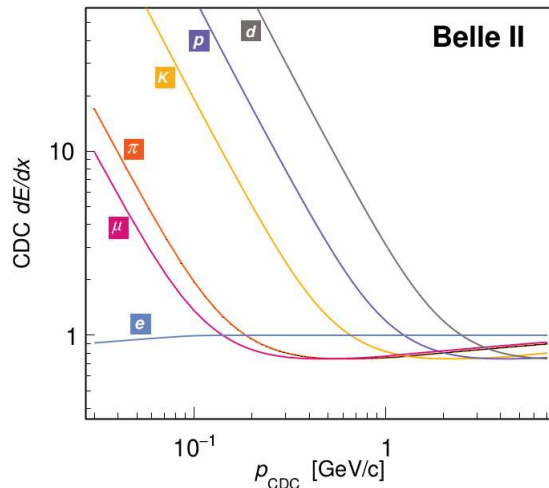
Figure 2.7: CDC-based dE/dx curve predictions for different charged particle species [27].

The dE/dx reconstruction algorithms in CDC allow us to determine likelihood values using information from individual hits. A likelihood value is determined for each charged particle hypothesis. These likelihoods are then further combined with information from other sub-detectors for particle identification. More details on PID from specific ionization measurements in CDC can be found in chapter 3.

13

2.3.3   Particle Identification (PID)

The main information on particle identification is provided by two dedicated sub-detectors, based on the Cherenkov effect: the Time Of Propagation counter (TOP) in the barrel region, and the Aerogel Ring-Imaging Cherenkov detector (ARICH) in the forward region.

Time Of Propagation counter (TOP)

The Time Of Propagation (TOP) sub-detector provides PID information in the barrel region of the Belle II detector, covering the polar angle region from 30° to 120°. This is a special kind of Cherenkov detector where 2D information of a Cherenkov ring image is given by the time of arrival and the impact position of Cherenkov photons at the photon detector placed at one end of a long quartz bar acting as radiator. The Cherenkov photons are produced by charged particles passing through one of the quartz radiator bars.

The TOP is composed of sixteen identical modules arranged around the CDC. A schematic for one module is given in Fig. 2.8. Each module is composed of four parts glued together: two fused silica bars of dimensions $125 \times 45 \times 2$ cm^3 acting as the Cherenkov radiator, a focusing mirror located at the forward end of the bar, and a 10 cm long prism that couples the bar to an array of MCP-PMTs (MicroChannel-Plate PhotoMultipliers Tubes). Thanks to the high average refractive index ($n = 1.44$ at 405 nm) of the fused silica, some of the Cherenkov radiation emitted by a charged particle crossing the radiator remains trapped by total internal reflection, propagating to the MCP-PMTs array. To propagate Cherenkov photons inside the radiator and preserve the Cherenkov ring image, the sides of the quartz bars must be flat and parallel to a very tight tolerance. The focusing mirror is spherical with a focal length corresponding to a radius of curvature of about 7 m.
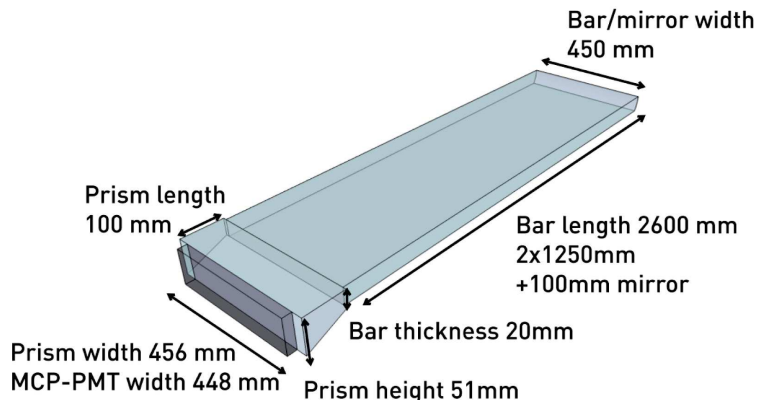
Figure 2.8: Schematic of one of modules of the Belle II TOP detector [5].

The MCP-PMTs provide excellent time resolution and fine granularity to measure also the position of arrival of the photons. They are suitable for the Belle II TOP detector which operates with small number of Cherenkov photons inside solenoid magnet. The photon detector array for each quartz module is composed of 32 MCP-PMTs, for a total of 512 MCP-PMTs. The array covers the face of the prism as fully as possible. A single MCP-PMT is square-shaped, the outer size is 27.6×27.6 mm^2 and the thickness is 15.6 mm, the effective area is 23.0×23.0 mm^2. The photon detector has a 4×4 anode array (5.3 mm wide and 0.3 mm gap), a multi-alkali photocathode (NaKSbCs) to have a high Quantum Efficiency (QE) at short wavelengths (QE of 18% at 400 nm), two 400 μm thick MCP plates with 10 μm pore size, and an aluminum layer on the second MCP to protect against ion feedback (Fig. 2.9). The Cherenkov photon is converted into an electron at the photocathode, then the produced electron is amplified by hitting the walls of the halls of two MCPs. The collection efficiency, which corresponds to the MCP aperture ratio, is $\sim 60\%$. The gain is of the order of 10^6 at a bias voltage around 3.4 kV, which enables single photon detection. The transit time spread is ~ 40 ps, which is essential for obtaining required PID performance. The performance loss in a magnetic field of 1.5 T, designed for the Belle II experiment, has been measured to be small.
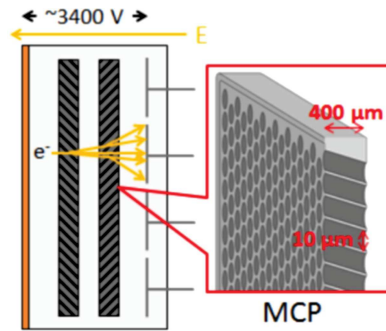
Figure 2.9: Schematic drawing of a Belle II TOP MCP-PMT [28].

The Cherenkov photon signal from the MCP-PMTs is read out using waveform sampling ASICs (Application Specific Integrated Circuits), to achieve a single photon timing resolution better than 100 ps at a nominal trigger rate of up to 30 kHz at the full projected luminosity of the SuperKEKB accelerator.

For a charged particle crossing the quartz bar at a given position and angle, the propagation of the Cherenkov light inside the quartz is a function of the Cherenkov emission angle θ_C (Fig. 2.10). The position of arrival of the photons on the array of MCP-PMTs, as well as the time of arrival, depends on the identification information of the incoming particle. The photon time is measured with respect to the bunch crossing that most likely produced the detected particles, therefore it can be decomposed into two parts: $t_{ph} = t_{tof} + t_{prop}$. The first parameter, t_{tof}, is the time of flight of the particle from the IP, proportional to $1/\beta$, while t_{prop} is the time spent by the Cherenkov photon inside the quartz bar. Comparing the distributions of time and position of the arrival of photons with the expected probability density functions (PDFs) corresponding to the particle hypotheses (e, μ, π, K, p, d), it is possible to obtain identification probabilities from the likelihood ratios. The TOP detector is mostly dedicated to K/π separation, but it can provide information also for other particles. More details on PID from the TOP will be presented in chapter 3.
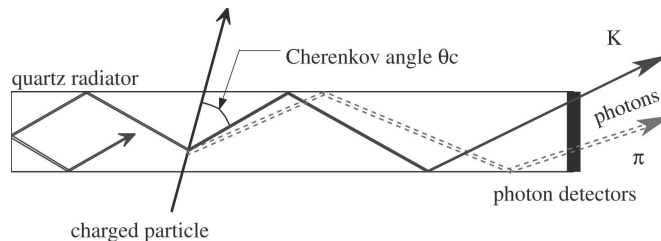
Figure 2.10: Schematic side view of Belle II TOP counter and internal reflecting Cherenkov photons [23].

Aerogel Ring-Imaging Cherenkov detector (ARICH)

The Aerogel Ring-Imaging Cherenkov detector (ARICH) covers the polar angle region from $17°$ to $30°$. This sub-detector is dedicated to pions and kaons separation and to discrimination between low momentum pions, electrons and muons.

The ARICH is composed of 4 cm aerogel radiator where the Cherenkov photons are emitted, and an expansion volume to allow the Cherenkov photons to form rings on the photon detector surface (Fig. 2.11) consisting of Hybrid Avalanche Photo Detectors (HAPDs). To achieve the necessary PID performance, enough photons (about 10) have to be detected for each ring image for at least one of the particle species. This requirement fixes the length of the aerogel radiator to several centimeters. The required resolution in the measurement of the Cherenkov angle is achievable only for an expansion gap of about 20 cm and a radiator thickness that does not exceed a few centimeters, with a photon detector granularity of a few mm. Two 2 cm thick layers with different refractive indices (1.045 and 1.055) are employed as Cherenkov radiators. The aerogel tiles are wedge shaped and have a size of about $18 \times 18\,\text{cm}^2$. The entire ARICH consists of 124 aerogel tiles per layer (a total of 248) arranged

in four concentric rings. 18 planar mirror plates are placed on the outer edge of the detector, where Cherenkov photons would miss the sensitive area, in order to maintain good performance.
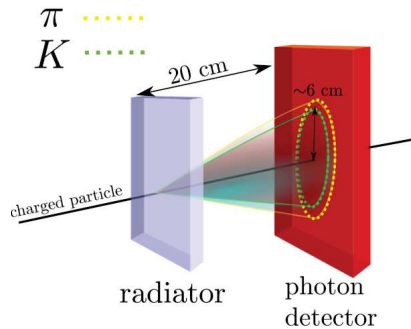
Figure 2.11: The concept of the Belle II ARICH [28].

HAPDs are used as photon detector. This type provides high single photon detection efficiency with the desired position resolution. The incident photon is converted into photo-electron by a bi-alkali photo-cathode, with peak QE of about 25%. The electron is then accelerated in a vacuum tube with high electric field towards the segmented Avalanche Photo-Diode (APD) with 144 pads of size 5.1×5.1 mm^2. The ARICH photon detectors consist of 420 HAPD modules arranged in seven concentric rings. Because of the large number of photon detector readout channels, ~ 80k in total, a dedicated front-end readout system based on ASICs is employed.

2.3.4 Electromagnetic Calorimeter (ECL)

A high resolution Electromagnetic Calorimeter (ECL) is a very important part of Belle II to detect photons with high efficiency and measure their energy in a wide range from 20 MeV to 4 GeV. Furthermore, ECL tasks are electron identification, generation of proper trigger signals and K_L detection together with the K_L and muon spectrometer.

The ECL detection principle is based on scintillation: the energy released by an incident particle causes a molecular excitation in the material with the passage of an electron from the valence band to the conduction band. The de-excitation of the electron to the valence band is associated with the emission of a photon, called scintillation light, that is usually difficult to detect and low-energetic. To improve the probability to emit a photon in the visible, impurities are added to create activator sites for the electrons with energy levels in the forbidden zone between the two bands. The scintillation light is then collected by photo-diodes. If the incident particle stops in the calorimeter, the total energy is measured.

Belle II ECL consists in a highly segmented array of thallium-doped caesium iodide CsI(Tl) crystals, a total of 8736, covering about 90% of the solid angle in the centre-of-mass system ($12.4° < \theta < 155.1°$ except for two gaps $\sim 1°$ wide between the barrel and endcaps). The crystals have a trapezoidal geometry with nominal cross-sectional area of 6×6 cm^2 and length of 30 cm, providing 16.1 X_0 of material. Each crystal is oriented to point in the direction of the collision point with a small tilt in θ. Crystals in the barrel region additionally have a small tilt in ϕ. The mechanical configuration of the ECL is shown in Fig. 2.12.

Scintillation light detection is achieved using two 10×20 mm^2 Hamamatsu Photonics S2744-08 photo-diodes glued to the rear of each crystal. The signals from the photo-diodes are integrated by a pre-amplifier mounted on the crystal before being sent to an external shaper module for processing. The support structure, CsI(Tl) crystals, and pre-amplifiers were re-purposed from the Belle experiment. The external shaper modules are new to Belle II and apply signal shaping via a CR-$(RC)^4$ filter before digitization at a sampling time of 567 ns.

In the ECL, photons and electrons are identified through their kinematics, shower shapes and timing information, as they have different (in shape and magnitude) energy losses with respect to charged hadrons. To separate electrons from photons, information from tracking detectors is correlated with the ECL signal.
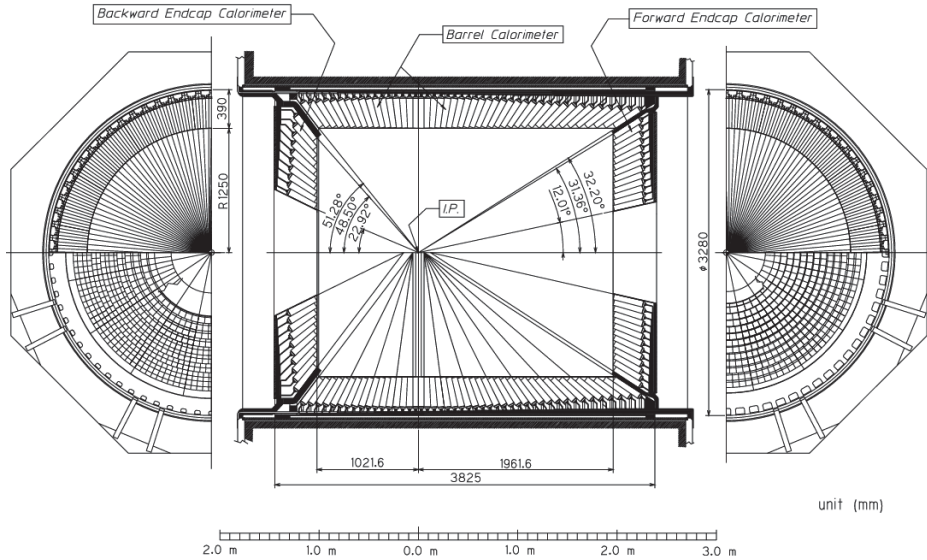
Figure 2.12: Belle II ECL overall configuration [29].

The photon energy resolution quantifies the precision that the reconstructed photon energy, E_γ^{rec}, predicts the true photon energy, E_γ^{true}. We consider the variance of the distribution of the photon energy residual $\delta_\gamma^{\text{true}} = E_\gamma^{\text{rec}} - E_\gamma^{\text{true}}$, FWHM is the Full Width at Half Maximum and $\Delta_\gamma^{\text{true}}$ is the distribution of the quantity $\delta_\gamma^{\text{true}}$ for a collection of events:

$$\sigma_\gamma = \frac{\text{FWHM}(\Delta_\gamma^{\text{true}})}{2.355}$$

Measuring the photon energy resolution in data requires a method to predict the value of E_γ^{true}. This can be achieved using the process $e^+ e^- \to \mu^+ \mu^- (\gamma)$: the recoil energy, $E_{\mu\mu}^{\text{recoil}}$, that is computed from the measured momentum of the muons, provides a proxy to predict the photon energy. The photon energy resolution of Belle II ECL obtained from fit of measured energy resolution as a function of $E_{\mu\mu}^{\text{recoil}}$ is

$$\frac{\sigma_\gamma}{E_\gamma} = \frac{(0.022 \pm 0.001)}{\sqrt{E_\gamma\,(\text{GeV})}} \oplus (0.010 \pm 0.001)$$

where $\oplus$ indicates addition in quadrature [30].

2.3.5 K_{L} and Muon detector (KLM)

The outermost sub-detector is the K_{L} and Muon (KLM) detector for the measurement of long-living particles, which travel large distances inside the Belle II detector. Muons and non-showering charged hadrons (that decay in flight or do not interact hadronically) with a momentum above 0.6 GeV/c traverse the KLM until they escape or range out due to electromagnetic energy deposition. They travel along nearly straight lines through the KLM. K_{L} mesons that interact in the ECL or the iron plates create a hadronic shower that can be detected in either the ECL alone, the KLM alone, or both.

The KLM is located outside of the superconducting solenoid and it is composed of 4.7 cm iron plates alternated with active detector elements. The iron plates serve as the magnetic flux return for the solenoid. They also provide 3.9 interaction lengths or more of material, beyond the 0.8 interaction lengths of the calorimeter, in which K_{L} mesons can shower hadronically. There are 15 detector layers and 14 iron plates in the barrel and 14 detector layers and 14 iron plates in each endcap. The barrel layers cover the polar angle range from 45° to 125°, while the endcaps extend this coverage from 20° to 155°.

The Belle KLM system, based on glass-electrode Resistive Plate Chambers (RPCs), demonstrated good performance during the entire data taking period of the Belle experiment. However, in some

Belle II KLM detector areas (both endcaps and the innermost layers in the barrel region) large background rates are expected due to neutrons that are mainly produced in electromagnetic showers from background reactions (i.e. radiative Bhabha scattering). The long dead time of the RPCs during the recovery of the electric field after a discharge significantly reduces the detection efficiency under such background fluxes. To mitigate this problem, the RPCs in the endcaps and the innermost layers in the barrel region have been replaced by layers of scintillator strips.

The barrel RPCs are rectangular in shape and vary in size from 2.2×1.5 m^2 to 2.2×2.7 m^2. The cross section of a super-layer is shown in Fig. 2.13: two RPCs are sandwiched between orthogonal pickup strips with ground planes for signal reference and proper impedance. RPCs have two parallel plate electrodes with high bulk resistivity ($\sim 5 \cdot 10^{12}$ Ω cm) separated by a gas-filled gap. We use a gas mixture of 62% HFC-134a, 30% argon, and 8% butane-silver. Each RPC is electrically insulated with a double layer of 0.125 mm thick mylar. Signals from both RPCs are picked up by copper strips above and below the pair of RPCs, providing a 3D space point for particle tracking. In streamer mode, an ionizing particle traversing the gap initiates a streamer in the gas that results in a local discharge of the plates. The discharge induces a signal on external pick up strips, which can be used to record the location and the time of the ionization.

The scintillator strips in the endcaps KLM are equipped with WaveLength Shifting (WLS) fibers readout by multipixel Silicon PhotoMultipliers (SiPMs) operating in the Geiger mode. SiPMs allow for compact detectors and uncompromised operation in strong magnetic fields. The whole system consists of 16800 scintillator strips, they have a cross section of $(7\text{-}10) \times 40$ mm^2 and a length of up to 2.8 m.

Ground plane		0.25 mm Mylar 0.035 mm Copper
Dielectric foam		7 mm
Cathode plane		0.035 mm Copper 0.25 mm Mylar
+HV		3.00 mm
Gas gap		2.00 mm
-HV		3.00 mm
Insulator		0.5 mm Mylar
+HV		3.00 mm
Gas gap		2.00 mm
-HV		3.00 mm
Cathode plane		0.25 mm Mylar 0.035 mm Copper
Dielectric foam		7 mm
Ground plane		0.035 mm Copper 0.25 mm Mylar
		31.6 mm total
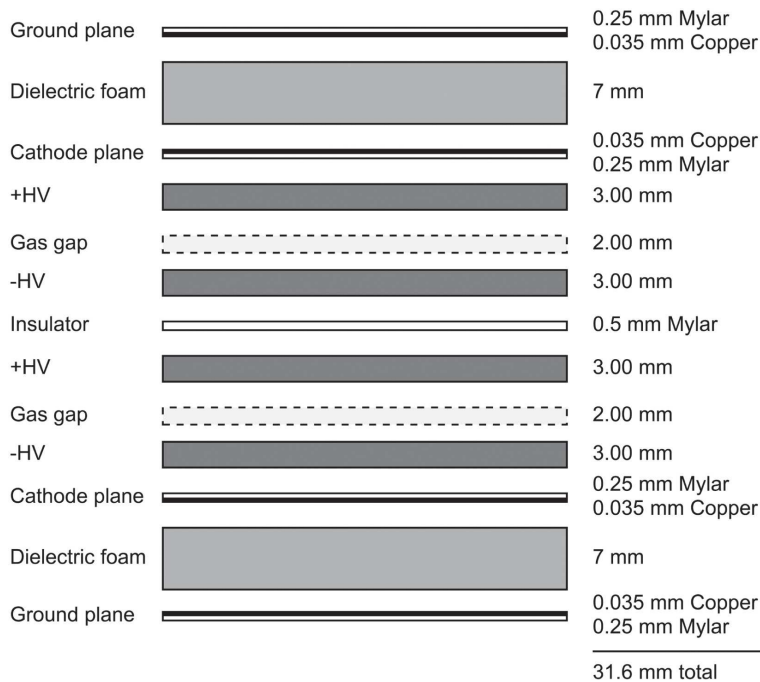
Figure 2.13: RPCs of the Belle II KLM system, structure of a super-layer [23].

2.3.6 Trigger

The main scope of the trigger system in Belle II is to identify events of interest and to reject background events that are not worth being acquired. The requirements for the trigger system are high efficiency for hadronic events from $\Upsilon(4S) \rightarrow B\bar{B}$ and for continuum background, a maximum average trigger rate of 30 kHz, a fixed latency of about 5 μs, a timing precision of less than 10 ns, a minimum two-event separation of 200 ns and a trigger configuration that is flexible and robust. To meet these requirements, the Belle triggering scheme with new technologies is adopted.

The Belle II trigger system is composed of two levels: a hardware-based Low-Level trigger (L1) and a software-based High-Level Trigger (HLT). The first one removes most of the background events with

the use of raw information from the faster sub-detectors, the second trigger refines the selection with a more exhaustive analysis.

In the Belle triggering scheme, the trigger system consists of sub-trigger systems and one final-decision logic. A sub-trigger system summarizes trigger information on its sub-system, and sends it to the final-decision logic, which then makes combinations of sub-triggers and fires a trigger when its criteria are satisfied. In Belle II, we use this concept but replace all components and connections with new technologies. Each component has a FPGA (Field Programmable Gate Array) so that the trigger logic is configurable rather than hard-wired. All data flow along high speed serial links, not parallel (ribbon) cables, which enables us to funnel a huge amount of information to one FPGA. The schematic overview of the Belle II L1 trigger system is shown in Fig. 2.14. The CDC sub-trigger provides the charged track information such as momentum, position, charge, multiplicity and so on. The ECL sub-trigger gives energy deposit information, energy cluster information, Bhabha identification, and cosmic-ray identification. The Barrel PID (BPID) sub-trigger gives precise timing and hit topology information. The Endcap PID (EPID) sub-trigger is expected to give precise timing information. The KLM sub-trigger gives muon track information. The Global Decision Logic (GDL) receives all of this sub-trigger information and makes the final decision. The total latency in the trigger system is about 5 µs. To achieve high efficiency for hadronic events, the L1 trigger mainly uses the information coming from the CDC and the ECL.
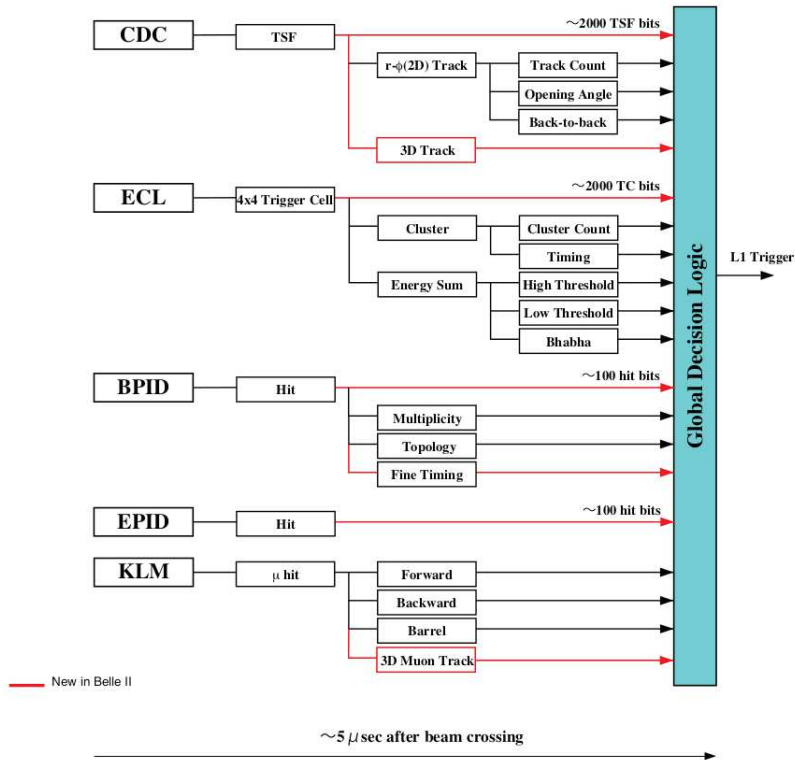
Figure 2.14: Schematic overview of the L1 trigger system. The output from the five sub-trigger systems are sent to the GDL, where the final trigger decision is made. The lines in red are newly added information paths in the Belle II trigger system [23].

The HLT consists of multiple units of PC clusters, that perform a full reconstruction of the event data. It uses all the sub-detectors (except the PXD) to perform an offline reconstruction of the event by applying physics requirements. The software trigger for the events uses the *basf2* framework (Belle II Analysis Software Framework) [31] to perform a reconstruction of the event to a physics level. It has the aim of reducing the trigger rate to 10 kHz for offline storage and assigning the skim flags, a set of selections for specific analysis (i.e. the hadronic tag from the HLT selects events in which a hadronic final state is produced, thus excluding Bhabha or $e^+e^- \to \mu^+\mu^-$ events).

Chapter 3

Charged particle identification

In this chapter, we will describe the method of charged particle identification at Belle II, which consists in determining particle identification probabilities as likelihood ratios for each charged particle hypothesis. Particle identification is then performed by applying a selection criterion to the identification probabilities.

3.1 Charged particle identification model

Effective and efficient charged particle identification (PID) is crucial to the physics goals of the Belle II experiment. Good PID information is necessary to isolate hadronic final states, reduce backgrounds, and allow for flavour-tagging techniques.

At Belle II charged PID is based on a combination of measurements from the various sub-detectors, namely:

- measurements of the particle's velocity ($\beta\gamma = p/mc$) via specific energy loss by ionisation (dE/dx) in the SVD and the CDC;

- measurements of the particle's velocity by measuring the Cherenkov angle of the radiation emitted by the charged particle passing through the TOP or the ARICH;

- measurements of the energy deposition in the CsI(Tl) scintillation crystals of the ECL and estimations of the shower shape;

- measurements in the KLM that account for the different penetration range and scattering of muons and hadrons.

In each sub-detector $d \in D = \{\text{SVD}, \text{CDC}, \text{TOP}, \text{ARICH}, \text{ECL}, \text{KLM}\}$ a likelihood $\mathcal{L}_\alpha^d$ is computed for each charged particle hypothesis $\alpha \in \{e, \mu, \pi, K, p, d\}$ from a PDF of a given set of observables. The PDFs are either predicted from simulation, extracted from data control samples with high purity, or determined analytically. Assuming that sub-detectors measurements of each of the identifying observables are independent, a global likelihood for each particle hypothesis is defined by

$$\mathcal{L}_\alpha = \prod_d^{d \in D} \mathcal{L}_\alpha^d \qquad (3.1)$$

where the product is over the active sub-detectors for the particle type of interest. It is useful to define the logarithm of the likelihood LL_α by summing the log-likelihoods from each sub-detector

$$LL_\alpha = \log \mathcal{L}_\alpha = \sum_d^{d \in D} \log \mathcal{L}_\alpha^d \qquad (3.2)$$

Typically the log-likelihood is used because it is more computationally stable. The resulting likelihoods (or log-likelihoods) are then used to compute likelihood ratios as given by

$$P_\alpha = \frac{\mathcal{L}_\alpha}{\sum_\beta \mathcal{L}_\beta} = \frac{\exp(\log \mathcal{L}_\alpha)}{\sum_\beta \exp(\log \mathcal{L}_\beta)} \tag{3.3}$$

where the sum over β is over all six charged particle hypotheses. The likelihood ratios are defined in such a way that the six ratios P_α will sum to one. Therefore, each likelihood ratio can be interpreted as a PID probability for the hypothesis α. We refer to this likelihood ratio as global PID probability.

PID for charged hadrons (π, K, p, d) depends primarily on likelihood information from CDC, TOP, and ARICH detectors. These sub-detectors also contribute to the PID of charged leptons (e, μ). Since this study is dedicated to PID performance of the TOP detector for K/π separation, the method to construct the likelihoods for this sub-detector is described in the following.

3.1.1 TOP likelihood

The TOP counter combines time-of-flight measurements with the Cherenkov ring-imaging technique. The dominant contribution to the resolution of this detector is the dispersion of light while propagating in the quartz bar. An extended likelihood method is used to determine log-likelihoods for the six stable charged particle types [32]. For a given charged particle hypothesis α, the extended log-likelihood probability is defined as

$$\log \mathcal{L}_\alpha = \sum_{j=1}^{N} \log \left(\frac{S_\alpha(x_j, y_j, t_j) + B_\alpha(x_j, y_j, t_j)}{N_e} \right) + \log P_N(N_e) \tag{3.4}$$

where $S_\alpha(x_j, y_j, t_j)$ is the signal distribution for the hypothesis α, $B_\alpha(x_j, y_j, t_j)$ is the distribution of background, and $N_e = N_\alpha + N_B$ is the expected number of detected photons, being the sum of the expected number of signal photons N_α and the expected number of background photons N_B. The second term in Eq. 3.4 is a Poisson probability of the mean N_e to obtain N photons. The normalization of $S_\alpha(x_j, y_j, t_j)$ and $B_\alpha(x_j, y_j, t_j)$ are

$$\sum_{j=1}^{n_{\text{ch}}} \int_0^{t_m} S_\alpha(x_j, y_j, t_j) \, dt = N_\alpha \tag{3.5}$$

$$\sum_{j=1}^{n_{\text{ch}}} \int_0^{t_m} B_\alpha(x_j, y_j, t_j) \, dt = N_B \tag{3.6}$$

where the sum runs over all channels n_{ch} of the photon detector array, x_j and y_j being the channel coordinates, and the integration is performed over the full range t_m of the time of arrival measurement. The ring-image of the TOP counter is a complicated pattern which, besides the Cherenkov angle, also depends on the particle impact position and the angles with respect to the quartz bar. The distribution for a particular detection channel j can be parameterized as a sum of Gaussian distributions

$$S_\alpha(x_j, y_j, t_j) = \sum_{k=1}^{m_j} n_{kj} \, g(t - t_{kj}; \sigma_{kj}) \tag{3.7}$$

where n_{kj} is the number of photons in the k-th peak of channel j, t_{kj} is the position along the time axis and σ_{kj} the width of the peak, then $g(t - t_{kj}; \sigma_{kj})$ is the normalized Gaussian distribution, and m_j counts the number of peaks in channel j for $t < t_m$. The quantities n_{kj}, t_{kj} and σ_{kj} are functions of the Cherenkov angle θ_C, the photon emission point (x_0, y_0, z_0) given by the particle impact position, the particle impact angles (θ, ϕ), and the unfolded channel coordinate $x_D = ka \pm x_j$, where k represents the number of internal reflections at the side walls and a the width of the quartz bar. Using the above input data, it is possible to solve for the unknown Cherenkov azimuthal angle ϕ_C^{kj} and thus determine

the photon directional vector. Once the photon direction is known, t_{kj} is obtained by ray-tracing. The number of photons in the peak is calculated with

$$n_{kj} = N_0 \ell \sin^2 \theta_C \frac{\Delta \phi_C^{kj}}{2\pi} \qquad (3.8)$$

where N_0 is the figure of merit of the Cherenkov counter, ℓ is the length of the particle trajectory in the quartz bar, and $\Delta \phi_C^{kj}$ is the range of the Cherenkov azimuthal angle covered by the measuring channel j. The peak width σ_{kj} is obtained by summing various contributions: photon emission point spread (parallax error), multiple scattering of the particle in the quartz, dispersion (chromatic error), channel size, and the transit time spread of the photon detector.

The PID information is extracted by comparing the distribution of the time of arrival of the photons in each of the TOP photon detector channels with the expected PDFs for the six stable charged particle hypotheses, where the PDFs are calculated analytically given the mass hypothesis and the particle's track parameters. This PID information is ultimately based on a measurement of the particle velocity, thus the separation power reduces as momentum increases.

Fig. 3.1 shows space-time distribution of the hits associated to a kaon candidate track. In the figure, the x-axis represents the position of the pixel along the transverse dimension of the bar, while the y-axis represents the detection time, with respect to the most probable bunch crossing. The black points represent the observed hits, while the smooth distribution is the expected PDF for a pion (left), a kaon (center) or a proton (right) of the same momentum and direction as the measured track.
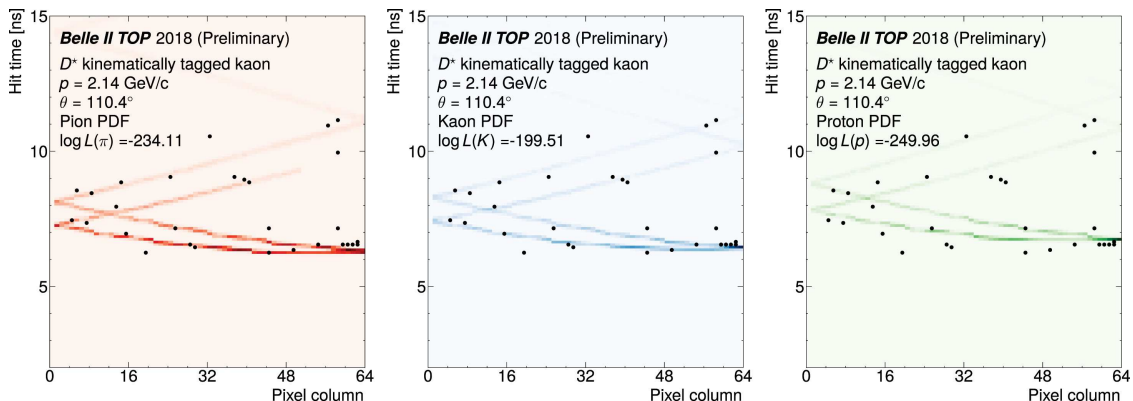
Figure 3.1: Comparison between the pion, kaon and proton TOP PDFs with the observed signal left by a kaon candidate [5].

3.1.2 Likelihood from other sub-detectors

In the following, we provide a concise description of the likelihood definition in the sub-detectors in addition to the TOP counter. All these likelihoods are combined together, also with the TOP likelihood, for a global PID.

SVD

Low-momentum charged particles having $p < 100$ MeV/c are unable to reach the CDC, the main tracking system of the experiment, owing to their highly curved trajectories. Even if the particles have a greater momentum and reach the CDC, the dE/dx values measured in the SVD can provide complementary information to that obtained from the main PID sub-detectors of Belle II. The dE/dx reconstruction algorithms construct likelihood values using information from each SVD hit for the particles. In order to combine the hit-level dE/dx information for reconstructing the track-level dE/dx value, a truncation is applied while combining the individual hit-level dE/dx measurements.

CDC

Likelihood values are constructed using information from individual hits in the CDC. Corrected values from each hit in the CDC are used to calculate a truncated mean: the lowest 5% and highest 25% of measurements on a given track are discarded and the remainder is averaged. Distributions of this measured truncated mean are used to calculate a normalised deviation for each hypothesis

$$\chi_\alpha = \frac{\Delta_{\mathrm{d}E/\mathrm{d}x}}{\sigma_{\mathrm{pred}}} = \frac{\mathrm{d}E/\mathrm{d}x_{\mathrm{means}} - \mathrm{d}E/\mathrm{d}x_{\mathrm{pred}}}{\sigma_{\mathrm{pred}}} \tag{3.9}$$

The expected spread $\Delta_{\mathrm{d}E/\mathrm{d}x}$, where $\mathrm{d}E/\mathrm{d}x_{\mathrm{means}}$ ($\mathrm{d}E/\mathrm{d}x_{\mathrm{pred}}$) is the measured (predicted) truncated mean, is parameterised as a function of track polar angle, the number of hits on track, and the $\mathrm{d}E/\mathrm{d}x$ value itself. The factor σ_{pred} at the denominator is the predicted resolution on $\Delta_{\mathrm{d}E/\mathrm{d}x}$, as obtained from a fit with a Gaussian function. The χ_α distributions are converted to likelihoods, assuming their PDF is well described by Gaussian functions.

ARICH

PID with the ARICH is based on the comparison between the observed spatial distribution of photons (hits) on the photon detector plane, and the PDF describing the expected distribution given the parameters of a track passing through the ARICH and the assumed charged particle hypothesis. The likelihood function for each hypothesis is constructed as a product of probabilities of individual pixels being in the observed state, i.e. *on* or *off* (we do not discriminate if a pixel is hit by a single or multiple photons).

ECL

The baseline method for charged PID in the ECL relies solely on E/p, namely the ratio of the particle's energy deposited in the calorimeter (E_{cluster}) and the reconstructed momentum of the topologically matched charged track (p_{lab}), where the latter measurement comes from the precision tracking systems. Templates of E/p are generated from simulated single-particle MC samples for each charged particle hypothesis, and PDFs are extracted via adaptive Gaussian Kernel Density Estimation (KDE). Variations in PDF shapes as a function of polar angle (θ_{lab}), momentum (p_{lab}), and charge (q) are taken into account. The charge dependency is introduced to account for differences in ECL energy deposition patterns due to charge asymmetry in hadronic interactions, which is mostly relevant for heavy hadron hypotheses (K and p).

KLM

Muon identification in the KLM relies on differences in longitudinal penetration depth and transverse scattering of tracks reconstructed in the inner detectors that are extrapolated to the KLM volume. The muon identification algorithm proceeds in two steps: track extrapolation, using the muon mass hypothesis for each charged particle candidate, and likelihood extraction for each of the six charged particle hypotheses. The track extrapolation proceeds step-wise through the detector geometry, starting at the outermost point of the reconstructed track's trajectory with its associated phase-space coordinates and track parameter covariance matrix. The likelihood values per hypothesis are the product of two terms: the longitudinal component and the transverse component, which are obtained from PDFs defined from MC for individual particle hypotheses, charge and extrapolation ending outcome.

3.1.3 Global, binary and re-weighted likelihood ratios

As described above, once the likelihoods $\mathcal{L}_\alpha$ for each charged particle hypothesis have been determined, we define the global PID probability for a particle hypothesis α (e, μ, π, K, p, d) as likelihood ratio (Eq. 3.3). However, many physics analyses only need information to distinguish between two particle hypotheses because the selection already reliably suppresses the others. Therefore, from the individual
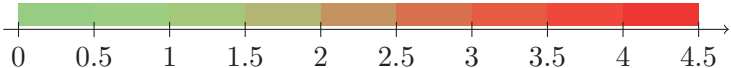
likelihoods it is also possible to define binary likelihood ratio discriminators between two hypotheses α and β

$$\mathrm{P}_{\alpha/\beta} = \frac{\mathcal{L}_\alpha}{\mathcal{L}_\alpha + \mathcal{L}_\beta} \tag{3.10}$$

Assuming that the PDFs of the sub-detectors are accurate and completely independent of each other, the global likelihood ratio is the theoretically optimal combination of the detector information. However, in some cases a particular sub-detector can degrade the overall PID performance of Belle II, indicating that our current combination of detector information is not optimal. To address the different contributions from the sub-detectors, their responses to different particle types and their different separation power, their coverage of the solid angle and the presence of blind spots (there are regions of the phase space where the sub-detectors measurements are not very reliable), and the different influence of background conditions, we add calibration weights to the computations performed for PID. We introduce a weighted likelihood $\tilde{\mathcal{L}}_\alpha$ for each particle hypothesis given by

$$\log \tilde{\mathcal{L}}_\alpha = \sum_d^{d \in D} w_{\alpha,d} \log \mathcal{L}_\alpha^d \tag{3.11}$$

where $\mathcal{L}_\alpha^d$ is the original likelihood and $w_{\alpha,d}$ is the PID calibration weight of particle type α and detector d. The tilde denotes that the quantity is derived from a weighted combination of detector information. In the case where all weights are unity, we exactly recover the standard PID quantities. The trained global calibration weights for each detector-hypothesis pair are shown in the following matrix

$$w = \begin{pmatrix} \begin{array}{cccccc} \text{SVD} & \text{CDC} & \text{TOP} & \text{ARICH} & \text{ECL} & \text{KLM} \\ 0.80897236 & 2.2702134 & 0.43437374 & 0.5949359 & 2.4642045 & 0.15824968 \\ 1.3629311 & 1.9584922 & 0.42866027 & 0.6106054 & 1.9623914 & 0.48289147 \\ 1.0637493 & 1.4330192 & 0.42645234 & 0.59815615 & 2.0181189 & 0.18273243 \\ 1.7925866 & 1.9082524 & 0.42230165 & 0.604627 & 1.662844 & 0.21941555 \\ 1.7190353 & 1.86573 & 0.41387647 & 0.64555 & 1.963171 & 0.22059433 \\ 1.264708 & 2.026095 & 0.38130292 & 0.6930469 & 4.0849915 & 0.14938803 \end{array} \end{pmatrix} \begin{array}{c} e \\ \mu \\ \pi \\ K \\ p \\ d \end{array} \tag{3.12}$$

In this matrix we can see that the weights for each sub-detector are largely shifted in unison. The relative importance of the sub-detector increases with the magnitude of the weights. Finally, we define the re-weighted PID probability as

$$\tilde{\mathrm{P}}_\alpha = \frac{\tilde{\mathcal{L}}_\alpha}{\sum_\beta \tilde{\mathcal{L}}_\beta} = \frac{\exp(\log \tilde{\mathcal{L}}_\alpha)}{\sum_\beta \exp(\log \tilde{\mathcal{L}}_\beta)} \tag{3.13}$$

By adding these weights, we expect to have a sufficient control over the relative contribution of each sub-detector to each particle type hypothesis, to obtain a powerful improvement over the standard PID performance at Belle II.

