Studying the hadron structure with $\overline{\text{P}}$ANDA and CLAS using machine learning techniques

PhD Thesis

II. Physikalisches Institut

Justus-Liebig-Universität Gießen, Germany

Áron Kripkó

Gießen, 2023

Zusammenfassung

Die Hadronenspektroskopie und -struktur sind derzeit sehr aktive Forschungsgebiete, um das nicht-perturbative Regime der Quantenchronodynamik zu untersuchen. Erstere untersucht die komplexe Struktur angeregter Hadronen, indem es ihre Zerfallsprodukte betrachtet, während letztere die Leptonstreuung an Nukleonen nutzt. Beide Methoden erfordern Rekonstruktionsalgorithmen mit hoher Effizienz, guter Teilchenidentifikation und Untergrundunterdrückung. Ziel dieser Arbeit ist es, diese bereitzustellen, indem entweder die bestehenden Methoden verbessert oder neue entwickelt werden.

Der erste Teil der Arbeit präsentiert eine Machbarkeitsstudie eines vorhergesagten hybriden Charmoniumzustands für das $\overline{\mathrm{P}}$ANDA-Experiment. Gitter-QCD-Berechnungen sagen voraus, dass das hybride Charmonium im Grundzustand spin exotisch mit Quantenzahlen von $J^{PC} = 1^{-+}$ bei einer Masse von etwa 4,3 GeV und einer Breite von etwa 20 MeV ist. Ein auf maschinellem lernen basierendes Datenanalyseschema wird vorgeschlagen, um die Signaleffizienz und die Untergrundunterdrückung weiter zu verbessern, zusammen mit Verbesserungen der Analysesoftware (PandaRoot), welche für diese Studie von entscheidender Bedeutung sind. Zu diesen Verbesserungen gehören ein überarbeiteter Clustering-Algorithmus für das elektromagnetische Kalorimeter (EMC) und ein optimiertes Monte-Carlo-Matching für neutrale Teilchen.

Im zweiten Teil dieser Arbeit geht es um die Untersuchung der Protonenstruktur. Eine mehrdimensionale Messung des Strukturfunktionsverhältnisses $\mathrm{F}_{\mathrm{LU}}^{\sin(\phi)}/\mathrm{F}_{\mathrm{UU}}$ wurde für $\mathrm{K}^{\pm}$ basierend auf der Messung von Strahl-Spin-Asymmetrien durchgeführt. Es verwendet Daten mit hoher Statistik, die mit dem CLAS12-Spektrometer am Jefferson Laboratory aufgezeichnet wurden. $\mathrm{F}_{\mathrm{LU}}^{\sin(\phi)}$ ist eine twist-3 Größe, die Auskunft über die Quark-Gluon-Korrelationen im Proton gibt. Diese Arbeit wird zum ersten Mal eine gleichzeitige Analyse von zwei Kaonkanälen über einen großen kinematischen Bereich in z, x_B, P_T und Q^2 mit Virtualitäten Q^2 von 1 GeV^2 bis zu 8 GeV^2 präsentieren wobei von Techniken des maschinellen lernens zur verbesserten Teilchenidentifizierung genutzt werden.

Abstract

The hadron spectroscopy and structure are currently very active fields of research to study the non-perturbative regime of quantum chronodynamics. The first one studies the complex structure of excited hadrons by looking at their decay products, while the latter uses lepton scattering on nucleons. Both methods require reconstruction algorithms with great efficiency and good particle identification and background rejection rates. This work aims to provide these by either improving the existing methods or developing new ones.

The first part of this document presents a feasibility study of a predicted hybrid charmonium state for the $\overline{\text{P}}$ANDA experiment. Lattice QCD calculations predict the ground state hybrid charmonium to be a spin exotic with quantum numbers of $J^{PC} = 1^{-+}$ at a mass of around 4.3 GeV with a width to be around 20 MeV. A machine learning based data analysis scheme is proposed to further improve the signal efficiency and the background reduction, alongside with improvements of the analysis software (PandaRoot), that are vital for this study. These improvements include a reworked clustering algorithm for the electromagnetic calorimeter (EMC) and an optimized monte carlo matching for neutral particles.

The second part of this document is about studying the proton structure. A multidimensional study of the structure function ratio $\text{F}_{\text{LU}}^{\sin(\phi)}/\text{F}_{\text{UU}}$ has been performed for $\text{K}^{\pm}$, based on the measurement of beam-spin asymmetries. It uses the high statistics data recorded with the CLAS12 spectrometer at Jefferson Laboratory. $\text{F}_{\text{LU}}^{\sin(\phi)}$ is a twist-3 quantity that provides information about the quark gluon correlations in the proton. This document will present for the first time a simultaneous analysis of two kaon channels over a large kinematic range of z, x_B, P_T and Q^2 with virtualities Q^2 ranging from 1 GeV^2 up to 8 GeV^2 using machine learning techniques for improved particle identification.

Contents

1 Introduction **4**
1.1 Hadron structure and the standard model . . . 4
1.2 Machine learning methods in hadron physics . . . 5

2 Feasibility study of the lowest lying hybrid charmonium state for $\overline{\text{P}}$ANDA **6**
2.1 Introduction and charmonium hybrid states . . . 6
2.2 The $\overline{\text{P}}$ANDA setup . . . 9
2.2.1 The FAIR accelerator facility . . . 9
2.2.2 The $\overline{\text{P}}$ANDA Detector . . . 10
2.3 The PandaRoot Framework . . . 12
2.3.1 Hit production, digitization and reconstruction for the EMC in PandaRoot . . . 12
2.3.2 Description of the most important Rho classes used for the data analysis . . . 13
2.4 Improvements of PandaRoot . . . 14
2.4.1 Reworked Clustering Algorithm of the EMC . . . 14
2.4.2 Improved Monte Carlo Matching . . . 17
2.5 Simulation . . . 21
2.5.1 Event generation . . . 21
2.5.2 Detector geometry and event reconstruction . . . 21
2.5.3 Single particle reconstruction efficiencies and kinematic distributions . . . 23
2.6 Determining the cuts using genetic algorithm . . . 27
2.7 The reconstructed exotic charmonium . . . 31
2.7.1 Reconstructed signal . . . 31
2.7.2 Reconstruction efficiency and background rejection . . . 32
2.7.3 Comparison with previous studies . . . 32
2.8 Beamtime estimation and conclusion . . . 36

3 A multidimensional study of charged kaon SIDIS beam spin asymmetry over a wide range of kinematics **38**
3.1 Introduction and physics motivation . . . 38
3.1.1 Definition of kinematic variables . . . 38
3.1.2 Deeply inelastic scattering, structure functions and the parton model . . . 39
3.1.3 Semi-inclusive deeply inelastic scattering and transverse momentum dependent functions 39
3.1.4 Beam spin asymmetry and SIDIS structure functions . . . 43
3.2 Experimental setup . . . 45
3.2.1 Accelerators . . . 45
3.2.2 The CLAS12 detector . . . 46
3.2.3 Data acquisition . . . 47
3.3 Preparation of the data set . . . 48
3.3.1 Analyzed data-set . . . 48
3.3.2 Quality checks and cuts . . . 48
3.3.3 Fiducial cuts . . . 48
3.3.4 Event Builder PID . . . 53
3.3.5 PID refinements . . . 54
3.3.6 Total survival rates of the fiducial cuts and PID refinements . . . 57

3.3.7 Selection of the topology . . . 58
3.4 Improving kaon-pion identification with Machine Learning Techniques . . . 59
3.4.1 Monte Carlo Matching . . . 59
3.4.2 Event selection . . . 59
3.4.3 Used variables . . . 60
3.4.4 Root TMVA . . . 61
3.4.5 Method selection and description . . . 61
3.4.6 Checks with simulations . . . 65
3.4.7 Checks with an exclusive channel . . . 66
3.4.8 Checks with the LTCC . . . 67
3.4.9 Checks with the RICH . . . 68
3.4.10 Negative kaons . . . 70
3.4.11 Conclusion on the machine learning PID . . . 72
3.5 Analysis of the Beam Spin Asymmetry . . . 73
3.5.1 Kinematic cuts . . . 73
3.5.2 Available statistics . . . 74
3.5.3 Particle distributions . . . 75
3.5.4 Kinematic distributions . . . 76
3.5.5 Comparison of the used datasets . . . 78
3.5.6 Binning scheme . . . 80
3.5.7 Extraction of the BSA . . . 82
3.6 Subtraction of the pion asymmetries . . . 90
3.6.1 Monte Carlo simulations . . . 90
3.6.2 Results using a more sophisticated RICH reconstruction . . . 96
3.6.3 Pion subtraction . . . 98
3.6.4 Extraction method of the final results . . . 103
3.7 Systematic uncertainties . . . 104
3.7.1 Uncertainty of the beam polarization . . . 104
3.7.2 Effect of the fiducial cuts . . . 106
3.7.3 Contamination of the electron and kaon samples . . . 108
3.7.4 Contamination of the SIDIS sample . . . 111
3.7.5 Acceptance effects . . . 115
3.7.6 Effect of the extraction method and higher order moments . . . 117
3.7.7 Bin-migration effects . . . 121
3.7.8 Radiative effects . . . 124
3.7.9 Effect of the pion subtraction and uncertainty of the contamination . . . 128
3.7.10 Total systematic uncertainty . . . 132
3.8 Final results and discussion . . . 135
3.8.1 Positive Kaons . . . 138
3.8.2 Investigation of the falling behavior at high z . . . 143
3.8.3 Negative Kaons . . . 145

4 Conclusion and summary **150**

Appendix **151**
I Results of the 1 dimensional binning . . . 151
i K^+ . . . 151
ii K^- . . . 153
II Results of the 3 dimensional binning . . . 154
i K^+ with 3 bins in $x_B - Q^2$ . . . 154
ii K^+ with 5 bins in $x_B - Q^2$ . . . 156
iii K^- . . . 158
III Results of the 4 dimensional binning . . . 160
i K^+ . . . 160

ii K^- 164

Bibliography **167**

Chapter 1

Introduction

1.1 Hadron structure and the standard model

The standard model of particle physics describes the known fundamental forces, except gravity and classifies all known elementary particles, although it does not explain the existence of neutrino masses, which are necessary to explain the observed neutrino oscillations, and it also does not explain the matter-antimatter asymmetry in the observable universe. It is a quantum field theory with the most general renormalizable Lagrangian that contains all the particle fields and that observes the global Poincaré symmetry and the local $\mathrm{SU}(3)\times\mathrm{SU}(2)\times\mathrm{U}(1)$ symmetry. It contains the Higgs sector with a complex scalar of group $\mathrm{SU}(2)_L$, the Yukawa sector coupling the Dirac fields (quarks and leptons) to the Higgs via Yukawa coupling, the electroweak sector, which is a Yang-Mills gauge theory with the symmetry group $\mathrm{U}(1)\times\mathrm{SU}(2)_L$ and the quantum chronodynamics (QCD) sector, which is a Yang-Mills gauge theory with the symmetry group SU(3).
QCD describes the interactions between quarks and gluons. Quarks are spinors in n_f flavors in the fundamental representation (triplet) of the SU(3) color gauge group, while gluons are vectors in the adjoint representation (octet) of the SU(3). As the gluons themselves carry color charge, they interact with each other, thus constraining color fields to flux tubes between two color charges. This means that while separating two color charges, at some point it becomes energetically more favorable for a new quark-antiquark pair to appear. In other words, free color charged particles can not be observed in normal conditions, only confined within colorless hadrons. The confining phase can be defined by the behavior of the action of the Wilson loop, which is simply the path in spacetime traced out by a quark-antiquark pair created at one point and annihilated at another point. The confinement scale (Λ_{QCD}) is the scale at which the running strong coupling constant ($\alpha_s(q^2)$) diverges. In terms of the momentum transfer (q), it can be written at one-loop accuracy as the following:

$$\alpha_s(q^2) = \frac{12\pi}{(33-2n_f)\ln\frac{q^2}{\Lambda_{QCD}}} \tag{1.1}$$

This separates the energy scale into two regions. At high energies, where the QCD can be solved perturbatively, with methods used for quantum electrodynamics (QED) calculations, the quarks and the gluons are the relevant degrees of freedom. At low energies, in the non-perturbative regime, the hadrons become the relevant degrees of freedom. The transition between the two regimes is a very interesting topic and is currently an active field of research. [1, 2]
One of the non-perturbative approaches is the lattice QCD. In this case, calculations are performed on a lattice, which is a discrete set of spacetime points, thus naturally regularizing the theory by introducing a momentum cutoff through a finite grid size. The results are then extrapolated towards the continuum limit to extract physical values. Some of these models predict special complex systems and excitations like glueballs (systems of bound gluons), hybrids (mesons with valence gluons) and multiquark systems. To study such complex hadronic systems and excitations, accelerators with a high luminosity and excellent beam quality are needed to get a sufficient production yield. In addition a fast, efficient and precise detector system, consisting of several subdetectors is needed to detect all decay products of these systems. The future $\overline{\mathrm{P}}$ANDA (AntiProton ANihilation at DArmstadt) experiment at the future FAIR (Facility for Antiproton and Ion Research) in Germany will meet these requirements and will be able to confirm or falsify these predictions. [3, 4]

Another possibility in the non-perturbative regime is to use effective theories. They keep the relevant symmetries of the original QCD and use the relevant degrees of freedom to describe a given problem (ie. certain range of energies, reactions, spectra, etc.). For example in case of heavy quarkonium, the system can be described by a non-relativistic potential to a certain extent as the charm and the bottom masses are large compared to Λ_{QCD} (pNRQCD). [3]
The proton spin puzzle also demonstrates the difficulties theoreticians experience due to the non-perturbative nature of QCD. As the EMC collaboration measured [5] only a fraction of the proton spin is carried by the constituent quarks and some of the missing spin is produced by the orbital angular momentum of the quarks [6]. Recent advances in theoretical models have led to a rigorous framework where information on the confined motion of the partons inside a fast moving nucleon is matched to transverse momentum dependent parton distribution functions (TMDs). [7, 8] To be able to further constrain these calculations, high precision fully multidimensional measurements of the 3D structure of the proton are needed. For this purpose beams and targets with high average polarization and detector systems with high acceptance are needed to provide high statistics over a wide range of kinematics. The CLAS12 (CEBAF Large Acceptance Spectrometer for experiments at 12 GeV) at Jefferson Lab satisfies these requirements.
With the help of these measurements, the partons motion and its spatial distribution in the transverse plane, perpendicular to the momentum of the parent hadron can be studied. To describe these distributions, two sets of non-perturbative functions have been introduced. On the one side, Transverse-momentum-dependent parton distributions (TMDs) are used to describe the longitudinal and transverse momentum distributions. On the other side, Generalized Parton Distributions (GPDs) are used to describe the longitudinal momentum distributions in correlation with the transverse positions of the partons. Both TMDs and GPDs can be derived from the most general objects describing the nucleon structure, called Wigner functions. It's Fourier-transform is called the Generalized Transverse Momentum Distribution (GTMD). By taking the forward limit in hadron momentum, one gets the TMDs, while by integrating over the transverse momentum, one gets the GPDs. If one integrates the GPDs over Bjorken x, then one ends up with the fragmentation functions (FFs). By studying the behavior of all of these functions in terms of fractional energy, Bjorken x, transverse momentum and photon virtualities Q^2 in electron scattering experiments, a better picture of the 3D structure of the proton can be achieved.