In principle, if there exist correlations between sub-detectors and since there are regions of the phase space where a sub-detector is more reliable in the measurements, the calibration weights may depend on other variables, namely track parameters (momentum p and polar angle θ). However, in this work we consider the calibration weights to be a single constant value per detector-hypothesis pair.

3.2 Efficiencies, fake rates and charge asymmetry

PID performance is discussed in terms of particle-ID efficiency and mis-identification rate (or fake rate). Both of these are determined as a ratio of signal yields in the studied decays after and before

identification selection. The PID selection criterion consists of a threshold cut on the global PID probability: $P_\alpha > P_{cut}$. Since this study is dedicated to kaons and pions identification, in the following we will mainly refer to these two types of charged particles. The K-ID and π-ID efficiencies are defined as

$$\varepsilon_K = \frac{\text{number of kaon tracks identified as kaon } (P_K(K) > P_{cut})}{\text{number of kaon tracks}} \tag{3.14}$$

$$\varepsilon_\pi = \frac{\text{number of pion tracks identified as pion } (P_\pi(\pi) > P_{cut})}{\text{number of pion tracks}} \tag{3.15}$$

Positive ε_+ and negative ε_- efficiencies are defined by selecting kaons or pions tracks of positive or negative charge. The mis-ID rates are determined as

$$f_K = \frac{\text{number of kaon tracks identified as pion } (P_\pi(K) > P_{cut})}{\text{number of kaon tracks}} \tag{3.16}$$

$$f_\pi = \frac{\text{number of pion tracks identified as kaon } (P_K(\pi) > P_{cut})}{\text{number of pion tracks}} \tag{3.17}$$

Particle-ID efficiencies and mis-ID rates can also be determined by applying PID selection to the re-weighted PID probability: $\tilde{P}_\alpha > \tilde{P}_{cut}$.

As said before in 3.1.3, some studies suggest that the global likelihoods do not combine the information from different sub-detectors in an optimal way. To investigate the contributions from different sub-detectors further, we study the binary likelihood ratio for individual sub-detectors, in particular for the TOP and the CDC. We define the binary likelihood ratios

$$P_{\alpha/\beta}^{TOP} = \frac{\mathcal{L}_\alpha^{TOP}}{\mathcal{L}_\alpha^{TOP} + \mathcal{L}_\beta^{TOP}} \qquad\qquad P_{\alpha/\beta}^{CDC} = \frac{\mathcal{L}_\alpha^{CDC}}{\mathcal{L}_\alpha^{CDC} + \mathcal{L}_\beta^{CDC}} \tag{3.18}$$

where α and β are π or K. We determine PID efficiencies of the individual sub-detectors by applying selection criteria to these additional PID probabilities.

Finally, the charge asymmetry in the identification of charged particles is defined as the normalized difference between the efficiencies for positive and negative charge

$$\mathcal{A}(\varepsilon) = \frac{\varepsilon_+ - \varepsilon_-}{\varepsilon_+ + \varepsilon_-} \tag{3.19}$$

If there are no differences in identifying a kaon or pion based on charge, the asymmetry will be zero. Instead, if the identification performance is different depending on whether the charge is positive or negative, the asymmetry will deviate from zero. It is also possible to define the charge asymmetry as the normalized difference between the fake rates for positive and negative charge

$$\mathcal{A}(f) = \frac{f_+ - f_-}{f_+ + f_-} \tag{3.20}$$

In the following chapters, we will present the analysed decay sample $D_s^+ \to \phi(K^+K^-)\,\pi^+$ (chapter 4) to conduct the study on the identification performance at Belle II, determining the K or π identification efficiencies and the charge asymmetry for both real data and MC simulation (chapter 5). Furthermore, we will compare the PID performance of the global identification variables with the other PID variables defined before (chapter 6 and chapter 7).

Chapter 4

Introduction to the analysis of the $D_s^+ \to \phi(K^+K^-)\,\pi^+$ decay

In this chapter, we will describe the data sets used for this analysis, introducing the signal and the main background sources. Moreover, we will present the model and fitting procedure for signal extraction.

4.1 Signal and background sources

At the SuperKEKB accelerator (chapter 2) the vast majority of the data are taken at the energy of the $\Upsilon(4S)$ resonance: $B\overline{B}$ meson pairs are produced via $e^+e^- \to \Upsilon(4S) \to B\overline{B}$ process, with $B = B^0$, $B^\pm$. However, the production of a $\Upsilon(4S)$ is not the most probable result of an e^+e^- collision at $\sqrt{s} = 10.58$ GeV. Cross sections for the most important physics processes, occurring at the default center-of-mass energy, are given in Tab. 4.1.

Table 4.1: Total production cross section for several physics processes from e^+e^- collision at $\sqrt{s} = 10.58$ GeV [9].

Physics process	Cross section [nb]
$\Upsilon(4S)$	1.110 ± 0.008
$u\bar{u}(\gamma)$	1.61
$d\bar{d}(\gamma)$	0.40
$s\bar{s}(\gamma)$	0.38
$c\bar{c}(\gamma)$	1.30
$e^+e^-(\gamma)$	300 ± 3 (MC stat.)
$\gamma\gamma(\gamma)$	4.99 ± 0.05 (MC stat.)
$\mu^+\mu^-(\gamma)$	1.148
$\tau^+\tau^-(\gamma)$	0.919
$\nu\bar{\nu}(\gamma)$	$0.25 \cdot 10^{-3}$
$e^+e^-e^+e^-$	39.7 ± 0.1 (MC stat.)
$e^+e^-\mu^+\mu^-$	18.9 ± 0.1 (MC stat.)

The signal for this study consists of events where a D_s^* meson has been detected, or more precisely, it has been reconstructed starting from the detection of its decay products. At the $\Upsilon(4S)$ resonance, the charm production cross section is approximately 2.4 nb: 1.3 nb from prompt $c\bar{c}$ production, and 1.1 nb from secondary $\Upsilon(4S) \to B \to D\,(D_s)$ production [9]. The non-resonant $e^+e^- \to q\bar{q}$ hadronic events ($q = u$, d, s) are the main background source for this analysis.

The D_s^* meson decays to $D_s\gamma$ (branching fraction $\mathcal{B} = (93.5 \pm 0.7)\%$ [8]), then the decay chain of interest is $D_s \to \phi\pi$ ($\mathcal{B} = (4.5 \pm 0.4)\%$ [8]), $\phi \to K^+K^-$ ($\mathcal{B} = (49.1 \pm 0.5)\%$ [8]). This sample provides

a natural source of kaons and pions. We will be able to discuss (chapter 5) the kaon identification performance for both positive and negative kaons independently of the charge asymmetries arising from the rest of the event. We also will study the pion identification performance.

4.2 Data and Monte Carlo samples

In this section the data and Monte Carlo (MC) samples used for this analysis are presented. Both data and MC are fully reconstructed from raw data and digitised simulated energy depositions with the Belle II software *basf2* [31]. Only events passing preliminary selection criteria have been considered.

Data

The data set used in this analysis was collected at Belle II from 2019 to 2022 and consists of two samples

- $\int \mathcal{L} \, dt = 362$ fb^{-1} collected at the $\Upsilon(4S)$ resonance,

- $\int \mathcal{L} \, dt = 42$ fb^{-1} off-resonance data set, i.e. data collected approximately 60 MeV below the $\Upsilon(4S)$ resonance.

These integrated luminosities are affected by an uncertainty of 0.7%. In the following, we will refer to the sample of all experimental data as `Data`.

Monte Carlo

The full simulation uses `KKMC` [33] for the generation of $q\bar{q}$ pairs from the e^+e^- collisions and `PYTHIA8` [34] for the hadronization. The decays of the produced hadrons are simulated using `EVTGEN` [35], while `GEANT4` [36] is used for the detector response. The simulated samples are described in Tab. 4.2. The MC analysed in this study was produced in the 15th campaign and incorporates simulated beam backgrounds and static detector conditions (run-independent campaign), `MC15ri`.

Table 4.2: MC samples used for the analysis.

MC sample	$\int \mathcal{L} \, dt$ [fb^{-1}]
$B^0 \overline{B}^0$	200
$B^+ B^-$	200
$q\bar{q}$ ($q = \{u, d, s, c\}$)	200

4.3 Reconstruction

In order to reconstruct the D_s meson candidates, two charged kaon tracks with opposite sign and one charged pion track are required from the interaction point. We constrain the trajectory of the charged particles by requiring $|d_0| < 1$ cm and $|z_0| < 4$ cm, where d_0 is the signed distance on the transverse plane of the point of closest approach (POCA) of the track to the origin of the coordinate system, z_0 is the z coordinate of the POCA. The tracks are selected in the momentum ranges 0.3 GeV/$c < p(\pi^\pm) < 5.0$ GeV/c and 0.3 GeV/$c < p(K^\pm) < 3.5$ GeV/c. Two opposite charged kaon tracks are combined to reconstruct ϕ candidates requiring their invariant mass to be 1015 MeV/$c^2 < M_\phi < 1025$ MeV/c^2 (true value $M_\phi = 1019.461 \pm 0.016$ MeV/c^2 [8]). We require that at least one kaon track has a value of global PID probability $P_K > 0.5$. Then, the D_s candidates are reconstructed from the ϕ and charged pion track with invariant mass 1920 MeV/$c^2 < M_{D_s} < 2010$ MeV/c^2 (true value $M_{D_s} = 1968.35 \pm 0.07$ MeV/c^2 [8]) and $\cos\alpha > -0.2$, where α is the angle between the two daughters of the D_s meson. The charge of the D_s meson is defined by the charge of the pion.
To reduce the background contribution we also reconstruct the D_s^* meson candidates from the D_s and a photon with $E_\gamma > 0.25$ GeV. The D_s^* invariant mass is required to be 2000 MeV/$c^2 < M_{D_s^*} <$

2200 MeV/c^2 (true value $M_{D_s^*} = 2112.2 \pm 0.4$ MeV/c^2 [8]) and the energy released in the decay is required to be 0.1 GeV $< Q <$ 0.17 GeV. Each photon is required to have `clusterTiming` < 200 ns and `clusTimingSignif` < 2, where the variable `clusterTiming` is the difference between the photon timing and the event time, the variable `clusTimingSignif` is defined as the ratio between `clusterTiming` and `clusterErrorTiming` which is ECL cluster's timing uncertainty that contains 90% of true photons. Photons produced at the interaction point should have `clusterTiming` close to zero.

All cuts and selection intervals are chosen to optimize the figure of merit FoM $= S/\sqrt{S+B}$ using simulated data, where S and B are the number of signal and background events in the signal region, respectively. Thanks to the constraints on the ϕ and D_s^* masses, in the analysis range for M_{D_s} we observe a distribution with a well-defined peak and a linear background contribution. A further constraint on the events could be the difference in mass between the D_s^* and D_s mesons. However, by imposing 136 MeV/$c^2 < \Delta M < 150$ MeV/c^2 (true value $\Delta M = 143.8 \pm 0.4$ MeV/c^2 [8]) there is an increase of the FoM, but the background distribution is no longer linear. The choice is not to apply constraints to ΔM in favour of greater simplicity of the shape of the D_s invariant mass distribution. Furthermore, for some events more than one D_s candidate can be reconstructed from the selected tracks. To choose the best candidate, we require that the photon to combine with the D_s to obtain the D_s^* is the one with the highest energy. This selection criterion does not affect the background distribution, instead the choice of the candidate D_s with the minimum difference between the reconstructed value and the true value of the invariant mass M_{D_s} modifies the background distribution and this should be taken into account in the model for the analysis. The event selection is optimized and validated using simulation.

Fig. 4.1 shows the invariant mass distribution of the D_s candidates after the reconstruction of the particles of interest in the analysed sample and event selection. On the left there is the distribution for the MC, on the right that for the real data.
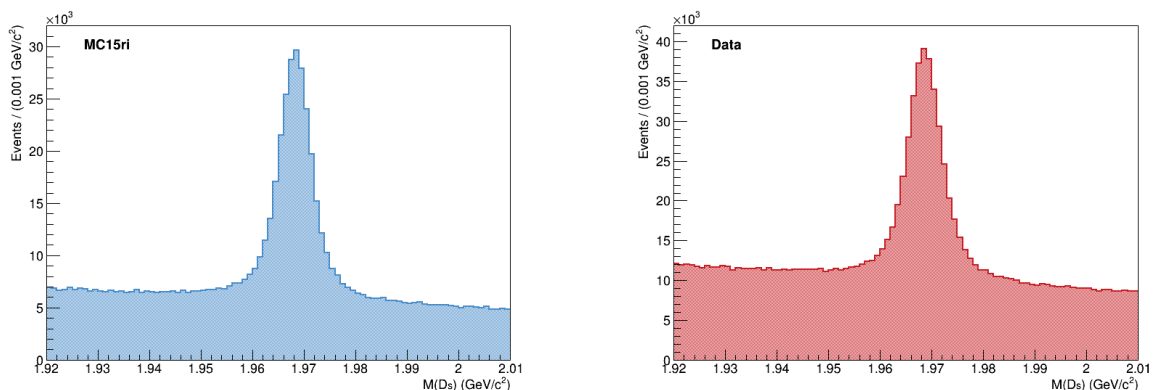
Figure 4.1: Invariant mass distribution of the D_s candidates: MC (left) and real data (right).

4.4 Signal extraction

The total number of D_s candidates both in data and MC is determined by an unbinned maximum likelihood fit of the invariant mass distribution M_{D_s}. The number of signal events in MC is also obtained through counting of truth-matched D_s candidates, the agreement with the estimated number provides a test of the model for the fit on the MC. The fit is performed using the `RooFit` [37] toolkit, the minimization is based on `Minuit` [38] using the `MIGRAD` minimizer. The error computation is performed with the `HESSE` algorithm.

4.4.1 Fit model

The signal peak in the D_s invariant mass distribution is modelled by three Gaussians centered around the true value $M_{D_s} = 1968.35 \pm 0.07$ MeV/c^2 [8], while the background is modelled by a first order Chebyshev polynomial

$$\text{PDF} = f_S \times [f_1 \times \text{Gauss}(\mu_1, \sigma_1) + f_2 \times \text{Gauss}(\mu_2, \sigma_2) + (1 - f_1 - f_2) \times \text{Gauss}(\mu_3, \sigma_3)]$$
$$+ (1 - f_S) \times \text{Chebyshev}(c_1) \tag{4.1}$$

In the fit model expression, f_S represent the fraction of signal events, so we define $N_{\text{sig}} = f_S \cdot N$ and $N_{\text{bkg}} = N - N_{\text{sig}}$ as the signal and background yields, respectively, where N is the total number of events. Each of the three Gaussians is defined by its own parameters, the mean μ and the width σ, numbered from 1 to 3. The coefficients f_1 and f_2 are the fractions of the contribution of each Gaussian component to the signal peak. The parameter c_1 represents the weight of the first order Chebyshev polynomial. The choice of the PDF is determined by seeking a good balance between the simplicity of the PDF shape (linear background, well-defined signal peak) and the total error associated with the fitting procedure (both statistical and systematic due to constraints on the parameters of the fitting model). The parameters of the first Gaussian are kept fixed: $\mu_1 = 1.968$ (very close to the true value of the D_s mass) and $\sigma_1 = 0.05$ (it is a large width to describe the tails of the peak), also f_1 is fixed. All the other PDF parameters are left free to float.

4.4.2 Fit validation

As a first step, we validate the fit model on simulated data. Next, we use the same PDF function to fit the real data and extract the number of signal events. The model is tested by fitting the entire data set, i.e. no PID selections are applied but only selections for the reconstruction of the D_s meson candidates. Fig. 4.2 presents the fit results for the invariant mass distribution of the D_s candidates. The fitted function is that described above (Eq. 4.1) and we set $f_1 = 0.088$. In Tab. 4.3 we summarize all the parameters and the resulting values of the floating parameters after the fit.

On the left in Fig. 4.2, we show the best fitted PDF (blue line) on MC, the red peak corresponds to the estimated signal component. To evaluate the goodness of the fit model we look if, for the MC, the pulls are normally distributed and we calculate the agreement between the signal yield estimated with the fitting procedure and that using MC truth information. The results of the fit to the pull distribution with Gaussian function are shown in Fig. 4.3. As can be seen from the plot on the left in Fig. 4.3, the pulls for the MC are distributed around zero, although the width of the fitted Gaussian is slightly greater than one and this result may indicate an underestimation of the statistical uncertainties. In Tab. 4.4 we report the number of total events and the estimated signal yield with its relative statistical
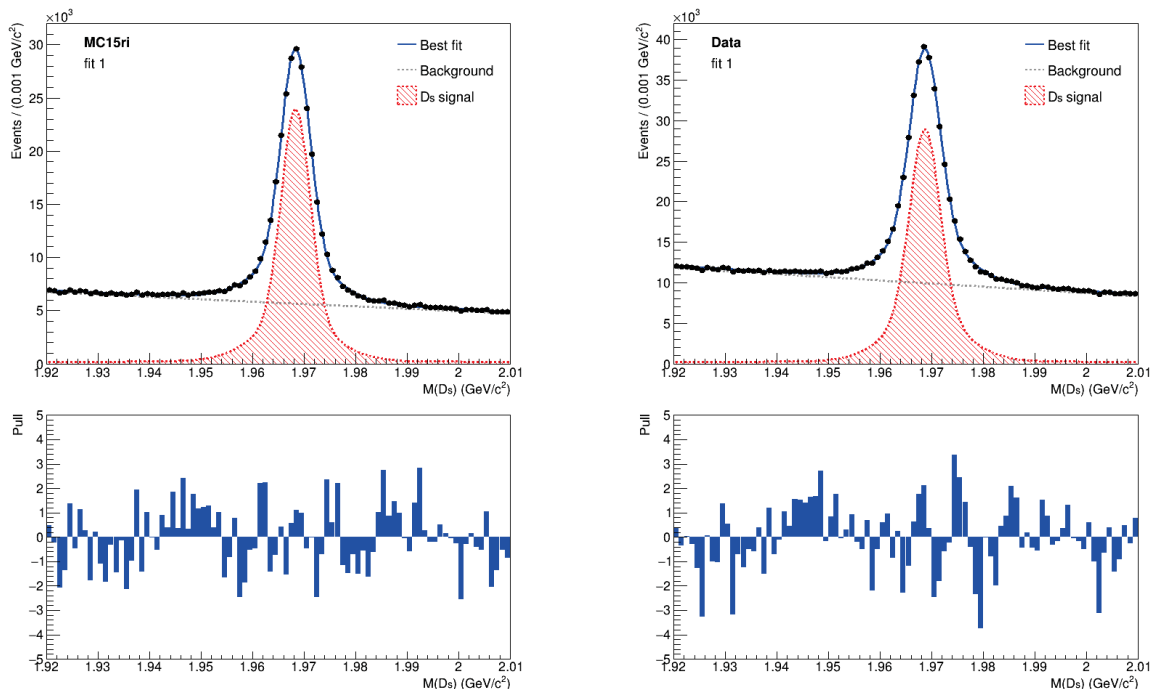

Figure 4.2: Fitted PDF for the D_s invariant mass distribution and relative pulls: MC (left) and real data (right). The red peak corresponds to the estimated signal yield.

Table 4.3: PDF model parameter list and resulting values after the fitting procedure.

Parameter		MC15ri	Data
	μ_1	1.968	1.968
Gauss1	σ_1	0.05	0.05
	f_1	0.088	0.088
	μ_2	1.96784 ± 0.00009	1.96881 ± 0.00008
Gauss2	σ_2	0.0074 ± 0.0002	0.0072 ± 0.0002
	f_2	0.298 ± 0.009	0.34 ± 0.01
Gauss3	μ_3	1.96834 ± 0.00001	1.96866 ± 0.00002
	σ_3	0.00280 ± 0.00002	0.00291 ± 0.00003
Chebyshev	c_1	-0.175 ± 0.003	-0.174 ± 0.002
	f_S	0.307 ± 0.001	0.244 ± 0.001
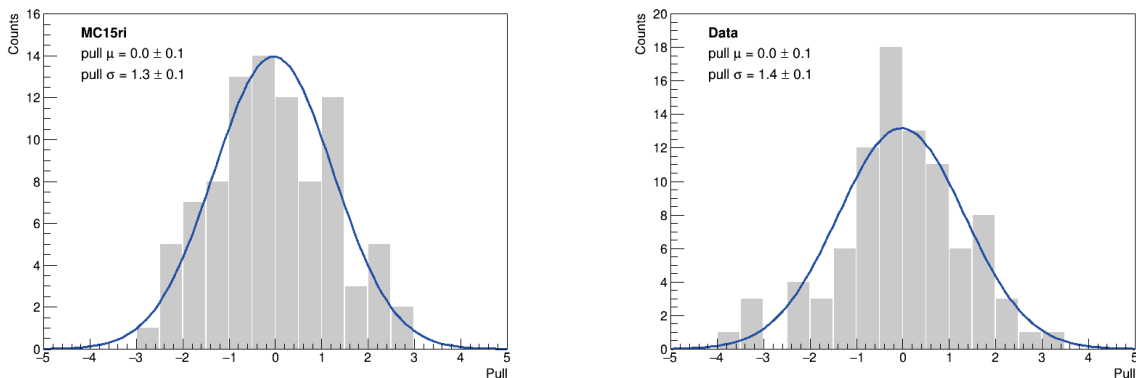
Figure 4.3: Pull distribution: MC (left) and real data (right).

Table 4.4: Total number of events and estimated number of signal events for both MC and data. True number of signal events for MC and its agreement with the fit result.

Sample	N	N_{sig}	$N_{\text{sig}}^{\text{true}}$	λ
MC15ri	745212	229000 ± 1000	228850	0.02
Data	1202501	293000 ± 1000	-	-

error, for both MC and real data. The error associated with N_{sig} is calculated by propagation, the error on f_S is determined in the fitting procedure, while the error associated with the total number of events is equal to $\sqrt{N}$. For the simulated data we also report the true number of signal events $N_{\text{sig}}^{\text{true}}$ and its agreement with the number of signal events obtained from the fit. We express the agreement with the parameter

$$\lambda(N_{\text{sig}}) = \frac{|N_{\text{sig}} - N_{\text{sig}}^{\text{true}}|}{\sqrt{\sigma^2(N_{\text{sig}}) + \sigma^2(N_{\text{sig}}^{\text{true}})}} \tag{4.2}$$

The lower the λ value, the greater the compatibility. The excellent agreement ($\lambda \ll 1$) between the estimated and real signal yields leads us to consider the PDF function used to fit the simulated data as valid.

After validating the model on simulation, we fit the experimental data and the resulting function is

shown on the right in Fig. 4.2. The parameters that have been set fixed in the fit on MC are kept set to the same value in the fit on real data. Also the pull distribution for the data is approximately a Gaussian centered around zero but its width is not one (Fig. 4.3, right).

4.5 PID efficiency and mis-ID rate

We do not expect any changes in the shape of the signal peak and the distribution of the background by applying PID selection, therefore we use the same PDF model to fit the invariant mass distribution of the D_s candidates before and after PID selection.

The parameter of interest for the calculation of PID efficiencies is f_S from which we get N_{sig}. Indeed, the efficiency is defined as the ratio

$$\varepsilon = \frac{N_{\mathrm{sig}}^{\mathrm{ID}}}{N_{\mathrm{sig}}} \tag{4.3}$$

where $N_{\mathrm{sig}}^{\mathrm{ID}}$ is the number of D_s candidates which pass a given PID selection criterion applied to the kaon or pion track (Eq. 3.14 and 3.15), N_{sig} is the number of all signal candidates. Since the process of applying PID selection criterion is binomial (the event can either pass or fail the criterion with some probability), the estimated statistical uncertainty is

$$\sigma_\varepsilon = \sqrt{\frac{\varepsilon(1-\varepsilon)}{N_{\mathrm{sig}}}} \tag{4.4}$$

In the following (chapter 5) we will report the K/π identification efficiencies for separate charges, estimated through the fit model described above. The charged particle tracks will be divided in bins of momentum p or cosine of the polar angle θ that ranges in the geometrical acceptance of the TOP detector.

To account for the systematic uncertainty from the PDF model due to the choice of fixed parameter values, we determine PID efficiencies with two fitting procedures that differ in the Gaussian fraction f_1 set at different values. In the first fitting strategy (fit1) the parameter f_1 is fixed at a constant value equal for positive and negative particle tracks, equal for the fit of the D_s mass distribution before and after applying PID selection criterion, and equal for each bin. The PID efficiencies estimated with this first fitting procedure are the reference efficiencies for our study. In the second fitting strategy (fit2) f_1 is fixed at different constant value depending on the charge, PID selection, bin. The parameter is fixed in order to obtain a good convergence of the fit, the choice of f_1 is tested on MC sample by comparing the fit results by applying PID selection to the global probability with the signal yields using MC truth information. As another systematic uncertainty due to the fitting procedure, we consider the difference between the efficiency estimated with the simulated data and the true efficiency determined using MC truth information. On the real data we consider the same contribution of systematic uncertainty of possible discrepancy due to the fits that we have for the MC. The total systematic uncertainty is obtained by taking the two contributions in quadrature.

We determine also the mis-ID rate as the ratio

$$f = \frac{N_{\mathrm{sig}}^{\mathrm{mis\text{-}ID}}}{N_{\mathrm{sig}}} \tag{4.5}$$

where $N_{\mathrm{sig}}^{\mathrm{mis\text{-}ID}}$ is the number of D_s candidates for which kaon (pion) tracks are mis-identified, namely kaon tracks pass π-ID selection criterion (Eq. 3.16) and vice versa pion tracks pass K-ID selection criterion (Eq. 3.17). We expect a smaller signal peak by applying mis-ID selection, but no alteration in the shape of the signal peak and background distribution. Thus, we use the same PDF model to fit the invariant mass distribution of the D_s candidates. However, in the PDF fitting the distribution with mis-ID selection we fix more parameters to have a good convergence of the procedure. Statistical and systematic uncertainties on fake rate are calculated in the same way as those on efficiency.

Chapter 5

Global PID performance and charge asymmetries

We will present the standard PID performance using the decay sample $D_s^+ \to \phi(K^+K^-)\pi^+$ and the measured charge asymmetries for kaon and pion identification. We will compare the results for MC and experimental data. The measurement of the identification efficiencies (mis-ID rates) and their uncertainties are estimated by following the method outlined in sections 4.4 and 4.5.

5.1 Kaon identification efficiencies and pion mis-ID rates

We present the kaon identification performance based on global PID probability selections. We compare the signal yield in the D_s invariant mass plot when PID selection criteria are applied to both kaon tracks, to the signal yield when criteria are applied to one kaon track only. Namely, in the preselection of our decay sample (section 4.3) we required $P_K > 0.5$ for at least one of the two kaons. To estimate the PID efficiency for K^+ we apply the preselection $P_K > 0.5$ to the opposite charge kaon tracks (K^-). Then, we apply the PID selection $P_K > \alpha$ to the positively charged kaons, where α is a generic threshold cut on the global PID probability, to study the PID performance. Vice versa, to estimate the PID efficiency for K^- we only consider events for which $P_K > 0.5$ for K^+ tracks, then we apply the PID selection $P_K > \alpha$ to the negatively charged kaons. In the following, we will always refer to PID selection for kaons as the one applied to the charged kaon tracks of interest. We will assume that the cut on the opposite charge kaon tracks is an implicit preselection in the sample of charged kaons. In formulas, we write

$$\varepsilon_{K^+} = \frac{N_{\text{sig}}(K^+ \, ; \, P_K(K^+) > \alpha)}{N_{\text{sig}}(K^+)} \tag{5.1}$$

$$\varepsilon_{K^-} = \frac{N_{\text{sig}}(K^- \, ; \, P_K(K^-) > \alpha)}{N_{\text{sig}}(K^-)} \tag{5.2}$$

In this work, we evaluate the K-ID performance by applying $P_K > 0.5$. To obtain the number of signal events before and after applying kaon identification selection to charged kaon tracks, we fit the invariant mass distribution of the D_s candidates with the PDF model defined in Eq. 4.1. We set fixed the parameters $\mu_1 = 1.968$, $\sigma_1 = 0.05$ and f_1. The different values of f_1 chosen in each fit are shown in Tab. 5.1 together with the signal yields resulting from the fitting procedures, the associated error is calculated by propagation. In the same table we summarize the results of K-ID efficiencies for K^+ and K^-, for both MC and data. The statistical uncertainty on the efficiency is calculated as in Eq. 4.4. As systematic uncertainties we consider, in quadrature, the difference between the efficiencies determined with the two fitting procedures (`fit1` and `fit2`) that differ by the value of f_1, and the difference between the estimated and true efficiencies in the MC sample. Moreover, in

the table we report the average efficiencies for both simulation and real data, the uncertainties are calculated by propagation. For MC we also report the true efficiencies, determined through counting of truth-matched D_s candidates before and after PID selections. We observe that the K-ID efficiencies on real data are slightly lower than those on MC.

Fig. 5.1 and Fig. 5.2 show the D_s invariant mass distributions and the fitted functions on simulated data, before and after the PID selection on positive and negative kaons, respectively. Fig. 5.3 and Fig. 5.4 show the distributions and the fitted functions on real data. The pulls are distributed around zero, however the width of the Gaussian fitting the pull distributions is slightly greater than one.

To evaluate the K/π separation performance, we also determine the pion mis-identification rate. We estimate by separate charge the signal yields before and after K-ID selection criterion on charged pion tracks. The applied PID selection consists of $P_K > \alpha$. In formulas, we write

$$f_\pi^+ = \frac{N_{\text{sig}}(\pi^+ \,;\, P_K(\pi^+) > \alpha)}{N_{\text{sig}}(\pi^+)} \tag{5.3}$$

$$f_\pi^- = \frac{N_{\text{sig}}(\pi^- \,;\, P_K(\pi^-) > \alpha)}{N_{\text{sig}}(\pi^-)} \tag{5.4}$$

In this study, we choose $\alpha = 0.5$. We perform the fit to the mass distribution of the D_s candidates to determine the number of signal events before and after PID selection on charged pion tracks. We use the same fit model as we did for efficiencies. In the fit without PID selection we set fixed the parameters μ_1 and σ_1 as before, f_1 is chosen to have a good convergence of fit on MC. Instead, in the PDF fitting the distribution with mis-ID selection we set fixed $\mu_2 = 1.968$ and $\sigma_2 = 0.01$ in addition to the other fixed parameters. We report the resulting pion fake rates for both MC and real data in Tab. 5.1. The average π mis-ID rate on real data is slightly higher than that on MC.

In Fig. 5.5 and Fig. 5.6 we show the D_s invariant mass distributions and the fitted functions on MC, in Fig. 5.7 and Fig. 5.8 the distributions and the fitted functions on real data, before and after the mis-identification selection on positive and negative pions. The pulls are normally distributed. As can be seen from the right plots in the figures, the K mis-ID selection on the pion tracks dramatically reduces the peak in the mass distribution of the D_s candidates. Indeed, the identification is performing well and the number of signal events for which there is an incorrect identification is very low.

The results of K-ID efficiencies and π mis-ID rates based on the global PID selection $P_K > 0.5$ are summarized in Tab. 5.2 for both MC and real data, the uncertainties are the sum in quadrature of statistical and systematic errors. The dashed ID observables indicate the average value obtained from the results for positive and negative charge. For simulated data, we report also the results using MC truth information and their statistical uncertainty. Finally, we compute the charge asymmetries of both efficiencies and fake rates according to Eq. 3.19 and Eq. 3.20. The uncertainties are obtained by propagation of the statistical and systematic uncertainties on the efficiencies (fake rates). Since the goal of this study is to investigate any significant differences in the charge asymmetries observed in experimental data and simulation, we quantify the discrepancy with the parameter

$$\lambda(\mathcal{A}) = \frac{|\mathcal{A}_{\text{MC15ri}} - \mathcal{A}_{\text{Data}}|}{\sqrt{\sigma^2(\mathcal{A}_{\text{MC15ri}}) + \sigma^2(\mathcal{A}_{\text{Data}})}} \tag{5.5}$$

The agreement between the charge asymmetries is reported in Tab. 5.2. The charge asymmetries of the efficiencies show good agreement ($\lambda < 1$), there is no significant discrepancy in the PID performance for oppositely charged kaons. Both charge asymmetries of the efficiencies $\mathcal{A}(\varepsilon)$ for simulation and experimental data are very close to zero. Instead, the charge asymmetry of the fake rates $\mathcal{A}(f)$ for real data is slightly larger than for MC, however the agreement is still good ($\lambda \sim 1$).

Sample			before PID		after PID ($P_K > 0.5$)		Efficiency	
			f_1	N_{sig}	f_1	N_{sig}	ε (%)	$\varepsilon^{\text{true}}$ (%)
MC15ri	K^+	fit1	0.08	198400 ± 800	0.08	171100 ± 800	$86.24 \pm 0.08 \pm 0.81$	85.68 ± 0.08
		fit2	0.085	199400 ± 800	0.078	170800 ± 600		
	K^-	fit1	0.08	199600 ± 800	0.08	171100 ± 800	$85.71 \pm 0.08 \pm 0.65$	85.28 ± 0.08
		fit2	0.084	200400 ± 800	0.078	170800 ± 600		
	average						$85.97 \pm 0.06 \pm 0.52$	85.48 ± 0.06
Data	K^+	fit1	0.08	248000 ± 1000	0.08	208800 ± 800	$84.22 \pm 0.07 \pm 0.80$	-
		fit2	0.085	249000 ± 1000	0.078	208400 ± 800		
	K^-	fit1	0.08	251000 ± 1000	0.08	208800 ± 800	$83.02 \pm 0.07 \pm 0.64$	-
		fit2	0.084	252000 ± 1000	0.078	208400 ± 800		
	average						$83.62 \pm 0.05 \pm 0.51$	-

Sample			before PID		after PID ($P_K > 0.5$)		Fake rate	
			f_1	N_{sig}	f_1	N_{sig}	f (%)	f^{true} (%)
MC15ri	π^+	fit1	0.08	115100 ± 700	0.15	6400 ± 300	$5.55 \pm 0.07 \pm 0.34$	5.32 ± 0.07
		fit2	0.09	116300 ± 700	0.115	6200 ± 300		
	π^-	fit1	0.08	111800 ± 700	0.15	6400 ± 300	$5.73 \pm 0.07 \pm 0.44$	6.04 ± 0.07
		fit2	0.087	112600 ± 700	0.205	6800 ± 300		
	average						$5.64 \pm 0.05 \pm 0.28$	5.68 ± 0.05
Data	π^+	fit1	0.08	148700 ± 900	0.15	7800 ± 300	$5.28 \pm 0.06 \pm 0.33$	-
		fit2	0.09	150200 ± 900	0.115	7600 ± 300		
	π^-	fit1	0.08	142000 ± 800	0.15	9300 ± 300	$6.53 \pm 0.07 \pm 0.47$	-
		fit2	0.087	143000 ± 800	0.205	9800 ± 400		
	average						$5.90 \pm 0.04 \pm 0.29$	-

Table 5.1: Signal yields before and after K-ID selection ($P_K > 0.5$) on K and π tracks using both MC and data. K-ID efficiencies and π mis-ID rates (positive, negative and average). The uncertainties are statistical, the second uncertainty in the efficiency/fare rate column is the systematic uncertainty.

Figure 5.1: Invariant mass distribution of the D_s candidates in MC sample: before (left) and after (right) applying $P_K > 0.5$ criterion on K^+ tracks. The blue line is the best fitted PDF, the red peak corresponds to the estimated signal yield. The relative pulls are under the corresponding plot.

Figure 5.2: Invariant mass distribution of the D_s candidates in MC sample: before (left) and after (right) applying $P_K > 0.5$ criterion on K^- tracks. The blue line is the best fitted PDF, the red peak corresponds to the estimated signal yield. The relative pulls are under the corresponding plot.
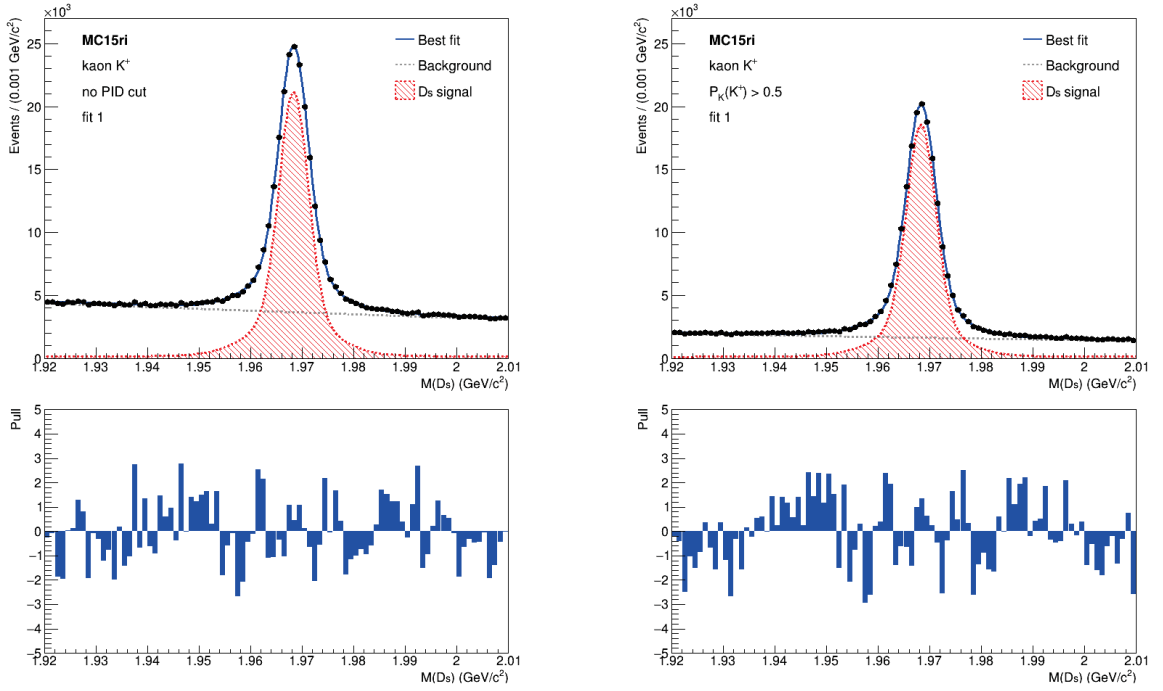
Figure 5.3: Invariant mass distribution of the D_s candidates in data sample: before (left) and after (right) applying $P_K > 0.5$ criterion on K^+ tracks. The blue line is the best fitted PDF, the red peak corresponds to the estimated signal yield. The relative pulls are under the corresponding plot.