1.2 Machine learning methods in hadron physics

In both the $\overline{\text{P}}$ANDA and the CLAS experiments there are a lot of optimization problems. They appear during reconstruction and data analysis, where the goal is to achieve the best possible efficiency, resolution, particle identification and background rejection rate. As stated above this is very important to be able to measure exotic states with $\overline{\text{P}}$ANDA and to achieve a clean sample with high statistics to better measure the observables sensitive to the structure of the proton with CLAS. Due to the recent trends in computational science and due to the rapid increase of the computers computational capability the toolkit applicable for optimization problems is broadened with new machine learning tools. The neural networks can now incorporate more hidden layers and the fitting methods used during their training and also for fitting other machine learning methods, like boosted decision trees or genetic algorithms for example, becoming more efficient. This enables these methods to be used for more complex problems and to achieve better results. As most of the problems presented in this work are classification problems in which these machine learning methods excel, methods implementing these tools were developed beside traditional ones.
This work will compare the results achieved by traditional and AI methods developed for data analysis in experimental hadron physics. The first part will present the developments implemented in the analysis software of the $\overline{\text{P}}$ANDA experiment. The improvements mostly focus on the reconstruction and simulation of the neutral particles. The impact of the improvements will be demonstrated through a feasibility study of the predicted lowest lying hybrid charmonium state. The second part will present a multidimensional study of charged kaon semi-inclusive deep inelastic scattering beam spin asymmetry over a wide range of kinematics with CLAS. This enables a better measurement of the observables sensitive to the structure of the proton. As a clean particle identification and a good reconstruction efficiency is a very important aspect of this analysis, the main focus of this part will be the kaon-pion identification and the estimation of the pion contamination in the kaon sample. Throughout the whole work the $c = 1$ (speed of light in vacuum) convention is used for simplicity. Also the angles are measured in radians, except stated otherwise.

Chapter 2

Feasibility study of the lowest lying hybrid charmonium state for $\overline{\mathrm{P}}$ANDA

2.1 Introduction and charmonium hybrid states

This chapter presents a feasibility study of a predicted hybrid charmonium state for the $\overline{\mathrm{P}}$ANDA experiment. It also presents a proposed machine learning based data analysis scheme to further improve the signal efficiency and the background reduction, alongside with improvements of the analysis software (PandaRoot), that are vital for this study.
One of the main goals of $\overline{\mathrm{P}}$ANDA [9] is the detailed investigation of the spectrum of charmonia and charmonium like hybrids [10]. The main observables are the masses, widths, quantum numbers and decay properties. Compared to the light meson spectrum, the charmonium spectrum is well separated (shown in Figure 2.1), providing a clean environment to search for non-conventional states. By using proton-antiproton collisions all non-exotic quantum numbers can be accessed in formation reactions, unlike using electrons and positrons. States with exotic J^{PC} can be made in associated formation processes: $p\bar{p} \rightarrow mh$, where m is a light meson (π, η) and h is a charmonium hybrid.[11]

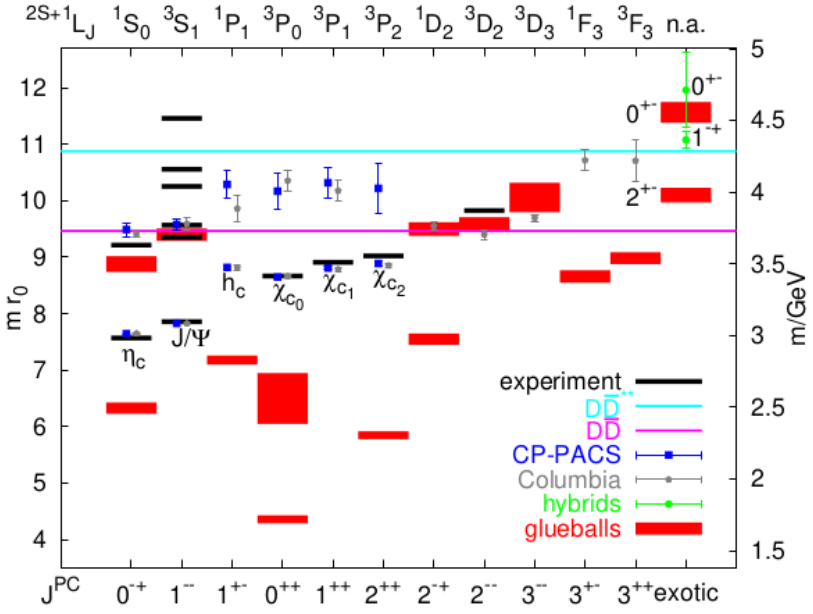


Figure 2.1: Charmonium spectrum from lattice QCD based on [3]. [9]

Lattice QCD calculations predict the ground state hybrid charmonium to be a spin exotic with quantum numbers of $J^{PC} = 1^{-+}$ at a mass of around 4.3 GeV (∼4.15-4.43 GeV) [3,12–14]. Its Breit-Wigner width (Γ) is expected to be small (around 20 MeV/c^2) due to the dynamical suppression of its decay into open charm [3,15,16]. In the following this state will be called $\tilde{\eta_{c1}}$. Since it has a low production cross section and most of its decay channels contain many neutral particles in the final state, its feasibility study provides a great study case for testing machine learning techniques for optimizing the physics analysis procedure. It can also serve as a benchmark for improving the reconstruction and simulation software package of the $\overline{\text{P}}$ANDA experiment.
Based on flux-tube models, the lowest excited gluonic field with angular momentum L=1 and negative CP couples to the heavy charm spins. This gives eight different hybrid states. Out of these states the $J^{PC} = 1^{-+}$ has the lowest mass. [3] For this study, the mass of the charmonium hybrid was chosen to be 4.29 GeV, the average of the smallest and largest predictions with relatively small uncertainties.
Table 2.1 shows the possible decay channels of $\tilde{\eta_{c1}}$ predicted by [17]. In the flux-tube model a symmetry argument gives the selection rule that the $\tilde{\eta_{c1}}$ cannot decay into two states with the same non-relativistic structure and zero internal angular momentum. This gives the $\bar{D}_1^0 D^0$ as the first open charm channel with a threshold at 4286 MeV. In this case the width and the decay modes of the hybrid depends sensitively on the mass. For a low mass hybrid the width could be as narrow as 20 MeV. [15]
With open charm decay forbidden or at least suppressed a slow decay into hidden charm is most probable. Assuming a non-relativistic decay (i.e. conservation of the heavy quark spin and angular momentum) the daughter charmonium has quantum numbers $J^{PC} = 1^{++}$, corresponding to the χ_{c1}. This state is reached by emission of light hadrons where scalars are preferred to isoscalars. The lightest scalar alternative is a two pion state in a relative s-wave. The χ_{c1} can then decay electromagnetically to a J/Ψ and a photon. Then finally the J/Ψ decays into a lepton pair. [15] The decay channel was studied with this lepton pair being an electron or a muon pair. As the reconstruction efficiency was found to be better for the muon case, this was used for further studies. Beside this decay channel, 3 other possible decay channels were investigated: $\tilde{\eta_{c1}} \to \chi_{c1}\eta$, $\tilde{\eta_{c1}} \to J/\Psi\Phi$ and $\tilde{\eta_{c1}} \to J/\Psi\omega$. The complete decay trees are shown if Figures 2.2 and 2.3.

Open charm	Hidden charm	Light hadrons
D^*D	$\chi_{c\{1,2,3\}}(\pi\pi)_S$	$a_{\{0,1,2\}}a_{\{0,1,2\}}$
D^*D^*	$\eta_c\{f_{\{1,2\}}, \eta^{(\prime)}\}$	$f_{\{0,1,2\}}f_{\{0,1,2\}}$
	$\chi_{c\{1,2\}}\eta$	$\{\rho,\gamma\}\{\rho, b_1\}$
	$\{h_c, J/\Psi\}\}\{\omega, h_1, \Phi, \gamma\}$	$\{\omega, h_1, \Phi, \gamma\}\{\omega, h_1, \Phi, \gamma\}$

Table 2.1: Predicted decay channels of $\tilde{\eta_{c1}}$ [17]

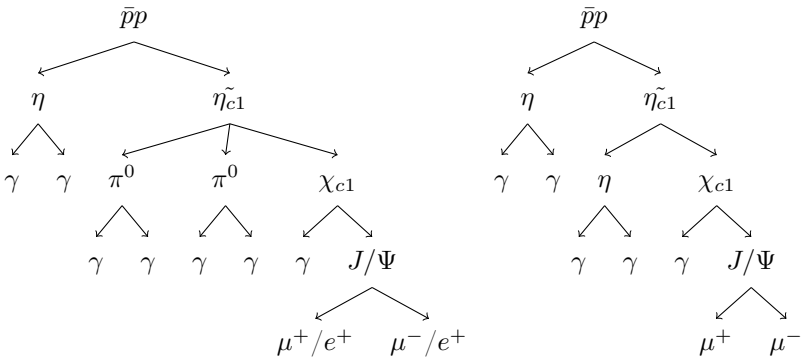


Figure 2.2: Possible decay channels of $\tilde{\eta_{c1}}$ by decaying to χ_{c1} and 2 π^0 (left) or η (right) investigated in this study

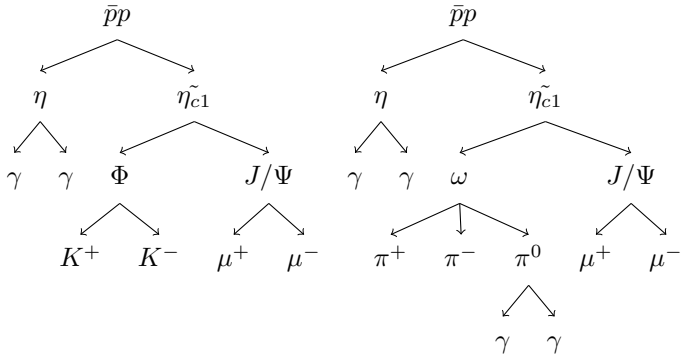


Figure 2.3: Possible decay channels of $\tilde{\eta_{c1}}$ by decaying to J/Ψ and Φ or ω investigated in this study

In proton-antiproton annihilations all non-exotic quantum numbers are directly accessed whereas spin-exotics can be produced only in association with other particles. The future $\overline{\text{P}}$ANDA experiment at GSI-FAIR will let an antiproton beam (p_{lab} =1.5-15 GeV) impinge on a dense proton target. The $\tilde{\eta_{c1}}$ could at p_{lab} =15 GeV be produced in association with a π^0 or an η, as $p\bar{p}$ is a superposition of I=0 and I=1, thus both isospins are allowed. [15] Previous studies [11, 15] showed that both cases are very similar in terms of expected cross-section and reconstruction efficiency. Since the reconstruction efficiency of the channels with η were slightly better, these channels were used during the studies.

2.2 The $\overline{\mathrm{P}}$ANDA setup

The $\overline{\mathrm{P}}$ANDA (Anti-Proton Annihilations at Darmstadt) experiment will be a part of the future FAIR (Facility for Anti-Proton and Ion Research), an extension of the existing GSI (Gesellschaft für Schwerionenforschung) Helmholtz Centre for Heavy Ion Research at Darmstadt, Germany. The layout of the existing GSI and of the planned FAIR facilities are shown in Figure 2.4.

2.2.1 The FAIR accelerator facility

Protons will be accelerated to 70 MeV using a new linear accelerator and will be injected into the currently existing SIS18 synchrotron, which will accelerate them to 2 GeV. The SIS100 synchrotron will then accelerate the protons to 29 GeV. The beam will then be directed into a nickel antiproton production target. The antiprotons created here will be collected and pre-cooled in the Collector Ring (CR). Finally the antiprotons will be transferred to the High Energy Storage Ring (HESR) through the RESR accumulator ring. The HESR is dedicated to supply $\overline{\mathrm{P}}$ANDA with high quality antiproton beams over a wide momentum range of 1.5-15 GeV. Two different operation modes are planned: the high resolution (HR) and high luminosity (HL) modes. These modes are detailed in Table 2.2. [18]

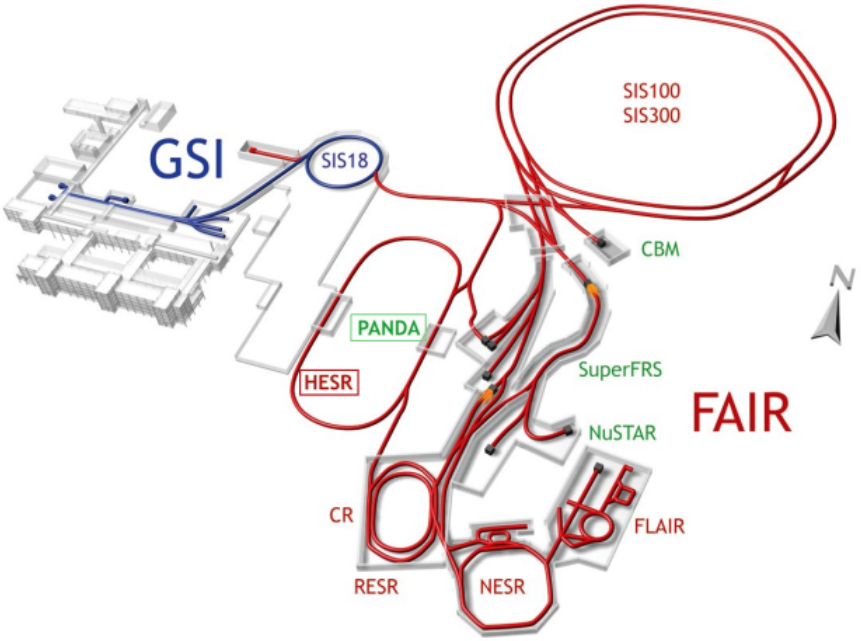


Figure 2.4: Layout of the existing GSI (blue) and of the planned FAIR (red) facilities. Some of the planned experiments are also sown in green. The $\overline{\mathrm{P}}$ANDA experiment and the HESR are highlighted. [18]

Operation mode	High resoultion (HR)	High luminosity (HL)
Peak luminosity	$2 \cdot 10^{31} \frac{1}{cm^2 s}$ for 10^{10} $\bar{p}$	$2 \cdot 10^{32} \frac{1}{cm^2 s}$ for 10^{11} $\bar{p}$
RMS momentum spread	$\sigma_p/p \leq 4 \cdot 10^{-5}$	$\sigma_p/p \sim 10^{-4}$
Momentum range	$1.5 - 9$ GeV	$1.5 - 15$ GeV

Table 2.2: Operation modes of the High Energy Storage Ring (HESR)

2.2.2 The $\overline{\mathrm{P}}$ANDA Detector

The main objectives of the design of the $\overline{\mathrm{P}}$ANDA experiment are to achieve 4π acceptance, high resolution for tracking, particle identification and calorimetry, high rate capabilities and a versatile readout and event selection. To obtain a good momentum resolution the detector will be composed of two magnetic spectrometers: the Target Spectrometer (TS), based on a superconducting solenoid magnet surrounding the interaction point, which will be used to measure at large angles and the Forward Spectrometer (FS), based on a dipole magnet, for small angle tracks. A silicon vertex detector will surround the interaction point. In both spectrometer parts, tracking, charged particle identification, electromagnetic calorimetry and muon identification will be available to allow to detect the complete spectrum of final states relevant for the $\overline{\mathrm{P}}$ANDA physics objectives. [18]

The Target Spectrometer

The Target Spectrometer is shown in Figure 2.5. It will surround the interaction point and measure charged tracks in a solenoidal field of 2T. It will contain detectors in an onion shell configuration, arranged in 3 parts. The barrel covering angles between 22°and 140°, the forward end cap extending the angles down to 5°and 10°in the vertical and horizontal planes, respectively, and the backward end cap covering the region between about 145°and 170°. [18]
3 different target types are foreseen for $\overline{\mathrm{P}}$ANDA: pellet target, cluster jet target and fixed nuclear targets. The first 2 complementary methods are capable of providing sufficient densities with hydrogen or heavier gases at the interaction point. For non-gaseous targets the whole upstream end cap geometry will be modified. [18]
The Micro-Vertex Detector (MVD), the closest to the interaction point, will be used to measure secondary vertices and to improve the transverse momentum resolution of the particles measured by the Central Tracker, which is planned to be a Straw Tube Tracker (STT). These detectors can handle high particle fluxes and have a very good momentum resolution ($\frac{\delta p}{p} \sim 1\%$). In the forward angles 3 Gas Electron Multiplier (GEM) planes will be placed to cover the angles not covered by the STT. The charged particle identification will be done with the aid of the Detection of Internally Reflected Cherenkov light (DIRC) and Time-Of-Flight (TOF) system. [19] The Electromagnetic Calorimeter (EMC) will consist of 11360 (Barrel), 3600 (Forward End Cap) and 562 (Backward End Cap) crystals. The material of the crystals was chosen to be lead tungsten ($PbWO_4$), which is a high density inorganic scintillator with a good energy and time resolution. The crystals will be 20 cm long, which corresponds to approximately to 22 radiation length, which limits the longitudinal leakage. The energy resolution at 1 GeV will be below 2%. To increase the light yield the crystals will be cooled down to -25°C. The outermost detectors are the Muon Filters. [18]

The Forward Spectrometer

The Forward Spectrometer is shown in Figure 2.6. It will measure charged particle tracks in a dipole magnetic field and it will cover vertical angles below 5°and horizontal angles below 10°. [18]
The Forward Tracker (FT) will consist of three pairs of planar tracking stations. In contrast to the EMC, the Forward Electromagnetic Calorimeter is planned to be a Shashlyk-type sampling calorimeter (Shashlyk). These detectors will be accompanied by a Ring Imaging Cherenkov Detector (RICH), a Time-Of-Flight Wall (TOF Wall) and the Forward Muon Detector. The detector at the lowest angle will be the Luminosity Monitor, which will measure the luminosity by detecting elastically scattered antiprotons at very low angles. [18]

Data acquisition

Every sub-detector system is a self-triggering entity. Signals are detected autonomously by the sub-systems and are preprocessed. Only the physically relevant information is extracted and transmitted. This requires hit-detection, noise-suppression and cluster finding at the readout level. The data related to a particle hit, with a substantially reduced rate in the preprocessing step, is marked by a precise time stamp and buffered for further processing. The trigger selection finally occurs in computing nodes which access the buffers via a high-bandwidth network fabric. The new concept provides a high degree of flexibility in the choice of trigger algorithms. It makes trigger conditions available which are outside the capabilities of the standard approach. [18]

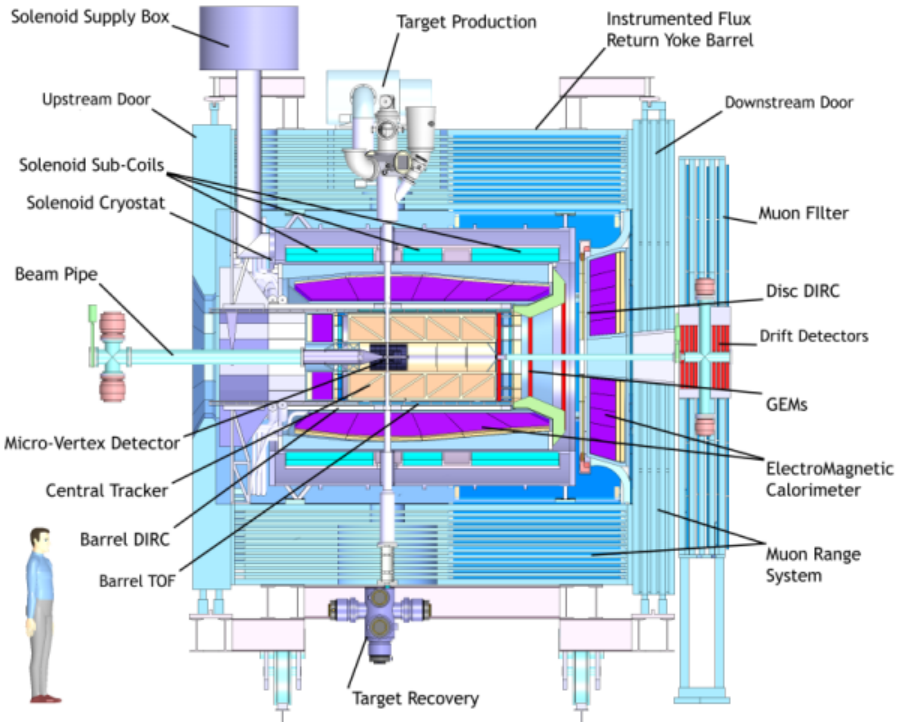


Figure 2.5: Schematic side view of the Target Spectrometer of $\overline{\text{P}}$ANDA [20]

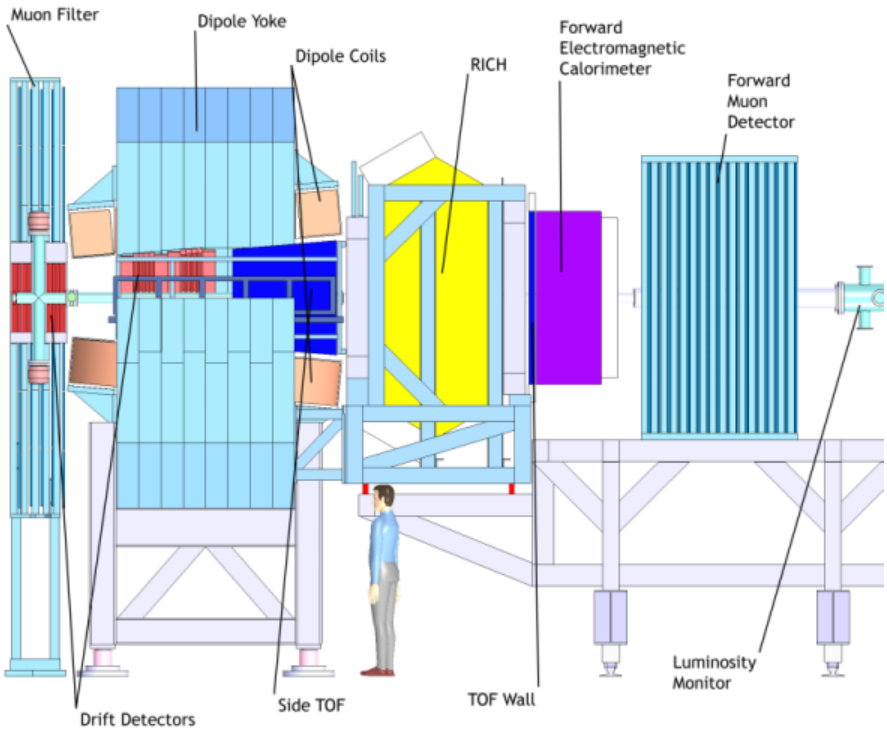


Figure 2.6: Schematic side view of the Forward Spectrometer of $\overline{\text{P}}$ANDA [20]

2.3 The PandaRoot Framework

PandaRoot [21] is the official software framework of the $\overline{\text{P}}$ANDA experiment written mostly in C++. It is based on FairRoot, which is the common software framework for all future FAIR experiments ($\overline{\text{P}}$ANDA, R3B, CBM, ...). It is based on CERN's ROOT and the Virtual MonteCarlo interface for Geant3 and Geant4. The structure of the framework is shown in Figure 2.7. Several event generators are integrated into the PandaRoot framework: EvtGen, DPM, UrQMD, Pythia, Fluka, ... It is used to simulate detector performances, to check the efficiencies of different reconstruction algorithms and to study the feasibility of different physics processes. It contains accurate geometry descriptions of the $\overline{\text{P}}$ANDA detector's different configurations. It has different reconstruction algorithms for tracking and particle identification. In the central tracker a first track fit is performed using a conformal map transformation based on a helix assumption, then the track is used as input for a Kalman Filter (package genfit), using GEANE as track follower. The track is then correlated to the pid detectors to evaluate a global particle identification probability, using a Bayesian approach or multivariate methods. Further implemented packages in PandaRoot are: the analysis tools framework Rho, the kinematic fitter package for vertex and mass constraint fits, and a fast simulation code based upon parameterized detector responses. [21]

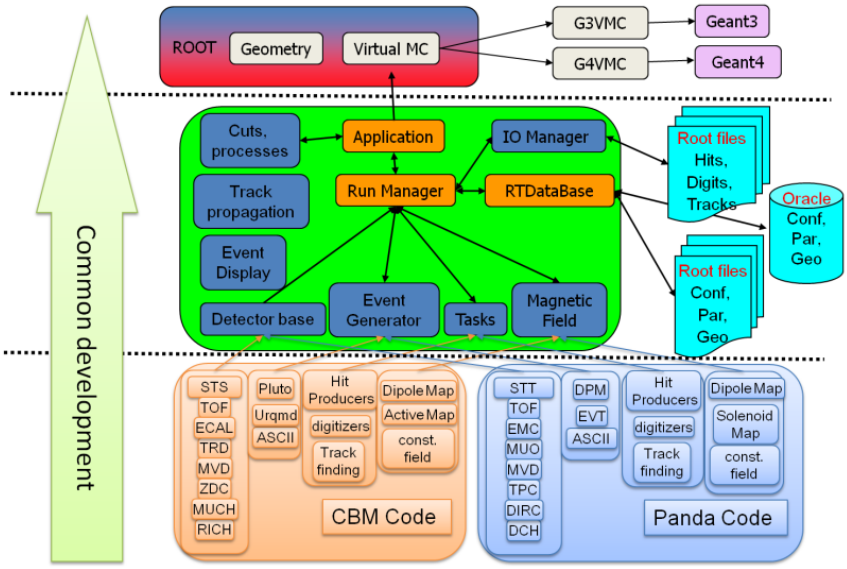


Figure 2.7: Structure of the software framework. At the top level stay ROOT and the Virtual Montecarlo. On a step below, the FairRoot framework manages the general infrastructure with simulation and tasks. Detector specifics and reconstruction stay at the bottom level. [21]

PandaRoot also contains a large collection of macros for various purposes. The most important ones are the macro containing the event generation (with or without an event-filter), the particle transport and the digitization, the macro responsible for the tracking and particle identification and an example skeleton macro for the analysis. These macros also have a version written to be used on the Virgo cluster at GSI. The Virgo cluster is build with the incentive to align its software stack with the cluster computing infrastructure build by a majority of scientific research institutions. The hosting infrastructure is build of components supported by the OpenHPC Community. As operating system CentOS is used. Slurm is used as workload management system and supports the allocation of Linux Containers. All containerized applications use Singularity as container run-time engine. [23]

2.3.1 Hit production, digitization and reconstruction for the EMC in PandaRoot

As explained above, PandaRoot uses CERN's ROOT for creating the geometry of the detectors, for the particle transport and for data storage and access trough interfaces implemented in the FairRoot framework. The hit production, digitization and reconstruction for each detector is implemented in PandaRoot. As the main focus

of this work is the EMC, these processes are described in the following in more details for this case, which is very similar for most of the detectors.
When running a simulation, during the particle transport, the individual energy depositions inside the EMC's active volume can be accessed from PandaRoot as points. These points are then grouped together by the hit producer to create hits. These hits represent energy deposition inside a single crystal. The hits are then digitized during the digitization process. This is either done directly by using smearing functions for a fast simulation or through wave-functions for a realistic digitization. The second method simulates the processes in a calorimeter instead of using simple smearing functions. The created digis have exactly the same data-structure and are fed to exactly the same reconstruction algorithms as the real measured data will. The only exemption is that they also contain the Monte Carlo information. They represent the energy deposition measured in a crystal with a realistic energy and time resolution.
During the reconstruction, the digis are processed by a cluster finding algorithm to create clusters, as usually the showers created by the primary particles span across multiple crystals. The clusters are then split into two or more bumps if they satisfy certain criteria. This procedure is implemented to identify particles with overlapping showers, for example two photons from a decay of a high energetic π^0. These two algorithms are described in much more details in section 2.4.1.
The clusters and bumps are then used as input for the global tracking and PID algorithms. If they can be associated to any charged track, they are assigned to them. From all the remaining ones, the neutral particles are created.
Everything else happens during the analysis, which is managed by an analysis class, also providing a user interface for the analyzer. It creates the list of particles with a given mass hypotheses and according to the requested PID probability per event. It also does the MC matching in case of simulated data. For charged particles the information from the tracking detectors is used. For neutral particles it is based only on the clusters, disregarding the bumps. This is described in more details in section 2.4.2. [22]

2.3.2 Description of the most important Rho classes used for the data analysis

The lists of reconstructed particles created by the analysis class are part of the Rho framework, which is a collection of data containers representing different sets of reconstructed particles and of analysis tools which can be used to manipulate these containers. The basis is the rho candidate, providing data access to all analysis relevant information, including basic properties like PID, mass, charge, etc. and detector information, like Zernike moments of the assigned reconstructed shower in the calorimeter. A collection of these candidates is the rho candidate list. These lists can be filtered based on certain kinematic criteria, like a given mass or energy range or based on PID probabilities, if it is available. The framework also deals with the combinatorics and overlaps when reconstructing decay trees. This is done through a bit marker concept: a unique bit set is created for every reconstructed particle and composite particles get the logical or of all of their daughters. This way it is easy to identify overlapping composite candidates. The MC matching of composite particles is also done at this stage. This is described in more details in section 2.4.2.
Another major part of this framework is the kinematic fitting, which can be used to improve the quality of the reconstructed objects and to reject background. Most of the commonly used fitters are implemented, like mass, vertex and four constraint fits, using a common interface. These fitters fit the reconstructed four momenta of the final state particles so that the masses of narrow resonances or the initial four momenta of the initial $\bar{p}p$ system is set to their known value or that two charged tracks originate from the same vertex. [22]

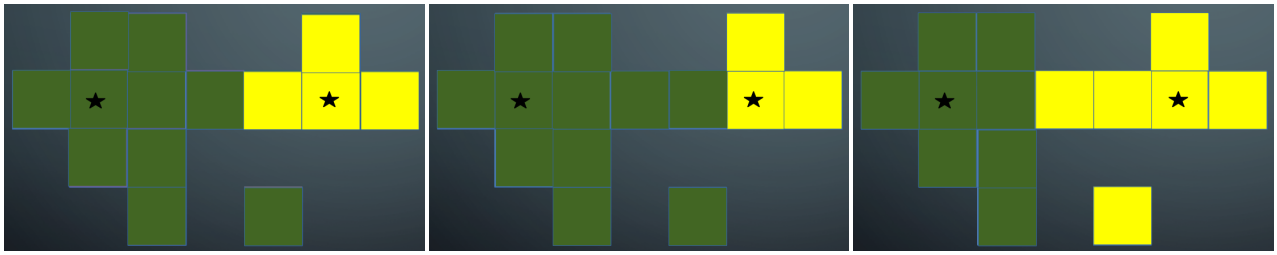
2.4 Improvements of PandaRoot

Since the accurate estimation of the photon detection efficiency is a very crucial part of this analysis, the photon reconstruction and the MonteCarlo matching of PandaRoot were investigated in details and a few improvements were implemented. The clustering algorithm of the EMC was reworked. As a result, the combinatorial background coming from secondary clusters was highly reduced and the energy of the reconstructed photons is closer to the simulated energies. Another important improvement was the improvement of the Monte Carlo matching. For beam-time estimations it is essential to accurately estimate the efficiency by correctly matching the simulated and the reconstructed particles.