Figure 5.4: Invariant mass distribution of the D_s candidates in data sample: before (left) and after (right) applying $P_K > 0.5$ criterion on K^- tracks. The blue line is the best fitted PDF, the red peak corresponds to the estimated signal yield. The relative pulls are under the corresponding plot.
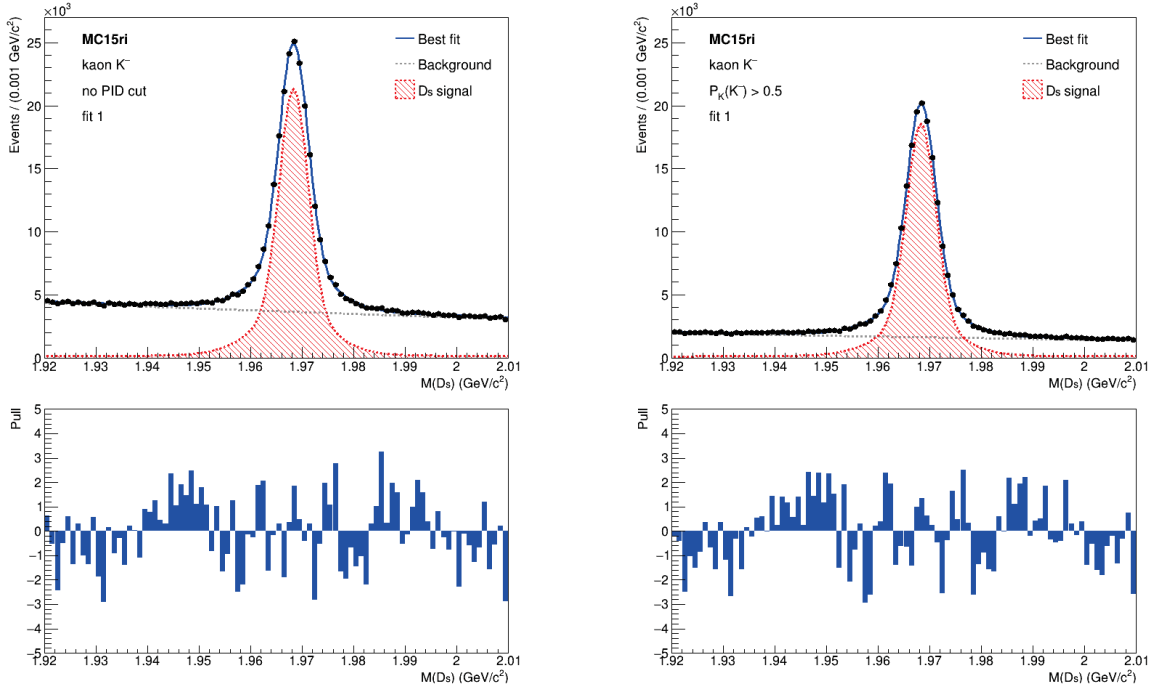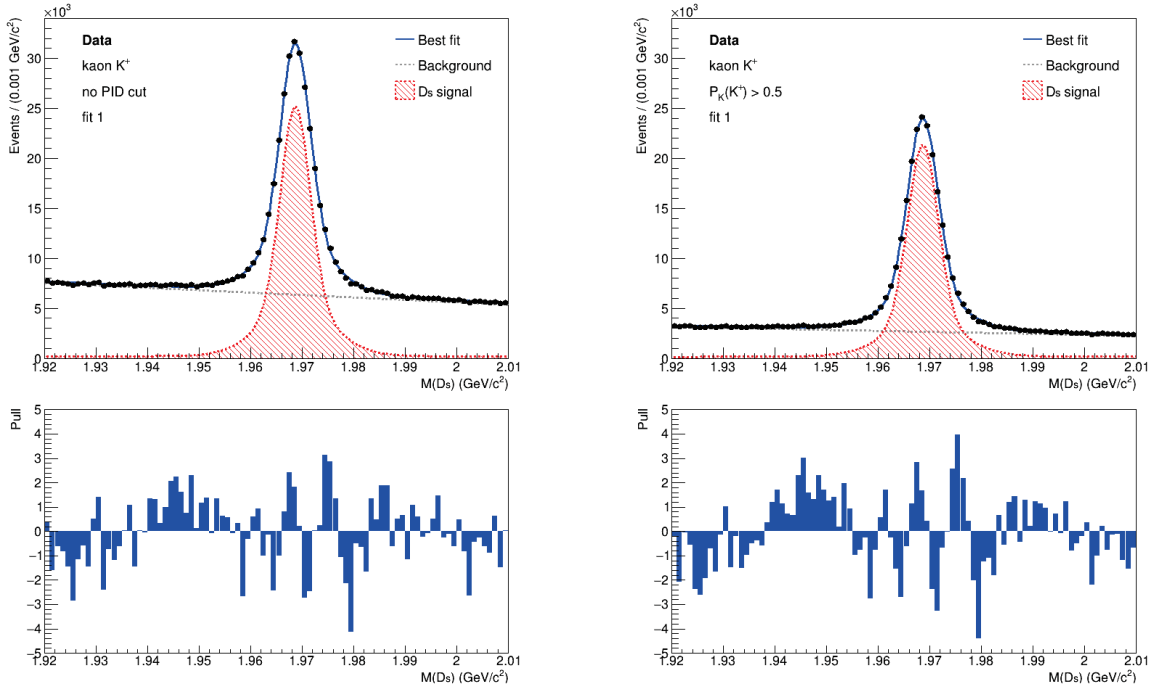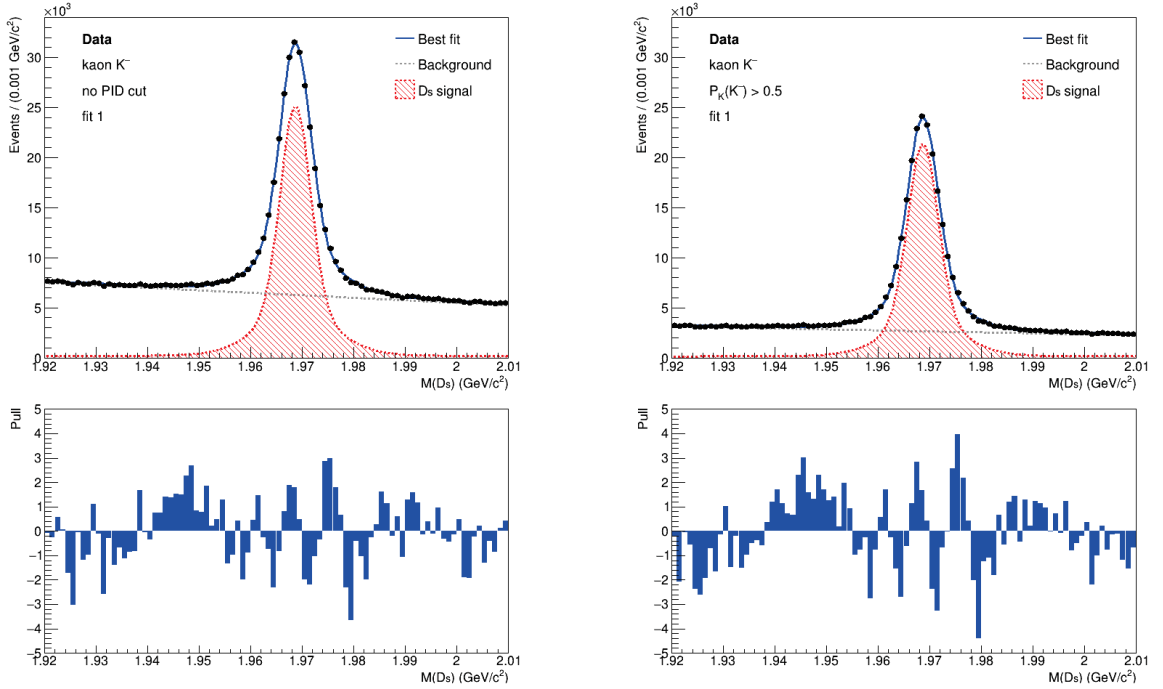
Figure 5.5: Invariant mass distribution of the D_s candidates in MC sample: before (left) and after (right) applying $P_K > 0.5$ criterion on π^+ tracks. The blue line is the best fitted PDF, the red peak corresponds to the estimated signal yield. The event count scale is very different between the two plots, on the right it is much lower. The relative pulls are under the corresponding plot.
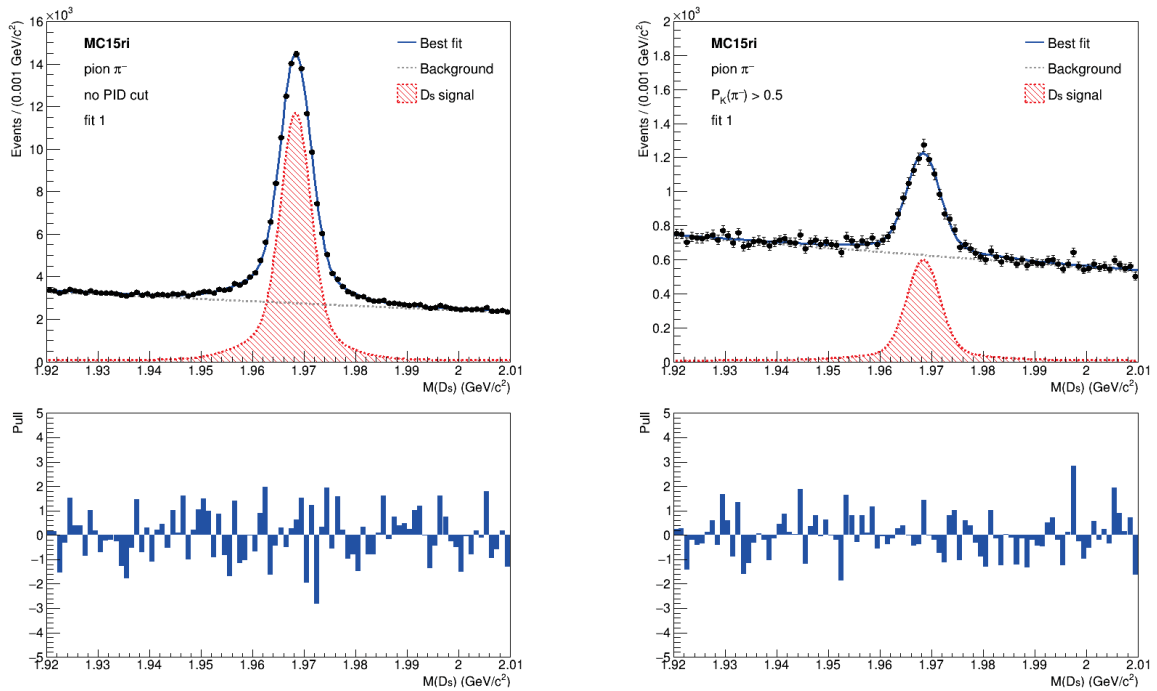
Figure 5.6: Invariant mass distribution of the D_s candidates in MC sample: before (left) and after (right) applying $P_K > 0.5$ criterion on π^- tracks. The blue line is the best fitted PDF, the red peak corresponds to the estimated signal yield. The event count scale is very different between the two plots, on the right it is much lower. The relative pulls are under the corresponding plot.
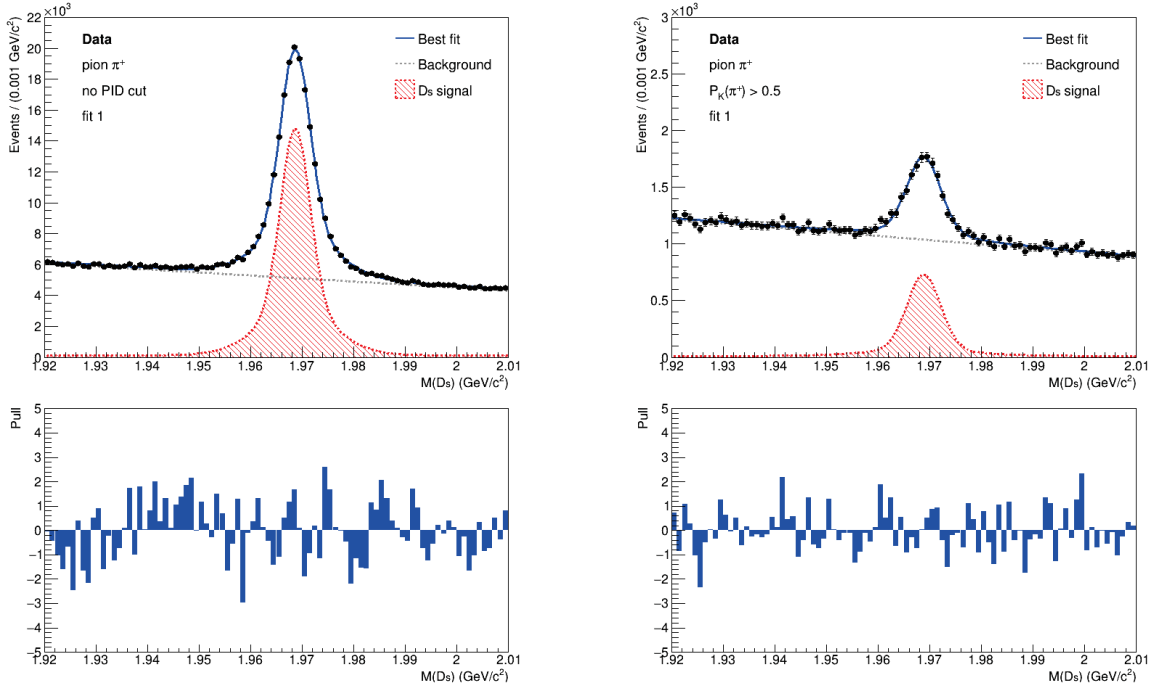
Figure 5.7: Invariant mass distribution of the D_s candidates in data sample: before (left) and after (right) applying $P_K > 0.5$ criterion on π^+ tracks. The blue line is the best fitted PDF, the red peak corresponds to the estimated signal yield. The event count scale is very different between the two plots, on the right it is much lower. The relative pulls are under the corresponding plot.
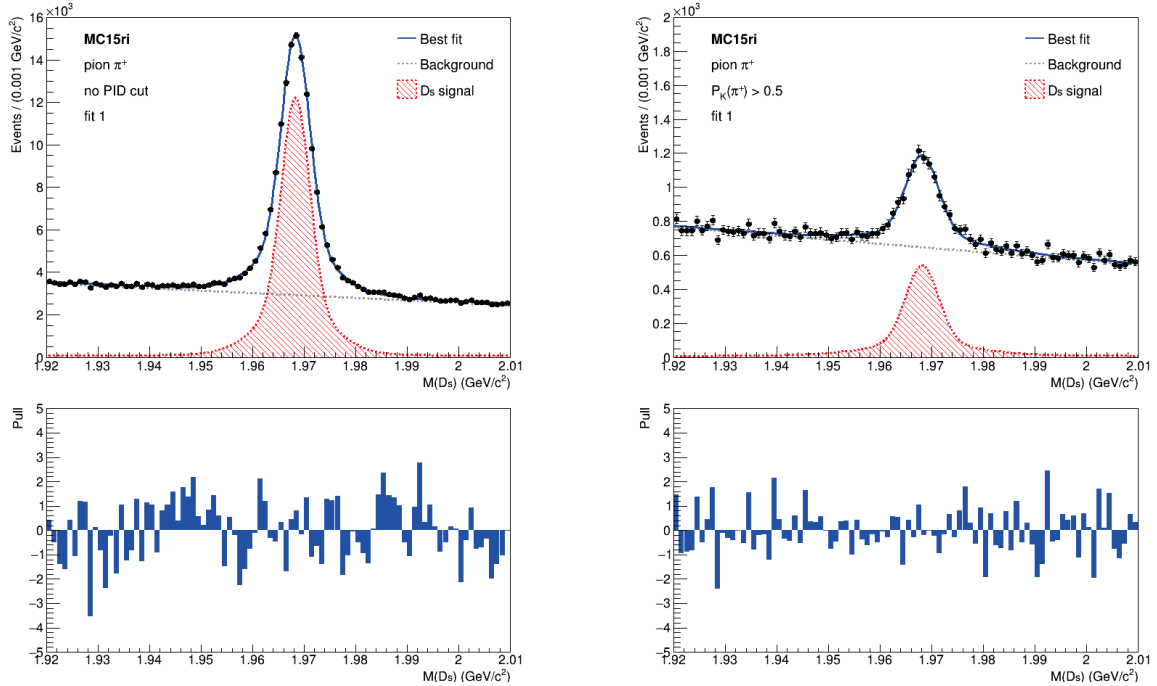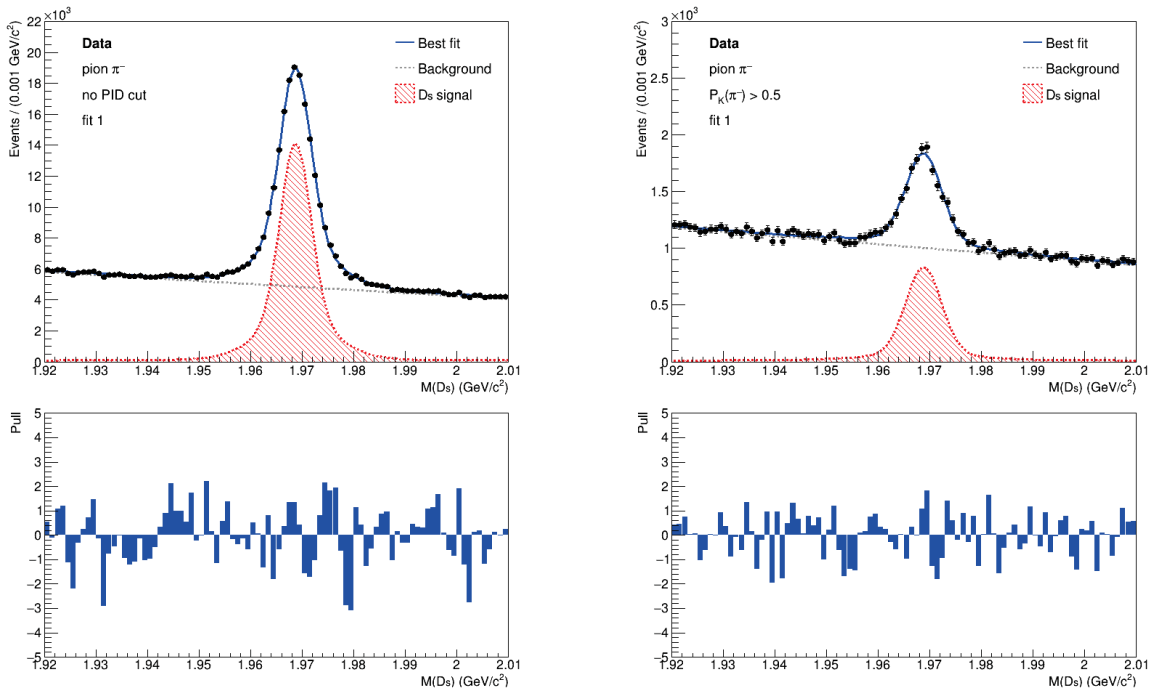

Figure 5.8: Invariant mass distribution of the D_s candidates in data sample: before (left) and after (right) applying $P_K > 0.5$ criterion on π^- tracks. The blue line is the best fitted PDF, the red peak corresponds to the estimated signal yield. The event count scale is very different between the two plots, on the right it is much lower. The relative pulls are under the corresponding plot.

Table 5.2: Summary of results for the K-ID performance based on the selection $P_K > 0.5$. K-ID efficiencies, π mis-ID rates and charge asymmetries. The measurements are reported with their total uncertainty. λ quantifies the agreement between the charge asymmetries in MC and real data.

	MC15ri	MC truth	Data	λ
$\varepsilon_K{}^+$ (%)	86.2 ± 0.8	85.68 ± 0.08	84.2 ± 0.8	-
$\varepsilon_K{}^-$ (%)	85.7 ± 0.6	85.28 ± 0.08	83.0 ± 0.6	-
$\bar{\varepsilon}_K$ (%)	86.0 ± 0.5	85.48 ± 0.06	83.6 ± 0.5	-
$\mathcal{A}(\varepsilon)$	0.003 ± 0.006	0.0023 ± 0.0007	0.007 ± 0.006	0.5
$f_\pi{}^+$ (%)	5.6 ± 0.4	5.32 ± 0.07	5.3 ± 0.3	-
$f_\pi{}^-$ (%)	5.7 ± 0.5	6.04 ± 0.07	6.5 ± 0.5	-
$\bar{f}_\pi$ (%)	5.6 ± 0.3	5.68 ± 0.05	5.9 ± 0.3	-
$\mathcal{A}(f)$	-0.02 ± 0.05	-0.064 ± 0.008	-0.11 ± 0.05	1.3

5.2 Pion identification efficiencies and kaon mis-ID rates

We present the pion identification performance based on global PID probability selections. We divide the sample by selecting only events with positively charged pions or negatively charged pions. Then we compare the signal yield in the D_s invariant mass plot when PID selection criteria are applied to charged pion track to the signal yield when no criteria are applied. The PID selection criterion consists of a threshold cut on the global PID probability: $P_\pi > \alpha$. In formulas, we take the ratios

$$\varepsilon_\pi{}^+ = \frac{N_{\mathrm{sig}}(\pi^+ \,;\, P_\pi(\pi^+) > \alpha)}{N_{\mathrm{sig}}(\pi^+)} \tag{5.6}$$

$$\varepsilon_\pi{}^- = \frac{N_{\mathrm{sig}}(\pi^- \,;\, P_\pi(\pi^-) > \alpha)}{N_{\mathrm{sig}}(\pi^-)} \tag{5.7}$$

We conduct the study by applying $P_\pi > 0.5$. We follow the same method described in section 5.1 to determine the π-ID efficiencies with fits of the D_s invariant mass plot before and after applying PID selection to oppositely charged pion tracks. We summarize the results in Tab. 5.3: the values of the parameter f_1, the signal yields resulting from the fitting procedures and the PID efficiencies for both simulation and data. The π-ID efficiencies on real data are slightly lower than those on MC.
In Fig. 5.9 and Fig. 5.10 we can see the distributions of the invariant mass of the D_s candidates before and after the PID selection and the fitted functions on simulation for positive and negative pion tracks, respectively. In Fig. 5.11 and Fig. 5.12 we show the same but for real data with PID selection on charged pion tracks. The pulls for each fitting procedure are normally distributed.

We determine the kaon mis-identification rate by applying π-ID selection criterion ($P_\pi > \alpha$) to positive or negative kaon tracks. As said in the previous section, a preselection $P_K > 0.5$ is applied to opposite charge kaons. In formulas, we write

$$f_K{}^+ = \frac{N_{\mathrm{sig}}(K^+ \,;\, P_\pi(K^+) > \alpha)}{N_{\mathrm{sig}}(K^+)} \tag{5.8}$$

$$f_K{}^- = \frac{N_{\mathrm{sig}}(K^- \,;\, P_\pi(K^-) > \alpha)}{N_{\mathrm{sig}}(K^-)} \tag{5.9}$$

In this study, we choose $\alpha = 0.5$. We determine the signal yields before and after applying mis-ID selection with fit of the D_s mass distribution, the PDF model is defined in Eq. 4.1. We follow the same method described for the determination of the pion fake rates. We summarize the resulting K mis-ID rates for both MC and real data in Tab. 5.3. We can observe that the fake rates on data are slightly higher than those on MC, as expected.
In Fig. 5.13 and Fig. 5.14 we show the D_s invariant mass distributions and the fitted functions on MC, in Fig. 5.15 and Fig. 5.16 the distributions and the fitted functions on real data, before and after the mis-identification selection on positive and negative kaons. The mis-ID selection on kaon tracks significantly reduces the peak in the D_s mass distribution. The pulls for the fits after the PID selection are normally distributed, instead the pulls for the fits before the PID selection are distributed according to a Gaussian with center at zero and width greater than one.

In Tab. 5.4 we report the measured π-ID efficiencies and K mis-ID rates based on the global PID selection $P_\pi > 0.5$ for both MC and real data. The uncertainties are the sum in quadrature of statistical and systematic errors. We report also the results using MC truth information and their statistical uncertainty for simulated data. Then, we determine the charge asymmetries of both efficiencies and fake rates according to Eq. 3.19 and Eq. 3.20, the uncertainties are obtained by propagation. From the table, there is excellent agreement between the measured charge asymmetries of both efficiencies and mis-ID rates ($\lambda < 1$). Both charge asymmetries of the efficiencies $\mathcal{A}(\varepsilon)$ for simulation and experimental data are very close to zero. The charge asymmetry of the fake rates $\mathcal{A}(f)$ for MC is negative while that for real data is positive, but both are close to zero.

| Sample | | | before PID | | after PID ($P_\pi > 0.5$) | | Efficiency | |
			f_1	N_{sig}	f_1	N_{sig}	ε (%)	$\varepsilon^{\mathrm{true}}$ (%)
MC15ri	π^+	fit1	0.08	115100 ± 700	0.06	88800 ± 500	$77.1 \pm 0.1 \pm 0.7$	76.6 ± 0.1
		fit2	0.09	116300 ± 700	0.063	89000 ± 500		
	π^-	fit1	0.08	111800 ± 700	0.06	85600 ± 500	$76.5 \pm 0.1 \pm 0.8$	76.0 ± 0.1
		fit2	0.087	112600 ± 700	0.06	85600 ± 500		
	average						$76.80 \pm 0.09 \pm 0.53$	76.3 ± 0.1
Data	π^+	fit1	0.08	148700 ± 900	0.06	110500 ± 600	$74.3 \pm 0.1 \pm 0.7$	-
		fit2	0.09	150200 ± 900	0.063	110800 ± 600		
	π^-	fit1	0.08	142000 ± 800	0.06	104200 ± 600	$73.4 \pm 0.1 \pm 0.7$	-
		fit2	0.087	143000 ± 800	0.06	104200 ± 600		
	average						$73.83 \pm 0.08 \pm 0.52$	-

| Sample | | | before PID | | after PID ($P_\pi > 0.5$) | | Fake rate | |
			f_1	N_{sig}	f_1	N_{sig}	f (%)	f^{true} (%)
MC15ri	K^+	fit1	0.08	198400 ± 800	0.08	7300 ± 300	$3.70 \pm 0.04 \pm 0.01$	3.71 ± 0.04
		fit2	0.085	199400 ± 800	0.085	7400 ± 300		
	K^-	fit1	0.08	199600 ± 800	0.08	8000 ± 300	$3.98 \pm 0.04 \pm 0.37$	3.73 ± 0.04
		fit2	0.084	200400 ± 800	0.01	7400 ± 300		
	average						$3.84 \pm 0.03 \pm 0.19$	3.72 ± 0.03
Data	K^+	fit1	0.08	248000 ± 1000	0.08	10900 ± 500	$4.39 \pm 0.04 \pm 0.01$	-
		fit2	0.085	249000 ± 1000	0.085	10900 ± 500		
	K^-	fit1	0.08	251000 ± 1000	0.08	10800 ± 300	$4.31 \pm 0.04 \pm 0.26$	-
		fit2	0.084	252000 ± 1000	0.01	11000 ± 400		
	average						$4.35 \pm 0.03 \pm 0.13$	-

Table 5.3: Signal yields before and after π-ID selection ($P_\pi > 0.5$) on π and K tracks using both MC and data. π-ID efficiencies and K mis-ID rates (positive, negative and average). The uncertainties are statistical, the second uncertainty in the efficiency/fare rate column is the systematic uncertainty.

Figure 5.9: Invariant mass distribution of the D_s candidates in MC sample: before (left) and after (right) applying $P_\pi > 0.5$ criterion on π^+ tracks. The blue line is the best fitted PDF, the red peak corresponds to the estimated signal yield. The relative pulls are under the corresponding plot.

Figure 5.10: Invariant mass distribution of the D_s candidates in MC sample: before (left) and after (right) applying $P_\pi > 0.5$ criterion on π^- tracks. The blue line is the best fitted PDF, the red peak corresponds to the estimated signal yield. The relative pulls are under the corresponding plot.
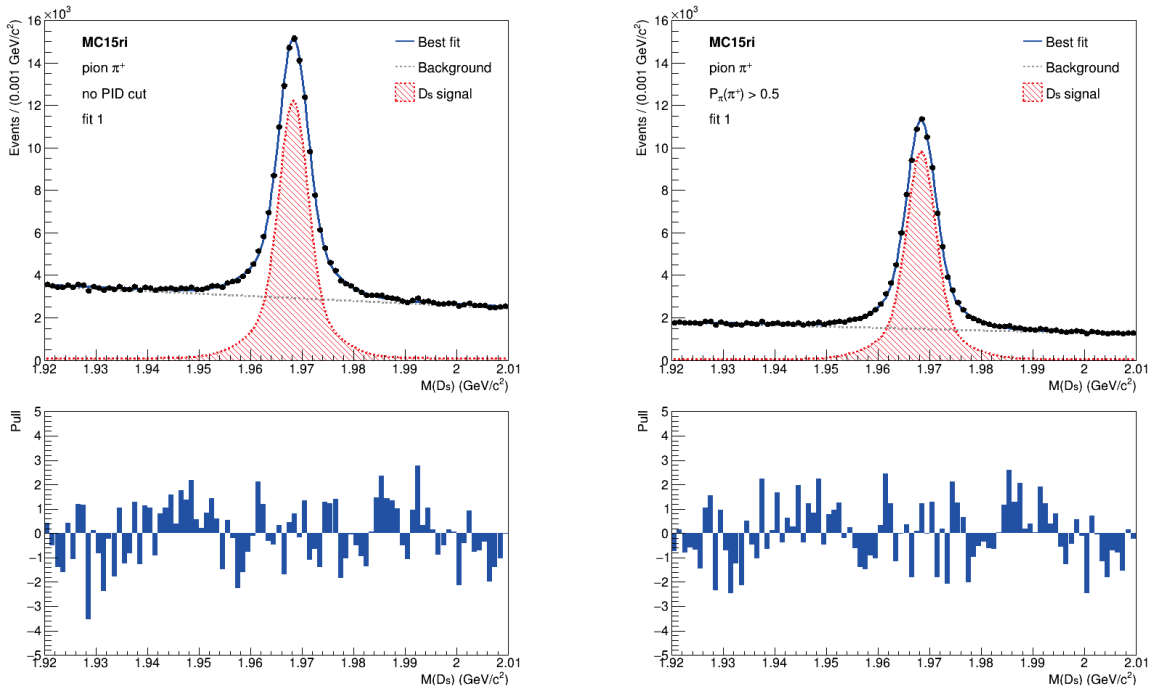
Figure 5.11: Invariant mass distribution of the D_s candidates in data sample: before (left) and after (right) applying $P_\pi > 0.5$ criterion on π^+ tracks. The blue line is the best fitted PDF, the red peak corresponds to the estimated signal yield. The relative pulls are under the corresponding plot.

Figure 5.12: Invariant mass distribution of the D_s candidates in data sample: before (left) and after (right) applying $P_\pi > 0.5$ criterion on π^- tracks. The blue line is the best fitted PDF, the red peak corresponds to the estimated signal yield. The relative pulls are under the corresponding plot.
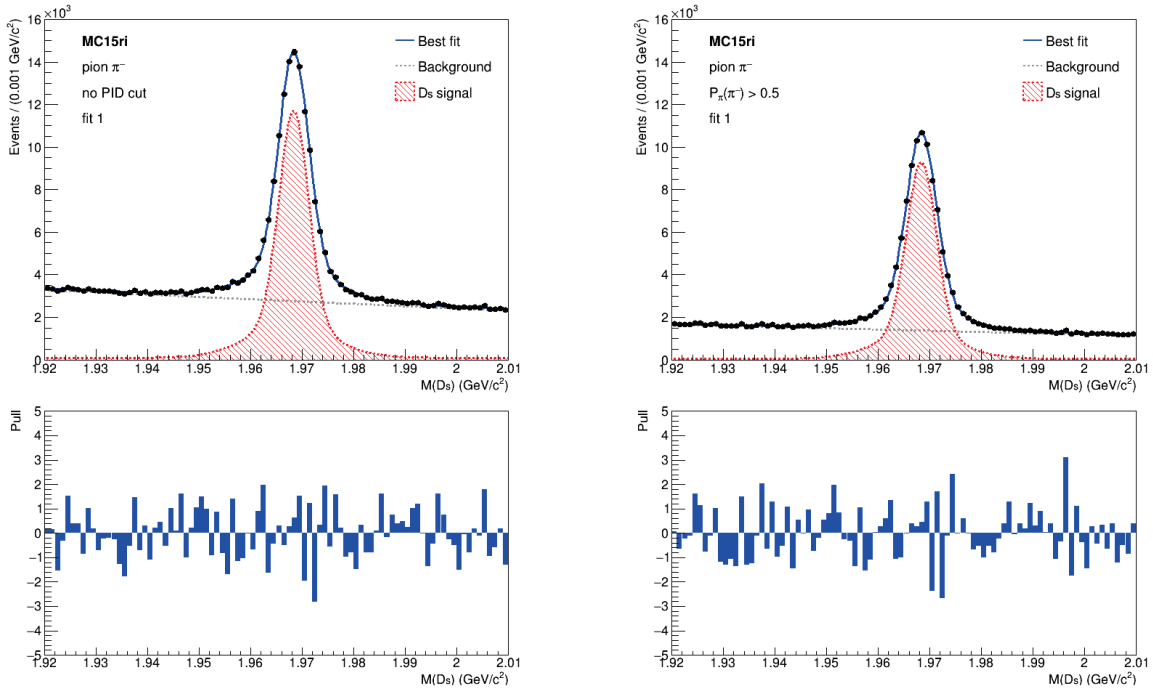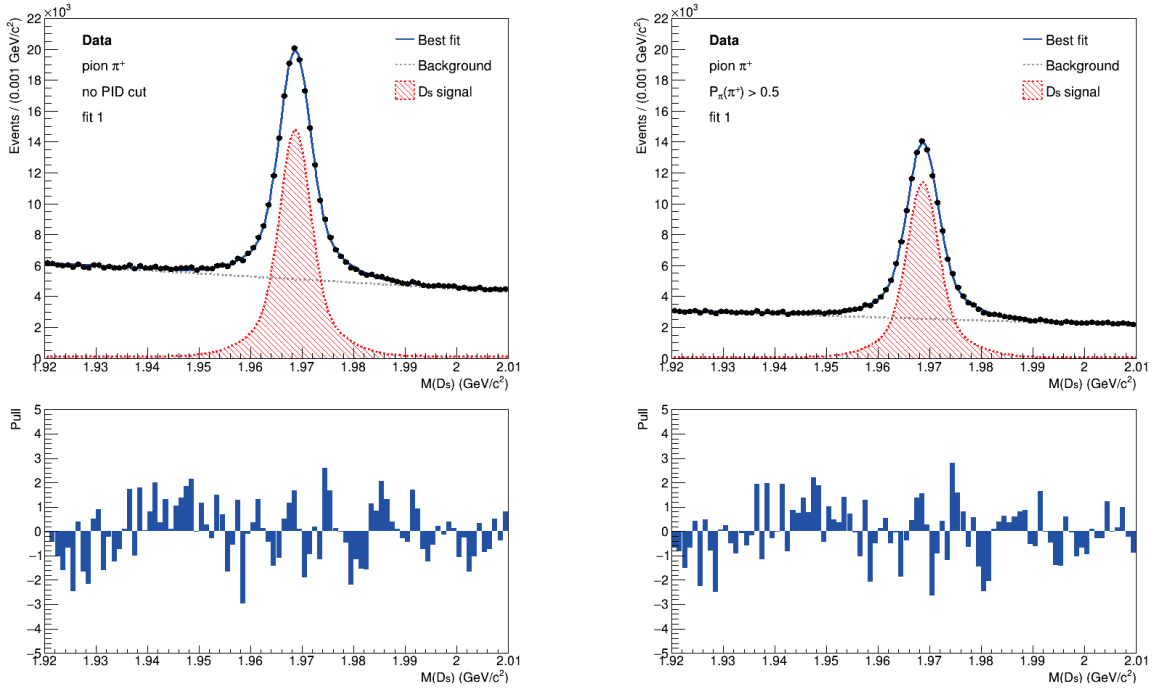
Figure 5.13: Invariant mass distribution of the D_s candidates in MC sample: before (left) and after (right) applying $P_\pi > 0.5$ criterion on K^+ tracks. The blue line is the best fitted PDF, the red peak corresponds to the estimated signal yield. The event count scale is very different between the two plots, on the right it is much lower. The relative pulls are under the corresponding plot.
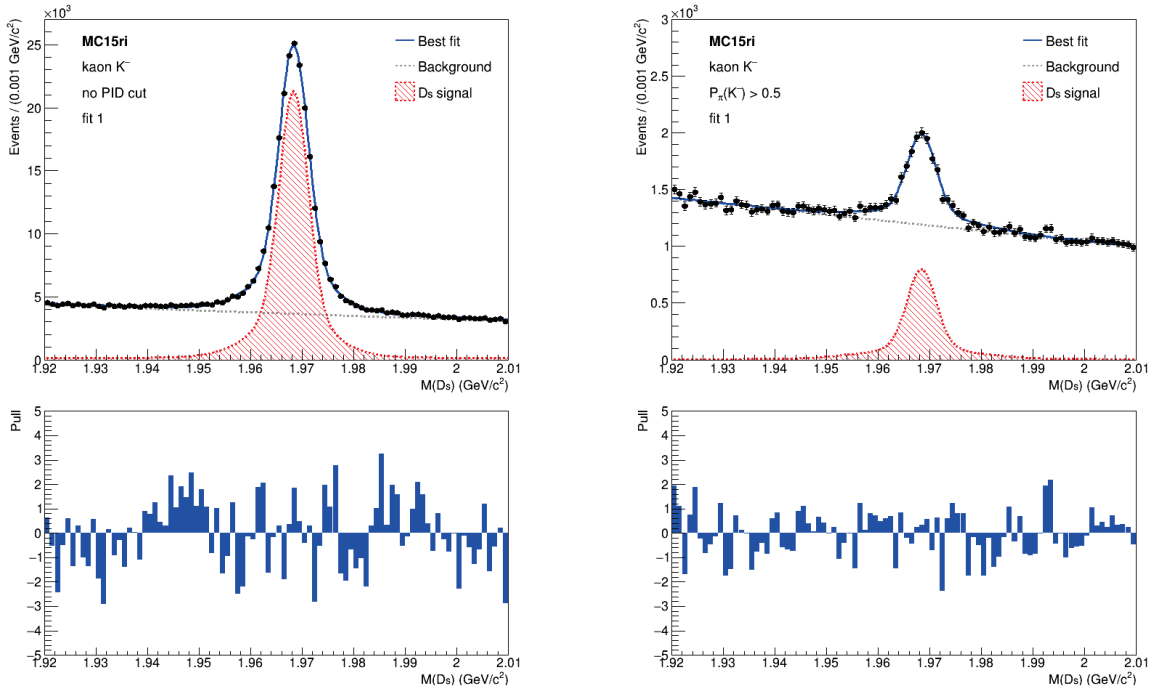
Figure 5.14: Invariant mass distribution of the D_s candidates in MC sample: before (left) and after (right) applying $P_\pi > 0.5$ criterion on K^- tracks. The blue line is the best fitted PDF, the red peak corresponds to the estimated signal yield. The event count scale is very different between the two plots, on the right it is much lower. The relative pulls are under the corresponding plot.
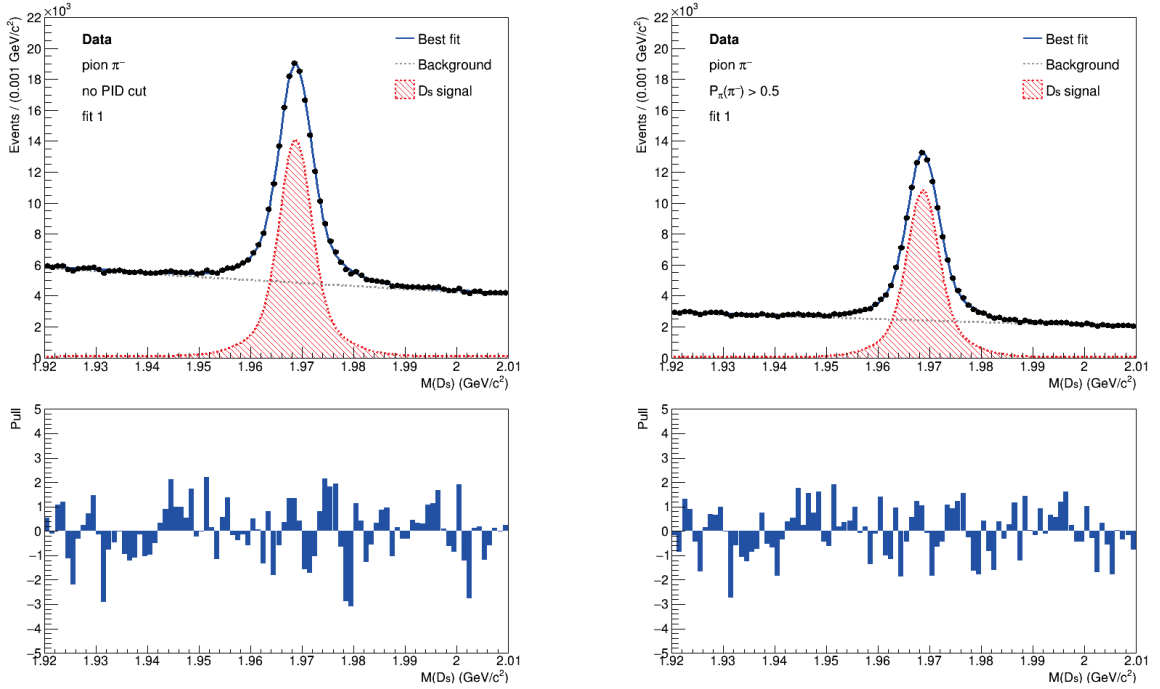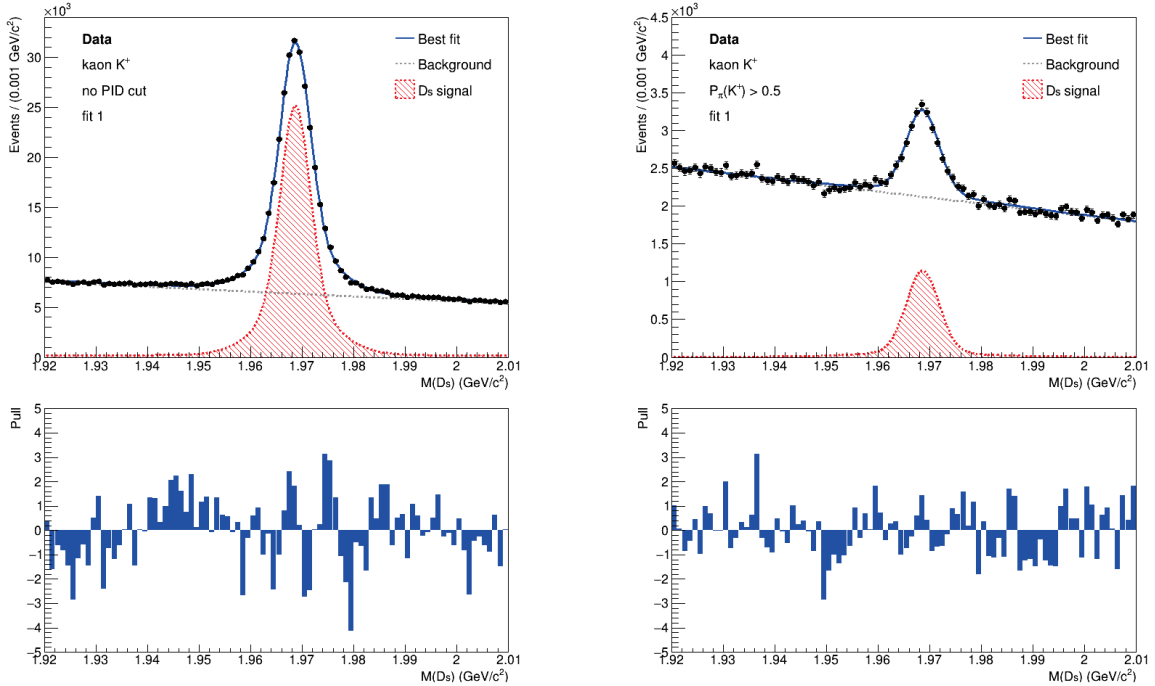
Figure 5.15: Invariant mass distribution of the D_s candidates in data sample: before (left) and after (right) applying $P_\pi > 0.5$ criterion on K^+ tracks. The blue line is the best fitted PDF, the red peak corresponds to the estimated signal yield. The event count scale is very different between the two plots, on the right it is much lower. The relative pulls are under the corresponding plot.
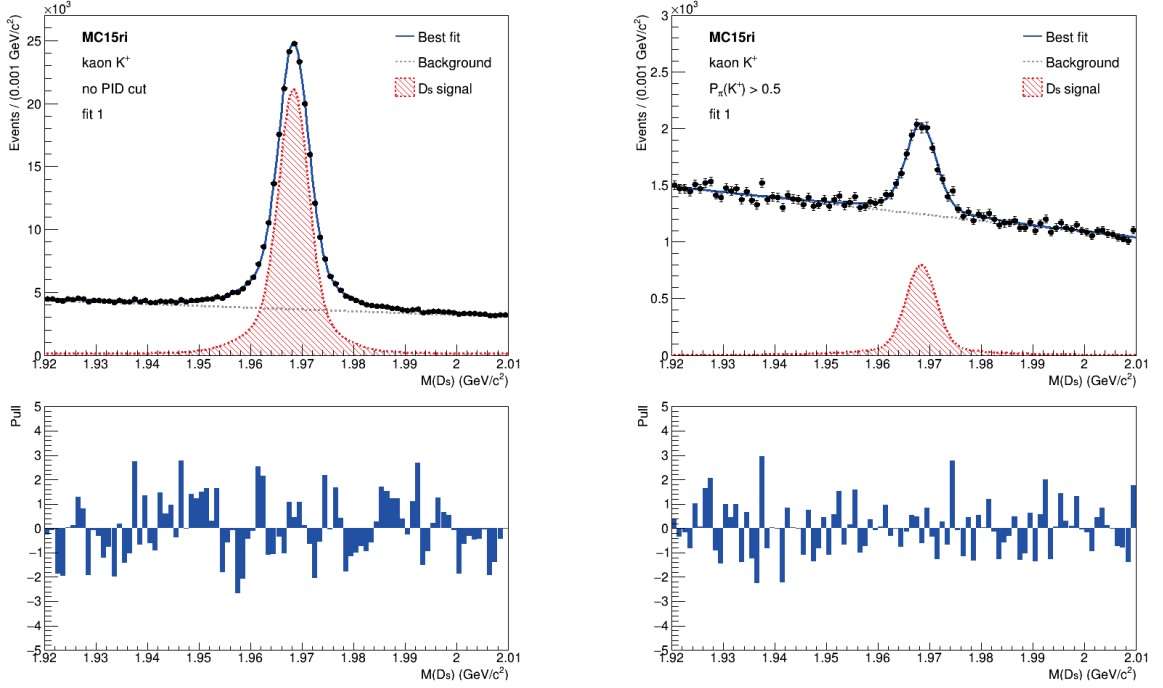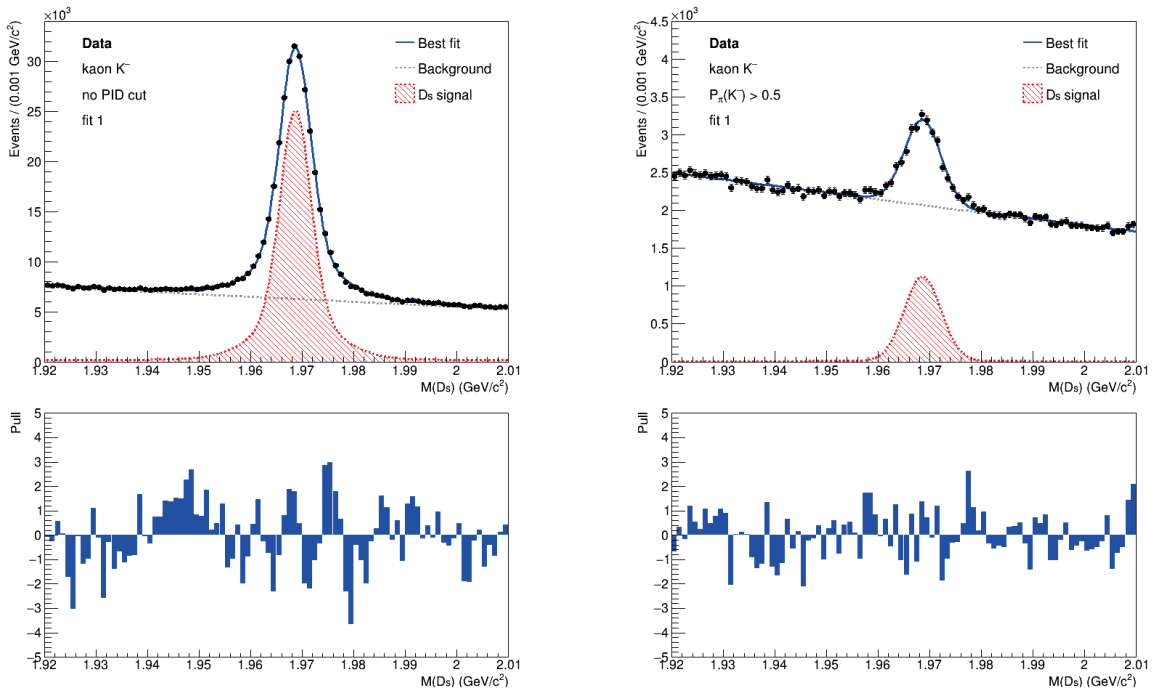

Figure 5.16: Invariant mass distribution of the D_s candidates in data sample: before (left) and after (right) applying $P_\pi > 0.5$ criterion on K^- tracks. The blue line is the best fitted PDF, the red peak corresponds to the estimated signal yield. The event count scale is very different between the two plots, on the right it is much lower. The relative pulls are under the corresponding plot.

Table 5.4: Summary of results for the π-ID performance based on the selection $P_\pi > 0.5$. π-ID efficiencies, K mis-ID rates and charge asymmetries. The measurements are reported with their total uncertainty. λ quantifies the agreement between the charge asymmetries in MC and real data.

	MC15ri	MC truth	Data	λ
$\varepsilon_\pi{}^+$ (%)	77.1 ± 0.7	76.6 ± 0.1	74.3 ± 0.7	-
$\varepsilon_\pi{}^-$ (%)	76.5 ± 0.8	76.0 ± 0.1	73.4 ± 0.7	-
$\bar{\varepsilon}_\pi$ (%)	76.8 ± 0.5	76.3 ± 0.1	73.8 ± 0.5	-
$\mathcal{A}(\varepsilon)$	0.004 ± 0.007	0.004 ± 0.001	0.006 ± 0.007	0.3
$f_K{}^+$ (%)	3.70 ± 0.04	3.71 ± 0.04	4.39 ± 0.04	-
$f_K{}^-$ (%)	4.0 ± 0.4	3.73 ± 0.04	4.3 ± 0.3	-
$\bar{f}_K$ (%)	3.8 ± 0.2	3.72 ± 0.03	4.3 ± 0.1	-
$\mathcal{A}(f)$	-0.04 ± 0.05	-0.002 ± 0.008	0.01 ± 0.03	0.8

5.3 PID performance and charge asymmetries in bins of kinematic variables

The PID performance discussed in the previous sections is the performance averaged over the kinematic distribution of the samples used for the analysis. In this section we investigate the PID performance as a function of two of the main kinematic variables: momentum p in the laboratory frame and cosine of the polar angle θ.

In the preselection of our decay sample we required kaon tracks with $p \in [0.3, 3.5]$ GeV/c and pion tracks with $p \in [0.3, 5.0]$ GeV/c. Fig. 5.17 and Fig. 5.18 show the distribution of momentum for charged kaons and pions, respectively. For the MC sample, the p distribution of total events and that counting only truth-matched D_s candidates are shown on the left. We observe that the p distribution for real data (in the figures on the right) has the same shape as that for simulated data. To study the PID performance as a function of momentum, we divide our charged particle samples in 5 bins of momentum. The criterion for bin division is to ensure that the true signal yield in each bin, after applying a PID selection to kaon or pion tracks, is roughly uniform. Therefore, the bin selection is determined based on simulated data and the same division is applied to the experimental data set. The binning depends on the identification or mis-identification criterion applied to K or π tracks. Tab. 5.5 reports the momentum bins chosen by particle (K or π) and by PID selection: we show the low and high p values for each bin, and the number of true signal events in MC sample for positive and negative charge.

To study the PID performance as a function of the polar angle we divide in 5 bins of $\cos \theta$ across the geometrical acceptance of the TOP detector ($30° < \theta < 120°$). In Fig. 5.19 and Fig. 5.20 we can see the distribution of $\cos \theta$ for charged kaons and pions, respectively. For the simulated data we show the distribution of all the events, with the distribution of true signal events overlaid (left), for the real data we show the distribution of all the events (right). We point out that no constraints on the polar angle were imposed in the preselection, therefore the $\cos \theta$ distribution of our charged particle samples covers a wider range than the TOP acceptance. Since our goal is to study the identification performed by the TOP detector, we restrict our analysis of PID performance as a function of $\cos \theta$ in the range between -0.5 and 0.85. We divide into 5 bins following the same criterion described above for the momentum bins. Also in this case the binning depends on the identification or mis-identification criterion applied to K or π tracks. In Tab. 5.6 we report the chosen binning, indicating the low and wigh $\cos \theta$, and the number of true signal events in MC sample for positive and negative charge.

In the following subsections we present the PID performance for kaons and pions in bins of kinematic variables. We determine the signal yield in each bin before and after applying selection to the global PID probability. We fit the distribution of the invariant mass of the D_s candidates with the PDF model in Eq. 4.1. We follow the same strategy described in the previous sections (5.1 and 5.2) to measure the efficiencies and mis-ID rates (and their uncertainties) as a function of p and $\cos \theta$ for MC and real data.
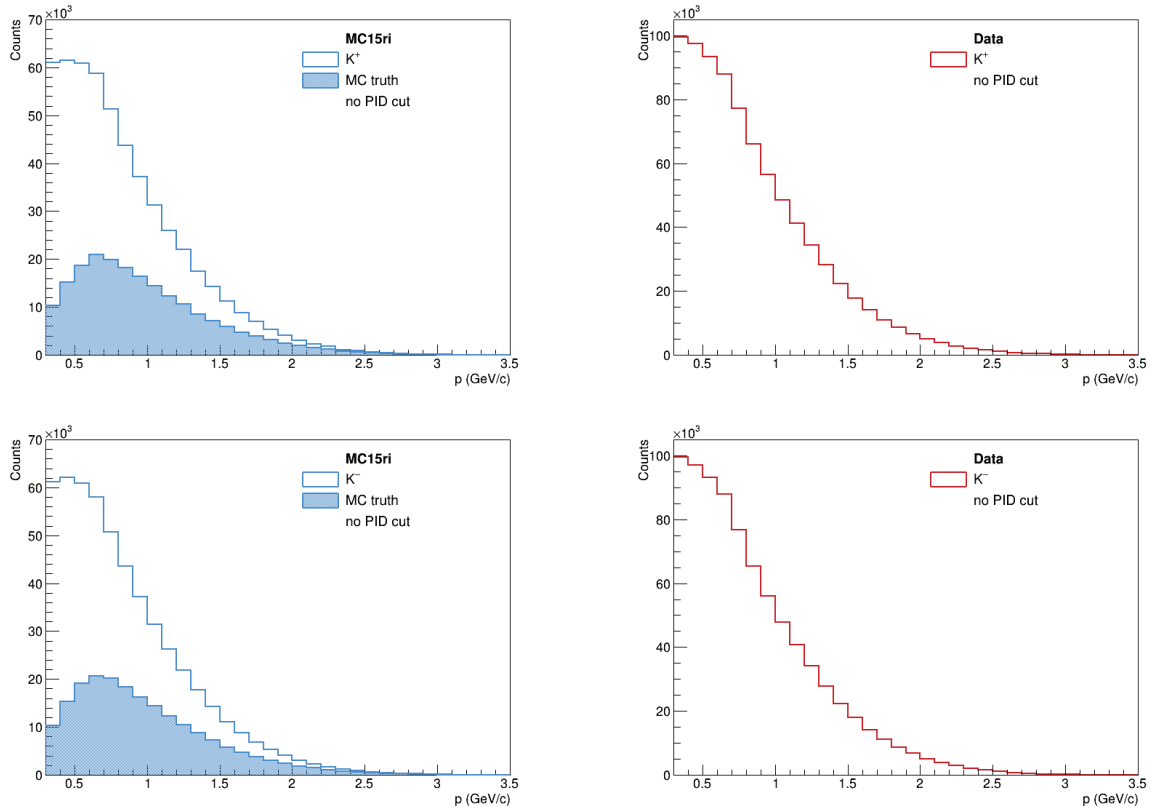
Figure 5.17: Distribution of the momentum p of the positive (top) and negative (bottom) charged kaon tracks. MC (left) and real data (right).
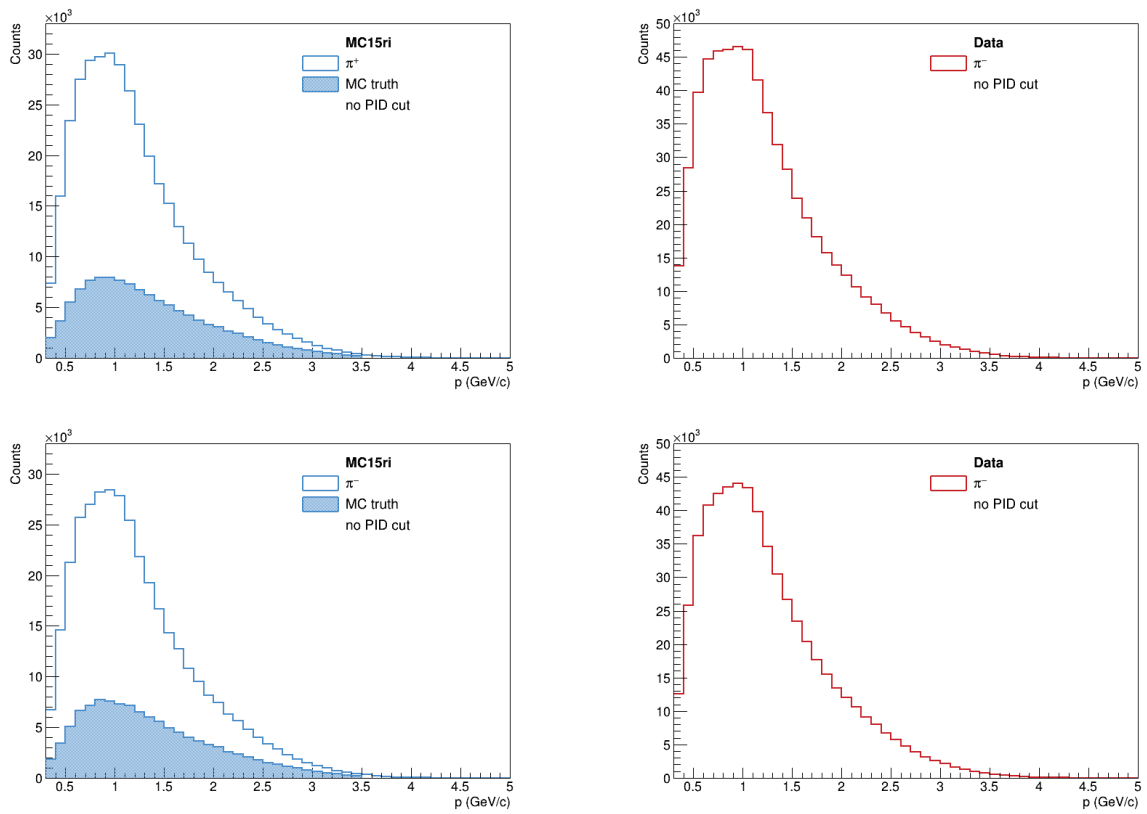
Figure 5.18: Distribution of the momentum p of the positive (top) and negative (bottom) charged pion tracks. MC (left) and real data (right).

Table 5.5: Momentum bins for kaons and pions depending on the PID selection. For each bin, low and high p and number of true signal events in MC sample for positive and negative charge.

Particle	PID selection	Bin	Low p [GeV/c]	High p [GeV/c]	$N_{\text{sig}}^{\text{true}}(+)$	$N_{\text{sig}}^{\text{true}}(-)$
		1	0.3	0.6	40924	41200
		2	0.6	0.8	36052	35838
K	$P_K > 0.5$	3	0.8	1.0	29516	29192
		4	1.0	1.3	30842	30627
		5	1.3	3.5	33482	33959
		1	0.3	0.8	1567	1580
		2	0.8	1.1	1562	1713
π	$P_\pi > 0.5$	3	1.1	1.5	1493	1462
		4	1.5	2.0	1488	1344
		5	2.0	5.0	1289	1371
		1	0.3	0.9	18335	16899
		2	0.9	1.1	17998	17312
K	$P_\pi > 0.5$	3	1.1	1.3	20411	19668
		4	1.3	1.6	16361	15888
		5	1.6	3.5	15910	15777
		1	0.3	0.9	1149	1234
		2	0.9	1.1	1203	1206
π	$P_K > 0.5$	3	1.1	1.3	1207	1367
		4	1.3	1.8	1440	1683
		5	1.8	5.0	1181	1315

Figure 5.19: Distribution of the cosine of the polar angle θ of the positive (top) and negative (bottom) charged kaon tracks. MC (left) and real data (right).

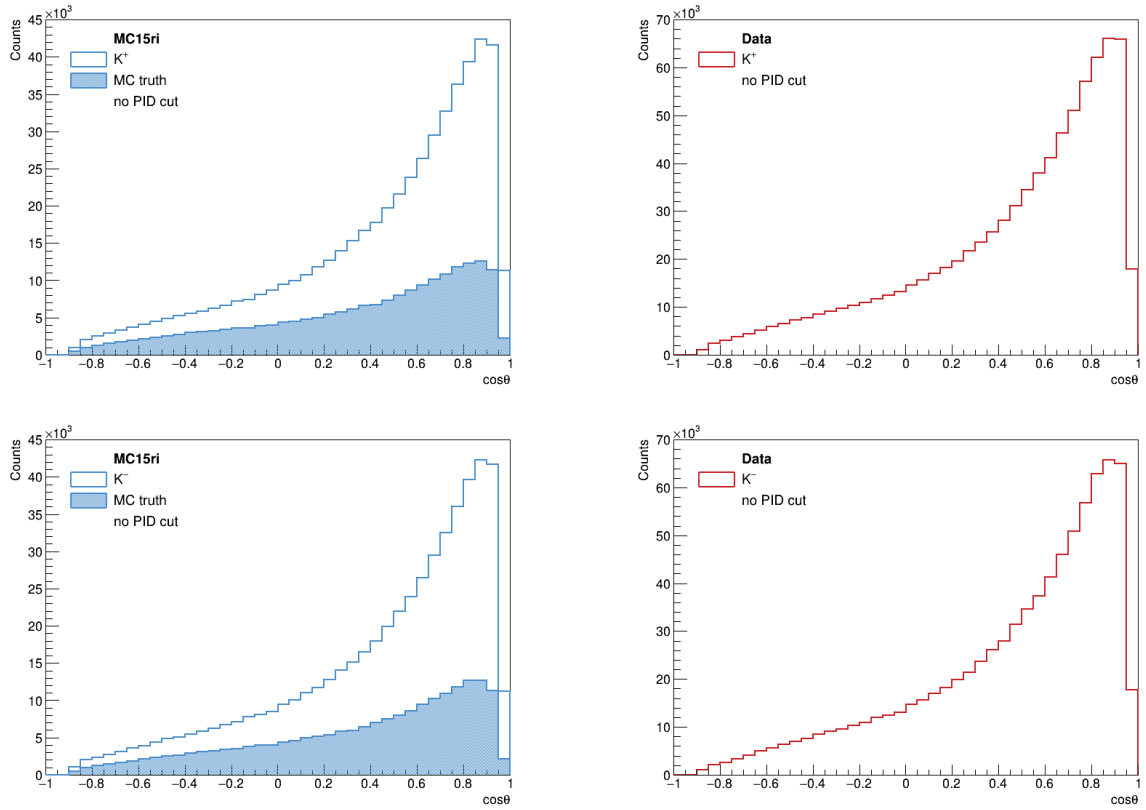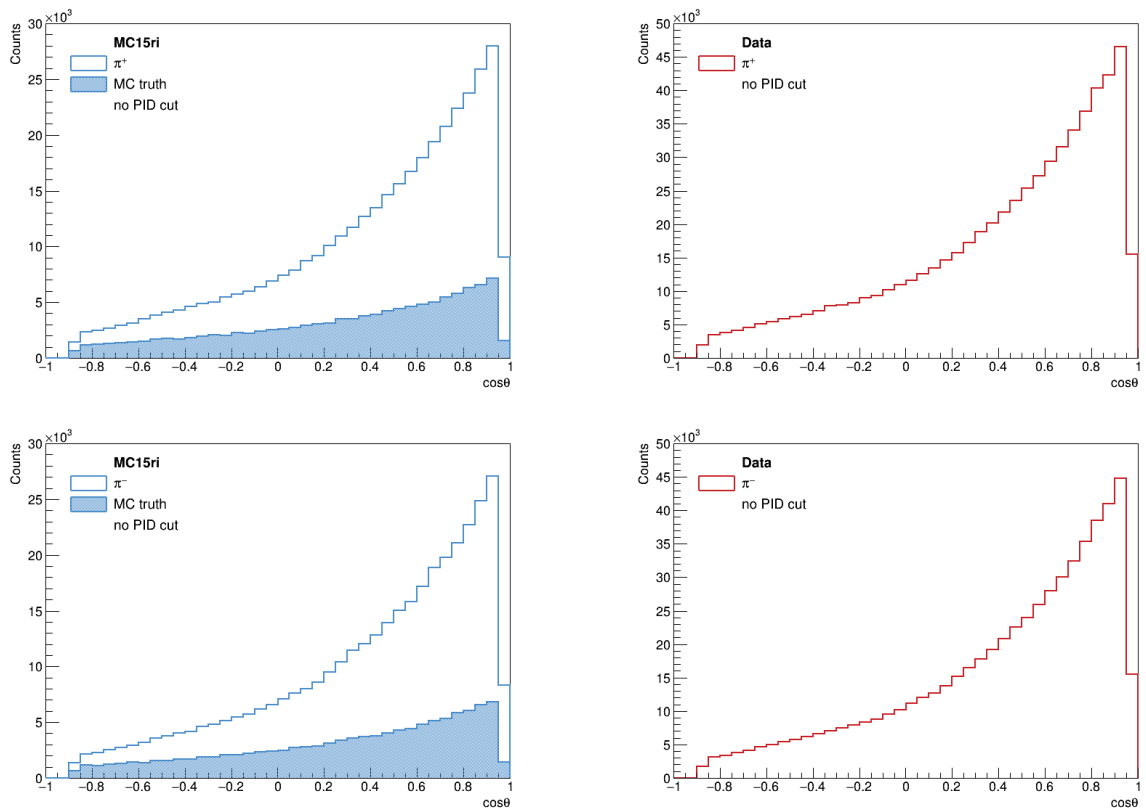Figure 5.20: Distribution of the cosine of the polar angle θ of the positive (top) and negative (bottom) charged pion tracks. MC (left) and real data (right).

Table 5.6: Cosine of polar angle bins for kaons and pions depending on the PID selection. For each bin, low and high $\cos\theta$ and number of true signal events in MC sample for positive and negative charge.

Particle	PID selection	Bin	Low $\cos\theta$	High $\cos\theta$	$N_{\text{sig}}^{\text{true}}(+)$	$N_{\text{sig}}^{\text{true}}(-)$
K	$\text{P}_K > 0.5$	1	-0.5	-0.1	22569	22497
		2	-0.1	0.2	23441	23436
		3	0.2	0.45	25932	25775
		4	0.45	0.65	28055	28347
		5	0.65	0.85	38049	38011
π	$\text{P}_\pi > 0.5$	1	-0.5	-0.1	12948	12341
		2	-0.1	0.2	13651	13006
		3	0.2	0.45	14895	14560
		4	0.45	0.65	15063	14398
		5	0.65	0.85	17924	17649
K	$\text{P}_\pi > 0.5$	1	-0.5	0.1	1351	1418
		2	0.1	0.35	1286	1303
		3	0.35	0.5	1265	1204
		4	0.5	0.7	1505	1520
		5	0.7	0.85	1156	1297
π	$\text{P}_K > 0.5$	1	-0.5	0.15	660	752
		2	0.15	0.4	631	767
		3	0.4	0.6	737	763
		4	0.6	0.75	716	778
		5	0.75	0.85	557	636

5.3.1 K-ID performance and charge asymmetries in bins of momentum

We calculate the kaon ID efficiencies and pion mis-ID rates separately for positive and negative tracks divided into momentum bins. We apply the PID selection criterion $P_K > 0.5$. The efficiencies and the corresponding charge asymmetries for both MC and real data are shown in Tab. 5.7, the fake rates and the corresponding charge asymmetries are shown in Tab. 5.8. For each measurement we report both statistical (first) and systematic (second) uncertainties. Moreover, we estimate the agreement between the charge asymmetries for MC and experimental data according to Eq. 5.5.

Table 5.7: Kaon ID efficiencies (positive and negative) and charge asymmetries for the PID selection $P_K > 0.5$, in bins of momentum for kaon tracks. λ quantifies the agreement between the charge asymmetries.

	Bin	MC15ri	Data	λ
	1	$94.0 \pm 0.1 \pm 1.9$	$94.2 \pm 0.1 \pm 1.9$	-
	2	$88.0 \pm 0.2 \pm 0.5$	$86.7 \pm 0.2 \pm 0.5$	-
ε_{K^+} (%)	3	$85.9 \pm 0.2 \pm 0.7$	$82.1 \pm 0.2 \pm 0.7$	-
	4	$83.3 \pm 0.2 \pm 0.8$	$80.3 \pm 0.2 \pm 0.8$	-
	5	$79.2 \pm 0.2 \pm 0.8$	$78.1 \pm 0.2 \pm 0.8$	-
	1	$92.4 \pm 0.1 \pm 0.7$	$91.1 \pm 0.1 \pm 0.7$	-
	2	$87.95 \pm 0.16 \pm 0.06$	$86.15 \pm 0.16 \pm 0.06$	-
ε_{K^-} (%)	3	$84.30 \pm 0.20 \pm 0.03$	$80.95 \pm 0.19 \pm 0.03$	-
	4	$83.4 \pm 0.2 \pm 1.4$	$78.6 \pm 0.2 \pm 1.3$	-
	5	$79.4 \pm 0.2 \pm 0.2$	$77.9 \pm 0.2 \pm 0.3$	-
	1	$0.0086 \pm 0.0009 \pm 0.0108$	$0.0166 \pm 0.0009 \pm 0.0108$	0.5
	2	$0.000 \pm 0.001 \pm 0.003$	$0.003 \pm 0.001 \pm 0.003$	0.6
$\mathcal{A}(\varepsilon)$	3	$0.009 \pm 0.002 \pm 0.004$	$0.007 \pm 0.002 \pm 0.004$	0.3
	4	$-0.001 \pm 0.002 \pm 0.010$	$0.011 \pm 0.002 \pm 0.010$	0.8
	5	$-0.001 \pm 0.002 \pm 0.005$	$0.001 \pm 0.002 \pm 0.006$	0.3

Table 5.8: Pion mis-ID rates (positive and negative) and charge asymmetries for the PID selection $P_K > 0.5$, in bins of momentum for pion tracks. λ quantifies the agreement between the charge asymmetries.

	Bin	MC15ri	Data	λ
	1	$3.8 \pm 0.1 \pm 0.5$	$3.71 \pm 0.09 \pm 0.49$	-
	2	$8.1 \pm 0.2 \pm 0.6$	$6.9 \pm 0.2 \pm 0.5$	-
f_{π^+} (%)	3	$9.4 \pm 0.2 \pm 1.1$	$5.4 \pm 0.2 \pm 0.9$	-
	4	$5.7 \pm 0.1 \pm 0.2$	$5.4 \pm 0.1 \pm 0.2$	-
	5	$4.9 \pm 0.1 \pm 0.6$	$6.3 \pm 0.1 \pm 0.6$	-
	1	$3.4 \pm 0.1 \pm 0.6$	$4.3 \pm 0.1 \pm 0.7$	-
	2	$8.0 \pm 0.2 \pm 0.2$	$8.5 \pm 0.2 \pm 0.2$	-
f_{π^-} (%)	3	$9.1 \pm 0.2 \pm 1.4$	$8.9 \pm 0.2 \pm 1.4$	-
	4	$6.5 \pm 0.2 \pm 0.3$	$7.6 \pm 0.2 \pm 0.3$	-
	5	$5.2 \pm 0.1 \pm 0.4$	$6.5 \pm 0.1 \pm 0.5$	-
	1	$0.05 \pm 0.02 \pm 0.11$	$-0.07 \pm 0.02 \pm 0.10$	0.8
	2	$0.01 \pm 0.02 \pm 0.04$	$-0.10 \pm 0.02 \pm 0.04$	1.8
$\mathcal{A}(f)$	3	$0.02 \pm 0.02 \pm 0.10$	$-0.25 \pm 0.02 \pm 0.11$	1.8
	4	$-0.07 \pm 0.02 \pm 0.03$	$-0.16 \pm 0.01 \pm 0.03$	2.0
	5	$-0.03 \pm 0.02 \pm 0.07$	$-0.01 \pm 0.01 \pm 0.06$	0.2

The results of the K-ID performance as a function of p are shown in Fig. 5.21. In the two plots (positive charge top, negative charge bottom) the blue triangles represent the efficiencies (fake rates) determined on MC, while the red triangles represent those measured from the data. The error bars represent the contribution of both statistical and systematic uncertainties. Moreover, in the two plots we also show the true efficiencies (fake rates) calculated using MC truth information (black triangles), the error associated with them is only statistical. The horizontal black bars represent the width of the p bins. We can see good agreement between the true ID observables and those determined with the fitting procedure on MC. The ID efficiencies on real data are slightly lower than those on simulation, while the resulting fake rates for MC and data are about equal.
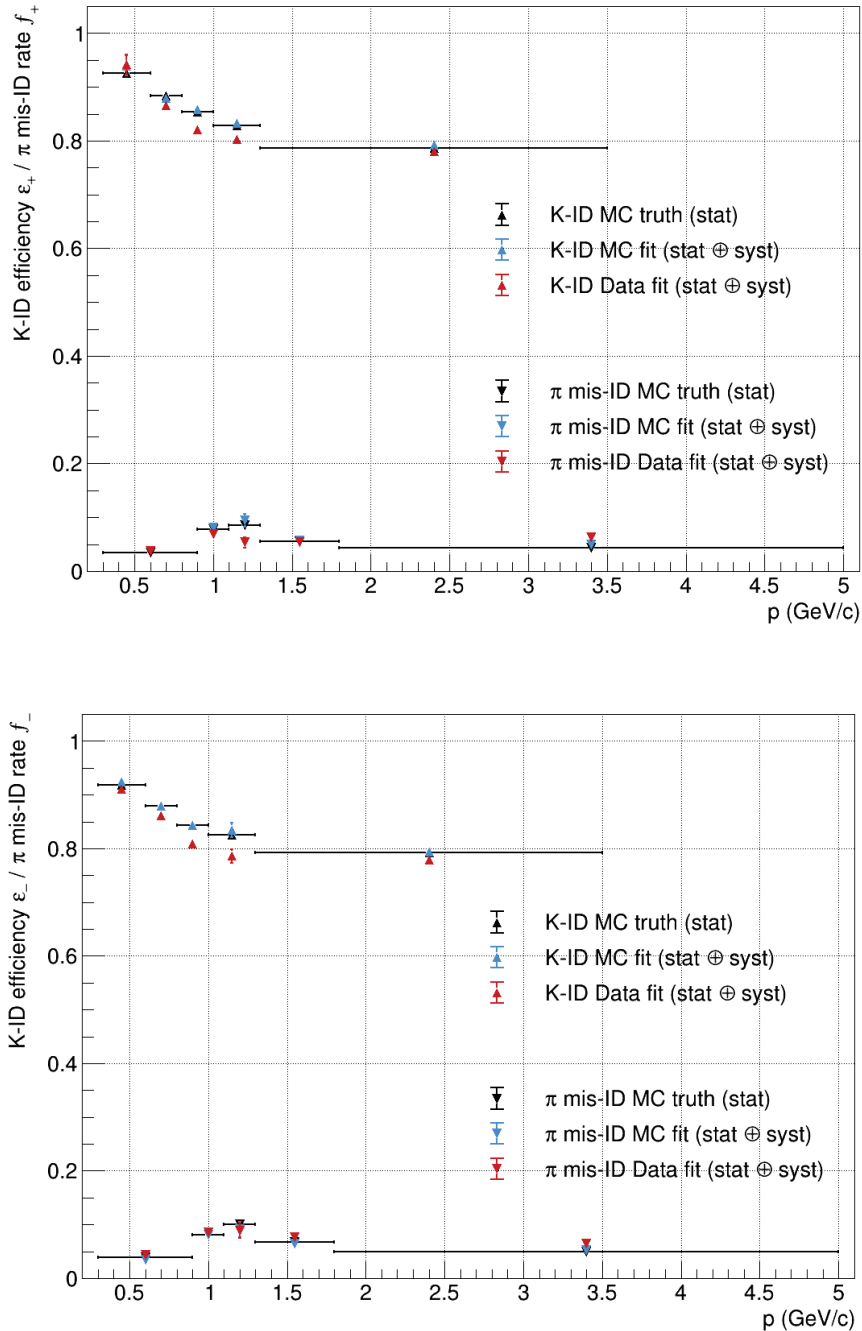

Figure 5.21: Kaon ID efficiency and pion mis-ID rate for the PID selection $P_K > 0.5$, in bins of momentum: positively charged particles (top) and negatively charged particles (bottom).

In Fig. 5.22 we show the charge asymmetries measured in momentum bins. In the top plot we compare the charge asymmetries of the efficiencies. The asymmetries estimated by fitting the simulated data well reproduce the known charge asymmetries using MC truth information. We do not observe a significant discrepancy between the measured asymmetries for MC and experimental data, in all bins the agreement is good with $\lambda < 1$. In the bottom plot we show the charge asymmetries of the fake rates. In some bins we notice an overestimation of the charge asymmetries resulting from the fits on MC, however the error associated with the measurements covers the difference. Furthermore, there is a discrepancy between the charge asymmetries for simulated and real data, however a not bad agreement is observed between the measurements.
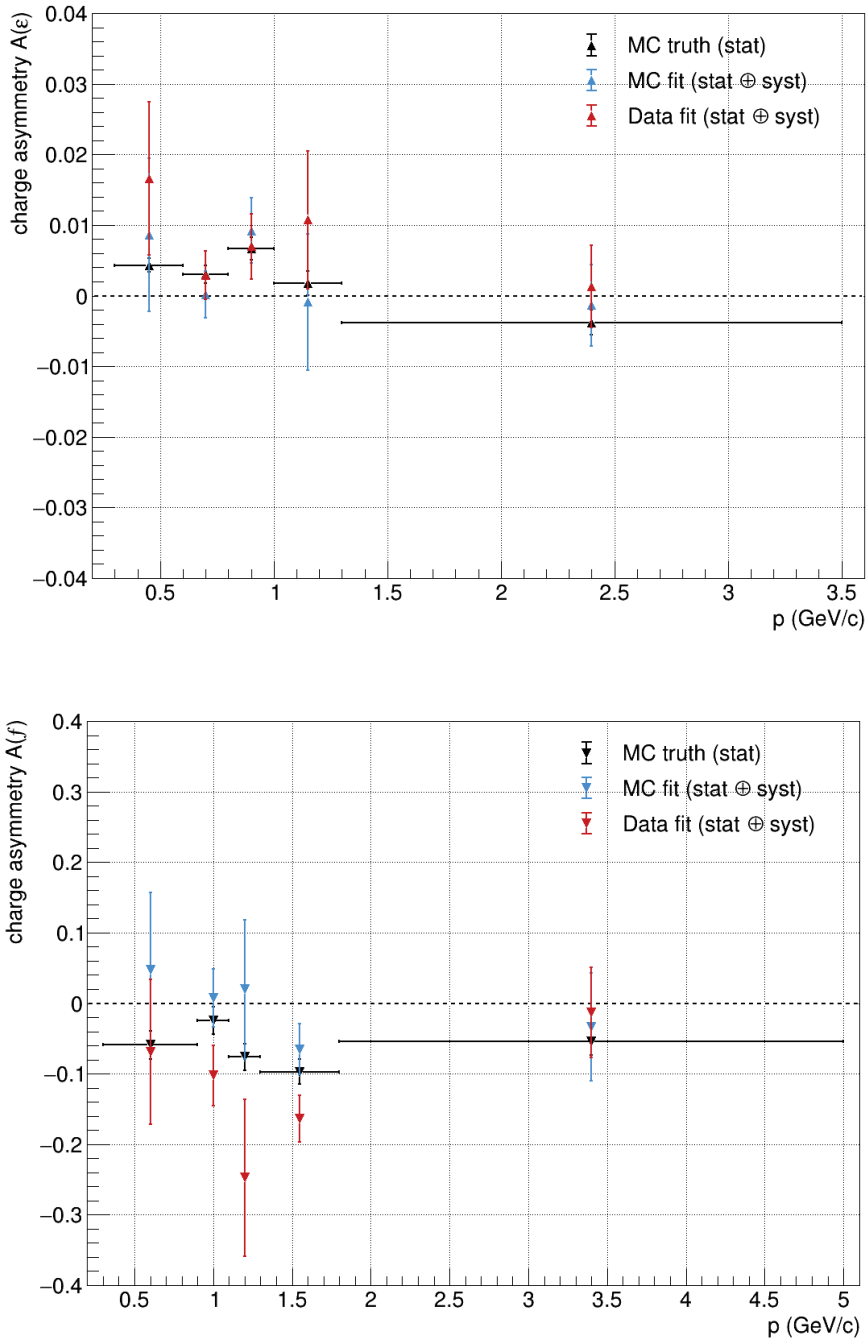

Figure 5.22: Charge asymmetries for the PID selection $P_K > 0.5$, in bins of momentum: asymmetry of efficiencies (top) and asymmetry of fake rates (bottom).