2.4.1 Reworked Clustering Algorithm of the EMC

When reconstructing electromagnetic showers, there are some cases, when a secondary maximum is created a few crystals away from the impinging point of the primary photon. These secondaries are called split-offs. Due to the wrong reconstruction, the combinatorial background is high. This complicates the reconstruction of events, especially those which consist of many low energy photons. To improve the reconstruction, 3 new clustering algorithms were developed and tested. The first algorithm is based on the island clustering method, while the second is a cellular automaton. The third algorithm, which is based on the algorithm currently in use, merges low energetic clusters fulfilling certain criteria.
The default algorithm loops over all the digis and if one of its neighbors is in a cluster, it puts it into the cluster. If not, then a new cluster is created. If a digi belongs to two cluster, then those clusters are merged. Only digis above a certain energy threshold are considered. This threshold can be set for the different parts of the calorimeter individually. This threshold is currently 3 MeV for the barrel and 8 MeV for the shashlyk. If the reconstruction is running time-based (i.e. similar to the planned real data scenario, since no trigger will be used), then each cluster has a limited lifetime and only non-expired clusters are considered during the cluster finding. The lifetime of the clusters is set to 20 ns by default. After the clusters are created, they are split by the bump splitter algorithm. If there are multiple local maxima inside a cluster and these maxima have an energy higher then a given threshold (20 MeV) and the ratio of their energy and the total energy of their neighbors is higher than an other threshold, which is also defined in the parameter database: $2.5 + \text{Number of neigbours}/2$. The newly implemented merging merges clusters to the closest ones if they will not be removed by the bump splitter, i.e. the energy of their highest energetic digi is less than 20 MeV. More than 90% of these clusters were actually split offs based on the MC information, in the studied cases. As these clusters could also be hadrons or other neutrals of interest with low energy deposition in the calorimeter, this merging can be easily turned off by the analyzer. After the merging, clusters with a total energy under 30 MeV are removed, as they are also split-offs in most cases and they cannot be assigned to any other cluster reliably.
The island algorithm start with a search of seed digis. The seeds are local maxima with an energy higher than the maximum energy threshold of the bump splitter (20 MeV). The algorithm then continues to add all neighboring digis to the seeds, until a 5% increase in energy or another seed is found by scanning in ϕ and θ. The scan is started with the highest energetic seed and then it is performed for all the other seeds ordered by decreasing energy. The scan is performed in the following pattern for each seed: increasing in ϕ, decreasing in ϕ, increasing in θ and decreasing in θ. For every θ step, a complete ϕ scan is done. As these ϕ scans are completely independent in every θ steps, they run fully parallelized, making this algorithm the fastest one. The remaining unassigned digis are then merged to the closest clusters at the end of the algorithm.
As the island algorithm can introduce an anisotropy due to the scan pattern, the cellular automaton algorithm was developed. It finds the same seeds as the island algorithm and it assigns a cluster ID to each of them. The cellular automaton then loops through all digis randomly and in every iteration each digi "infects" its neighbors with their cluster ID, if it does not have one already. If there aren't any changes during an iteration, then the algorithm stops and the remaining digis are assigned to the closest clusters. As this algorithm requires the most iterations and data access, it is the slowest of them.
Figure 2.8 demonstrates the differences between the algorithms. They show the resulting clusters in different colors in case of an example set of digis. The digis that are above the 1 crystal energy threshold are represented as squares. The ones marked with stars satisfy the bump splitting criteria and also the criteria to be considered as seeds for the island and cellular automaton algorithms. The colors (green and yellow) represent the clusters after the cluster finding. The different assignation of the digis in the middle row illustrates the shortcomings of the algorithms: the anisotropy of the island algorithm, the random assignation of the digis on the borders of

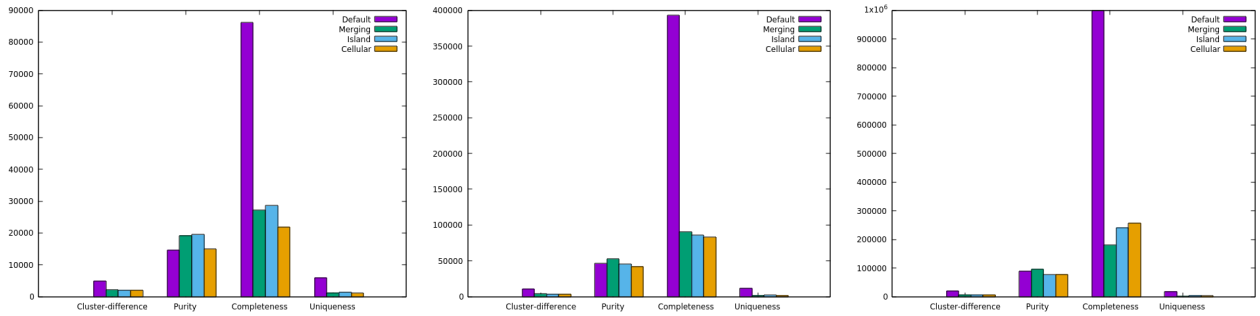
two clusters and the not optimal application of the criteria in case og the merging algorithm. In this example the merged split off is not assigned to the closest cluster, as it is done before the bump splitting. All of these issues remedied by the OnMer algorithm, which will be described later.



Merging algorithm after bump splitting | Island algorithm | Cellular automaton

Figure 2.8: Comparison of the 3 new algorithms in case of an example set of digis. The digis that are above the one crystal energy threshold are represented as squares. The ones marked with stars satisfy the bump splitting criteria and also the criteria to be considered as seeds for the island and cellular automaton algorithms. The colors (green and yellow) represent the reconstructed clusters after the cluster finding.

New measures were defined, which give detailed insight into the capabilities of the algorithms by using the MonteCarlo information. In case of an ideal clustering algorithm if a track created a digi (measured energy deposition in a crystal), then all digis created by this track or its daughter tracks should be in one cluster and that cluster should contain no other digis. By assigning the track ID of the primary track that contributed the most (it and all of its xth granddaughters have the highest number of energy depositions in the given digi) to the digi. A primary track is defined as a track that deposited energy in the calorimeter but none of it's xth grandmothers did. After assigning an ID to every digi, an ID can also be assigned to a cluster by finding the most frequent digi ID in a given cluster. By using the digi and cluster IDs 4 different measures were defined. The first one is the difference between the number of reconstructed clusters and the number of generated photons in a given event. The second is the purity which is the number of digis with other IDs in a given cluster. The sum of the individual cluster purities describe the overall impurity in a given event. The third is the completeness which is the number of digis with the same ID but assigned to any other cluster. The sum of the individual cluster completenesses describe the overall incompleteness in a given event. The last one is the uniqueness which is the number of clusters which have the same ID. Figure 2.9 compares the 3 newly developed methods with the old method using these measures. These studies clearly showed that all 3 new algorithms perform better than the old one. The cellular algorithm performs the best at low energies, but it is the worst at higher ones and it is very slow. The island algorithm, although the fastest one, seems to perform the worst at lower energies and only slightly better than the cellular automaton at high energies. This is due to the fact that it privileges high energetic seeds. Since the performance of the merging algorithm doesn't depend on the photon energy as much as the other algorithms and it is fast, it was chosen to be used as a basis for further improvements.



1 GeV beam energy | 5 GeV beam energy | 15 GeV beam energy

Figure 2.9: Comparison of the 3 new algorithms with the previous one using the 4 measures defined above (smaller is better). The data was obtained from simulations using a photon gun with different energies.

Based on the PhD Thesis of Marcel Tiemens [24] a clustering algorithm developed for online purposes was

developed for offline use as well (online algorithm). It is very similar to the default algorithm, but it uses a different loop structure. First it lists all neighboring digis for each digi. Based on that information, it assigns a pre-cluster number to every digi and then during a second loop it creates the clusters. Due to this double loop structure, it performs less read operations and comparisons, thus it is faster than the original algorithm. It also has better results despite using the same criteria for creating clusters, as they are applied in a different order: the pre-clusters already have to fulfill some criteria, this way less final clusters are created from noise, split-offs, etc. Adding on top the merging to this algorithm (OnMer algorithm) is a further improvement as shown in Figure 2.10. This method is now used as a default in PandaRoot and the results shown later were acquired using this reconstruction algorithm.

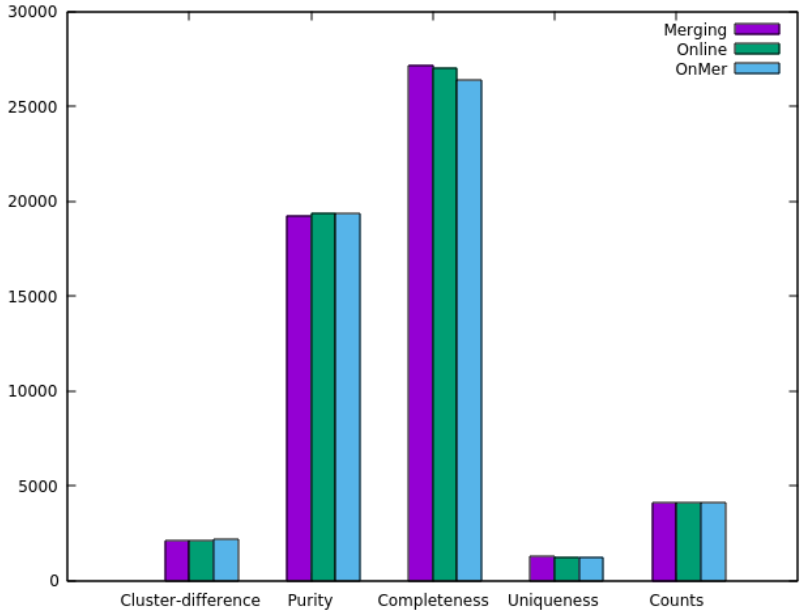


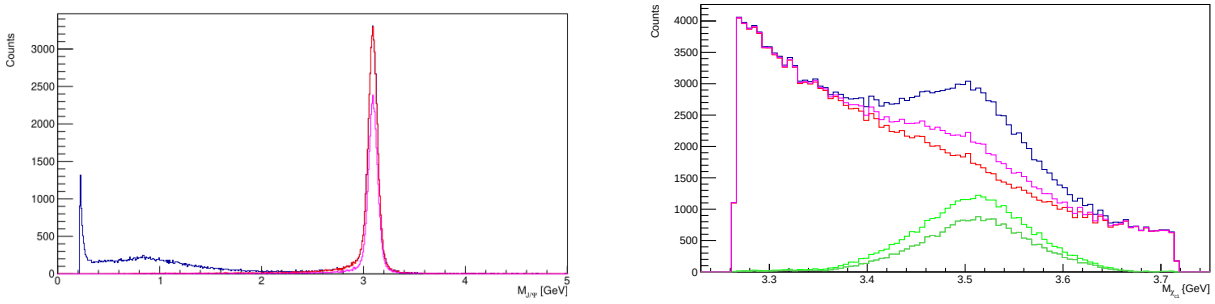
Figure 2.10: Comparison of the 3 best algorithms using the 4 measures defined above (smaller is better). It also compares the reconstructed matched counts in the $\tilde{\eta_{c1}}$ peak. (higher is better) The data was obtained from simulations using the decay channel of $\tilde{\eta_{c1}}$ decaying to χ_{c1} and 2 π^0 with 15 GeV beam energy (see Figure 2.2).

Beside these conventional algorithms, machine learning methods were also tested. Pairs of digis were fed to a deep neural network to decide if they belong to a cluster. The ϕ and θ coordinates and the energies of the two digis were used as input parameters and the output was high if the digis should belong to one cluster. For the training different Monte Carlo files were combined using the photon gun with energies between 0.5 GeV and 15 GeV and with 1-5 photons per event with angles covering 4π. To determine if the two digis belong to the same cluster, the same MC matching method was used as described in the next section, but this time instead of matching the clusters to the generated tracks, the digis were matched to the tracks.
A fully connected deep neural network was created consisting of 4 hidden layers. The first layer consists of 800 neurons, while the hidden layers consists of 400, except the last one, which have only 200. All neurons have selu activation function, except the last one, which have a sigmoid one. The weights were were initialized using the Xavier initialization method. They were optimized using the ADAM optimizer with binary cross-entropy loss function, with a batch size of 8000 and with a learning rate of 10^{-5}. The input variables were normalized to 0-1 interval. The training stopped after 1500 steps, where the training and validation errors started to rise. 5% of the input data was used for validation.
The actual clustering algorithm evaluates the neural network for every digi pairs in a given event, orders the pairs in descending order in respect to the DNN output value and then it loops trough the pairs. If the return value is bigger than 0.5, then the two digis are put into the same cluster.
As this algorithm doesn't provide significantly better results than the OnMer algorithm, but it is much slower and it provides a large overhead the framework, it is not used. Although the fact, that it provides clusters almost identical to the ones created by the OnMer algorithm, provides additional validation for the chosen OnMer method.

2.4.2 Improved Monte Carlo Matching

Matching of composite particles

One of the most crucial points in estimating the beam time is the accurate determination of the reconstruction efficiency. For this we match the reconstructed particles to the generated ones. This is done in 2 steps in case of reconstructed composite particles. First the final state particles which were reconstructed by the detectors are matched. Then based on these matches, the whole decay tree is checked. There are 3 different levels implemented in PandaRoot. The most basic on is only checking the PID of the final particles, which means that the second step is completely missing. The next level checks if the daughter-mother relations are the same in the generated and in the reconstructed trees. In addition to this the third level also checks the PID of all the intermediate states. In both cases the check is done by starting from the top of the generated decay tree and then looping through the daughters at the beginning of every event during the event loop.
An alternative fourth method was developed. In this case we start by getting the matches of the final state particles and then we compare the PIDs of the mothers in both trees. After these checks we check for all sisters if their mctruth mother is the same object. So, in comparison with the previous methods, the loop is done from bottom to top and not from top to bottom. This way also those decay trees are matched which contained additional daughters. In some cases an additional low energetic photon is generated as an additional daughter in the simulation. Because of this an otherwise correctly reconstructed event was not mcmatched. The effect of the new reconstruction is demonstrated by Figure 2.11. On the left plot the invariant mass spectrum of all combined muon pairs are showed in blue. In purple and red the mcmatched J/Ψ are shown with the old and new matching respectively. We can see that the we would lose approximately 10% using the old matching, while the by using the new one, the whole peak is matched. Which contains the number of counts which we expect based on the single muon detection efficiency. On the right, the invariant mass spectrum of the accepted J/Ψ candidates combined with a photon candidate in blue. In purple and red the non-mcmatched χ_{c1} are shown with the old and new matching, while in dark green and green the mcmatched χ_{c1} are shown with the old and new matching respectively. Here we can observe that some signal is left in the background using the old matching, but this is not the case by using the new one.



J/Ψ invariant mass

χ_{c1} invariant mass

Figure 2.11: On the left, the invariant mass spectrum of all combined muon pairs are showed in blue. In purple and red the mcmatched J/Ψ are shown with the old and new matching respectively. On the right, the invariant mass spectrum of the accepted J/Ψ candidates combined with a photon candidate in blue. In purple and red the non-mcmatched χ_{c1} are shown with the old and new matching, while in dark green and green the mcmatched χ_{c1} are shown with the old and new matching, respectively.