5.3.2 K-ID performance and charge asymmetries in bins of $\cos\theta$

We present the kaon ID efficiencies and pion mis-ID rates separately for positive and negative tracks divided into $\cos\theta$ bins. We apply the PID selection criterion $P_K > 0.5$ to both charged kaon and pion tracks. We show the efficiencies and the corresponding charge asymmetries for both MC and real data in Tab. 5.9. For each measurement the first quoted error is the statistical one, while the second is the systematic error. We report the agreement λ (Eq. 5.5) between the charge asymmetries for simulation and data. Then, we show the fake rates and the corresponding charge asymmetries in Tab. 5.10.

Table 5.9: Kaon ID efficiencies (positive and negative) and charge asymmetries for the PID selection $P_K > 0.5$, in bins of $\cos\theta$ for kaon tracks. λ quantifies the agreement between the charge asymmetries.

	Bin	MC15ri	Data	λ
ε_{K^+} (%)	1	$90.6 \pm 0.2 \pm 1.2$	$90.2 \pm 0.2 \pm 1.2$	-
	2	$88.3 \pm 0.2 \pm 0.2$	$86.3 \pm 0.2 \pm 0.2$	-
	3	$84.8 \pm 0.2 \pm 0.2$	$82.0 \pm 0.2 \pm 0.2$	-
	4	$84.1 \pm 0.2 \pm 0.2$	$82.4 \pm 0.2 \pm 0.2$	-
	5	$85.0 \pm 0.2 \pm 0.8$	$83.1 \pm 0.2 \pm 0.7$	-
ε_{K^-} (%)	1	$89.2 \pm 0.2 \pm 0.2$	$88.9 \pm 0.2 \pm 0.2$	-
	2	$87.0 \pm 0.2 \pm 0.7$	$84.6 \pm 0.2 \pm 0.7$	-
	3	$84.4 \pm 0.2 \pm 0.2$	$80.6 \pm 0.2 \pm 0.2$	-
	4	$85.4 \pm 0.2 \pm 1.3$	$82.5 \pm 0.2 \pm 1.3$	-
	5	$83.6 \pm 0.2 \pm 0.5$	$80.7 \pm 0.2 \pm 0.4$	-
$\mathcal{A}(\varepsilon)$	1	$0.008 \pm 0.001 \pm 0.007$	$0.007 \pm 0.001 \pm 0.007$	0.05
	2	$0.007 \pm 0.002 \pm 0.004$	$0.010 \pm 0.002 \pm 0.004$	0.5
	3	$0.003 \pm 0.002 \pm 0.002$	$0.008 \pm 0.002 \pm 0.002$	1.6
	4	$-0.008 \pm 0.002 \pm 0.008$	$-0.001 \pm 0.002 \pm 0.018$	0.6
	5	$0.008 \pm 0.001 \pm 0.005$	$0.015 \pm 0.001 \pm 0.005$	0.8

Table 5.10: Pion mis-ID rates (positive and negative) and charge asymmetries for the PID selection $P_K > 0.5$, in bins of $\cos\theta$ for pion tracks. λ quantifies the agreement between the charge asymmetries.

	Bin	MC15ri	Data	λ
f_{π^+} (%)	1	$2.71 \pm 0.10 \pm 0.51$	$2.38 \pm 0.08 \pm 0.49$	-
	2	$3.5 \pm 0.1 \pm 0.3$	$5.1 \pm 0.1 \pm 0.4$	-
	3	$4.9 \pm 0.2 \pm 0.8$	$6.0 \pm 0.2 \pm 0.9$	-
	4	$4.8 \pm 0.2 \pm 0.2$	$7.3 \pm 0.2 \pm 0.2$	-
	5	$4.56 \pm 0.19 \pm 0.04$	$5.6 \pm 0.2 \pm 0.6$	-
f_{π^-} (%)	1	$2.46 \pm 0.09 \pm 0.35$	$3.35 \pm 0.31 \pm 0.40$	-
	2	$4.4 \pm 0.2 \pm 0.3$	$6.2 \pm 0.2 \pm 0.4$	-
	3	$4.2 \pm 0.2 \pm 0.6$	$5.7 \pm 0.2 \pm 0.7$	-
	4	$5.12 \pm 0.18 \pm 0.09$	$8.0 \pm 0.2 \pm 0.1$	-
	5	$5.0 \pm 0.2 \pm 0.4$	$8.4 \pm 0.2 \pm 0.5$	-
$\mathcal{A}(f)$	1	$0.05 \pm 0.03 \pm 0.12$	$-0.17 \pm 0.02 \pm 0.12$	1.3
	2	$-0.11 \pm 0.03 \pm 0.06$	$-0.09 \pm 0.02 \pm 0.05$	0.2
	3	$0.08 \pm 0.02 \pm 0.11$	$0.03 \pm 0.02 \pm 0.10$	0.4
	4	$-0.03 \pm 0.03 \pm 0.02$	$-0.05 \pm 0.02 \pm 0.02$	0.5
	5	$-0.05 \pm 0.03 \pm 0.04$	$-0.20 \pm 0.02 \pm 0.06$	1.9

The results of the K-ID performance as a function of $\cos\theta$ for positive and negative charge are shown in Fig. 5.23, respectively in the top and bottom plots. We show the efficiencies and fake rates for both simulation (blue triangles) and experimental data (red triangles) with their uncertainties (statistical and systematic in quadrature). In the same plots we show also the true ID observables (black triangles), the horizontal black bars represent the width of the $\cos\theta$ bins. We observe a very good agreement between the efficiencies determined with the MC sample (the true ones and those estimated with the fits). Also for the mis-ID rates we see excellent agreement. The ID efficiencies on real data are slightly lower than the efficiencies on simulation, while the fake rates are higher for $\cos\theta$ close to 1 (small polar angles).
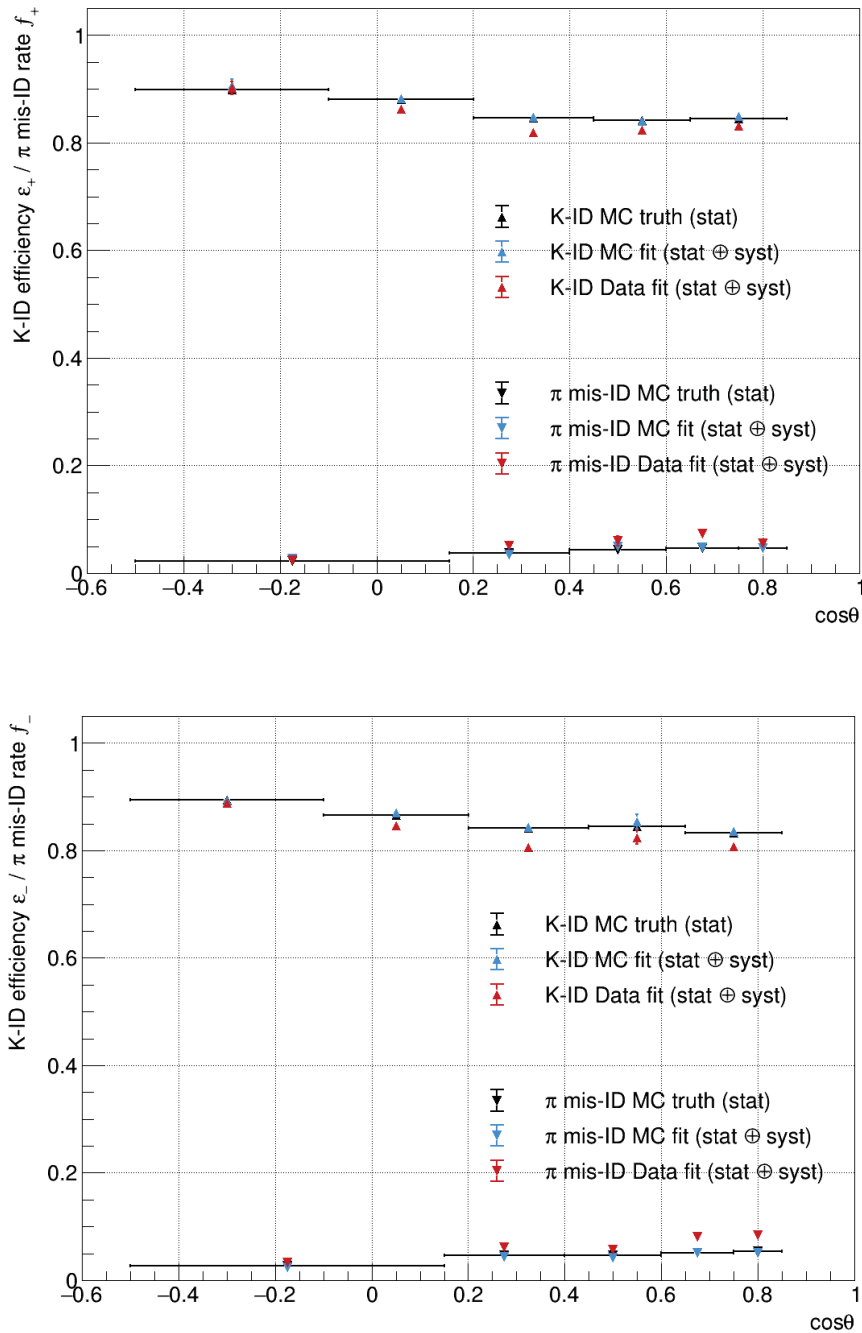
Figure 5.23: Kaon ID efficiency and pion mis-ID rate for the PID selection $P_K > 0.5$, in bins of $\cos\theta$: positively charged particles (top) and negatively charged particles (bottom).

We show the measured charge asymmetries of the efficiencies $\mathcal{A}(\varepsilon)$ and those of the fake rates $\mathcal{A}(f)$ in Fig. 5.24, respectively in the top and bottom plots. In both cases the charge asymmetries estimated with the fits on simulated data are in good agreement with those known using MC truth information. Overall, the real data do not exhibit charge asymmetries larger than those determined from the MC simulation. Only in some bins the error bars do not overlap, however the agreement between the charge asymmetries is still good with $\lambda < 1$ or in the worst cases $\lambda < 2$. We observe that, as a function of $\cos\theta$, the asymmetries of the efficiencies tend to be positive, while those of the fake rates tend to be negative. There is also a similar trend depending on p. As a result, positively charged kaons are better identified.
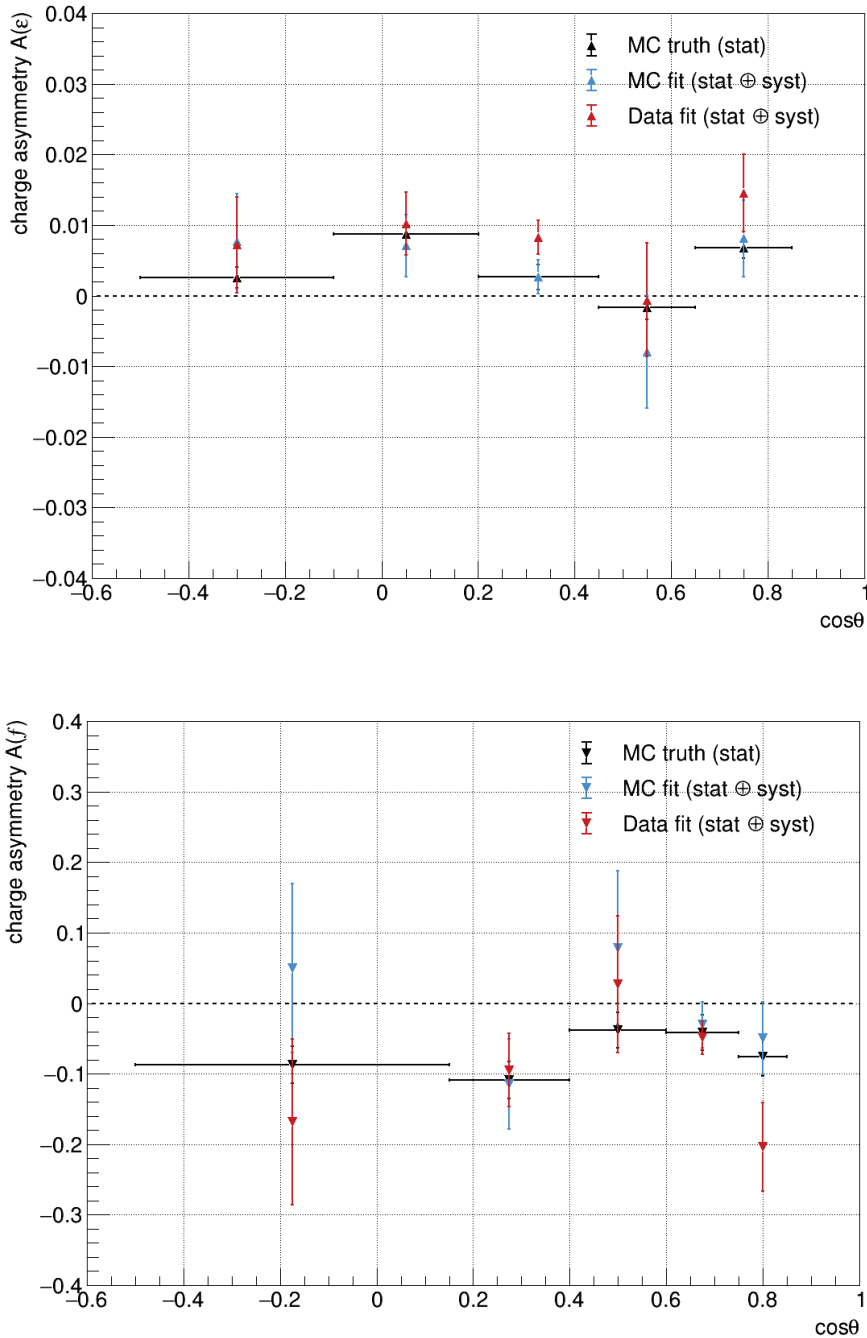

Figure 5.24: Charge asymmetries for the PID selection $P_K > 0.5$, in bins of $\cos\theta$: asymmetry of efficiencies (top) and asymmetry of fake rates (bottom).

5.3.3 π-ID performance and charge asymmetries in bins of momentum

We present the pion ID efficiencies and kaon mis-ID rates separately for positive and negative tracks divided into momentum bins. We apply the PID selection criterion $P_\pi > 0.5$ to both charged pion and kaon tracks. We calculate the ID efficiencies, mis-ID rates and charge asymmetries for both MC and real data, all the results are reported in Tab. 5.11 and Tab. 5.12. For each measurement the first quoted error is the statistical one, while the second is the systematic error. We estimate the agreement between the charge asymmetries according to Eq. 5.5.

Table 5.11: Pion ID efficiencies (positive and negative) and charge asymmetries for the PID selection $P_\pi > 0.5$, in bins of momentum for pion tracks. λ quantifies the agreement between the charge asymmetries.

	Bin	MC15ri	Data	λ
$\varepsilon_\pi{}^+$ (%)	1	$73.6 \pm 0.3 \pm 2.6$	$72.3 \pm 0.3 \pm 2.5$	-
	2	$78.8 \pm 0.3 \pm 3.2$	$78.5 \pm 0.2 \pm 3.2$	-
	3	$80.6 \pm 0.2 \pm 2.7$	$79.0 \pm 0.2 \pm 2.7$	-
	4	$79.3 \pm 0.3 \pm 2.9$	$76.7 \pm 0.3 \pm 2.8$	-
	5	$80.5 \pm 0.3 \pm 1.8$	$73.9 \pm 0.2 \pm 1.7$	-
$\varepsilon_\pi{}^-$ (%)	1	$71.8 \pm 0.3 \pm 2.7$	$70.1 \pm 0.3 \pm 2.7$	-
	2	$78.3 \pm 0.3 \pm 2.6$	$78.3 \pm 0.3 \pm 2.7$	-
	3	$80.3 \pm 0.3 \pm 3.2$	$77.2 \pm 0.2 \pm 3.2$	-
	4	$79.1 \pm 0.3 \pm 1.9$	$75.0 \pm 0.3 \pm 1.9$	-
	5	$80.2 \pm 0.3 \pm 2.7$	$73.4 \pm 0.3 \pm 2.6$	-
$\mathcal{A}(\varepsilon)$	1	$0.012 \pm 0.003 \pm 0.026$	$0.009 \pm 0.003 \pm 0.026$	0.09
	2	$0.003 \pm 0.002 \pm 0.026$	$0.001 \pm 0.002 \pm 0.027$	0.05
	3	$0.001 \pm 0.002 \pm 0.026$	$0.011 \pm 0.002 \pm 0.027$	0.3
	4	$0.001 \pm 0.003 \pm 0.022$	$0.011 \pm 0.002 \pm 0.022$	0.3
	5	$0.002 \pm 0.002 \pm 0.020$	$0.004 \pm 0.002 \pm 0.021$	0.07

Table 5.12: Kaon mis-ID rates (positive and negative) and charge asymmetries for the PID selection $P_\pi > 0.5$, in bins of momentum for kaon tracks. λ quantifies the agreement between the charge asymmetries.

	Bin	MC15ri	Data	λ
$f_K{}^+$ (%)	1	$1.43 \pm 0.04 \pm 0.01$	$1.56 \pm 0.04 \pm 0.11$	-
	2	$5.5 \pm 0.1 \pm 0.6$	$6.7 \pm 0.1 \pm 0.6$	-
	3	$6.8 \pm 0.2 \pm 0.4$	$7.6 \pm 0.2 \pm 0.5$	-
	4	$6.8 \pm 0.2 \pm 0.1$	$7.8 \pm 0.2 \pm 0.1$	-
	5	$6.0 \pm 0.2 \pm 0.2$	$7.0 \pm 0.1 \pm 0.2$	-
$f_K{}^-$ (%)	1	$1.74 \pm 0.04 \pm 0.28$	$2.07 \pm 0.04 \pm 0.30$	-
	2	$6.4 \pm 0.1 \pm 1.0$	$6.8 \pm 0.1 \pm 1.0$	-
	3	$6.6 \pm 0.2 \pm 0.2$	$8.6 \pm 0.2 \pm 0.3$	-
	4	$6.5 \pm 0.2 \pm 0.5$	$7.6 \pm 0.2 \pm 0.5$	-
	5	$6.8 \pm 0.2 \pm 0.5$	$8.2 \pm 0.2 \pm 0.5$	-
$\mathcal{A}(f)$	1	$-0.10 \pm 0.02 \pm 0.09$	$-0.14 \pm 0.01 \pm 0.08$	0.3
	2	$-0.08 \pm 0.02 \pm 0.09$	$-0.01 \pm 0.01 \pm 0.09$	0.5
	3	$0.02 \pm 0.02 \pm 0.04$	$-0.06 \pm 0.01 \pm 0.03$	1.5
	4	$0.02 \pm 0.02 \pm 0.04$	$0.02 \pm 0.01 \pm 0.03$	0.08
	5	$-0.06 \pm 0.02 \pm 0.04$	$-0.08 \pm 0.01 \pm 0.03$	0.3

We show in Fig. 5.25 the results of the π-ID performance as a function of p for positive (top) and negative (bottom) charge. We compare the efficiencies (fake rates) determined with the simulated data (blue triangles) to those measured from the experimental data (red triangles). Then, the black triangles represent the true efficiencies (fake rates) calculated using MC truth information, their uncertainties are only statistical. The horizontal black bars represent the width of the p bins. We observe a very good agreement between the efficiencies determined with the MC sample (the true ones and those estimated with the fits). The same observation is valid for the mis-ID rates. The ID efficiencies on data are slightly lower than the efficiencies on MC, while the fake rates are equal or slightly higher.
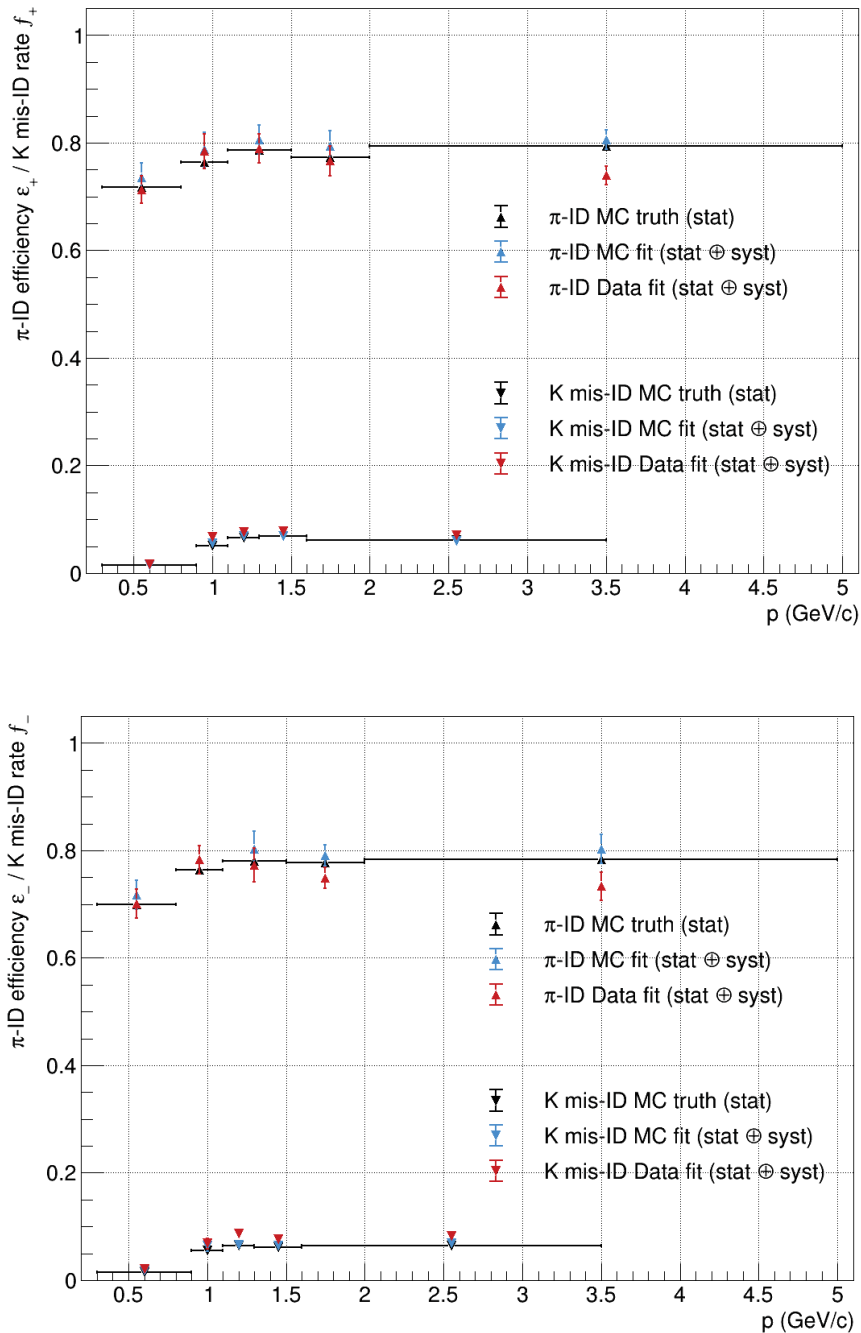

Figure 5.25: Pion ID efficiency and kaon mis-ID rate for the PID selection $P_\pi > 0.5$, in bins of momentum: positively charged particles (top) and negatively charged particles (bottom).

In Fig. 5.26 we show the charge asymmetries measured in momentum bins: the true charge asymmetries and those measured in this study for both MC and experimental data. The good agreement obtained in the results of the MC data (true in all cases in the analysis) allows us to validate the fitting strategy. In the top plot we compare the charge asymmetries of the efficiencies and we do not observe discrepancies between the results for simulated and real data. The agreement in each bin of momentum is very good with $\lambda < 1$. In the bottom plot we compare the charge asymmetries of the fake rates, also in this case there is overall good agreement between the charge asymmetries obtained for simulated and experimental data. The asymmetries of the efficiencies are all positive, while those of the fake rates show oscillating signs.
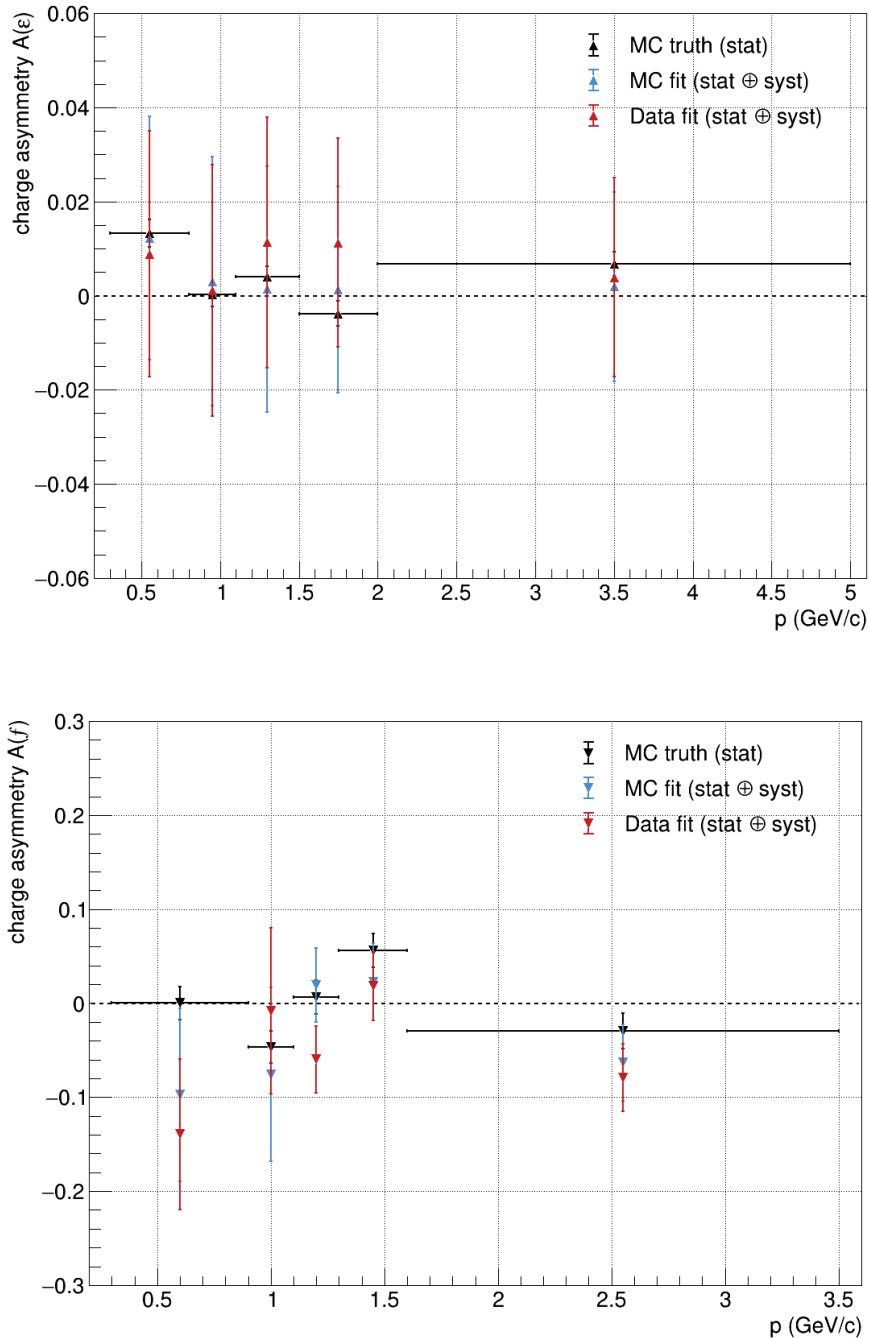

Figure 5.26: Charge asymmetries for the PID selection $P_\pi > 0.5$, in bins of momentum: asymmetry of efficiencies (top) and asymmetry of fake rates (bottom).

5.3.4 π-ID performance and charge asymmetries in bins of $\cos\theta$

We report the pion ID efficiencies and kaon mis-ID rates for positive and negative charge, dividing the charged tracks of K and π into $\cos\theta$ bins. We apply the PID selection criterion $P_\pi > 0.5$. In Tab. 5.13 and Tab. 5.14 we summarize the ID observables, efficiencies and fake rates respectively, and the charge asymmetries. For each measurement we report both statistical (first) and systematic (second) uncertainties. Then, we report the agreement λ (Eq. 5.5) between the charge asymmetries for simulation and data.

Table 5.13: Pion ID efficiencies (positive and negative) and charge asymmetries for the PID selection $P_\pi > 0.5$, in bins of $\cos\theta$ for pion tracks. λ quantifies the agreement between the charge asymmetries.

	Bin	MC15ri	Data	λ
	1	$82.2 \pm 0.3 \pm 0.2$	$79.1 \pm 0.3 \pm 0.2$	-
	2	$85.2 \pm 0.3 \pm 1.6$	$80.0 \pm 0.3 \pm 1.5$	-
$\varepsilon_\pi{}^+$ (%)	3	$85.0 \pm 0.3 \pm 1.9$	$80.4 \pm 0.3 \pm 1.9$	-
	4	$82.99 \pm 0.28 \pm 0.01$	$79.71 \pm 0.26 \pm 0.01$	-
	5	$80.4 \pm 0.3 \pm 1.7$	$76.9 \pm 0.2 \pm 1.6$	-
	1	$82.7 \pm 0.3 \pm 0.9$	$79.0 \pm 0.3 \pm 0.9$	-
	2	$83.7 \pm 0.3 \pm 0.1$	$79.4 \pm 0.3 \pm 0.1$	-
$\varepsilon_\pi{}^-$ (%)	3	$85.4 \pm 0.3 \pm 3.0$	$80.3 \pm 0.2 \pm 3.0$	-
	4	$82.8 \pm 0.3 \pm 1.0$	$78.9 \pm 0.3 \pm 1.0$	-
	5	$80.0 \pm 0.3 \pm 1.7$	$76.7 \pm 0.2 \pm 1.7$	-
	1	$-0.003 \pm 0.003 \pm 0.006$	$0.001 \pm 0.003 \pm 0.006$	0.4
	2	$0.009 \pm 0.002 \pm 0.009$	$0.003 \pm 0.002 \pm 0.010$	0.4
$\mathcal{A}(\varepsilon)$	3	$-0.002 \pm 0.002 \pm 0.021$	$0.001 \pm 0.002 \pm 0.022$	0.09
	4	$0.001 \pm 0.002 \pm 0.006$	$0.005 \pm 0.002 \pm 0.006$	0.4
	5	$0.002 \pm 0.002 \pm 0.015$	$0.001 \pm 0.002 \pm 0.015$	0.07

Table 5.14: Kaon mis-ID rates (positive and negative) and charge asymmetries for the PID selection $P_\pi > 0.5$, in bins of $\cos\theta$ for kaon tracks. λ quantifies the agreement between the charge asymmetries.

	Bin	MC15ri	Data	λ
	1	$3.02 \pm 0.08 \pm 0.28$	$3.21 \pm 0.08 \pm 0.41$	-
	2	$4.78 \pm 0.13 \pm 0.04$	$4.99 \pm 0.12 \pm 0.03$	-
$f_K{}^+$ (%)	3	$6.8 \pm 0.2 \pm 0.7$	$8.0 \pm 0.2 \pm 0.7$	-
	4	$4.2 \pm 0.1 \pm 0.1$	$4.6 \pm 0.1 \pm 0.1$	-
	5	$3.5 \pm 0.1 \pm 0.3$	$4.2 \pm 0.1 \pm 0.3$	-
	1	$3.68 \pm 0.04 \pm 0.41$	$3.31 \pm 0.04 \pm 0.39$	-
	2	$4.85 \pm 0.13 \pm 0.08$	$6.13 \pm 0.1 \pm 0.08$	-
$f_K{}^-$ (%)	3	$6.0 \pm 0.2 \pm 0.3$	$6.6 \pm 0.2 \pm 0.3$	-
	4	$4.0 \pm 0.1 \pm 0.3$	$5.3 \pm 0.1 \pm 0.3$	-
	5	$4.3 \pm 0.1 \pm 0.7$	$5.6 \pm 0.1 \pm 0.7$	-
	1	$-0.10 \pm 0.02 \pm 0.07$	$-0.02 \pm 0.02 \pm 0.09$	0.7
	2	$-0.007 \pm 0.019 \pm 0.009$	$-0.103 \pm 0.016 \pm 0.008$	3.5
$\mathcal{A}(f)$	3	$0.06 \pm 0.02 \pm 0.06$	$0.09 \pm 0.02 \pm 0.05$	0.7
	4	$0.03 \pm 0.02 \pm 0.04$	$-0.06 \pm 0.01 \pm 0.03$	1.7
	5	$-0.10 \pm 0.02 \pm 0.09$	$-0.14 \pm 0.01 \pm 0.07$	0.3

The results of π-ID performance in bins of $\cos\theta$ for positive and negative charge are shown in Fig. 5.27. In the top (bottom) plot we compare the ID observables for positively (negatively) charged tracks determined with fits on the MC sample (blue triangles) and on the data sample (red triangles), the error bars represent the contribution of all uncertainties. Moreover, we also show the true efficiencies (fake rates) calculated with the MC truth (black triangles). The horizontal black bars represent the width of the $\cos\theta$ bins. There is a very good agreement between the efficiencies (fake rates) on MC determined with the fitting procedure and the true ones. Overall, the efficiencies on real data are lower than those on simulation, while the resulting fake rates for MC and data are about equal.
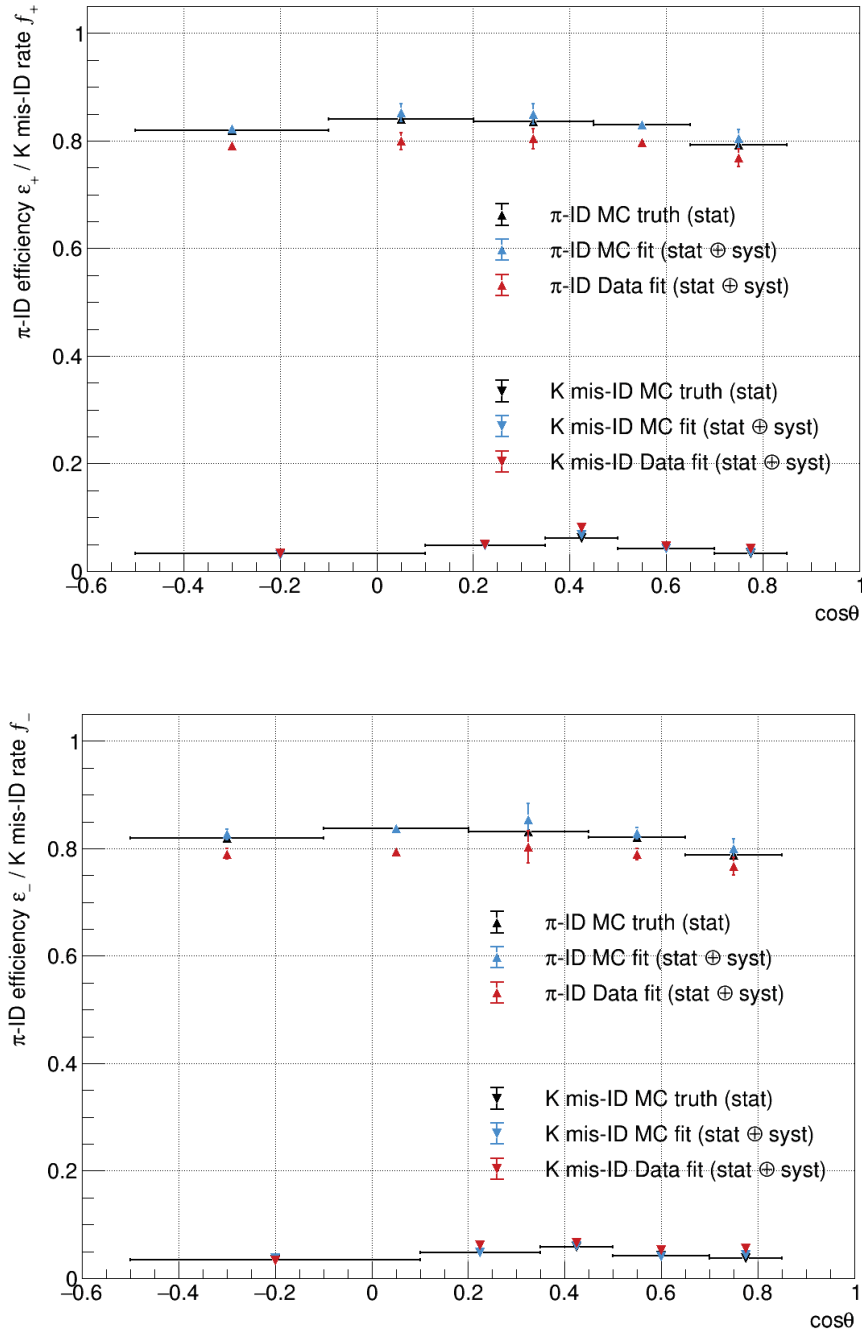

Figure 5.27: Pion ID efficiency and kaon mis-ID rate for the PID selection $P_\pi > 0.5$, in bins of $\cos\theta$: positively charged particles (top) and negatively charged particles (bottom).

In Fig. 5.28 we show the charge asymmetries as a function of $\cos\theta$, the asymmetries $\mathcal{A}(\varepsilon)$ in the top plot and the asymmetries $\mathcal{A}(f)$ in the bottom plot. The asymmetries estimated by fitting the simulated data well reproduce the known charge asymmetries using MC truth information. In the top plot we do not observe a significant discrepancy between the charge asymmetries of the efficiencies for MC and data ($\lambda < 1$). In the bottom plot we compare the charge asymmetries of the fake rates for both simulation and real data. In the second bin we observe a discrepancy between the measurements. However, it should be noted that the error associated with the charge asymmetries in the second bin is much lower than the error in the other bins. In all other bins we have good agreement between the two asymmetries. Overall, the charge asymmetries are close to zero.
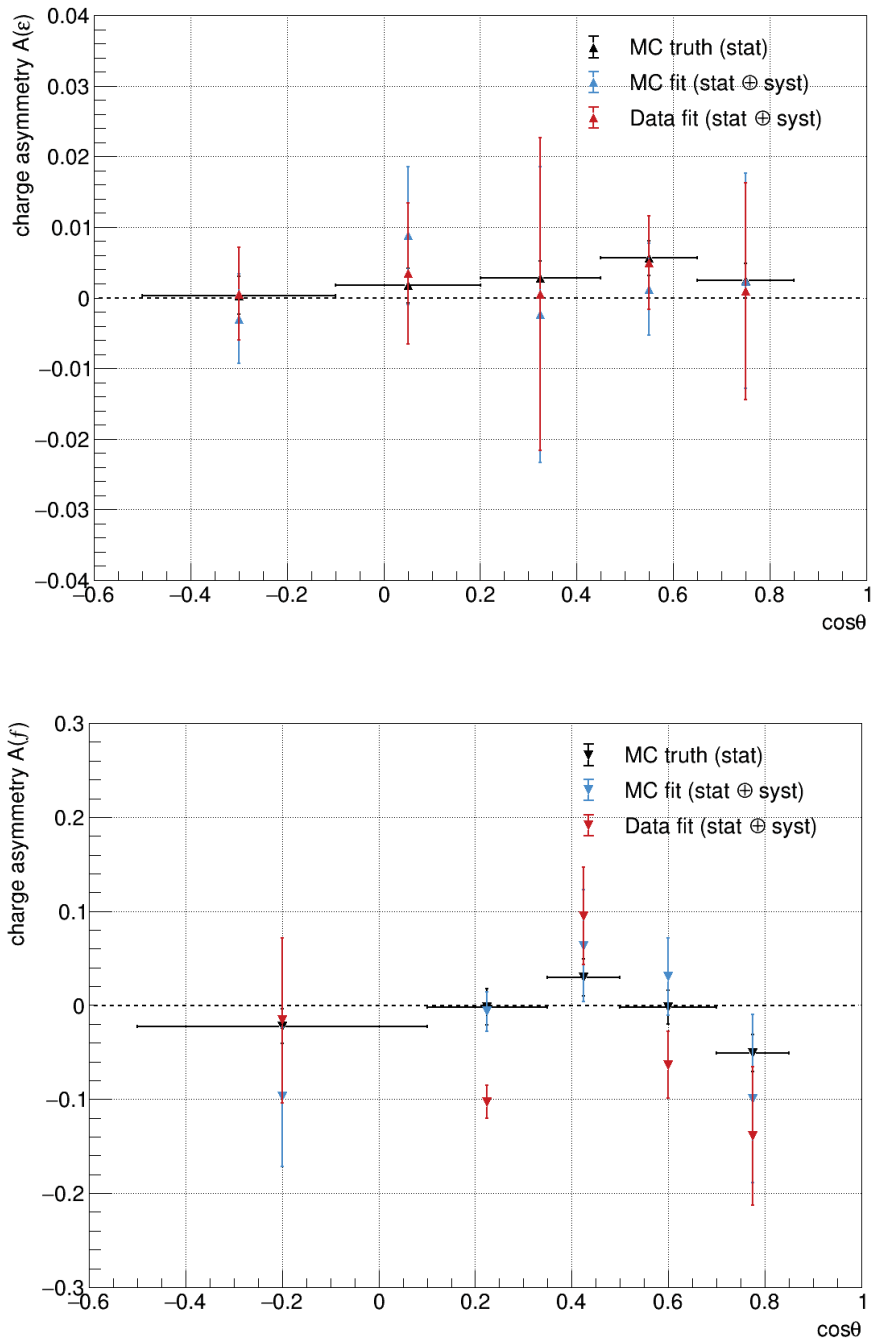

Figure 5.28: Charge asymmetries for the PID selection $P_\pi > 0.5$, in bins of $\cos\theta$: asymmetry of efficiencies (top) and asymmetry of fake rates (bottom).

5.4 Identification performance for different PID criteria

Finally, we study the identification performance of charged kaons and pions as a function of the threshold cut applied to the global PID probability. In the previous sections we showed the results for the PID selection $P_K > 0.5$ or $P_\pi > 0.5$, now we present the results for a generic threshold cut α.

The Receiver Operating Characteristic (ROC) curve is a performance measurement for classification problems at various thresholds settings, namely it is a plot of the true positive rate (TPR) against the false positive rate (FPR) for different possible threshold cuts. The closer the ROC curve gets to the upper-left corner of the space TPR vs. FPR, the better the classification. The closer the curve is to the diagonal, the worse the classification. In our case we study K/π separation performance and the ROC curve shows the ID efficiency vs. mis-ID rate. We determine the average identification efficiencies and average mis-identification rates by following the strategy outlined previously. We consider the samples of charged kaon and pion tracks separately, then we estimate the signal yields before and after applying PID selection with fit of the invariant mass distribution of the D_s candidates. We use the PFD model defined in Eq. 4.1. We set the parameters as in the previous sections, but we choose a different value of the parameter f_1 depending on the threshold cut. In this part we consider only the statistical uncertainty on the identification observables.

In Fig. 5.29 we show the ROC curves for the K-ID performance, i.e. the efficiency to identify a kaon vs. the rate with which pions are mis-identified as kaons. The performance is evaluated on both MC and real data samples, the applied selection criteria are $P_K > 0.5, 0.55, 0.6, 0.65, 0.7, 0.75, 0.8, 0.85, 0.9, 0.95, 0.96, 0.97, 0.98, 0.99$. We observe for real data a slightly worse identification performance, in fact the ROC curve for data is below the curve for simulation. The K-ID efficiency for real data is 3-7% lower than that for MC for a fixed fake rate.
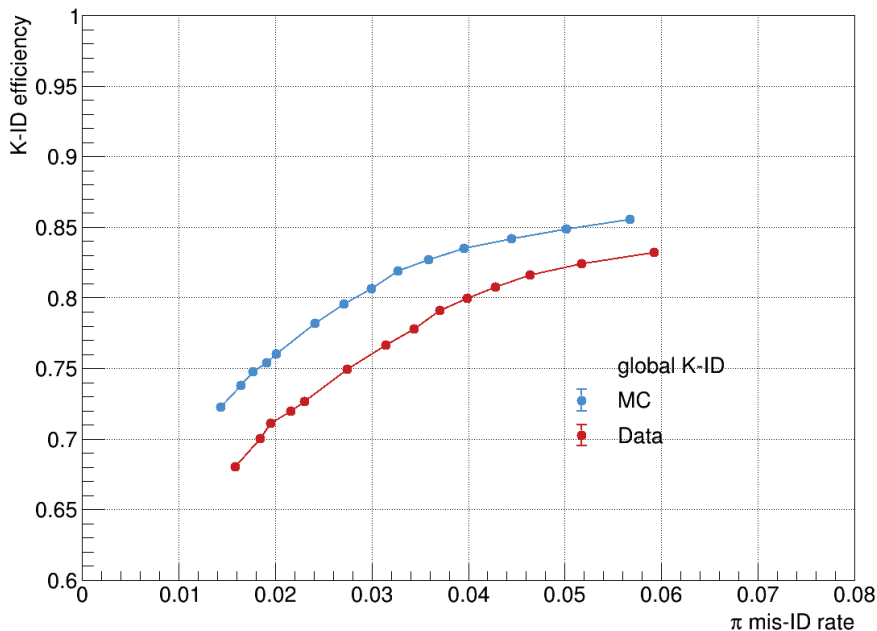

Figure 5.29: ROC curves for kaon identification: K-ID efficiencies vs. π mis-ID rates for different PID criteria.

In Fig. 5.30 we show the ROC curves π-ID efficiency vs. K mis-ID rate. We consider the PID selection criteria $P_\pi > 0.2, 0.25, 0.3, 0.35, 0.4, 0.45, 0.5, 0.55, 0.6, 0.65, 0.7, 0.75, 0.8, 0.85, 0.9$ applied to both simulated and experimental data. Also in this case, the identification performance is slightly worse for experimental data with respect to that for MC. The π-ID efficiency for real data is 5-10% lower than that for MC for a fixed fake rate.
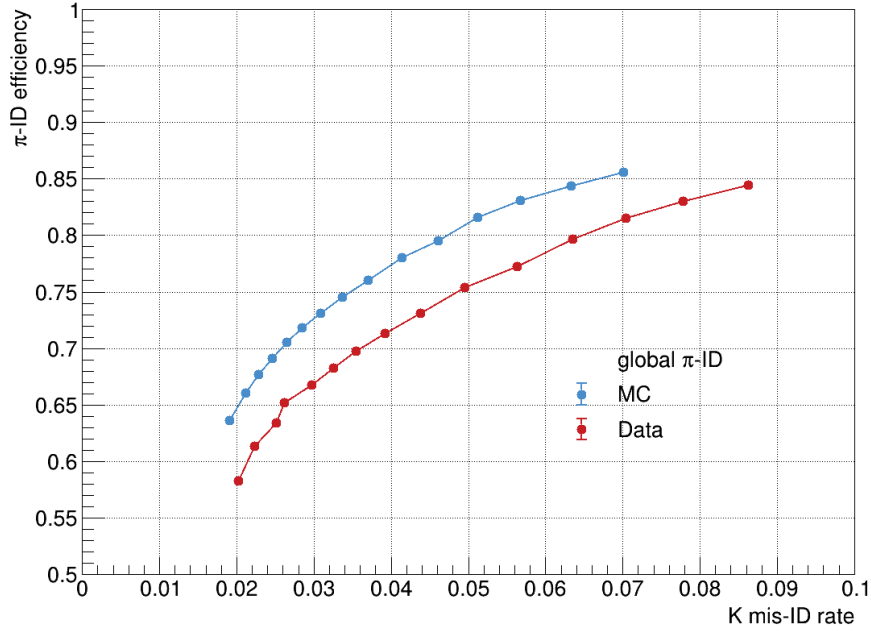
Figure 5.30: ROC curves for pion identification: π-ID efficiencies vs. K mis-ID rates for different PID criteria.

5.5 Summary

To summarise, we have studied the kaon and pion identification performance by applying PID selection criteria to the global PID probabilities: $P_K > 0.5$ or $P_\pi > 0.5$. Both in the analysis averaged over the kinematic distribution of the samples of charged particles and in the analysis in bin of momentum or $\cos\theta$, the real data do not exhibit charge asymmetries in the identification of charged kaons or pions that are significantly larger than those determined from the simulation.

Chapter 6

PID performance with weights and charge asymmetries

In this chapter, we will present the PID performance based on re-weighted PID probability selections and the measured charge asymmetries for kaon and pion identification. We will compare the results with those obtained in chapter 5 by applying PID selection criteria to the global probability.

6.1 Average identification performance

As discussed in chapter 3, the identification of charged particles relies on the determination of the likelihoods for six charged particle hypotheses thanks to the measurements from six Belle II sub-detectors. The simple combination of these likelihoods allows to define the global PID probability (Eq. 3.3), the selection criteria on this probability define the standard PID method. Adding calibration weights to the likelihood combination allows us to define a re-weighted PID probability (Eq. 3.13). In this chapter we evaluate the identification performance of kaons and pions based on re-weighted PID probability selections. To compare the results for the two different PID variables, we choose a threshold cut on $\tilde{P}$ so that the average identification efficiency, separately for kaons and pions, is the same as the average efficiency obtained with the global PID selection. To obtain the number of signal events before and after applying the PID selection, we fit the invariant mass distribution of the D_s candidates with the PDF model defined in chapter 4. We set fixed the parameters as in the previous chapter, in particular for the fits we choose the same values of the parameter f_1 used in the analysis with the global PID, both for `fit1` and `fit2`. We estimate the systematic uncertainties as described in the section 4.5.

We show in Tab. 6.1, both for MC and real data, the cuts chosen for the two PID variables, the K-ID efficiencies and π mis-ID rates, and the corresponding charge asymmetries. We report each measurement with its total error given by the sum in quadrature of the statistical and systematic uncertainties. Finally, we report the agreement λ (Eq. 5.5) between the charge asymmetries for MC simulation and real data obtained by applying the selection criteria to the re-weighted PID probability. In Tab. 6.2 we report the same for the π-ID performance. A good agreement is observed between the charge asymmetries determined with the simulated and experimental data when re-weighted PID selection is applied. Moreover, the charge asymmetries measured with the two PID variables have excellent agreement between them. It is important to observe the resulting fake rates by applying the PID selection with weights: in the case of pions mis-identified as kaons the fake rate is lower with respect to the fake rate obtained by applying the standard PID selection, instead for kaons mis-identified as pions the fake rate is higher. Therefore, the addition of calibration weights increases the K-ID performance, instead there is a degradation of the π-ID performance.

Table 6.1: Comparison of the K-ID performance by applying selection criteria to standard probability and re-weighted probability. The uncertainties are the sum in quadrature of statistical and systematic errors. λ quantifies the agreement between the charge asymmetries for the PID selection on $\tilde{P}_K$, in MC and real data.

| | MC15ri | | Data | | |
	global P_K	re-weighted $\tilde{P}_K$	global P_K	re-weighted $\tilde{P}_K$	λ
Cut	0.5	0.45	0.5	0.3	-
$\varepsilon_K{}^+$ (%)	86.2 ± 0.8	86.0 ± 0.8	84.2 ± 0.8	84.5 ± 0.8	-
$\varepsilon_K{}^-$ (%)	85.7 ± 0.6	85.4 ± 0.7	83.0 ± 0.6	83.4 ± 0.7	-
$\bar{\varepsilon}_K$ (%)	86.0 ± 0.5	85.7 ± 0.5	83.6 ± 0.5	84.0 ± 0.5	-
$\mathcal{A}(\varepsilon)$	0.003 ± 0.006	0.003 ± 0.006	0.007 ± 0.006	0.007 ± 0.006	0.4
$f_\pi{}^+$ (%)	5.6 ± 0.4	3.2 ± 0.2	5.3 ± 0.3	4.1 ± 0.2	-
$f_\pi{}^-$ (%)	5.7 ± 0.5	3.1 ± 0.5	6.5 ± 0.5	4.4 ± 0.5	-
$\bar{f}_\pi$ (%)	5.6 ± 0.3	3.1 ± 0.2	5.9 ± 0.3	4.3 ± 0.3	-
$\mathcal{A}(f)$	-0.02 ± 0.05	0.03 ± 0.08	-0.11 ± 0.05	-0.04 ± 0.06	0.7

Table 6.2: Comparison of the π-ID performance by applying selection criteria to standard probability and re-weighted probability. The uncertainties are the sum in quadrature of statistical and systematic errors. λ quantifies the agreement between the charge asymmetries for the PID selection on $\tilde{P}_\pi$, in MC and real data.

| | MC15ri | | Data | | |
	global P_π	re-weighted $\tilde{P}_\pi$	global P_π	re-weighted $\tilde{P}_\pi$	λ
Cut	0.5	0.8	0.5	0.8	-
$\varepsilon_\pi{}^+$ (%)	77.1 ± 0.7	77.5 ± 0.8	74.3 ± 0.7	75.0 ± 0.8	-
$\varepsilon_\pi{}^-$ (%)	76.5 ± 0.8	76.8 ± 0.6	73.4 ± 0.7	73.9 ± 0.6	-
$\bar{\varepsilon}_\pi$ (%)	76.8 ± 0.5	77.1 ± 0.5	73.8 ± 0.5	74.5 ± 0.5	-
$\mathcal{A}(\varepsilon)$	0.004 ± 0.007	0.005 ± 0.007	0.006 ± 0.007	0.007 ± 0.007	0.2
$f_K{}^+$ (%)	3.70 ± 0.04	4.09 ± 0.10	4.39 ± 0.04	5.33 ± 0.10	-
$f_K{}^-$ (%)	4.0 ± 0.4	4.4 ± 0.4	4.3 ± 0.3	5.6 ± 0.4	-
$\bar{f}_K$ (%)	3.8 ± 0.2	4.3 ± 0.2	4.3 ± 0.1	5.4 ± 0.2	-
$\mathcal{A}(f)$	-0.04 ± 0.05	-0.04 ± 0.04	0.01 ± 0.03	-0.02 ± 0.04	0.3

6.2 PID performance and charge asymmetries in bins of kinematic variables

In this section we investigate the PID performance with weights as a function of the two kinematic variables: momentum p and $\cos\theta$. We divide the tracks according to the binning chosen in section 5.3. We perform the fits of the D_s invariant mass distribution before and after applying PID selection, then we estimate the ID observables according to the equations in chapter 5. We apply the threshold cuts on $\tilde{P}_K$ and $\tilde{P}_\pi$ reported in Tab. 6.1 and Tab. 6.2, respectively. We calculate the charge asymmetries for both MC and real data and we evaluate their agreement. Furthermore, we compare these charge asymmetries as a function of the kinematic variables with those obtained by applying PID selection to the global probability.

6.2.1 K-ID performance and charge asymmetries in bins of kinematic variables

In Fig. 6.1 we show the kaon ID efficiencies and pion mis-ID rates separately for positively (left) and negatively (right) charged tracks divided into momentum bins. In Fig. 6.2, on the left we show the charge asymmetries of the efficiencies as a function of p for both MC (blue triangles) and real data (red triangles), on the right the charge asymmetries of the fake rates. Then, in Fig. 6.3 and Fig. 6.4 we show the ID observables and charge asymmetries as a function of $\cos\theta$, respectively.

Both in the analysis in momentum bins and in $\cos\theta$ bins, the K-ID efficiencies calculated with selection on the re-weighted PID probability are about the same as the efficiencies with selection on the global PID probability (Fig. 5.21 and Fig. 5.23). Instead, the π mis-ID rates obtained by considering selection criteria on the new PID variable with weights are lower.
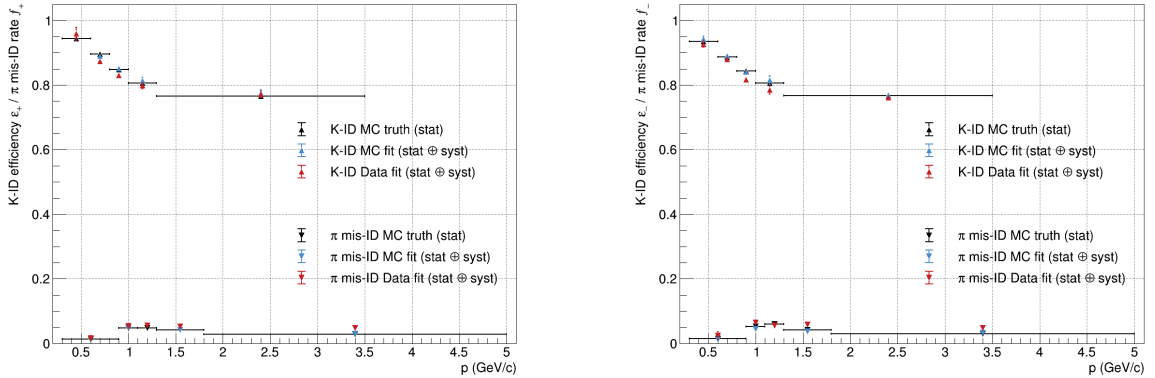

Figure 6.1: Kaon ID efficiency and pion mis-ID rate for the PID selection on $\tilde{P}_K$, in bins of momentum: positively charged particles (left) and negatively charged particles (right).
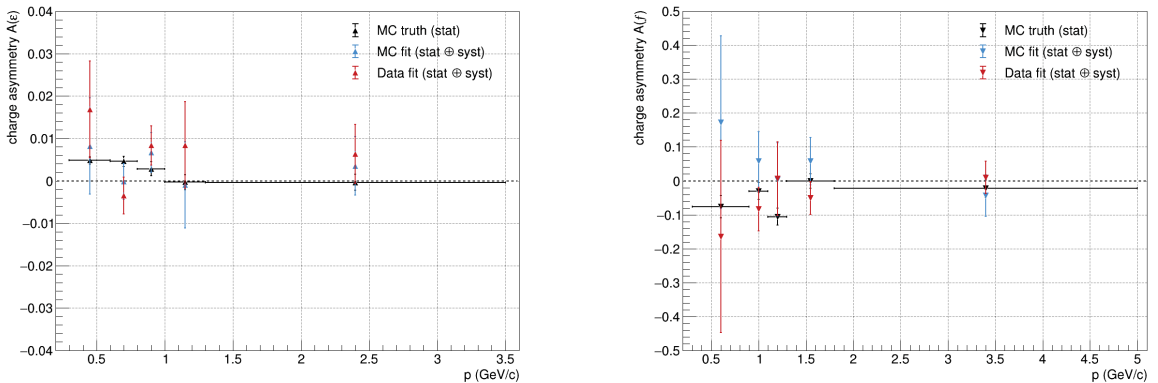

Figure 6.2: Charge asymmetries for the PID selection on $\tilde{P}_K$, in bins of momentum: asymmetry of efficiencies (left) and asymmetry of fake rates (right).
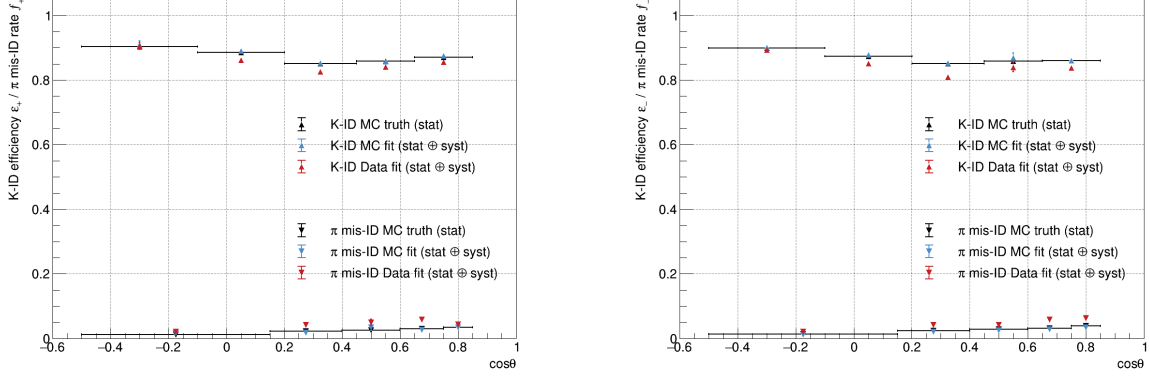
Figure 6.3: Kaon ID efficiency and pion mis-ID rate for the PID selection on $\tilde{P}_K$, in bins of $\cos\theta$: positively charged particles (left) and negatively charged particles (right).
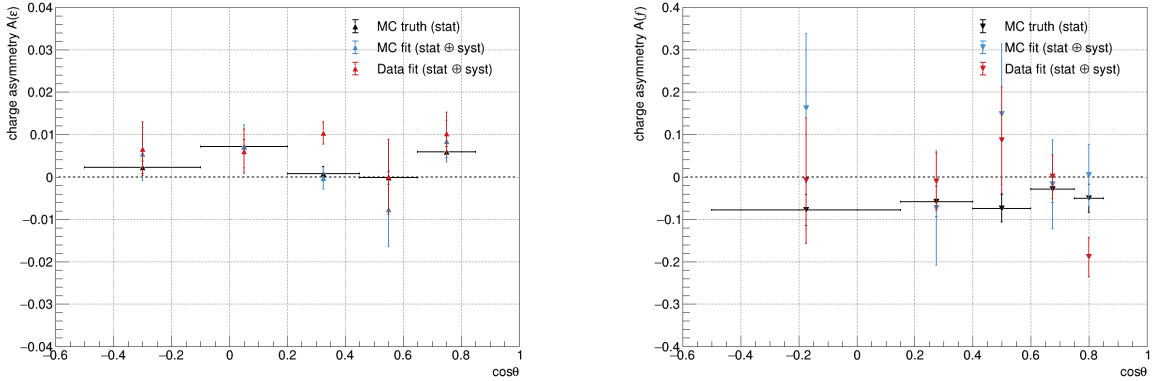

Figure 6.4: Charge asymmetries for the PID selection on $\tilde{P}_K$, in bins of $\cos\theta$: asymmetry of efficiencies (left) and asymmetry of fake rates (right).

Overall, the real data do not exhibit charge asymmetries in the identification of kaons that are significantly larger than those from the simulation. This conclusion is valid after splitting the tracks in momentum bins or $\cos\theta$ bins. Only in few cases we observe discrepancies between the charge asymmetries for simulated and experimental data (the error bars do not overlap), however there is no net difference in the measurements as a function of the kinematic variables. Finally, the charge asymmetries of the efficiencies and those of the fake rates by applying PID selections to the re-weighted PID probability have good agreement with the charge asymmetries measured with the standard PID. We can compare the plots in Fig. 6.2 with those in Fig. 5.22, and the plots in Fig. 6.4 with those in Fig. 5.24.

6.2.2 π-ID performance and charge asymmetries in bins of kinematic variables

We present the pion ID efficiencies and kaon mis-ID rates as a function of momentum or $\cos\theta$, for positive and negative particle tracks. In Fig. 6.5 we show the ID observables (positive charge left, negative charge right) as a function of p, in Fig. 6.6 the charge asymmetries of the efficiencies (left) and the charge asymmetries of the fake rates (right) for both MC (blue triangles) and real data (red triangles). In Fig. 6.7 and Fig. 6.8 we show the ID observables and charge asymmetries as a function of $\cos\theta$, respectively.

We evaluate the PID performance with weights by comparing the results presented in this section with those reported in Fig. 5.25 and Fig. 5.27: the π-ID efficiencies are about the same as the efficiencies with selection on the global PID probability, while the K mis-ID rates are slightly higher. We do not observe significant differences between the charge asymmetries of the ID efficiencies for MC and real data as a function of p, the agreement between the measurements is good. The same also applies to

the charge asymmetries of the mis-ID rates. Moreover, there is a good agreement between the charge asymmetries of the ID observables measured considering cuts on both re-weighted and global PID probabilities (Fig. 5.26). In the analysis in bins of $\cos\theta$ the agreement between the charge asymmetries of the efficiencies for MC and real data is very good, instead we observe a difference between the asymmetries of the fake rates in some bins. Even in the case of the standard identification method (Fig. 5.28), we observe a discrepancy between the asymmetries for simulated and real data in the same bins. However, there is no net difference between the asymmetries measured as a function of $\cos\theta$.
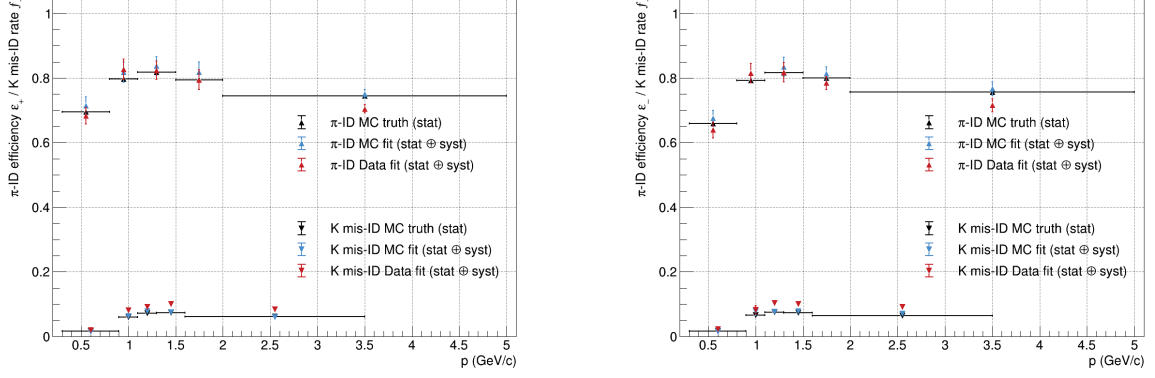

Figure 6.5: Pion ID efficiency and kaon mis-ID rate for the PID selection on $\tilde{P}_\pi$, in bins of momentum: positively charged particles (left) and negatively charged particles (right).
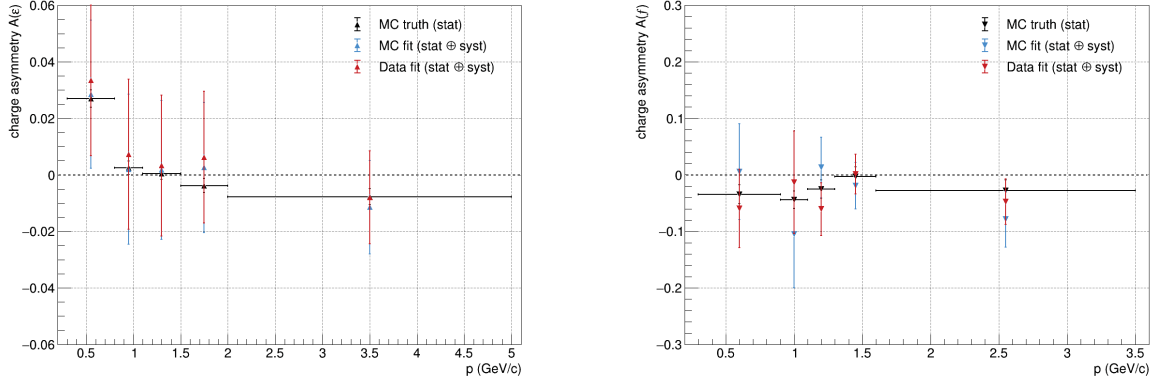

Figure 6.6: Charge asymmetries for the PID selection on $\tilde{P}_\pi$, in bins of momentum: asymmetry of efficiencies (left) and asymmetry of fake rates (right).
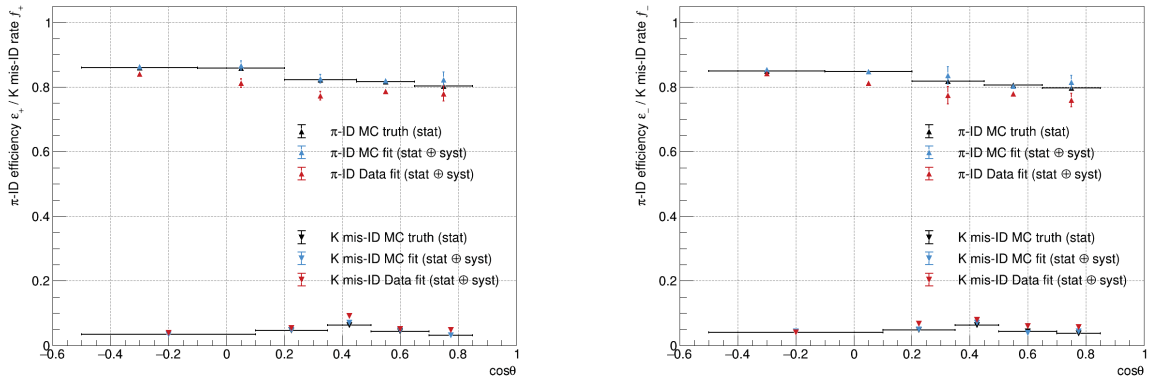

Figure 6.7: Pion ID efficiency and kaon mis-ID rate for the PID selection on $\tilde{P}_\pi$, in bins of $\cos\theta$: positively charged particles (left) and negatively charged particles (right).
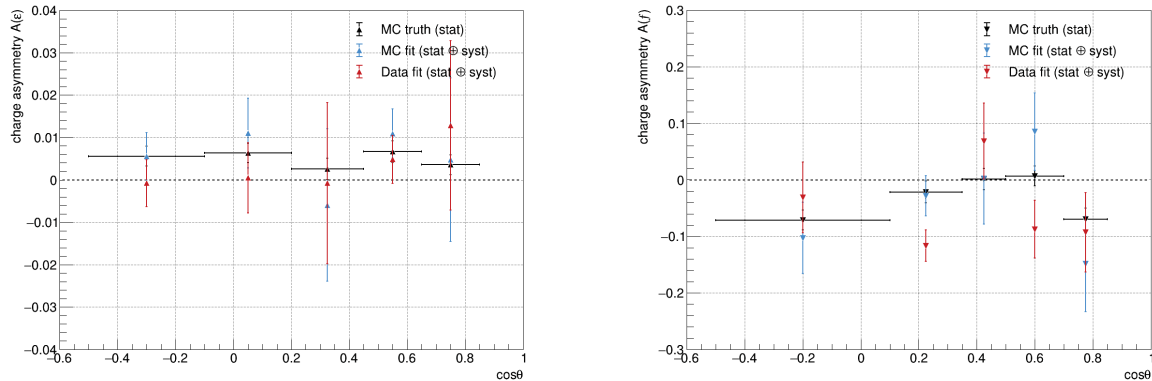
Figure 6.8: Charge asymmetries for the PID selection on $\tilde{P}_\pi$, in bins of $\cos\theta$: asymmetry of efficiencies (left) and asymmetry of fake rates (right).

6.3 Identification performance for different PID criteria

In this section we study the identification performance as a function of the threshold cut applied to the re-weighted PID probability: $\tilde{P} > \alpha$. We determine the average identification efficiencies and average mis-identification rates by varying α, each efficiency-fake rate pair defines a point of the ROC curve.

In Fig. 6.9 we show the ROC curve for the K-ID performance with weights (K-ID efficiency vs. π mis-ID rate). The applied selection criteria are $\tilde{P}_K > 0.2$, 0.25, 0.3, 0.35, 0.4, 0.45, 0.5, 0.55, 0.6, 0.65, 0.7, 0.75, 0.8, 0.85, 0.9 on both the MC and data samples. In the same figure we show also the ROC curve for the standard performance and we can compare the results. Overall, the re-weighted PID performs better than the global PID on both simulated and experimental data. Indeed, for a fixed fake rate the K-ID efficiency with the re-weighted PID is higher than that with the global PID: an improvement of 3-6% for MC and 2-7% for real data is observed. For real data the identification performance is slightly worse with both PID variables.
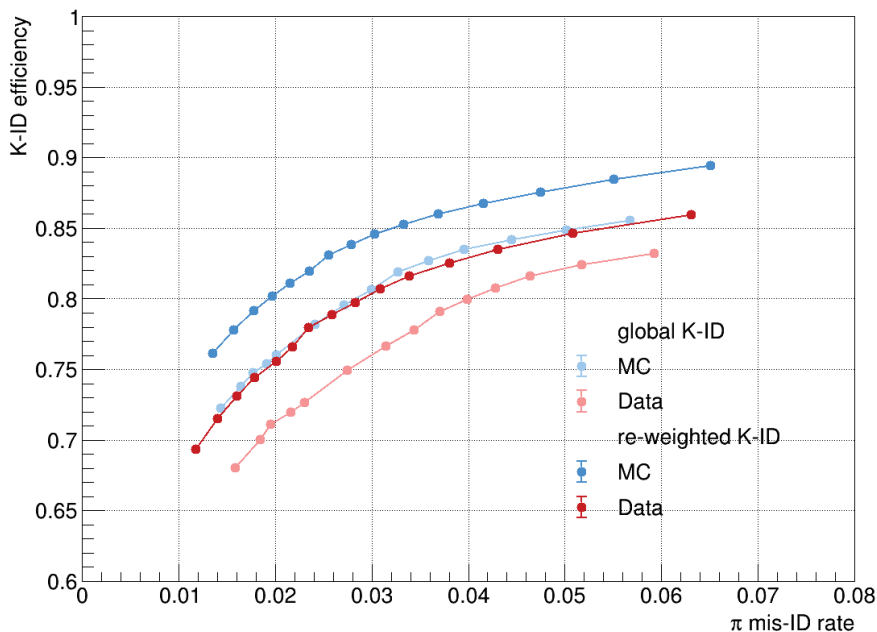

Figure 6.9: ROC curves for kaon identification: K-ID efficiencies vs. π mis-ID rates for different PID criteria. Comparison of the ROC curves for the identification based on the re-weighted and global PID selections.

In Fig. 6.10 we show the ROC curve for the π-ID performance with weights (π-ID efficiency vs. K mis-ID rate). The applied selection criteria are $\tilde{P}_\pi > 0.4, 0.45, 0.5, 0.55, 0.6, 0.65, 0.7, 0.75, 0.8, 0.85, 0.9, 0.95$. We compare the ROC curves for both MC and data with the two curves for the standard PID performance. Contrary to what is observed for the identification of kaons, in this case of pion identification performance the addition of calibration weights does not produce improvements, but rather we observe a degradation, especially for real data (2-3%). For simulation the two ROC curves are very close, the one for the re-weighted PID is only slightly lower than the other.
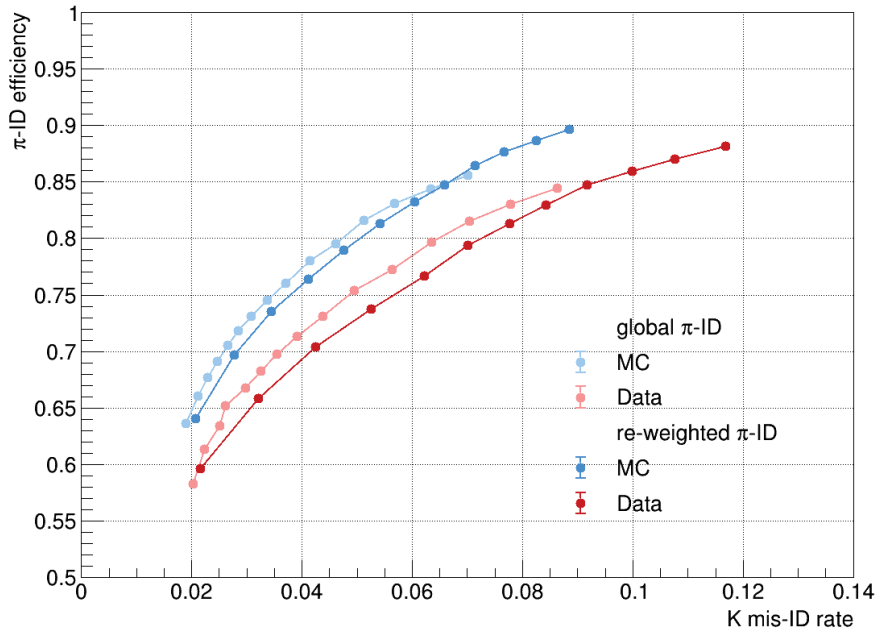

Figure 6.10: ROC curves for kaon identification: π-ID efficiencies vs. K mis-ID rates for different PID criteria. Comparison of the ROC curves for the identification based on the re-weighted and global PID selections.

6.4 Summary

To summarise, we have studied the kaon and pion identification performance by applying PID selection criteria to the re-weighted PID probabilities. We have chosen the threshold cuts to have average identification efficiencies for MC and data similar to those with the standard PID. Adding weights improves the K identification performance, while the π identification performance decreases slightly. We do not observe significant differences between the charge asymmetries for both MC and real data.

Chapter 7

PID performance with binary likelihood ratios and charge asymmetries

In this chapter, we will present the PID performance based on the binary likelihood ratio discriminators between the two hypotheses K and π. We will study the performance for individual sub-detectors, in particular for the TOP and the CDC. The goals are to evaluate the PID performance of the TOP detector alone and to investigate possible effects arising from external sources such as the tracking of charged particles in the CDC. We will measure the efficiencies (fake rates) and the corresponding charge asymmetries for kaon and pion identification, then we will compare the results with those obtained in chapter 5 by considering the global PID.

7.1 Likelihood-based selection

The global PID probability for a particle hypothesis is defined as likelihood ratio (Eq. 3.3), the likelihoods are calculated from measurements of the identifying observables in the active sub-detectors for a particle track. However, from the likelihoods for an individual sub-detector it is possible to build binary likelihood ratios between two particle hypotheses (in our case π and K) and study the PID performance of the individual sub-detector. In this work we want to investigate the contribution of TOP and CDC detectors to particle identification, therefore we define two additional PID probabilities: binary likelihood ratio (LR) for the TOP and binary likelihood ratio for the CDC (Eq. 3.18).

In order to compare the kaon (pion) ID performance with binary LR for the TOP and that with binary LR for the CDC, we select only tracks for which it has been possible to calculate both TOP and CDC likelihoods for K and π hypotheses. In fact, it is not always possible to determine for a track the likelihoods of both individual sub-detectors. Furthermore, we require that the tracks are in the acceptance of the TOP detector. Thus, in this chapter we restrict our analysis to tracks from the decay sample $D_s^+ \to \phi(K^+ K^-)\,\pi^+$ that fulfil the selection criteria given in Tab. 7.1.

Table 7.1: Likelihood-based selection criteria for individual sub-detectors, $\beta = K, \pi$.