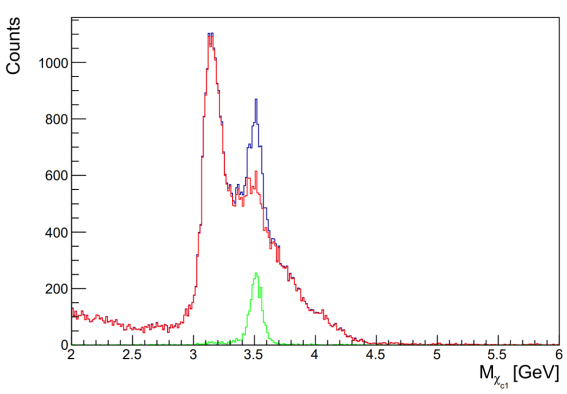
Matchig of final state neutral particles

The matching of the final state particles is very straightforward in case of charged particles. Since they usually create just a few hits in the tracking detectors, it is easy to link the reconstructed particles with the generated tracks. This is not the case when we have neutral particles. They only interact with the calorimeter by creating electromagnetic showers. This means that in the simulation every neutral track can create thousands of other tracks, which could all create hits. In some cases these showers overlap and the hits belonging to two or more different particles are incorrectly put into the same cluster. This usually happens with high energetic pions and

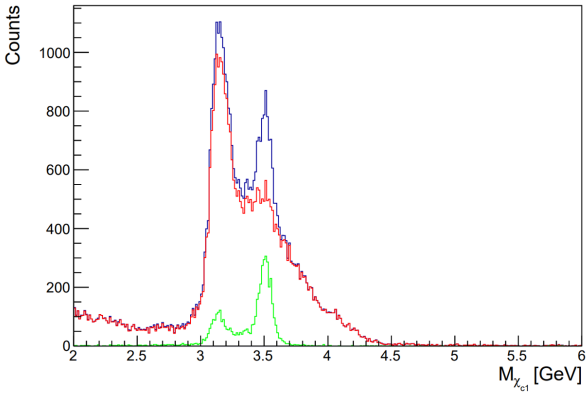
etas decaying into two photons. The bump-splitter algorithm finds most of these merged clusters and splits them into bumps, but this technique is not working perfectly as demonstrated in the previous section. Another challenge for the matching is the pre-shower phenomenon. If the primary particle interacts with the material before the calorimeter, then one of its daughters creates the first hit in the cluster, but the reconstructed properties is in most cases are accurate.
The currently default matching loops trough all reconstructed neutral particles, which are basically the bumps which have no charged tracks associated with them during the reconstruction. Then the matching algorithm falls back to clusters, which means that in case of overlapping photon pairs, the same generated photon is assigned to both reconstructed ones incorrectly. Then the track with the smallest trackID which has a non-zero depleted energy is assigned as the mcmatch. In not completely pure clusters this could also be a problem in some cases. Then to reconstruct the pre-shower events correctly, the mother of the assigned track is assigned if the distance between the first interaction point in the calorimeter and the point where the assigned track was created is smaller than a given distance. Figure 2.12a shows the missing mass spectrum of the χ_{c1} candidates from the decay tree which contain two η-s in blue, the matched ones in green and the non-matched ones in red with this reconstruction.
Working together with Manav Bilakhia, a new matching method was developed and tested for neutral particles. The algorithm starts with the bumps. It loops through all the tracks that contributed to the given bump and then finds the first mother-track that had an energy deposition in anywhere in the calorimeter. As a next step the mother-track which was found most frequently is selected as a match, since we expect that the particle that had the most contribution to a given bump has the largest shower with the most daughters. To consider pre-showers, then the algorithm loops backwards until the first track generated by the eventgenerator. The results are shown in Figure 2.12b. It can be observed that an improvement was achieved at the signal peak, but false matches were also created below it.
This is due to the phenomena that one generated neutral particle sometimes creates multiple bumps. To solve this, the bump with the highest total energy deposit was chosen as a match for the given neutral particle and for all other ones the most frequent mother track was assigned as the mcmatch if it was different from the eventgenerated track. If the latter happened then the track with the smallest track ID in the bump was assigned. If that was still the same as an already assigned ID, then no mctrack was assigned to this bump. Figure 2.12c shows that a great reduction was achieved by this method in the number of the the incorrectly matched neutrals.
We see that a lot of neutrals create more than one bump, so it is a valid question if the assigned bump is a valid match and the reconstructed properties are close to the generated ones. The difference of the relative energy and the polar and azimuth angles were checked. Based on these studies we found that if relative energy difference is small, then the angular differences are also small. The relative energy difference is shown on Figure 2.13 along with a Gaussian fit on the peak around 0. The cut on it was determined to be $3\sigma = 0.007$ GeV. If the difference is bigger than this, then the same fall-backs to previous methods are done as described above. According to our findings, most of these events were due to pair-production just before the calorimeter. And since the energy distribution of the electrons/positrons are asymmetric, using the higher energetic one is actually a good reconstruction of the original photon. Figure 2.12d shows the spectra of with the final matching. It can be observed that the false matches below the signal has disappeared and compared to the original version, more matches are found at the signal peak.
Table 2.3 compares the single photon reconstruction efficiencies obtained at different points in the matching process. We can see that the photon reconstruction efficiency is considerably lower than expected according to the default matching. After the first step the efficiency is two high due the double counting, while the second efficiency also considers photons, which were detected, but due to pre-showers, split-offs or detector inefficiencies only a fraction of their energy was reconstructed.

	Efficiency	Efficiency without double counting
Default	63%	61%
New	117%	89%
New, unambiguous	89%	89%
New, correctly reconstructed	76%	74%

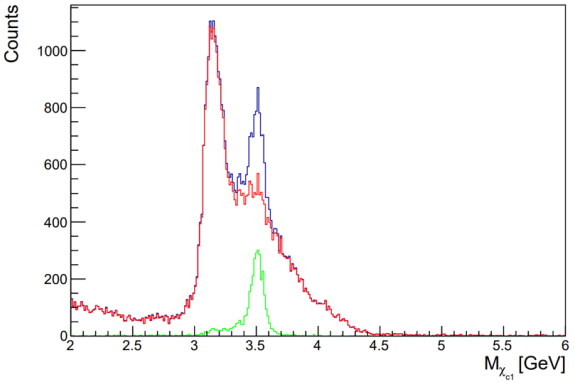
Table 2.3: Reconstruction efficiencies of neutral particles according to different mcmatching versions



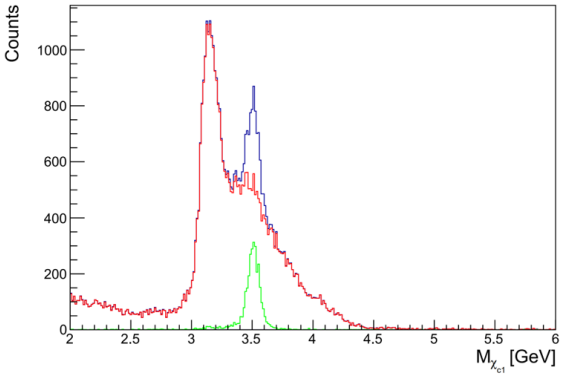
(a) Default MC matching



(b) New MC matching, looping back to the first track



(c) New MC matching, but unambiguously



(d) New MC matching, but only keep good matches

Figure 2.12: The invariant mass spectrum of all χ_{c1} candidates are shown. The matched candidates are shown in green, while the unmatched candidates are shown in red.

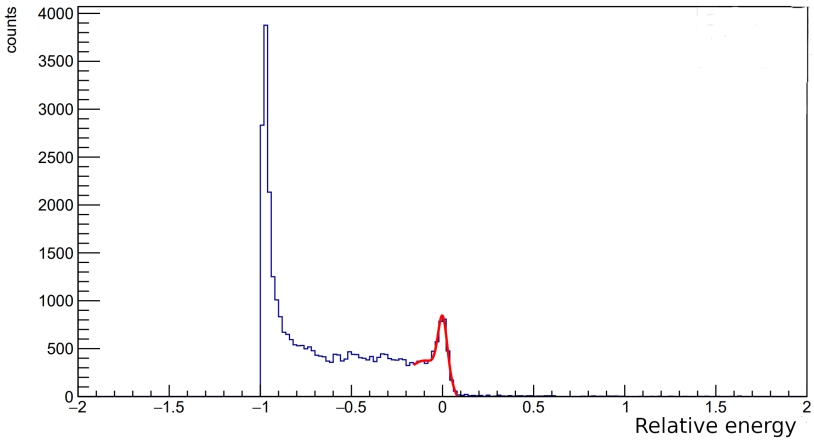
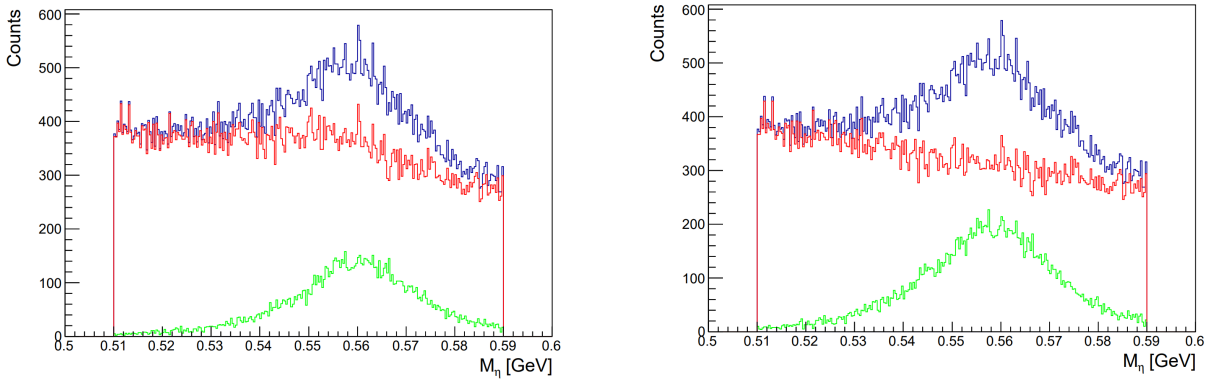


Figure 2.13: Ratio of the difference of the reconstructed and generated and the generated energy of neutral particles ($\frac{E_{rec}-E_{gen}}{E_{gen}}$) and the fitted Gaussian around zero.

Additional checks were done with other decay channels to check the new matching and we found that the performance was better in most cases or the same. To provide an example, Figure 2.14 compares the missing mass spectra of $\eta \to 2\gamma$ in a photon rich environment, which causes a high background. We can observe that some signal remains unmatched by using the old matching method, but with the new one we get a nice and smooth background.



Default MC

New MC matching

Figure 2.14: Invariant mass spectra of $\eta \to 2\gamma$ in a photon rich simulation. All reconstructed candidates are showed in blue, the matched ones in green and the non-matched ones in red.

2.5 Simulation

While all the macros were developed and tested on a local machine, the large scale simulations were performed on the Virgo Slurm Cluster [25]. All software ran within the Debian 10 Virtual Application Environment. The latest production versions of Fairsoft and FairRoot were used, versions apr22 and 18.6.8 respectively, which can be found preinstalled on the cluster. The PandaRoot installation was the latest development version that contains all the new features described in Section 2.4.

2.5.1 Event generation

All four decay channels shown in Figures 2.2 and 2.3 were simulated using the PandaRoot framework. In all cases the beam momentum was chosen to be 15 GeV and the EvtGen [26] event generator was used with Photos on and using the PHSP model for every decay. This means that the final state radiation was included in the simulation and that every decay was calculated according to the phase-space, so no special model was used. Table 2.4 summarizes the number of signal events simulated and the different dedicated background channels for each decay channel. The background channels taken into consideration have a very similar final state and decay structure compared to the signal channel, but they do not contain the $\tilde{\eta_{c1}}$. In case of the decay channel $\tilde{\eta_{c1}} \to 2\pi\chi_{c1}$, 100000 signal events were also generated with the J/Ψ decaying to an electron pair.

Decay channel	$\tilde{\eta_{c1}} \to 2\pi\chi_{c1}$	$\tilde{\eta_{c1}} \to \eta\chi_{c1}$	$\tilde{\eta_{c1}} \to \Phi J/\Psi$	$\tilde{\eta_{c1}} \to \omega J/\Psi$
Number of signal events	100000	100000	100000	100000
Number of background events per channel	1000000	500000	1000000	1000000
Dedicated background channels	$\bar{p}p \to \chi_{c0}\pi^0\pi^0\eta$ $\bar{p}p \to \chi_{c1}\pi^0\pi^0\pi^0\eta$ $\bar{p}p \to \chi_{c1}\pi^0\eta\eta$ $\bar{p}p \to J/\Psi\pi^0\pi^0\pi^0\eta$	$\bar{p}p \to \chi_{c1}\eta\eta$ $\bar{p}p \to J/\psi\eta\eta\pi^0$	$\bar{p}p \to \eta\Phi J/\Psi$	$\bar{p}p \to \eta\omega J/\Psi$

Table 2.4: Number of simulated events and dedicated background channels.

To estimate a general background, the DPM generator [27] was used. 250 million events were simulated using the event filter. Only those events were considered which contained at least 2 charged particles with opposite charges and their combined invariant mass was between 2.5 and 4 GeV with muon mass hypothesis. If these criteria were not met at the generator level, then a new event was generated. The ratio of the number of generated events and the number of events accepted by the filter is 5.1.

2.5.2 Detector geometry and event reconstruction

The default geometry was used during the simulation, which means that all detectors were fully modeled and the detector responses were simulated as realistically as possible with the current software. The recently updated GEM geometry was used for the final simulations. This is important, since the older version overestimated the material budget and thus highly reduced the photon detection efficiency in the kinematic region, where most of the photons can be found in case of these simulations.
During the event reconstruction all reconstructed neutral particles and all reconstructed charged particles, reconstructed with muon mass hypothesis, are considered. So no PID cuts are used. It was found that due to the low reconstruction efficiency and small cross-section of the process combined with the great background rejection efficiency of the kinematic fits and mass cuts used during the analysis, the highest signal to background ratio is achieved by not using any PID algorithm for the muons.
The same procedure is used for the study of the $J/\Psi \to e^+e^-$ case. The only difference there is that the the reconstructed electrons are Bremsstrahlung corrected. [28] Although this correction significantly decreased the width of the J/Ψ peak and increased the reconstruction efficiency, they are still worse for the e^+e^- case. This is illustrated by Figure 2.15. If exactly the same analysis procedure is followed, then the ratio of reconstruction efficiency for the final charmonium hybrid is: $\frac{Eff_{e^+e^-}}{Eff_{\mu^+\mu^-}} = 0.091$. Based on this result, only the $J/\Psi \to \mu^+\mu^-$ is used for any further studies.

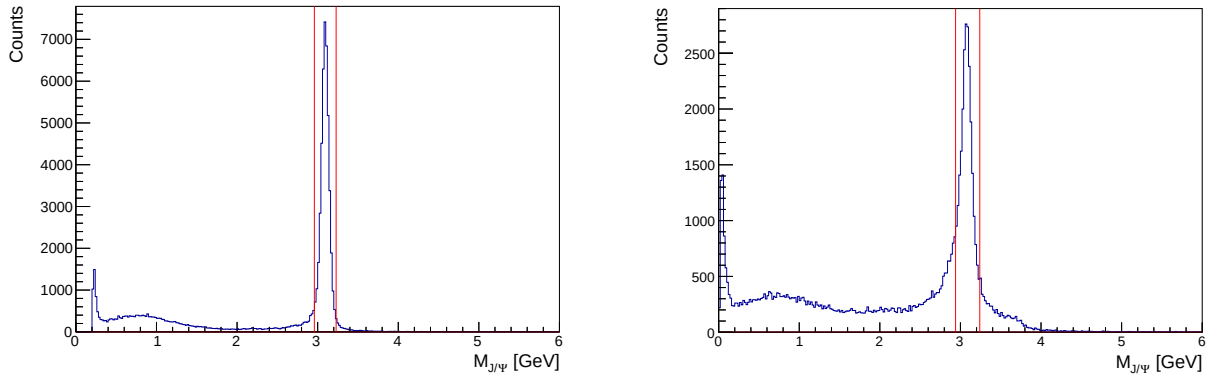


Figure 2.15: Invariant mass distribution of J/Ψ candidates in case of $J/\Psi \to \mu^+\mu^-$ (left) and $J/\Psi \to e^+e^-$ (right) of decay channel $\tilde{\eta_{c1}} \to 2\pi\chi_{c1}$ in blue. The $\pm 1\sigma$ limits are shown in red for the MC matched $J/\Psi \to \mu^+\mu^-$ candidates to highlight the difference in resolution.

By combining all possible charged tracks with opposite charges, with muon mass hypotheses, the J/Ψ candidates are created. Only those candidates are accepted which have an invariant mass around the nominal J/Ψ mass, within a window width of 0.27 GeV ($M\epsilon[M_{J/\Psi} - 0.27/2$ GeV$; M_{J/\Psi} + 0.27/2$ GeV$]$). A vertex fit with a mass constraint on the J/Ψ mass is then performed on these candidates. To create the η candidates 2 photon candidates are combined. Only those candidates are kept, which fall within a window width of 63 MeV around the η mass of 554 MeV. The invariant mass distribution of all combined photon candidates together with the mass cuts to selected the π^0 and the η candidates are shown on Figure 2.16.

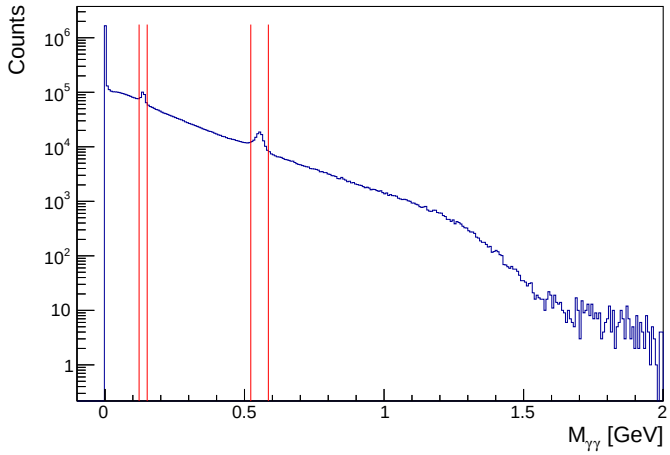


Figure 2.16: Invariant mass distribution of all combined photon candidates in blue and the mass cuts to selected the π^0 and the η candidates in red.

In the following steps all of the intermediate states are reconstructed:

$\tilde{\eta_{c1}} \to 2\pi\chi_{c1}$

In this case the π^0 candidates are created similarly to the η candidates, but the mass cut is placed around 136.9 MeV with a window width of 29 MeV. The J/Ψ candidates are combined with a photon candidate and these combinations are accepted as χ_{c1} candidates if their invariant mass is around 3.49 GeV with a window width of 0.45 GeV. The accepted candidates ($2\pi\chi_{c1}\eta$) are then used to reconstruct the initial $\bar{p}p$ system, which is accepted if its invariant mass is (5.1±0.6) GeV.

$\tilde{\eta_{c1}} \to \eta\chi_{c1}$

Since this decay channel is very similar to the previous one, most of the steps are the same. The only difference is that in this case no π^0 candidates are created and the initial $\bar{p}p$ system is reconstructed by using $\chi_{c1}2\eta$.

$\tilde{\eta_{c1}} \to \Phi J/\Psi$

All charged tracks with opposite charges, with kaon mass hypothesis are combined to create the Φ candidates. They are accepted if their mass is 1.02 GeV±25 MeV. These candidates are then combined with the J/Ψ and η candidates to reconstruct the $\bar{p}p$ system.

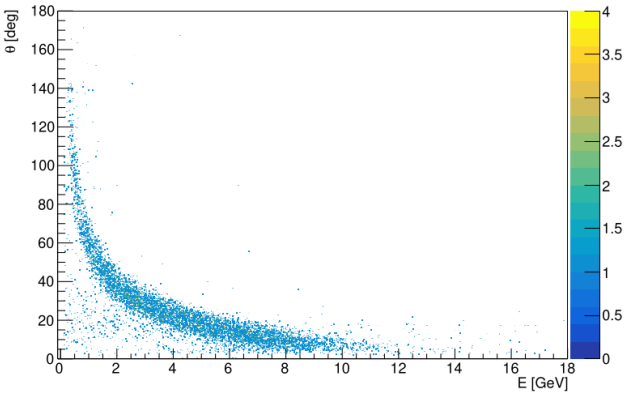
$\tilde{\eta_{c1}} \to \omega J/\Psi$

In this case the Φ from the previous channel is replaced with an ω. They are created by combining a π^0 candidate, which is reconstructed according to the same criteria as in the $\tilde{\eta_{c1}} \to 2\pi\chi_{c1}$ channel, with a positive and a negative charged track with pion mass hypothesis. The ω candidates are accepted if their mass is (783±50) MeV.

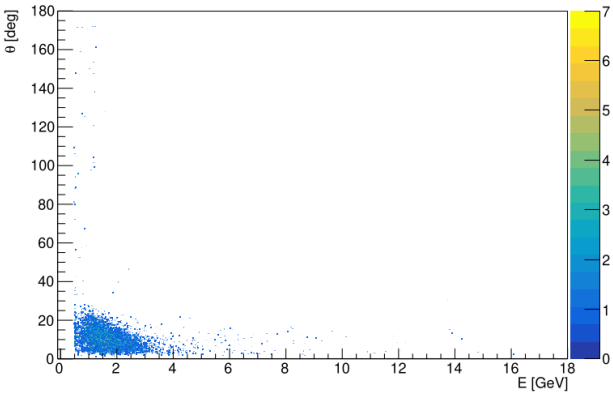
A 4-constraint fit is then performed on the reconstructed $\bar{p}p$ systems by constraining the 4-momentum of the system to its known value. During the fit the masses of the daughter particles are fixed. Also a mass constraint fit is performed on all intermediate states. The cuts applied on the probabilities from the kinematic fits are described in the next section. After all the cuts applied the $\tilde{\eta_{c1}}$ candidates are created.

2.5.3 Single particle reconstruction efficiencies and kinematic distributions

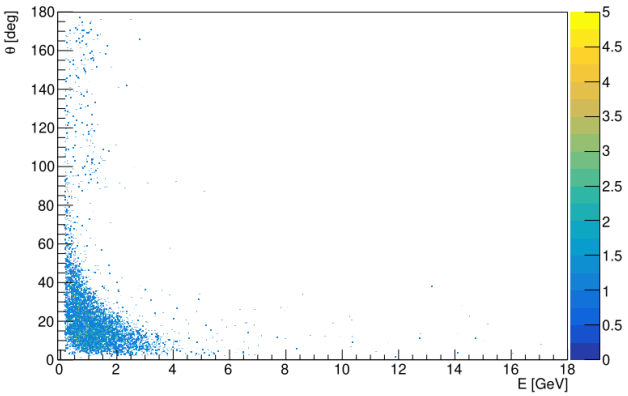
Figure 2.17 compares the 2 dimensional distribution of the generated and reconstructed μ^+-s, π^+-s, K^+-s and γ-s and Figure 2.18 for composite particles. Figure 2.19 shows the reconstruction efficiency as a function of energy, θ and p_T for μ^+-s and γ-s. All charged particles and photons for all decay channels have very similar behavior. In some bins we can see efficiencies higher than 1, which is a result of bin migration. The overall reconstruction efficiency is very close to 1, except of charged particles at low p_T, which is a known shortcoming of the current tracking algorithms and will be fixed in the near future. [29] We can also observe a drop in efficiency for neutral particles around $\theta = 20$ and a lot of photons reconstructed at lower energies in this region. This is caused by the high material budget of the GEM detector. [30] Overall we can conclude that the reconstruction efficiency for most particles are sufficient and agree with the values from the technical design reports using the newest version of PandaRoot.



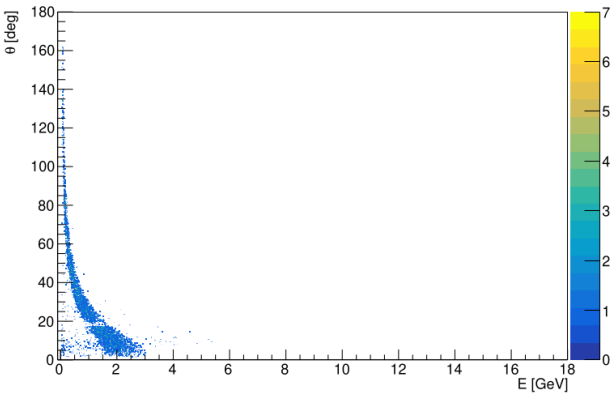
μ^+ from the channel $\eta\chi_{c1}$



K^+ from the channel $\Phi J/\Psi$

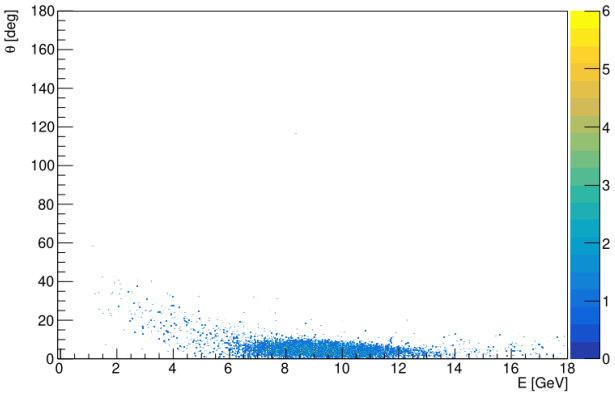


π^+ from the channel $\omega J/\Psi$

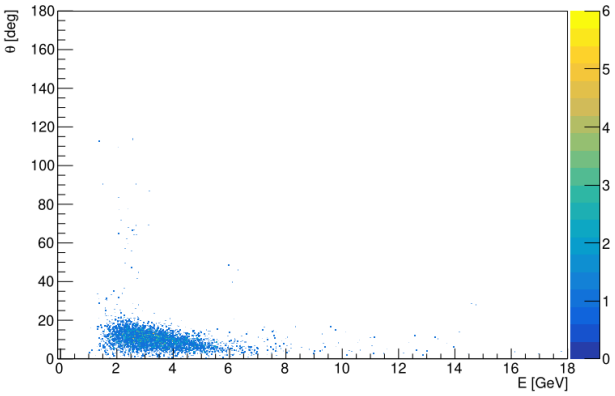


γ from the channel $\eta\chi_{c1}$

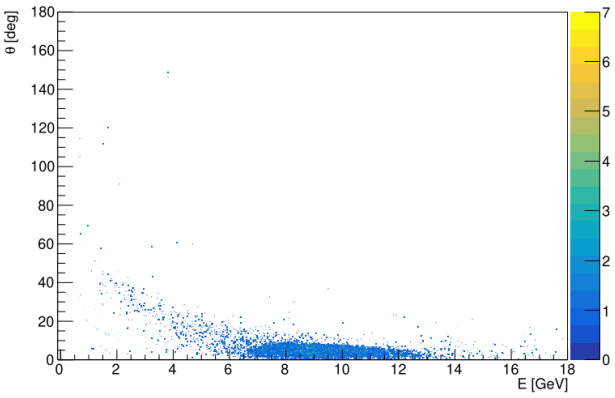
Figure 2.17: Kinematic distribution of single particles



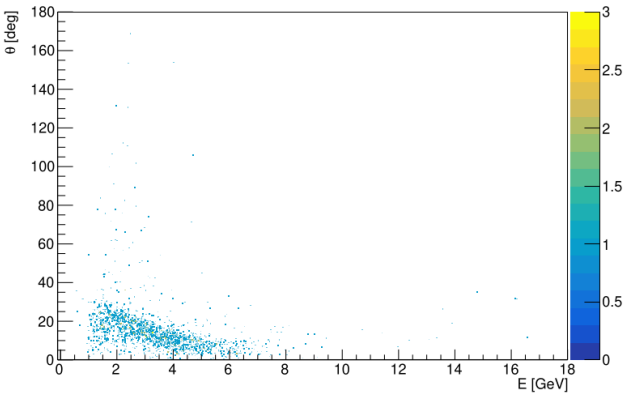
J/Ψ from the channel $\eta\chi_{c1}$



Φ from the channel $\Phi J/\Psi$

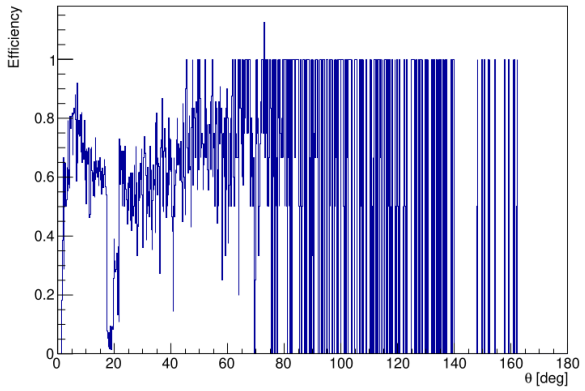


J/Ψ from the channel $\Phi J/\Psi$

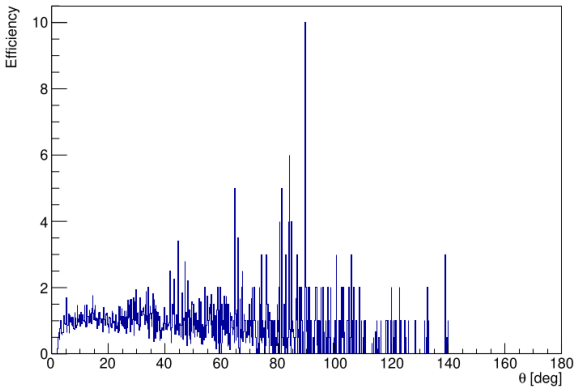


ω from the channel $\omega J/\Psi$

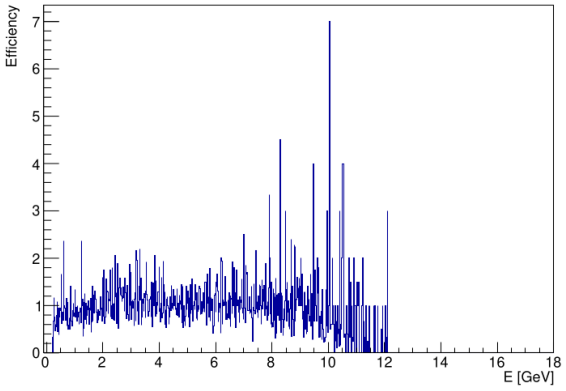
Figure 2.18: Kinematic distribution of reconstructed resonances



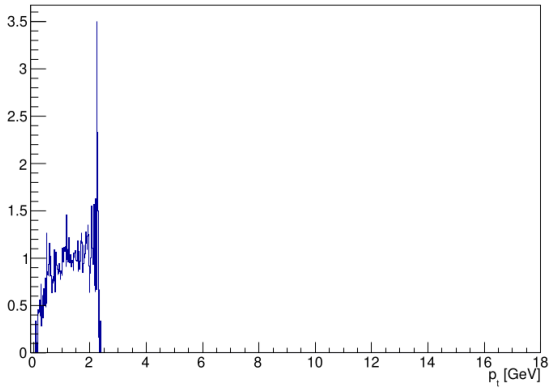
γ from the channel $\eta\chi_{c1}$ as a function of θ