Particle	Selection
K^+	$\mathcal{L}_\beta^{\mathrm{TOP}}(K^+) \neq \mathrm{NaN},\ \mathcal{L}_\beta^{\mathrm{CDC}}(K^+) \neq \mathrm{NaN},\ -0.5 < \cos\theta(K^+) < 0.866$
K^-	$\mathcal{L}_\beta^{\mathrm{TOP}}(K^-) \neq \mathrm{NaN},\ \mathcal{L}_\beta^{\mathrm{CDC}}(K^-) \neq \mathrm{NaN},\ -0.5 < \cos\theta(K^-) < 0.866$
π^+	$\mathcal{L}_\beta^{\mathrm{TOP}}(\pi^+) \neq \mathrm{NaN},\ \mathcal{L}_\beta^{\mathrm{CDC}}(\pi^+) \neq \mathrm{NaN},\ -0.5 < \cos\theta(\pi^+) < 0.866$
π^-	$\mathcal{L}_\beta^{\mathrm{TOP}}(\pi^-) \neq \mathrm{NaN},\ \mathcal{L}_\beta^{\mathrm{CDC}}(\pi^-) \neq \mathrm{NaN},\ -0.5 < \cos\theta(\pi^-) < 0.866$

7.2 Analysis by PID selection on the binary likelihood ratio for the TOP

Particle identification with the TOP is based on the measurement of the time and position where totally internally-reflected Cherenkov photons hit an array of photo-multipliers located at the end of a fused silica bar, Cherenkov light is emitted by the passage of charged particles through the bar (see chapters 2 and 3). From the individual likelihoods for kaon and pion hypotheses, we can define the binary likelihood ratios for the TOP as in Eq. 3.18

$$\mathrm{P}_{K/\pi}^{\mathrm{TOP}} = \frac{\mathcal{L}_K^{\mathrm{TOP}}}{\mathcal{L}_K^{\mathrm{TOP}} + \mathcal{L}_\pi^{\mathrm{TOP}}} \qquad\qquad \mathrm{P}_{\pi/K}^{\mathrm{TOP}} = 1 - \mathrm{P}_{K/\pi}^{\mathrm{TOP}} = \frac{\mathcal{L}_\pi^{\mathrm{TOP}}}{\mathcal{L}_K^{\mathrm{TOP}} + \mathcal{L}_\pi^{\mathrm{TOP}}}$$

We determine the identification efficiencies, separately for kaons and pions, as the ratio of the number of signal events after and before applying PID selection criteria to the binary LR for the TOP. We follow the same fitting strategy outlined in section 6.1.

In Tab. 7.2 and Tab. 7.3 we report, for both MC and real data, the chosen cut on the binary LR that provides an average ID efficiency similar to that obtained with the cut on the global PID probability. We observe that the threshold cut on the binary likelihood ratios is very high. Moreover, in the tables we show the identification efficiencies and fake rates (kaon ID in the first table, pion ID in the second one), and the corresponding charge asymmetries $\mathcal{A}(\varepsilon)$ and $\mathcal{A}(f)$. We show each measurement and its total uncertainty. In the last column of both tables we report the agreement λ (Eq. 5.5) between the charge asymmetries for MC and data obtained by applying selection criteria to the PID probability $\mathrm{P}^{\mathrm{TOP}}$: we observe good agreement in all cases ($\lambda < 1$).
By comparing the K-ID performance with the two PID variables, we see good agreement between the efficiencies and also between the charge asymmetries for both MC and experimental data. Instead, the π mis-ID rates in the identification performance of the TOP detector are lower than those measured with the standard PID. Therefore, if we consider the TOP detector alone we have a better K/π separation performance. A not bad agreement is observed between the charge asymmetries of the fake rates.
In the case of the π-ID performance, we note that the threshold cut is very close to 1 in order to have an average ID efficiency similar to that with global PID. Furthermore, the total uncertainty of the efficiencies is very large (of the order of percent), the largest contribution is given by the systematic uncertainty. In this case, the π/K separation performance of the TOP detector is worse than the performance with global PID, because the K mis-ID rates in the first case are higher than in the second one. The agreement between the charge asymmetries of the efficiencies measured by applying selections to the two PID variables is good, while the agreement between the asymmetries of the fake rates is not bad.

After the analysis averaged over the kinematic distribution of the samples of charged kaons and pions, in the following subsections we report the PID performance with binary LR for the TOP as a function of the two kinematic variables: momentum p and $\cos\theta$. We divide the tracks according to the binning chosen in section 5.3. We measure the charge asymmetries for both MC and real data and we evaluate their agreement. Then, we compare these charge asymmetries in the identification of charged kaons and pions performed by the TOP detector with those obtained by applying PID selection to the global probability.
Finally, we study the identification performance, separately for kaons and pions, as a function of the threshold cut applied to the LR for the TOP counter: $\mathrm{P}^{\mathrm{TOP}} > \alpha$. We determine the average identification efficiencies and average mis-identification rates by varying α, each efficiency-fake rate pair defines a point of the ROC curve.

Table 7.2: Comparison of the K-ID performance by applying selection criteria to standard probability and binary likelihood ratio for the TOP. The uncertainties are the sum in quadrature of statistical and systematic errors. λ quantifies the agreement between the charge asymmetries for the PID selection on $\mathrm{P}^{\mathrm{TOP}}_{K/\pi}$, in MC and real data.

| | MC15ri | | Data | | |
	global P_K	LR $\mathrm{P}^{\mathrm{TOP}}_{K/\pi}$	global P_K	LR $\mathrm{P}^{\mathrm{TOP}}_{K/\pi}$	λ
Cut	0.5	0.95	0.5	0.95	-
ε_{K^+} (%)	86.2 ± 0.8	87.0 ± 0.6	84.2 ± 0.8	84.6 ± 0.6	-
ε_{K^-} (%)	85.7 ± 0.6	85.8 ± 0.6	83.0 ± 0.6	82.5 ± 0.6	-
$\bar{\varepsilon}_K$ (%)	86.0 ± 0.5	86.4 ± 0.4	83.6 ± 0.5	83.6 ± 0.4	-
$\mathcal{A}(\varepsilon)$	0.003 ± 0.006	0.007 ± 0.005	0.007 ± 0.006	0.013 ± 0.005	0.8
f_{π^+} (%)	5.6 ± 0.4	3.3 ± 0.4	5.3 ± 0.3	4.3 ± 0.4	-
f_{π^-} (%)	5.7 ± 0.5	2.4 ± 0.6	6.5 ± 0.5	4.1 ± 0.6	-
$\bar{f}_\pi$ (%)	5.6 ± 0.3	2.9 ± 0.3	5.9 ± 0.3	4.2 ± 0.4	-
$\mathcal{A}(f)$	-0.02 ± 0.05	0.1 ± 0.1	-0.11 ± 0.05	0.02 ± 0.09	0.8

Table 7.3: Comparison of the π-ID performance by applying selection criteria to standard probability and binary likelihood ratio for the TOP. The uncertainties are the sum in quadrature of statistical and systematic errors. λ quantifies the agreement between the charge asymmetries for the PID selection on $\mathrm{P}^{\mathrm{TOP}}_{\pi/K}$, in MC and real data.

| | MC15ri | | Data | | |
	global P_π	LR $\mathrm{P}^{\mathrm{TOP}}_{\pi/K}$	global P_π	LR $\mathrm{P}^{\mathrm{TOP}}_{\pi/K}$	λ
Cut	0.5	0.999995	0.5	0.9997	-
ε_{π^+} (%)	77.1 ± 0.7	78 ± 2	74.3 ± 0.7	74 ± 2	-
ε_{π^-} (%)	76.5 ± 0.8	77 ± 2	73.4 ± 0.7	74 ± 2	-
$\bar{\varepsilon}_\pi$ (%)	76.8 ± 0.5	77 ± 1	73.8 ± 0.5	74 ± 1	-
$\mathcal{A}(\varepsilon)$	0.004 ± 0.007	0.00 ± 0.02	0.006 ± 0.007	0.00 ± 0.02	0.05
f_{K^+} (%)	3.70 ± 0.04	5.3 ± 0.2	4.39 ± 0.04	5.8 ± 0.2	-
f_{K^-} (%)	4.0 ± 0.4	6.3 ± 0.6	4.3 ± 0.3	7.1 ± 0.6	-
$\bar{f}_K$ (%)	3.8 ± 0.2	5.8 ± 0.3	4.3 ± 0.1	6.4 ± 0.3	-
$\mathcal{A}(f)$	-0.04 ± 0.05	-0.08 ± 0.05	0.01 ± 0.03	-0.10 ± 0.05	0.3

7.2.1 K-ID performance and charge asymmetries in bins of kinematic variables

Fig. 7.1 shows the kaon ID efficiencies and pion mis-ID rates separately for positively (left) and negatively (right) charged tracks divided into momentum bins, while Fig. 7.2 shows the charge asymmetries as a function of p for both MC and real data. In Fig. 7.3 and Fig. 7.4 we show the ID observables and charge asymmetries as a function of $\cos\theta$, respectively. The blue triangles represent the simulation results, while the red triangles represent the measurements from real data.

We compare the K-ID performance of the TOP detector with the performance with the global PID (Fig. 5.21 and Fig. 5.23). Both positive and negative ID efficiencies for low momenta are lower if we consider PID information only from the TOP, instead for higher momenta the efficiencies are higher. The π mis-ID rates are lower overall, therefore the K/π separation performance is better for $p > 0.6$ GeV/c. In the analysis as a function of $\cos\theta$, the K-ID efficiencies are higher for forward or backward tracks, while they are slightly lower for tracks with $50° < \theta < 90°$.

We observe that the charge asymmetries of the efficiencies for the TOP alone are slightly larger than the asymmetries with the global PID in many bins (Fig. 5.22 and Fig. 5.24). This suggests a better TOP PID performance for positively charged kaons than negatively charged kaons, regardless of kinematics. However, the real data do not exhibit charge asymmetries that are significantly larger than those from the simulation. On the other hand, for the charge asymmetries of the pion fake rates we note an overestimation of the asymmetries resulting from the fits on MC. Nevertheless, the error associated with the measurements is very large and the agreement between estimated and true asymmetries is not bad. Also in this case we do not observe a significant discrepancy between the asymmetries for MC and data. Finally, we note that the charge asymmetries of the fake rates with the global PID are overall more negative. The K/π separation power of the TOP is better for K^+.
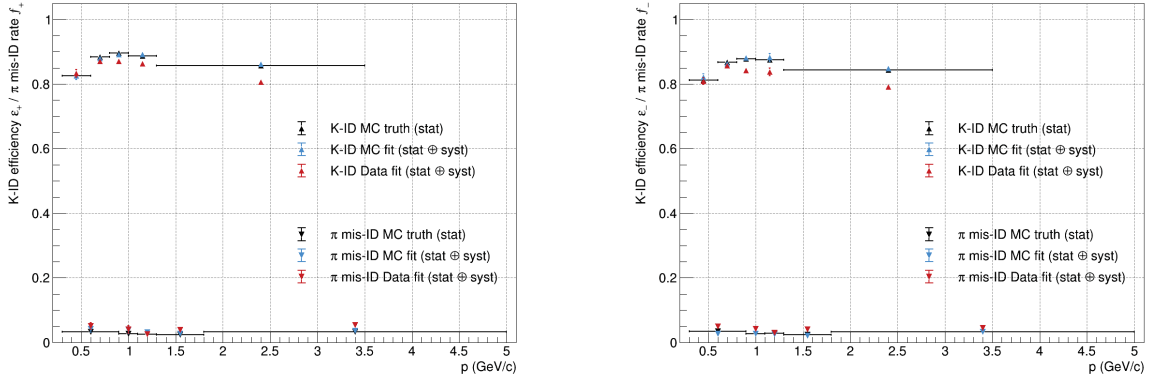

Figure 7.1: Kaon ID efficiency and pion mis-ID rate for the PID selection on $\mathrm{P}^{\mathrm{TOP}}_{K/\pi}$, in bins of momentum: positively charged particles (left) and negatively charged particles (right).

Figure 7.2: Charge asymmetries for the PID selection on $\mathrm{P}^{\mathrm{TOP}}_{K/\pi}$, in bins of momentum: asymmetry of efficiencies (left) and asymmetry of fake rates (right).

Figure 7.3: Kaon ID efficiency and pion mis-ID rate for the PID selection on $P_{K/\pi}^{TOP}$, in bins of $\cos\theta$: positively charged particles (left) and negatively charged particles (right).
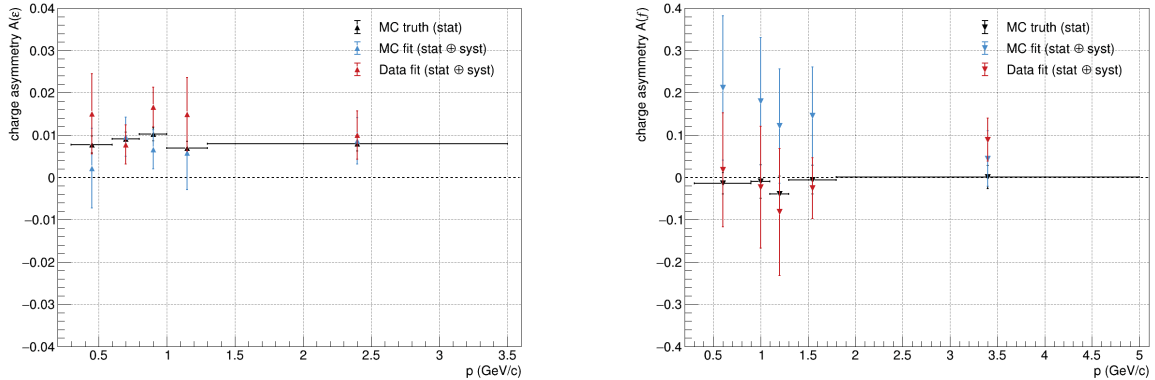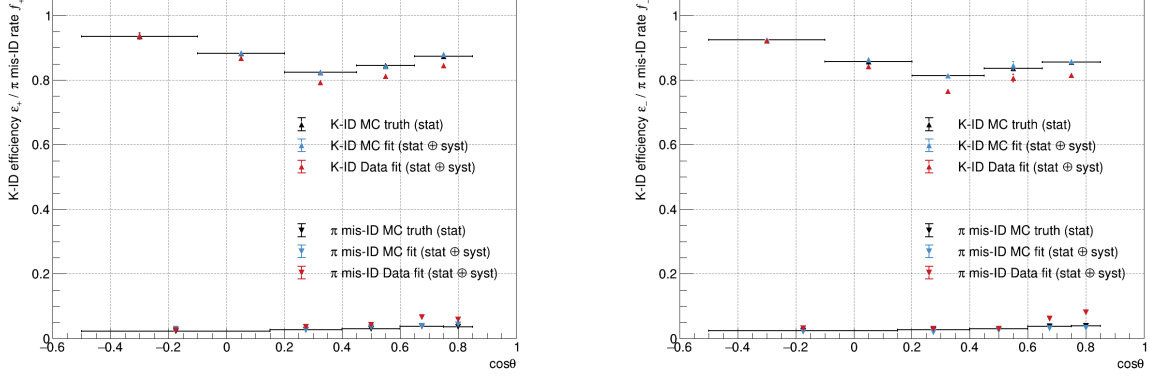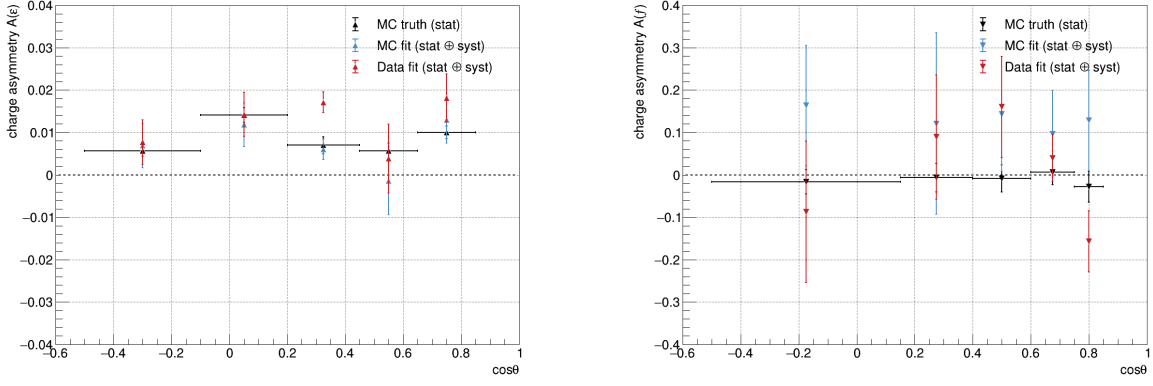

Figure 7.4: Charge asymmetries for the PID selection on $P_{K/\pi}^{TOP}$, in bins of $\cos\theta$: asymmetry of efficiencies (left) and asymmetry of fake rates (right).

7.2.2 π-ID performance and charge asymmetries in bins of kinematic variables

In this subsection, Fig. 7.5 shows the pion ID efficiencies and kaon mis-ID rates (positive charge left, negative charge right) as a function of momentum, Fig. 7.6 shows the charge asymmetries of the efficiencies (left) and the charge asymmetries of the fake rates (right) for both MC (blue triangles) and real data (red triangles). In Fig. 7.7 and Fig. 7.8 we show the ID observables and charge asymmetries as a function of $\cos\theta$, respectively.

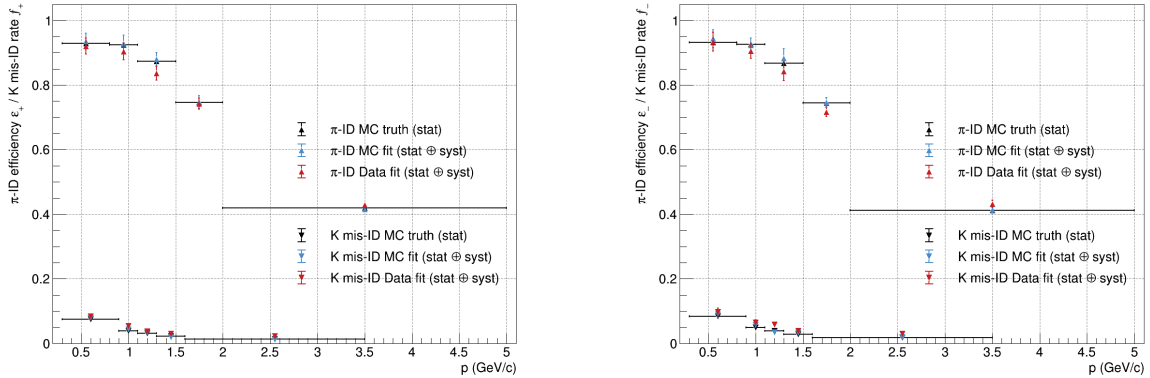We compare the π-ID performance of the TOP detector alone with the performance with the standard

Figure 7.5: Pion ID efficiency and kaon mis-ID rate for the PID selection on $P_{\pi/K}^{TOP}$, in bins of momentum: positively charged particles (left) and negatively charged particles (right).
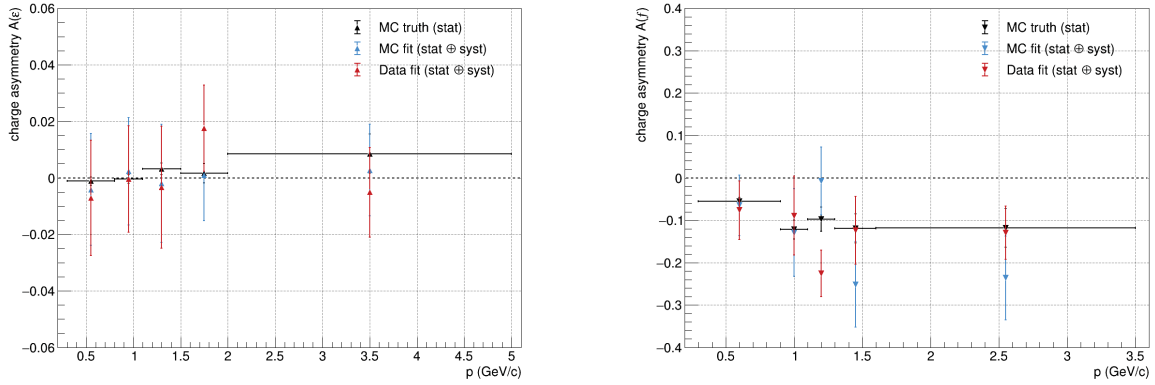
Figure 7.6: Charge asymmetries for the PID selection on $P_{\pi/K}^{TOP}$, in bins of momentum: asymmetry of efficiencies (left) and asymmetry of fake rates (right).
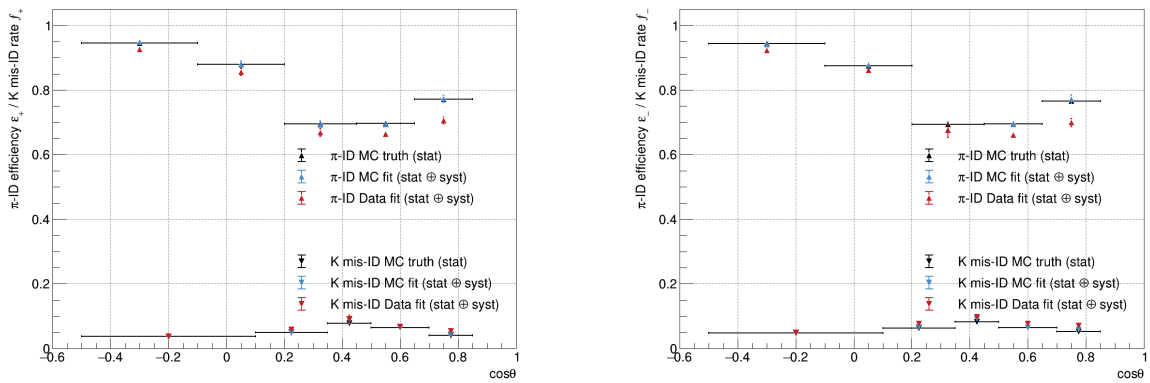

Figure 7.7: Pion ID efficiency and kaon mis-ID rate for the PID selection on $P_{\pi/K}^{TOP}$, in bins of $\cos\theta$: positively charged particles (left) and negatively charged particles (right).
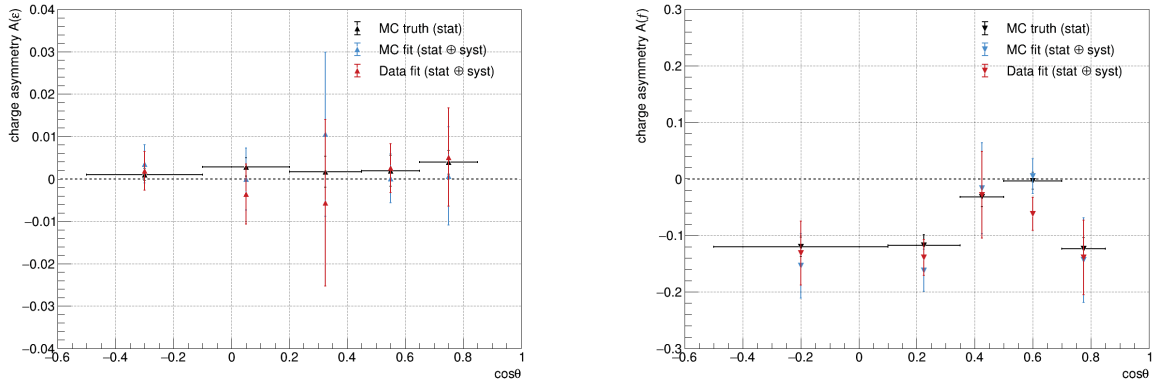

Figure 7.8: Charge asymmetries for the PID selection on $P_{\pi/K}^{TOP}$, in bins of $\cos\theta$: asymmetry of efficiencies (left) and asymmetry of fake rates (right).

PID (Fig. 5.25 and Fig. 5.27). For $p < 1$ GeV/c the identification efficiencies for both π^+ and π^- are much higher when considering only the TOP counter, however also the mis-ID rates are higher. In contrast, for high momenta ($p > 2$ GeV/c) there is a sharp degradation of efficiencies. Whereas with the global PID the pion identification efficiencies are more or less constant as a function of $\cos\theta$, the efficiencies based only on information from the TOP detector are higher for large θ angles, while they are lower for smaller angles.

We do not observe significant differences between the charge asymmetries of the efficiencies for MC

and real data as a function of both momentum and $\cos\theta$, the agreement between the measurements is good. In addition, we observe a good agreement between the asymmetries measured with the TOP only and global PID (Fig. 5.26 and Fig. 5.28). Instead, by looking at the plots on the right in Fig. 7.6 and Fig. 7.8, in some bins there is a difference between the charge asymmetries of the kaon fake rates for simulation and real data. However, even in this case the discrepancy is not very significant. Furthermore, the charge asymmetries $\mathcal{A}(f)$ measured by identification with the TOP tend to be more negative than the charge asymmetries with the standard PID. This indicates that the π/K separation power of the TOP is lower for negatively charged particles because the K mis-ID rates is higher for negative kaons mis-identified as pions.

7.2.3 Identification performance for different PID criteria

We want to study the K-ID performance of the TOP detector, therefore we build the ROC curve by determining the efficiency-fake rate pairs for the following threshold cuts on the binary LR for the TOP: $P_{K/\pi}^{\mathrm{TOP}} > 0.4, 0.45, 0.5, 0.55, 0.6, 0.65, 0.7, 0.75, 0.8, 0.85, 0.9, 0.95, 0.99, 0.995, 0.999$. The ROC curves for both MC and real data are shown in Fig. 7.9. In the same figure we show also the ROC curves for the standard performance and we can compare the results. By considering the results for the MC sample, the PID performance with the TOP detector is better than that with the global PID, indeed the ROC curve for the individual sub-detector is closer to the upper-left corner of the space K-ID efficiency vs. π mis-ID rate. The K-ID efficiency for MC with the TOP is 6-11% greater than that with the global PID for a fixed fake rate. We observe a better TOP identification performance for real data as well: the ID efficiency by applying PID selection to the binary LR for the TOP is higher for mis-ID rates between 2% and 6%. The improvement on data is 1-5%.
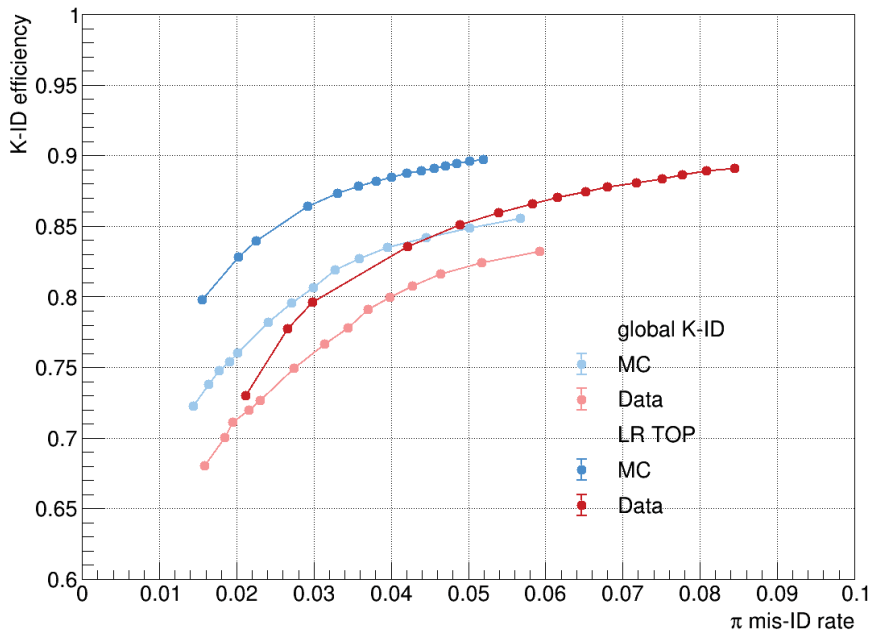

Figure 7.9: ROC curves for kaon identification: K-ID efficiencies vs. π mis-ID rates for different PID criteria. Comparison of the ROC curves for the identification based on the binary likelihood ratio for the TOP and global PID selections.

In Fig. 7.10 we show the ROC curves π-ID efficiency vs. K mis-ID rate for the TOP detector alone. The performance is evaluated on both MC and real data samples, the applied selection criteria are $P_{\pi/K}^{\mathrm{TOP}} > 0.5, 0.8, 0.9, 0.99, 0.999, 0.9999, 0.99999, 0.999999$. These curves are shifted to the right with respect to the ROC curves corresponding to standard PID performance. Moreover, we emphasize that to achieve efficiencies similar to those with selections on the global PID, the threshold cuts on

the binary LR must be very high. It is possible to have high pion identification efficiencies with the TOP, however the fake rates are also high: for $P_{\pi/K}^{TOP} > 0.8$ the π-ID efficiency for MC (real data) is close to 95% (90%), but the K mis-ID rate is about 10%. For $P_{\pi/K}^{TOP} > 0.99$, the ROC curves for the TOP pass below the curves for the standard PID. For the same fake rate, between 5% and 7%, the π-ID efficiency with the TOP only is lower than the other. Therefore, the PID of pions with the TOP counter alone does not perform better.
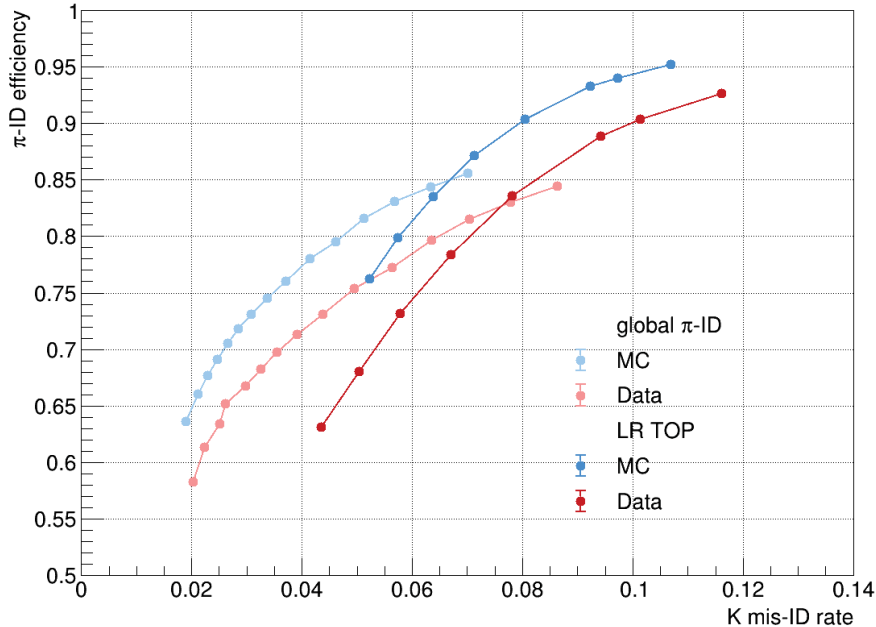

Figure 7.10: ROC curves for kaon identification: π-ID efficiencies vs. K mis-ID rates for different PID criteria. Comparison of the ROC curves for the identification based on the binary likelihood ratio for the TOP and global PID selections.

In summary, the TOP detector can identify charged kaons well, while it does not perform as well for charged pions. Although the overall K-ID performance with the TOP alone is better, in some regions of the phase space additional PID information from other sub-detectors is needed for improved K/π separation.

To ensure detector longevity and maintain good reconstruction performance, beam backgrounds must remain well controlled [39]. The most vulnerable sub-detectors are the TOP and the CDC. One key issue in the CDC is that pattern recognition of charged tracks becomes increasingly difficult as the wire-hit rate increases. Extra background hits, caused in particular by the injection background, progressively degrade the CDC performance. At present, the bias in the CDC measurements due to the injection background has not been properly calibrated. Consequently, the identification of kaons is overall worse with the global PID because, in addition to the likelihood information from the TOP, we also consider the likelihood information from the CDC (and other sub-detectors). In order to address the degradation of the global PID performance, it is necessary to evaluate and adjust the PID information from the individual sub-detectors. For example, if we compare the ROC curves for the K identification based on the binary likelihood ratio for the TOP and re-weighted PID selections, we observe that the curves for MC and real data, respectively, are much closer to each other. We can see the comparison of ROC curves in Fig. 7.11.

Furthermore, we recall that particle identification at Belle II uses only the pure likelihood information from the individual sub-detectors to formulate a classification variable for K/π separation. The inclusion of calibration weights for each detector-hypothesis pair already improves the PID performance. However, a neural-network (NN) based approach is being developed to improve the PID performance

by combining high-level information, such as log-likelihood values from each sub-detector and measured momentum of the particle track, into a classification variable for particle identification. In this machine learning based technique, the weights used to combine PID information are not static values, as in the definition of the weighted likelihood (Eq. 3.12). The inputs to the NN are also the charge of the track and its kinematics, i.e. p, $\cos\theta$, and ϕ. Therefore, the combination of the log-likelihoods depends on tracking information. The weights of the connections between the nodes of the network (including input, output and hidden nodes) are free parameters of the network and are determined during the training process. The NN returns the classification variables for the kaon and pion hypotheses, $C_{\mathrm{NN}}(K)$ and $C_{\mathrm{NN}}(\pi)$, respectively. A particle can be identified as kaon (pion) if $C_{\mathrm{NN}}(K)$ ($C_{\mathrm{NN}}(\pi)$) is above a given threshold cut, this threshold can be chosen according to the required efficiency and mis-identification rate. Overall, in early studies, the NN PID performs significantly better than the pure likelihood based PID. The prospects are allows to extend this approach in the future to the classification of all six particle species. This would result in a new common approach for hadron and lepton identification.
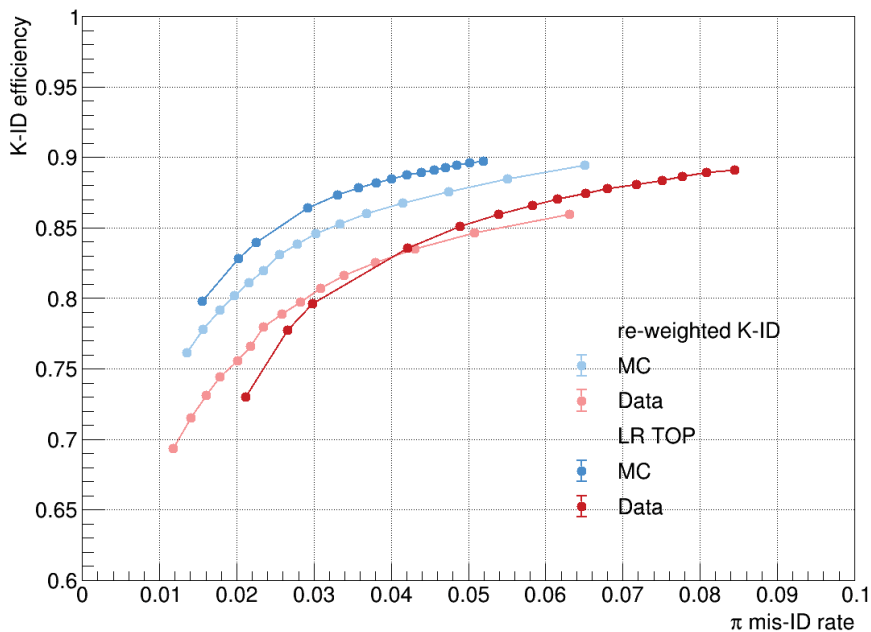

Figure 7.11: ROC curves for kaon identification: K-ID efficiencies vs. π mis-ID rates for different PID criteria. Comparison of the ROC curves for the identification based on the binary likelihood ratio for the TOP and re-weighted PID selections.

7.3 Analysis by PID selection on the binary likelihood ratio for the CDC

The CDC provides information for particle identification by measuring the particle energy loss by ionization ($\mathrm{d}E/\mathrm{d}x$). Low-momentum tracks, which do not reach the particle identification devices, can be identified using the CDC alone. Since we want to investigate the possible effects to K and π identification due to the main tracking device, we build the binary likelihood ratios for the CDC (Eq. 3.18)

$$\mathrm{P}^{\mathrm{CDC}}_{K/\pi} = \frac{\mathcal{L}^{\mathrm{CDC}}_K}{\mathcal{L}^{\mathrm{CDC}}_K + \mathcal{L}^{\mathrm{CDC}}_\pi} \qquad\qquad \mathrm{P}^{\mathrm{CDC}}_{\pi/K} = 1 - \mathrm{P}^{\mathrm{CDC}}_{K/\pi} = \frac{\mathcal{L}^{\mathrm{CDC}}_\pi}{\mathcal{L}^{\mathrm{CDC}}_K + \mathcal{L}^{\mathrm{CDC}}_\pi}$$

and we calculate the identification efficiencies as the ratio of the number of signal events after and before applying PID selection criteria to these PID discriminators. We follow the same fitting strategy outlined in section 6.1.

We compare the results obtained by applying PID selection to the binary LR for the CDC and those for global PID selection. Tab. 7.4 reports the chosen cut on $\mathrm{P}^{\mathrm{CDC}}$ and the results of K-ID performance averaged over the kinematic distribution: efficiencies, fake rates and charge asymmetries for both MC and real data. Tab. 7.5 reports the same for the π-ID performance with the CDC. In the last column of both tables we report the agreement λ between the charge asymmetries for simulated and experimental data obtained by considering only the PID information from the CDC: we observe good agreement in all cases ($\lambda < 1$).

In the first table we observe good agreement between the identification efficiencies of kaons with the CDC alone and those with standard PID, the agreement is also good for the charge asymmetries of the efficiencies. However, the fake rates are very high (above 40% for both MC and real data). Here, it will be important to evaluate the dependence of the PID performance on kinematic variables (especially momentum) for the sub-detector. Although the mis-ID rates are very different depending on the PID variables, the agreement between the charge asymmetries of the fake rates for the two variables is quite good.

In the case of the π-ID performance, the total uncertainty of the efficiencies is very large (of the order of percent), the largest contribution is given by the systematic uncertainty. Again, the fake rates are very high (almost 30%) if only PID information from the CDC is considered: pion identification with the CDC detector alone does not perform well. Anyway, we observe overall good agreement between the charge asymmetries measured with the two different discrimination probabilities between pions and kaons.

In the following subsections we report the PID performance with binary LR for the CDC as a function of the two kinematic variables: momentum p and $\cos\theta$. We divide the tracks according to the binning chosen in section 5.3. We measure the charge asymmetries for both MC and real data and we evaluate their agreement. We also compare the standard PID performance with that of the CDC alone.

Afterwards, we study the identification performance, separately for kaons and pions, as a function of the threshold cut applied to the LR for the CDC counter: $\mathrm{P}^{\mathrm{CDC}} > \alpha$. We determine the average identification efficiencies and average mis-identification rates by varying α, each efficiency-fake rate pair defines a point of the ROC curve.