μ^+ from the channel $\eta\chi_{c1}$ as a function of θ



μ^+ from the channel $\eta\chi_{c1}$ as a function of E



μ^+ from the channel $\eta\chi_{c1}$ as a function of p_T

Figure 2.19: Reconstruction efficiencies of single particles

2.6 Determining the cuts using genetic algorithm

Finding the optimal set of cuts is not obvious. Genetic algorithm is ideal for this optimization problem. It is a machine learning algorithm inspired by the process of natural selection. It doesn't use the gradient of the function to be optimized, so it is ideal for problems where the gradient is computationally heavy, unknown or doesn't exist. And also it has the ability to get out of local minima. Although it usually converges slower than a neural network, the results and the whole optimization process is transparent and can be easily interpreted.
During the optimization process a set of parameters is called an individual. A given number of individuals is generated at the beginning of the process with random parameters. As a first step, the fitness of each individual is calculated, which in our case is the signal to background ratio or the significance. After the initialization, the iteration steps start. At the beginning of a step, the best individuals are kept and the others are deleted. In their place new ones are generated by randomly selecting parameters from the individuals kept. Then some randomly selected parameters of the new individuals are randomly changed by a few percent. As a final step the worst individuals are deleted and completely new individuals are created in their place with random parameters. These iterations are running until a certain number is exceeded or there is no significant change in the best fitness.
This algorithm was implemented as a root macro and run on the GSI Linux farm (Kerberos). To speed up the iteration, the random parameters were generated within reasonable limits and one of the individuals was created with parameters optimized by hand. These limits and the parameters of the genetic algorithm are summarized in tables 2.5 and 2.6. Other parameters like other mass cuts and probabilities of different PID algorithms were included in the beginning, but they were removed, since didn't improve the final results, but significantly increased the optimization time. The mass cut parameters were generated according to a real uniform distribution within the given limits, while the probability cut parameters were generated according to an exponential distribution with a rate of $\lambda = 3.5$. The percentage of the mutation was determined according to a gamma distribution ($f(x;\alpha,\beta) = \frac{x^{\alpha-1}e^{-\beta x}\beta^{\alpha}}{\Gamma(\alpha)}$, where Γ is the gamma function: $\Gamma(z) = \int_0^{\infty} t^{z-1}e^{-t}\,dt$) with $\alpha = \beta = 3$. The probability of a positive/negative/no mutation was all 1/3. Based on some initial short runs, the algorithm converged faster with these settings. Since the goal is to maximize both the efficiency and the signal to background at the same time, the statistical significance was maximized, which incorporates both and defined as the following:

$$\text{Significance} = \frac{\text{Number of signal events}}{\sqrt{\text{Number of signal events} + \text{Number of background events}}} \tag{2.1}$$

The optimized parameters are shown in table 2.7. It can be observed, that compared to the reference parameters, the mass cuts are looser and the probability cuts are tighter. Also the mass cuts are not centered around the middle of the peaks, but shifted towards higher masses. This is reasonable since high background is expected under the lower tail, but by extending the cut toward higher masses, the efficiency can be increased without significantly increasing the background levels. This is demonstrated by Figure 2.20. Figure 2.21 compares the probability distribution of the 4C fit of the final $\bar{p}p$ system for the signal and one background channel of decay channel $\tilde{\eta_{c1}} \to \Phi J/\Psi$. It shows that a considerable background is cut away, while most of the signal is kept.

Parameter	Value
Population size	100
Max number of generations	10000
Newly generated individuals	10
Surviving individuals	10
Tournament population	7
η mass upper border min [GeV]	0.54
η mass upper border max [GeV]	0.61
η mass lower border min [GeV]	0.47
η mass lower border max [GeV]	0.54
χ_c mass upper border [GeV]	3.5
χ_c mass upper border max [GeV]	3.7
χ_c mass lower border min [GeV]	3.3
χ_c mass lower border max [GeV]	3.5
Φ mass upper border min [GeV]	1.02
Φ mass upper border max [GeV]	1.045
Φ mass lower border min [GeV]	0.995
Φ mass lower border max [GeV]	1.02
ω mass upper border min [GeV]	0.783
ω mass upper border max [GeV]	1.4
ω mass lower border min [GeV]	0.4
ω mass lower border max [GeV]	0.782
Kinematic fit probability cuts min	0
Kinematic fit probability cuts max	1

Table 2.5: Parameters of the genetic algorithm and the limits of the parameters to be optimised. Only the relevant parameters are used for each channel.

Channel	$\tilde{\eta_{c1}} \to 2\pi\chi_{c1}$	$\tilde{\eta_{c1}} \to \eta\chi_{c1}$	$\tilde{\eta_{c1}} \to \Phi J/\Psi$	$\tilde{\eta_{c1}} \to \omega J/\Psi$
$\eta^{\bar{p}p}$ mass upper border [GeV]	0.565	0.565	0.565	0.565
$\eta^{\bar{p}p}$ mass lower border [GeV]	0.53	0.53	0.53	0.53
$\chi_c/\Phi/\omega$ mass upper border [GeV]	3.53	3.53	1.03	0.87
$\chi_c/\Phi/\omega$ mass lower border [GeV]	3.49	3.49	1.01	0.7
Kinematic fit probability cuts	0.001	0.001	0.001	0.001

Table 2.6: Cuts determined by hand as a comparison and a starting point for the genetic algorithm for each channel

Channel	$\tilde{\eta_{c1}} \to 2\pi\chi_{c1}$	$\tilde{\eta_{c1}} \to \eta\chi_{c1}$	$\tilde{\eta_{c1}} \to \Phi J/\Psi$	$\tilde{\eta_{c1}} \to \omega J/\Psi$
$\eta^{\bar{p}p}$ mass upper border [GeV]	0.590273	0.583325	0.609009	0.576469
$\eta^{\bar{p}p}$ mass lower border [GeV]	0.528604	0.489714	0.472235	0.516278
$\chi_c/\Phi/\omega$ mass upper border [GeV]	3.61931	3.64309	1.04494	0.91273
$\chi_c/\Phi/\omega$ mass lower border [GeV]	3.45014	3.37777	0.995138	0.729029
$\bar{p}p$ 4C fit probability	0.539764	0.604857	0.136921	0.98463
$\chi_c/\Phi/\pi^{0\omega}$ mass constraint fit probability	0.427254	0.273101	2.39025e-07	0.001316
$\eta^{\bar{p}p}/\omega$ mass constraint fit probability	0.06143	0.0140402	-	8.12132e-9
$\pi^{0Eh}/\eta^{\chi_c}/\eta^{\bar{p}p}$ mass constraint fit probability	0.0394484	0.000147619	9.30564e-16	1.8756e-05
π^{0El} mass constraint fit probability	0.0223424	-	-	-

Table 2.7: Cuts optimized by the genetic algorithm for each channel

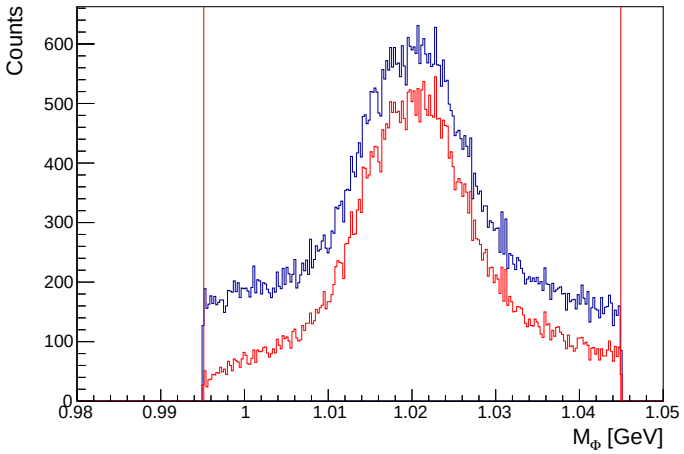


Figure 2.20: Invariant mass spectrum (blue) of Φ during the reconstruction of the decay channel $\tilde{\eta_{c1}} \to \Phi J/\Psi$. The MC matched Φ are shown in red.

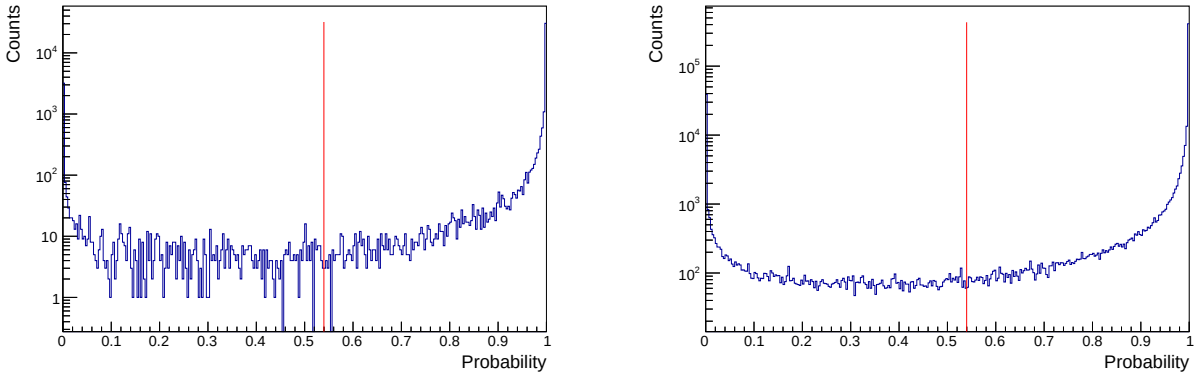


Figure 2.21: Probability distribution of the 4C fit of the final $\bar{p}p$ system for the signal (left) and one background channel (right) of decay channel $\tilde{\eta_{c1}} \to \Phi J/\Psi$ in blue. The cut placed on the probability is shown in red.

Figure 2.22 shows the best and the average fitness per generation during the optimization of the cuts of the decay channel $\tilde{\eta_{c1}} \to \omega J/\Psi$. Since the random introduction of individuals with low fitness causes spikes, the average fitness of 33 generations was also plotted for better visualization. It can be observed that the average fitness is increasing slowly during the whole process and the best fitness saturates after the 500th generation. This means that the maximum number of generations is enough to find the optimal set of cuts.

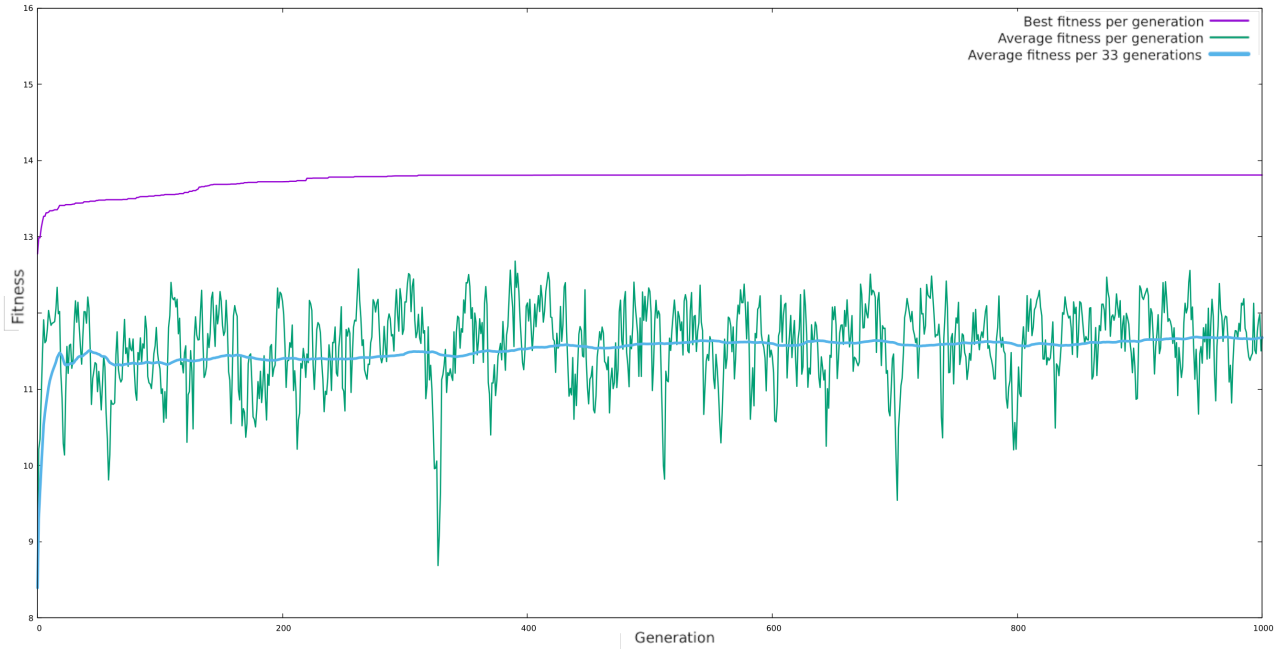


Figure 2.22: The best and the average fitness per generation during the optimisation of the cuts of the decay channel $\eta_{c1}^{\sim} \to \omega J/\Psi$.

2.7 The reconstructed exotic charmonium

2.7.1 Reconstructed signal

After all the cuts applied the $\tilde{\eta_{c1}}$ candidates are created. The reconstructed mass distributions, shown in Figure 2.23, are fitted with a Voigt profile, the convolution of a Lorentz and a Gaussian distribution:

$$V(x;\sigma,\Gamma,x_0) \equiv \int_{-\infty}^{\infty} \frac{e^{-x'^2/(2\sigma^2)}}{\sigma\sqrt{2\pi}} \frac{\Gamma}{\pi((x-x'-x_0)^2+\Gamma^2)}\, dx' \tag{2.2}$$

Here Γ is the extracted Breit-Wigner width, σ is the resolution and x_0 is the extracted mass. Due to the use of kinematic fits during the reconstruction, the resolutions are much smaller than the detector resolutions and they are non-physical. Table 2.8 shows the extracted results and compares them with the ones injected into the simulation. As the values show, the mass can be very well extracted during the analysis, but not the width. This is the result of fitting Γ and σ together. In case of real data, the resolution should be determined by a fit on a neighboring state with known width and it should be a fixed parameter in the final fit. But since there was no other resonance in the simulated data, both parameters were fit simultaneously.

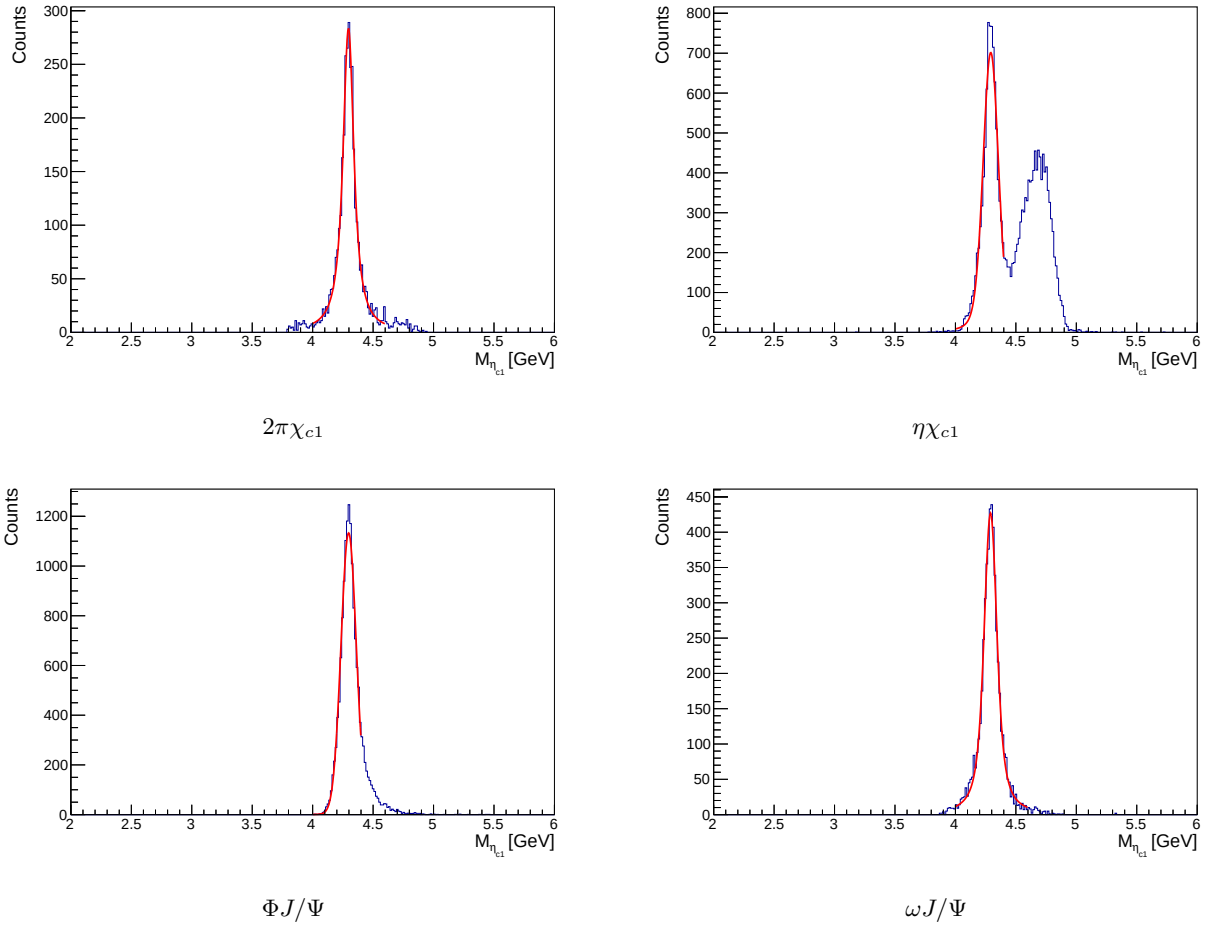


Figure 2.23: Mass distribution of the final reconstructed exotic charmonium for all 4 decay channels in blue and the Voigt fit in red.