Table 7.4: Comparison of the K-ID performance by applying selection criteria to standard probability and binary likelihood ratio for the CDC. The uncertainties are the sum in quadrature of statistical and systematic errors. λ quantifies the agreement between the charge asymmetries for the PID selection on $\mathrm{P}^{\mathrm{CDC}}_{K/\pi}$, in MC and real data.

| | MC15ri | | Data | | |
	global P_K	LR $\mathrm{P}^{\mathrm{CDC}}_{K/\pi}$	global P_K	LR $\mathrm{P}^{\mathrm{CDC}}_{K/\pi}$	λ
Cut	0.5	0.46	0.5	0.47	-
ε_{K^+} (%)	86.2 ± 0.8	85.9 ± 0.6	84.2 ± 0.8	84.3 ± 0.6	-
ε_{K^-} (%)	85.7 ± 0.6	85.8 ± 0.5	83.0 ± 0.6	83.2 ± 0.5	-
$\bar{\varepsilon}_K$ (%)	86.0 ± 0.5	85.9 ± 0.4	83.6 ± 0.5	83.7 ± 0.4	-
$\mathcal{A}(\varepsilon)$	0.003 ± 0.006	0.001 ± 0.005	0.007 ± 0.006	0.006 ± 0.004	0.9
f_{π^+} (%)	5.6 ± 0.4	45 ± 5	5.3 ± 0.3	42 ± 5	-
f_{π^-} (%)	5.7 ± 0.5	44 ± 4	6.5 ± 0.5	42 ± 4	-
$\bar{f}_\pi$ (%)	5.6 ± 0.3	44 ± 3	5.9 ± 0.3	42 ± 3	-
$\mathcal{A}(f)$	-0.02 ± 0.05	0.02 ± 0.08	-0.11 ± 0.05	0.00 ± 0.08	0.1

Table 7.5: Comparison of the π-ID performance by applying selection criteria to standard probability and binary likelihood ratio for the CDC. The uncertainties are the sum in quadrature of statistical and systematic errors. λ quantifies the agreement between the charge asymmetries for the PID selection on $\mathrm{P}^{\mathrm{CDC}}_{\pi/K}$, in MC and real data.

| | MC15ri | | Data | | |
	global P_π	LR $\mathrm{P}^{\mathrm{CDC}}_{\pi/K}$	global P_π	LR $\mathrm{P}^{\mathrm{CDC}}_{\pi/K}$	λ
Cut	0.5	0.48	0.5	0.49	-
ε_{π^+} (%)	77.1 ± 0.7	76 ± 2	74.3 ± 0.7	74 ± 2	-
ε_{π^-} (%)	76.5 ± 0.8	77 ± 2	73.4 ± 0.7	74 ± 1	-
$\bar{\varepsilon}_\pi$ (%)	76.8 ± 0.5	76 ± 1	73.8 ± 0.5	74 ± 1	-
$\mathcal{A}(\varepsilon)$	0.004 ± 0.007	-0.01 ± 0.01	0.006 ± 0.007	0.00 ± 0.01	0.1
f_{K^+} (%)	3.70 ± 0.04	29.8 ± 0.8	4.39 ± 0.04	27.9 ± 0.8	-
f_{K^-} (%)	4.0 ± 0.4	30 ± 3	4.3 ± 0.3	29 ± 2	-
$\bar{f}_K$ (%)	3.8 ± 0.2	30 ± 1	4.3 ± 0.1	28 ± 1	-
$\mathcal{A}(f)$	-0.04 ± 0.05	0.00 ± 0.04	0.01 ± 0.03	-0.02 ± 0.04	0.3

7.3.1 K-ID performance and charge asymmetries in bins of kinematic variables

Fig. 7.12 shows the kaon ID efficiencies and pion mis-ID rates (positive charge left, negative charge right) as a function of momentum, Fig. 7.13 shows the charge asymmetries of the efficiencies (left) and the charge asymmetries of the fake rates (right) for both MC (blue triangles) and real data (red triangles). Fig. 7.14 and Fig. 7.15 show the ID observables and charge asymmetries as a function of $\cos\theta$, respectively.
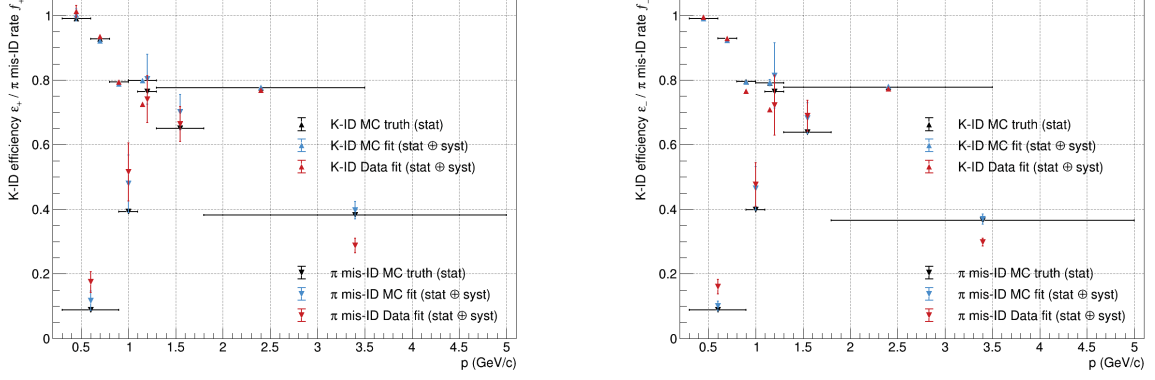

Figure 7.12: Kaon ID efficiency and pion mis-ID rate for the PID selection on $P_{K/\pi}^{CDC}$, in bins of momentum: positively charged particles (left) and negatively charged particles (right).
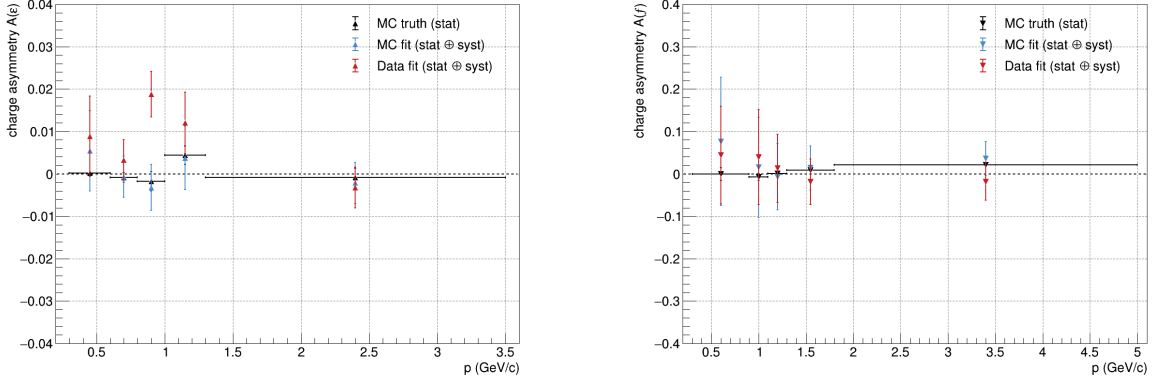

Figure 7.13: Charge asymmetries for the PID selection on $P_{K/\pi}^{CDC}$, in bins of momentum: asymmetry of efficiencies (left) and asymmetry of fake rates (right).
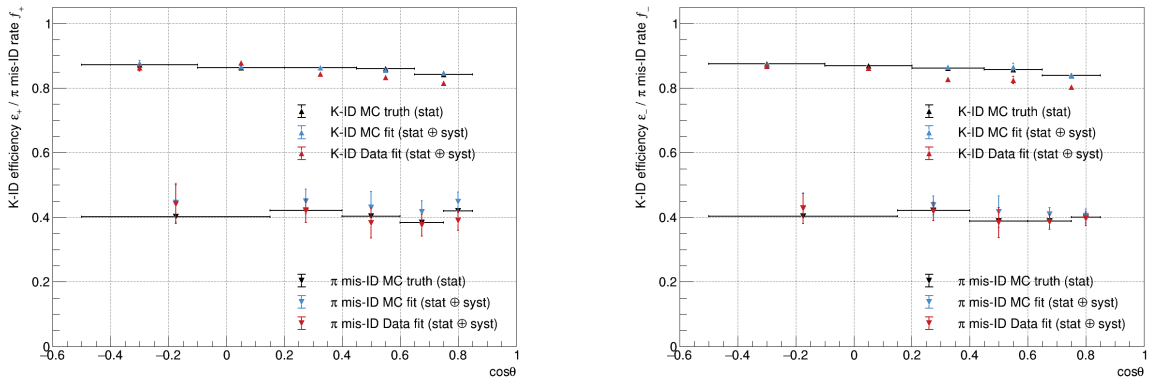

Figure 7.14: Kaon ID efficiency and pion mis-ID rate for the PID selection on $P_{K/\pi}^{CDC}$, in bins of $\cos\theta$: positively charged particles (left) and negatively charged particles (right).
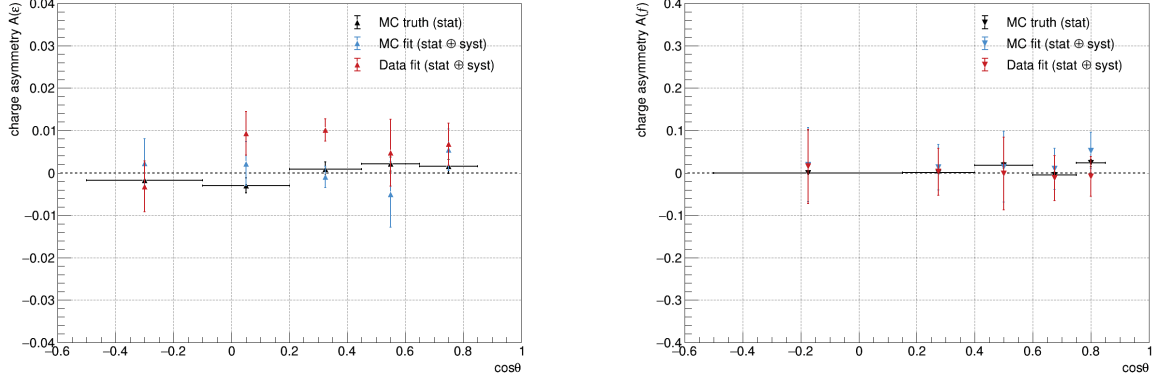
Figure 7.15: Charge asymmetries for the PID selection on $P_{K/\pi}^{CDC}$, in bins of $\cos\theta$: asymmetry of efficiencies (left) and asymmetry of fake rates (right).

We observe that both efficiencies and mis-ID rates based on identification with the CDC alone depend strongly on the momentum of the tracks: for low momenta there is high ID efficiency (close to 100%) and relatively low fake rate (about 10%), instead for momentum $p > 1$ GeV/c the K/π separation power is very poor. For $p \sim 1.2$ GeV/c, the π mis-ID rates are over 60%. This result is due to the dependence of the particle's energy loss on momentum. As can be seen in Fig. 2.7, the dE/dx measurement helps to separate low momentum K/π very well. However, for momentum around 1.2 GeV/c the intersection of the dE/dx curves as a function of p and the limited resolution of the CDC do not allow proper discrimination of kaons and pions. PID information from other devices compensates for this effect. The real data do not exhibit charge asymmetries in the identification of kaons that are significantly larger than those from the simulation (only in a momentum bin the error bars do not overlap). The charge asymmetries of both efficiencies and fake rates are very close to zero, the PID performance with the drift chamber shows no significant variation with respect to particle charge. There is good agreement with the asymmetries measured with the global PID (Fig. 5.22).

In the analysis as a function of $\cos\theta$, the K-ID efficiencies for both positive and negative charge are more or less constant, moreover they are very similar to those measured with global PID (Fig. 5.23). However, the fake rates are much higher and thus the PID performance with the CDC only is worse. The charge asymmetries for both MC and real data are quite close to zero and do not differ greatly from each other. Overall, we observe good agreement between the charge asymmetries measured for both PID variables.

7.3.2 π-ID performance and charge asymmetries in bins of kinematic variables

We present the π-ID performance of the CDC in bins of kinematic variables. In Fig. 7.16 and Fig. 7.17
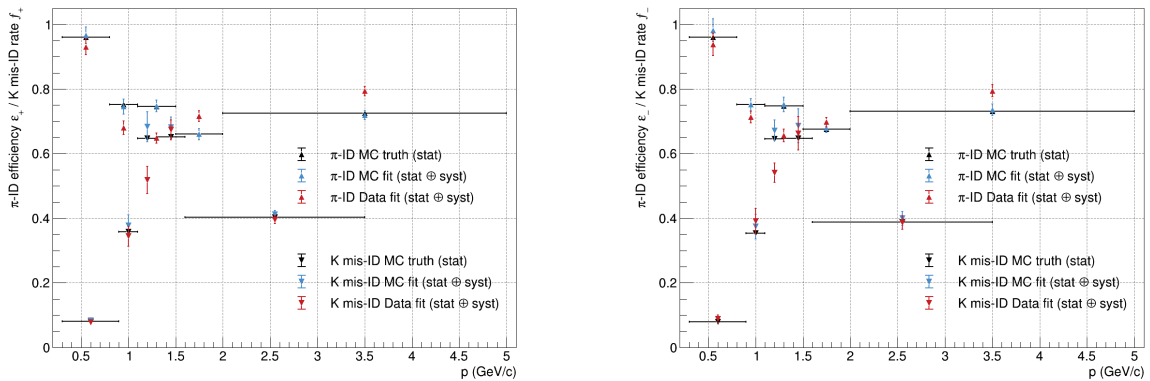

Figure 7.16: Pion ID efficiency and kaon mis-ID rate for the PID selection on $P_{\pi/K}^{CDC}$, in bins of momentum: positively charged particles (left) and negatively charged particles (right).
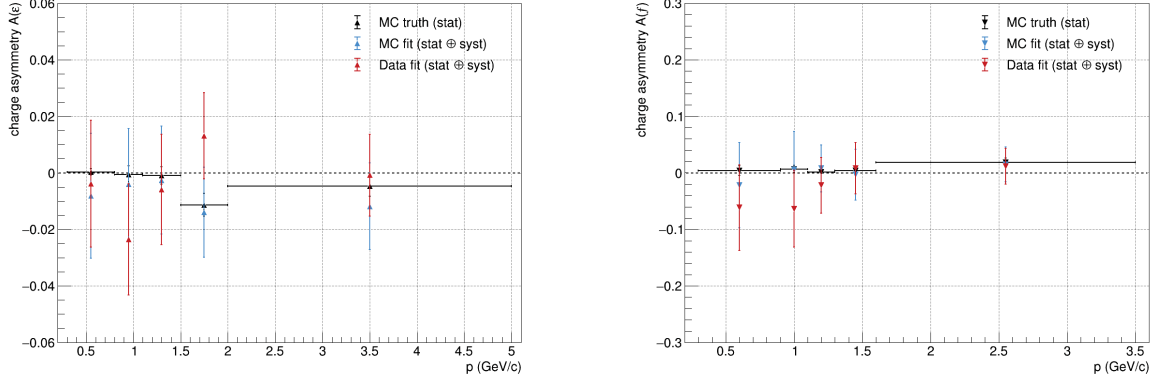
Figure 7.17: Charge asymmetries for the PID selection on $P_{\pi/K}^{CDC}$, in bins of momentum: asymmetry of efficiencies (left) and asymmetry of fake rates (right).
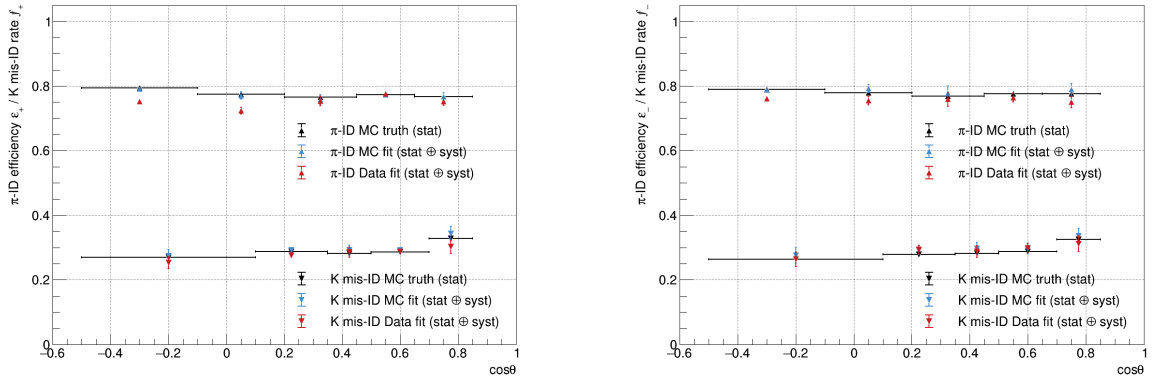

Figure 7.18: Pion ID efficiency and kaon mis-ID rate for the PID selection on $P_{\pi/K}^{CDC}$, in bins of $\cos\theta$: positively charged particles (left) and negatively charged particles (right).
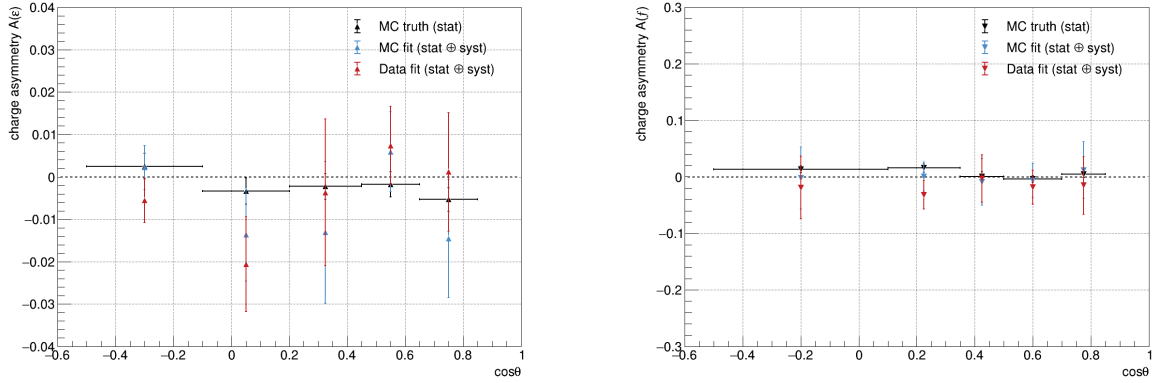

Figure 7.19: Charge asymmetries for the PID selection on $P_{\pi/K}^{CDC}$, in bins of $\cos\theta$: asymmetry of efficiencies (left) and asymmetry of fake rates (right).

we show the ID observables and charge asymmetries as a function of p, respectively. In Fig. 7.18 and Fig. 7.19 we show the ID observables and charge asymmetries as a function of $\cos\theta$. The blue triangles represent the results obtained from the MC, while the red triangles represent those obtained from the experimental data.

As seen above for kaons, also for pions there is good PID performance for low momentum (high efficiency and low fake rate), while for $p > 1$ GeV/c the fake rates are very high and it is not possible to separate pions and kaons well. The PID performance of the CDC as a function of p is very different

from that with the standard PID (Fig. 5.25). We do not observe large differences between the charge asymmetries of efficiencies for simulation and real data. The same is also true for the asymmetries of fake rates as a function of p. There is always good agreement with the charge asymmetries measured with the global PID (Fig. 5.26).

Both π-ID efficiencies and K mis-ID rates as a function of $\cos\theta$ are more or less constant. The efficiencies with the CDC alone are similar to the efficiencies with the global PID (Fig. 5.27), however the fake rates are much higher and, thus, the PID performance is overall less good. The real data do not exhibit charge asymmetries in the identification of pions that are significantly larger than those from the simulation. The agreement between the charge asymmetries measured for both PID variables is overall good (Fig. 5.28).

7.3.3 Identification performance for different PID criteria

To study the K-ID performance of the CDC detector, we build the ROC curves for both MC and real data for the following threshold cuts on the binary LR for the CDC: $P_{K/\pi}^{\mathrm{CDC}} > 0.45$, 0.47, 0.5, 0.52, 0.55, 0.57, 0.6, 0.62, 0.65, 0.67, 0.7, 0.72, 0.75, 0.77, 0.8. We show the ROC curves in Fig. 7.20. In the same figure we show the ROC curves for the standard performance and we can compare the results. It is clear from the comparison that considering all the tracks, with no constraints on momentum, the K/π separation performance with the CDC alone is not good: the ROC curves are shifted toward the diagonal of the efficiency vs. fake rate space. The identification efficiencies for threshold cuts such as to give fake rates between 2% and 6% are low, about 50%. We observe that for not too high cuts on the binary LR, the ID efficiencies with the CDC alone for real data are very similar to those for simulation. In contrast, for higher thresholds, the PID performance on data is slightly worse (for a fixed fake rate, the efficiency for data is slightly lower).

Figure 7.20: ROC curves for kaon identification: K-ID efficiencies vs. π mis-ID rates for different PID criteria. Comparison of the ROC curves for the identification based on the binary likelihood ratio for the CDC and global PID selections.

In Fig. 7.21 we show the ROC curves for the π-ID performance, i.e. the efficiency to identify a pion vs. the rate with which kaons are mis-identified as pions. The performance is evaluated on both MC and real data samples, the applied selection criteria to the binary LR for the CDC are $P_{\pi/K}^{\mathrm{CDC}} > 0.45$, 0.47, 0.49, 0.5, 0.55, 0.6, 0.65, 0.7, 0.75, 0.8. In the same figure we show also the ROC curves for the

standard performance. Even in the case of pion identification, we observe that the curves for the PID with the drift chamber are shifted to the right: for a fixed identification efficiency, the fake rates are much higher. The π-ID efficiencies corresponding to K mis-ID rates below 10% are very low, below 50%. Therefore, the identification performance with PID information from the CDC detector only is not good. We note that the ROC curves for the CDC (simulation and experimental data) are very close, almost overlapping.
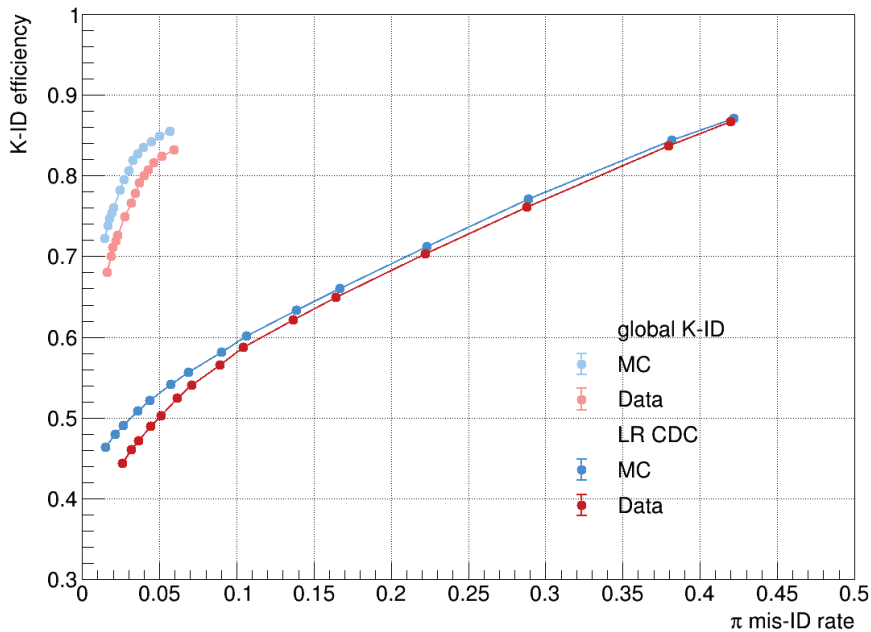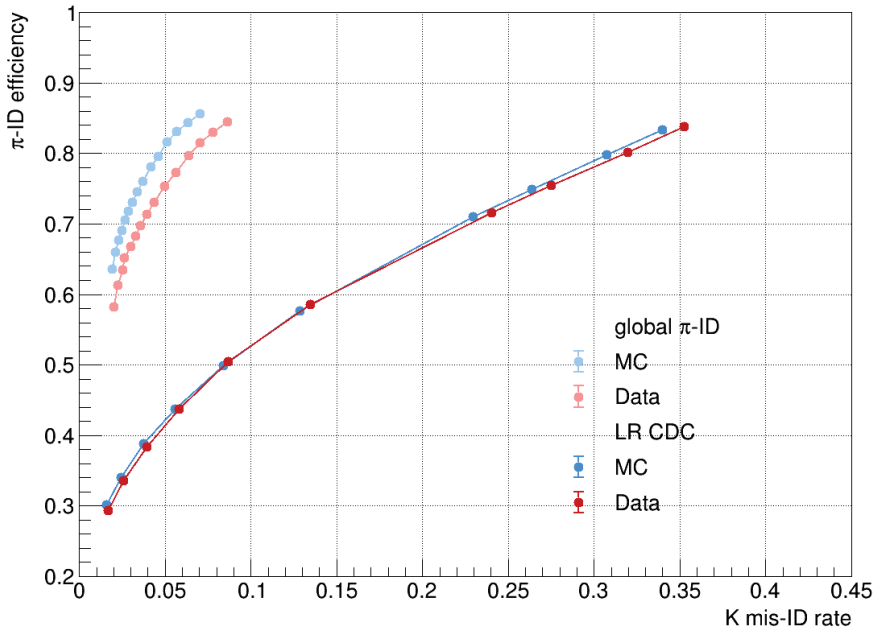

Figure 7.21: ROC curves for kaon identification: π-ID efficiencies vs. K mis-ID rates for different PID criteria. Comparison of the ROC curves for the identification based on the binary likelihood ratio for the CDC and global PID selections.

The CDC detector can provide good identification for particle tracks (both kaons and pions) with low momentum or very high momentum ($p > 2.5$ GeV/c). If we consider the PID performance averaged over the kinematic distribution of the samples of charged kaons and pions, the resulting fake rates are very high and the K/π separation power is poor.

7.4 Summary

The overall performance with the TOP detector alone in the identification of kaons and pions is better than the PID performance with the CDC alone. However, for low momentum tracks the CDC provides important PID information for optimal K/π separation. In contrast, the performance of the TOP alone is better for tracks with high momentum. The PID performance of the CDC is roughly constant as a function of $\cos\theta$, while for the TOP the identification efficiencies are higher for small or large angles in the detector acceptance. Comparison with the standard PID performance shows that, for individual sub-detectors there are improvements only in some regions of phase space where the device under investigation works best. Where the performance of one detector is lacking, PID information from other detectors is essential to achieve good K/π separation performance.

Overall, both for the selections applied to the LR for the TOP and for the selections applied to the LR of the CDC, the real data do not show charge asymmetries in the identification of charged kaons or pions that are significantly larger than those obtained from the simulation.

Chapter 8

Conclusions

The study of matter/antimatter asymmetries in the decays of B and D mesons requires excellent pion and kaon separation capabilities, as well as very firm control of the instrumental asymmetries involved in the detection or reconstruction of positive and negative particles.

In this work, we have studied the performance of the Belle II detector in identifying pions and kaons using the decay sample $D_s^+ \to \phi(K^+K^-)\pi^+$. We have analysed a collision data set of $\int \mathcal{L}\,dt = 404$ fb^{-1} collected between 2019 and 2022, and a MC sample with simulated run-independent beam background ($\int \mathcal{L}\,dt = 200$ fb^{-1}). We have determined the PID efficiencies (mis-ID rates) by fitting the invariant mass distribution of the D_s candidates both before and after applying PID selections, and by considering the ratio of signal yields. We have obtained the ID observables separately for positive and negative charges, and subsequently we have defined the charge asymmetries. Each measurement has been associated with both a statistical uncertainty and a systematic uncertainty due to the model used to fit the resonance distribution. The PID performance has been analysed by dividing the tracks into bins of the main kinematic variables: momentum p and $\cos\theta$.

We have presented the identification performance of charged kaons and pions by applying selection cuts to the global PID probability, which is a combination of all the identification information (likelihoods) from all the sub-detectors that are active in detecting a particle track. In our study, the real data do not exhibit charge asymmetries in the identification of K and π that are significantly larger than those determined from the simulation.

Next, we have examined the identification performance by adding calibration weights to the combination of the likelihoods from the individual sub-detectors, defining the re-weighted PID probability. The PID performance of kaons has improved, but that of pions has degraded. Even after re-weighting the PID probability, no significant discrepancies are observed between the measured charge asymmetries for both MC and real data.

Finally, we have evaluated the PID performance for individual sub-detectors, in particular for the TOP and the CDC. We have measured the efficiencies (fake rates) and charge asymmetries for kaon and pion identification by applying PID selections to the binary likelihood ratios. For low momentum tracks, the K/π separation performance of the CDC is very good, for higher momentum tracks, on the other hand, better performance is achieved with the TOP. No dependencies of note are observed as a function of $\cos\theta$. The analysis of individual sub-detectors shows no significant difference in charge asymmetries between real data and simulation.

Overall, the PID efficiencies as a function of kinematic variables behave as expected. No unknown effects have been observed due to the TOP and CDC sub-detectors. The charge asymmetries measured in the decay sample $D_s^+ \to \phi(K^+K^-)\pi^+$ for MC and experimental data are in agreement within the uncertainties estimated in this work. Since the charge asymmetries in K and π identification measured in control samples are directly applied to extract the physics asymmetries in measurements of CP violation phenomena, further investigation of the K and π PID performance is crucial in order to

achieve a better accuracy in the control of the instrumental asymmetries resulting from differences in interaction or reconstruction probabilities between opposite charge hadrons. A neural network is presently under development to enhance the PID performance at Belle II and improve the separation power between K/π. The novelty lies in the integration of weights in the log-likelihood combination that depend on both p and $\cos\theta$.

Bibliography

[1] A. D. Sakharov. "Violation of CP-invariance, C-asymmetry, and baryon asymmetry of the Universe." *In The Intermissions... Collected Works on Research into the Essentials of Theoretical Physics in Russian Federal Nuclear Center, Arzamas-16* (1998): 84-87.

[2] The ATLAS Collaboration, *et al.* "The ATLAS experiment at the CERN large hadron collider." *Journal of Instrumentation* 3 (2008): S08003.

[3] The CMS Collaboration, *et al.* "The CMS experiment at the CERN LHC." *Journal of Instrumentation* 3 (2008): S08004.

[4] The LHCb Collaboration, *et al.* "The LHCb detector at the LHC." *Journal of Instrumentation*, 3 (2008): S08005.

[5] U. Tamponi, on behalf of the Belle II TOP group. "The TOP counter of Belle II: Status and first results." *Nuclear Instruments and Methods in Physics Research* A 952 (2020).

[6] M. Kobayashi and T. Maskawa. "CP-violation in the renormalizable theory of weak interaction." *Progress of Theoretical Physics* 49.2 (1973): 652-657.

[7] L. Wolfenstein. "Parametrization of the Kobayashi-Maskawa matrix." *Physical Review Letters*, 51.21 (1983): 1945.

[8] R.L. Workman, *et al.* (Particle Data Group). "The Review of Particle Physics (2022)." *Progress of Theoretical and Experimental Physics* 2022 (2022): 083C01.
https://pdg.lbl.gov/.

[9] E. Kou, *et al.* "The Belle II physics book." *Progress of Theoretical and Experimental Physics* 2019.12 (2019): 123C01.

[10] C. S. Wu, *et al.* "Experimental test of parity conservation in beta decay." *Physical Review* 105.4 (1957): 1413.

[11] M. Goldhaber, *et al.* "Helicity of neutrinos." *Physical Review* 109.3 (1958): 1015.

[12] J. H. Christenson, *et al.* "Evidence for the 2π Decay of the K_2^0 Meson." *Physical Review Letters* 13 (1964): 138-140.

[13] K. Abe, *et al.*, Belle Collaboration. "Observation of large CP violation in the neutral B meson system." *Physical Review Letters* 87.9 (2001): 091802.

[14] B. Aubert, *et al.*, BaBar Collaboration. "Observation of CP violation in the B^0 meson system." *Physical Review Letters* 87.9 (2001): 091801.

[15] R. Aaij, *et al.*, LHCb Collaboration. "First Observation of CP Violation in the Decays of B_s^0 Mesons." *Physical Review Letters* 110 (2013): 221601.

[16] R. Aaij, *et al.*, LHCb Collaboration. "Observation of CP violation in charm decays." *Physical Review Letters* 122.21 (2019): 211803.

[17] J. Charles, Jérôme, *et al.* "Isospin analysis of charmless B-meson decays." *The European Physical Journal* C 77 (2017): 1-51.

[18] M. Gronau and J. L. Rosner. "Combining CP asymmetries in $B \to K\pi$ decays." *Physical Review* D 59.11 (1999): 113002.

[19] F. Abudinén, *et al*. "Measurement of the branching fractions and CP asymmetries of $B^+ \to \pi^+\pi^0$ and $B^+ \to K^+\pi^0$ decays in 2019-2021 Belle II data." *arXiv preprint arXiv:2209.05154* (2022).

[20] K. Akai, K. Furukawa, and H. Koiso. "SuperKEKB Collider." *Nuclear Instruments and Methods in Physics Research* A 907 (2018): 188-199.

[21] Y. Arimoto. "SuperKEKB MR status and upgrade during LS1." *43th B2GM* (2022). https://indico.belle2.org/event/7479/contributions/45962/.

[22] https://confluence.desy.de/display/BI/Belle+II+Luminosity.

[23] T. Abe, *et al*. "Belle II technical design report." *arXiv preprint arXiv:1011.0352* (2010).

[24] F. Mueller. "Some aspects of the Pixel Vertex Detector (PXD) at Belle II." *Journal of Instrumentation* 9.10 (2014): C10007.

[25] Y. Uematsu, *et al*. "The Silicon Vertex Detector of the Belle II experiment." *Nuclear Instruments and Methods in Physics Research* A 1033 (2022): 166688.

[26] T. V. Dong, *et al*. "Calibration and alignment of the Belle II central drift chamber." *Nuclear Instruments and Methods in Physics Research* A 930 (2019): 132-141.

[27] Belle II Lepton ID Group. "Muon and electron identification performance with 189 fb^{-1} of Belle II data." In preparation.

[28] E. Torassa, on behalf of the Belle II PID group. "Particle identification with the TOP and ARICH detectors at Belle II." *Nuclear Instruments and Methods in Physics Research* A 824 (2016): 152-155.

[29] I. Adachi, *et al*. "Detectors for extreme luminosity: Belle II." *Nuclear Instruments and Methods in Physics Research* A 907 (2018): 46-59.

[30] Belle II Neutrals Performance Group. "Photon and neutral pion performance in early Belle II data." In preparation.

[31] T. Kuhr, C. Pulvermacher, M. Ritter, T. Hauth, and N. Braun, Belle II Framework Software Group. "The Belle II Core Software." *Computing and Software for Big Science* 3.1 (2019): 1-12.

[32] M. Starič, *et al*. "Likelihood analysis of patterns in a time-of-propagation (TOP) counter." *Nuclear Instruments and Methods in Physics Research* A 595.1 (2008): 252-255.

[33] S. Jadach, B. F. L. Ward, and Z. Wąs. "The precision Monte Carlo event generator KK for two-fermion final states in e^+e^- collisions." *Computer Physics Communications* 130.3 (2000): 260-325.

[34] T. Sjöstrand, *et al*. "An introduction to PYTHIA 8.2." *Computer Physics Communications* 191 (2015): 159-177.

[35] D.J. Lange. "The EvtGen particle decay simulation package." *Nuclear Instruments and Methods in Physics Research* A 462.1-2 (2001): 152-155.

[36] S. Agostinelli, *et al*. "GEANT4-a simulation toolkit." *Nuclear Instruments and Methods in Physics Research* A 506.3 (2003): 250-303.

[37] W. Verkerke and D. P. Kirkby. "The RooFit toolkit for data modeling." *eConf* C0303241 (2003) MOLT007.

[38] F. James and M. Roos. "MINUIT function minimization and error analysis." *CERN program library long writeup* D 506 (1994): 1994.

[39] A. Natochii, *et al.* "Measured and projected beam backgrounds in the Belle II experiment at the SuperKEKB collider." *arXiv preprint arXiv:2302.01566* (2023).

Acknowledgements

Desidero dedicare questo spazio alle persone che hanno contribuito con il loro supporto e la loro presenza alla realizzazione di questo elaborato.

In primis, un ringraziamento speciale al mio relatore Alessandro Gaz, per la tua immensa pazienza, per i tuoi indispensabili consigli, per le conoscenze trasmesse durante tutto il percorso di tesi.

Ringrazio infinitamente la mia famiglia che mi ha sempre amata e sostenuta. Grazie a mia sorella Silvia, a te che sei per me come il sole del mattino, forte e vivo. Grazie per essere sempre accanto a me e per arricchire la mia vita di colori, ti voglio bene *pupi lupi* del mio cuore. Sii felice perché la vita è meravigliosa, la mia con te lo è ancora di più. Grazie a mio papà Fabio, per avermi incoraggiata e sostenuta in tutti questi anni. Caro *papi*, senza di te e i tuoi sacrifici non sarei mai potuta arrivare fino a qui. A mamma Emi, mi manchi tutti i giorni. Eppure, non ti cerco più perché ti trovo sempre. Grazie *mami* per tutto quello che mi hai insegnato. Grazie a mia nonna Adriana, per il tuo immenso amore e per avere sempre accolto i miei pensieri, e spesso i miei silenzi. Grazie a mia zia Giannina, la mitica *zia J*, per esserci sempre nei momenti di bisogno e per tutte le volte che mi hai atteso alla stazione dei treni. Grazie anche a Giliola (*zia Gilly*) per avere sempre creduto in me e nelle mie capacità, e grazie anche perché supporti ancora oggi i miei agguati dietro la porta. Grazie famiglia per riempire la mia vita di gioia.

Un ringraziamento particolare va a Nicola, perché sei parte della mia vita. Grazie per essermi stato accanto nei giorni di gioia e nei giorni più difficili, grazie per il tuo amore e per tutto quello che abbiamo condiviso. Grazie per avermi accompagnata in tutti questi anni di studio, per avere pazientemente ascoltato i miei "sproloqui" e per avere sempre rispettato i miei tempi di studio matto e disperatissimo. Grazie *moe*.

Grazie alla mia amica di sempre Maroua, la mia fidata compagna per formare *La coppia che scoppia!* Grazie perché conosci tutto di me ed io conosco tutto di te da una vita, sei la mia Amica. Grazie per ieri, per oggi e per domani.

Un pensiero è necessariamente rivolto ai miei amici del liceo, a voi che avete assistito alla nascita del mio amore per la Fisica per "colpa" di Zanoni. Grazie Michela, Nicolas, Carlo Alberto, Maria e Nicola perché eravate presenti all'ora e so che ci siete anche oggi. Soprattutto, grazie a Michela (*la Madonna*) per i cinque anni che abbiamo condiviso insieme come compagne di banco, e per gli anni successivi nei quali la nostra amicizia mai è mancata.

Grazie agli amici incontrati in università, ma in modo particolare grazie a Giacomo e Ilaria perché prima siete stati i migliori compagni di laboratorio in assoluto, e poi migliori amici. Siete arrivati in un momento un po' buio per me, e mi avete salvata. Tutti gli attimi trascorsi con voi li porterò sempre nel cuore. Grazie Giacomo per la tua gentilezza, per il tuo umorismo (che io adoro!) e per le tante passioni condivise insieme. Grazie Ilaria perché, oltre all'università, sei stata compagna di "avventure" nella casa in via Benavides. Grazie di cuore per la tua vicinanza, per il tuo affetto, per il tuo spirito tenace. Grazie anche a Daniele perché abbiamo vissuto insieme l'esperienza a Legnaro (e abbiamo odiato insieme la TPC).

Un ringraziamento specialissimo va alle mie amiche del collegio. Grazie ragazze perché insieme a voi ho passato momenti indimenticabili. Grazie per i pranzi e le cene in allegra compagnia, per le

serate film, per le partite a Machiavelli (nelle quali io sono stata la campionessa, seppur discussa), per i puzzle "psichedelici" miseramente schiacciati tra le porte dell'ascensore... Soprattutto, grazie per essere state al mio fianco in questo percorso intenso ed entusiasmante, nel bene e nel male. Grazie Marta, compagna nella stanza 314, per la serenità e l'affetto che mi hai donato, per tutte le chiacchiere e le risate, per i film di Wes Anderson, per le cartoline e per le passeggiate insieme per il centro di Padova. Grazie Angelica per i sorrisi e l'energia che mi hai trasmesso, e grazie perché con te è bellissimo condividere la passione per Harry Potter e i cartoni Disney (soprattutto le canzoni). Grazie Giulia per la tua dolcezza, per il tuo supporto nei momenti più bui, per le nostre colazioni al bar e le nostre confidenze. Grazie perché sei una vera amica ed io ci sarò sempre per te. Un grande grazie a Eleonora perché condividiamo entusiasmo e passioni (con nessun altro avrei mai potuto guardare le serie tv in turco), non conta la distanza fisica quando i cuori sono vicini. Mille mille grazie a Chiara perché nessuno mai mi ha fatta crepare dal ridere o di spavento come te, sei eccezionale baby. Grazie Serena perché da quando ti conosco mi hai sempre incoraggiata e ti auguro di realizzare presto tutti i tuoi sogni. Ringrazio in modo particolare Elisabetta, sei un'amica fidata e cara, sempre pronta ad offrire il proprio aiuto. Le tue parole sono una carezza per la mia anima spesso tormentata, grazie immensamente per tutto il tuo supporto. Grazie a Elena, le nostre strade si dovevano necessariamente incrociare e avrai sempre un posto speciale nel mio cuore.

Grazie alle mie amatissime Madri del collegio. Grazie a Madre Vincenzina, per avermi sempre ascoltata e per essere stata una guida nei primi anni di università. Grazie a Madre Domenica, per il tuo immenso affetto, per la tua dolcezza e per ricordarti sempre di me. Ti porto nel cuore. Ringrazio in modo particolare Madre Zita, per avermi accolta nella vostra casa e per avermi sempre fatto sentire a casa. Grazie per esserti presa cura del mio cuore, sei una amica speciale e ti voglio bene.

Grazie infinite a tutti voi perché siete tantissimi e mi ritengo molto fortunata ad avervi con me.

Con affetto,

<div align="right">Cecilia</div>