	$2\pi\chi_{c1}$	$\eta\chi_{c1}$	$\Phi J/\Psi$	$\omega J/\Psi$	Input
Constant	46±1	61±1	180±2	74±1	-
Mass [MeV]	4296±1	4289±1	4297.7±0.6	4290±1	4290
σ [MeV]	99±4	55±3	60±7	80±4	-
Γ [MeV]	12±6	37±5	18±9	32±2	20

Table 2.8: Extracted parameters

2.7.2 Reconstruction efficiency and background rejection

Figure 2.24 compares the final reconstructed signal with all the background channels combined and with the Monte Carlo matched reconstructed signal events. We can observe that most of the reconstructed signal events are Monte Carlo matched, which means that the cuts successfully remove most of the combinatorial background. These plots also demonstrate that the signal is clearly visible on top of the background, even assuming that the cross section of the dedicated background channels are at least 10 times bigger than of the signal channels. The background reduction is the best for the $2\pi\chi_{c1}$, since it has the most complex decay tree. It also contains the highest number of neutral final state particles, thus it has the lowest reconstruction efficiency and the highest ratio of non-matched signal events. The $\Phi J/\Psi$ channel is exactly the opposite as it only contains one η. In case of the $\eta\chi_{c1}$ there are 2 peaks reconstructed. This is due to the 2 η-s in the decay tree at different positions. In every event 2 charmonium candidates are reconstructed with different masses. The relative position of the 2 peaks are determined by the ratio of the beam energy and the mass of the charmonium. This means that in case of real data, the correct peak can be selected by measuring other decay channels or by varying the beam energy. Table 2.9 shows the background rejection rates and reconstruction efficiencies for each decay channels.

	$2\pi\chi_{c1}$	$\eta\chi_{c1}$	$\Phi J/\Psi$	$\omega J/\Psi$
Signal	3.5%	18.9%	15%	5.3%
MC Signal	1.6%	4.3%	14.2%	2.2%
Background 1	0.37%	18.4%	16.6%	5.3%
Background 2	0.03%	1.2%	-	-
Background 3	0.1%	-	-	-
Background 4	0.38%	-	-	-
DPM Background	$1.4{\cdot}10^{-5}$%	$< 10^{-6}$%	$< 10^{-6}$%	$< 10^{-6}$%

Table 2.9: Survival rates for the signal and background channels numbered according to their position in Table 2.4 using the genetic algorithm optimized cuts. The details of the general DPM backgound can be found in Section 2.5.1.

Figure 2.25 compares the individual contribution of the dedicated decay channels. For the $\eta\chi_{c1}$ channel, most of the background is coming from the $\bar{p}p \to \chi_{c1}\eta\eta$ channel, since it contains all the intermediate resonances as the signal, except the exotic charmonium and so it can not be removed by the applied cuts. But it gives an almost constant background in the region of interest and it is possible to spot the signal on top of it. In case of the other background channel and all the background channels of the other signal channels at least one intermediate state is missing or it is replaced with an other one with a different mass. In these cases the cuts work very well.
All the four channels together with their dedicated background channels were also analyzed using the cuts optimized by hand and listed in Table 2.6 to show the improvement achieved by the genetic algorithm. The achieved significances are compared in Table 2.10.

2.7.3 Comparison with previous studies

Two previous feasibility studies were performed for this state in $\overline{\text{P}}$ANDA, both only considering the decay channel $2\pi\chi_{c1}$. The first one [15] was before the start of the development of PandaRoot. Thus it only has a rough estimate of the $\overline{\text{P}}$ANDA detectors resolution and efficiency. It also doesn't contain any background estimation, but has a good MC matching. They used a few loose cuts and achieved 7.5% reconstruction efficiency. If no cuts are used in our analysis, then the reconstruction efficiency is 6%, which is very close to their value.

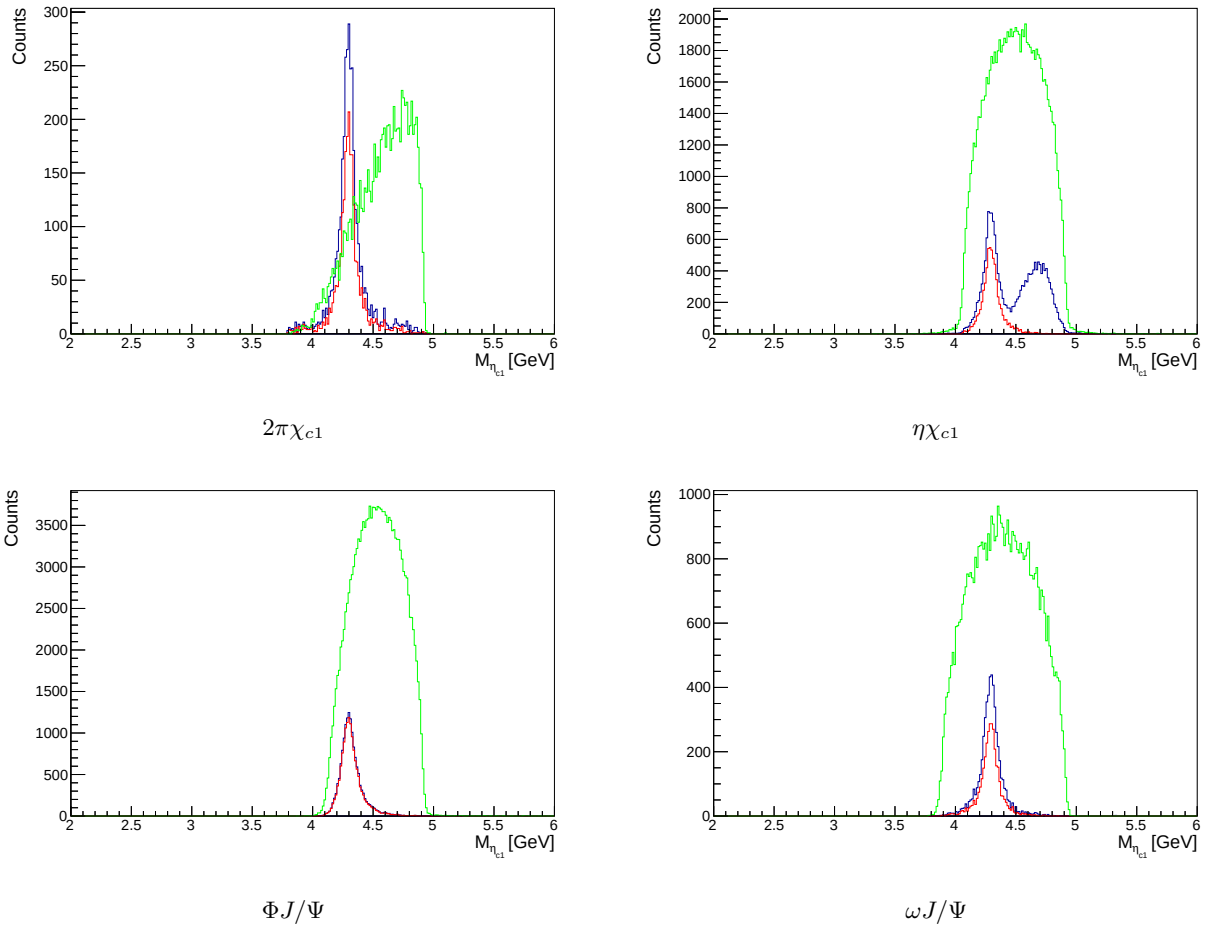


Figure 2.24: Mass distribution of the final reconstructed exotic charmonium for all 4 decay channels in blue, the Monte Carlo matched signal in red and all the background channels combined in green.

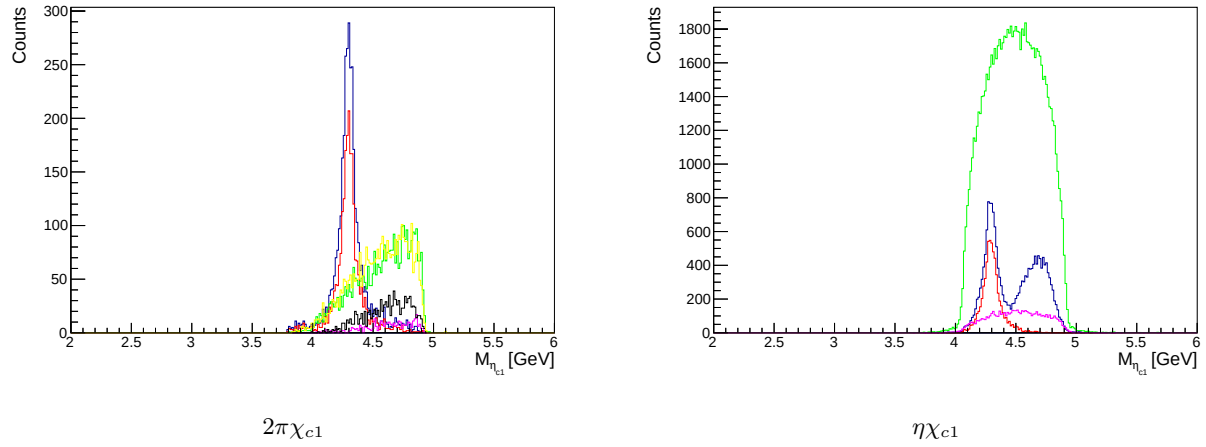


Figure 2.25: Mass distribution of the final reconstructed exotic charmonium in blue, the Monte Carlo matched signal in red and the individual contribution of the dedicated background channels in green, purple, black and yellow. The order of the colors are respective to the order of the channels listed in Table 2.4.

	Genetic Cuts	Conventional Cuts
$2\pi\chi_{c1}$	18.6	7.1
$\eta\chi_{c1}$	17.6	8.6
$\Phi J/\Psi$	288	209.5
$\omega J/\Psi$	120.6	109.1

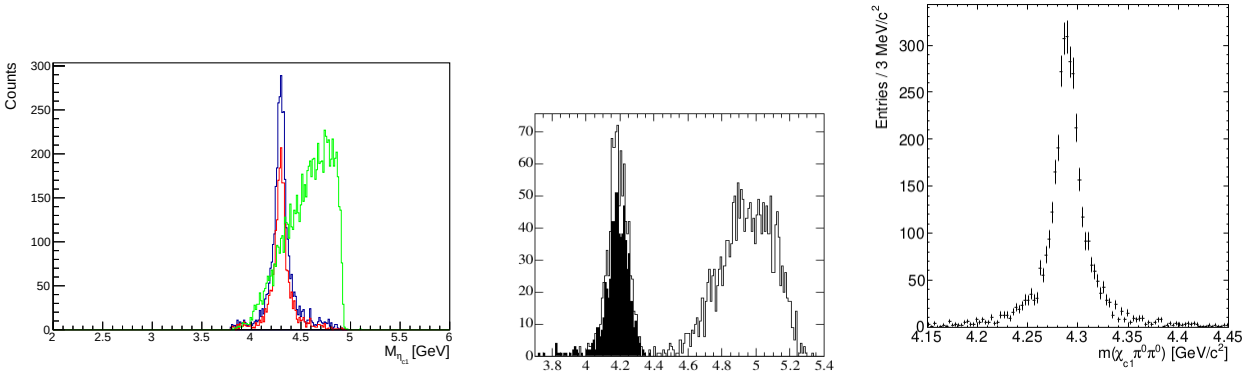
Table 2.10: Comparison of the achieved significance for each decay channel using genetic and conventional cuts.

Their MC matched signal to all reconstructed signal ratio was around 0.694, which is very close to this ratio in our study, which was found to be 0.7. They estimated the number of reconstructed events per day to be 7.5 without any cuts, which becomes 0.75 if we assume that 10% of the signal remains after applying stricter cuts. The other study [9] included the same background channels and use the same cuts and reconstruction process as described in Section 2.5.2 and Table 2.6. They obtained a Breait-Wigner width of $\Gamma = 15$ GeV, which is very close to the value reconstructed from our analysis, which is $\Gamma = (12 \pm 6)$ GeV. They defined a mass range of interest (4.24 GeV-4.33 GeV), where they determined the signal reconstruction efficiencies and the background rejection rates. Their values were reconstructed using the same cuts and are shown in Table 2.11 alongside with other values. The only exception is the signal reconstruction efficiency, which differs by a factor of 10. Most probably this is a typo from their side. They estimate 0.16 as a number of reconstructed events per day, which is very close to our estimate of 0.22.

	This study	Pre-PandaRoot study [15]	Early PandaRoot study [9]
MC signal reconstruction eff.	6%	7.5%	-
MC-all signal ratio	0.7	0.694	-
Signal reconstruction eff.	0.68%	-	6.83%
Background 1 rejection eff.	0.017%	-	0.02%
Background 2 rejection eff.	0.002%	-	0.001%
Background 3 rejection eff.	0.004%	-	0.004%
Background 4 rejection eff.	0.012%	-	0.01%
Breit-Wigner Γ [MeV]	12±6	-	15
Number of reconstructed events per day	0.22	∼0.75	0.16

Table 2.11: Comparison of the applicable results with 2 previous studies. The applied cuts were reproduced as good as possible. The red lines represent a significant change in the applied cuts and analysis method, so results should not be compared across those lines. The first set doesn't use any cuts, i.e. it is pure detector acceptance and performance. The second set uses the reconstruction procedure described in [9]. The last one combines the cross-section and the reconstruction efficiency estimation in the given analysis.

Figure 2.26 compares the reconstructed exotic charmonium mass spectra for the decay channel $2\pi\chi_{c1}$. In case of the pre-PandaRoot study only the spectrum for the $p\bar{p} \to \pi^0\tilde{\eta_{c1}}$ case was plotted. They claimed that the $p\bar{p} \to \eta\tilde{\eta_{c1}}$ spectrum was almost identical, except that it didn't have the 2 peak structure coming from the interchange of the π^0-s, which is similar to the $\eta\chi_{c1}$ channel in our case.



This analysis | Pre-PandaRoot study [15] | Early PandaRoot study [9]

Figure 2.26: Mass distribution of the final reconstructed exotic charmonium for the decay channel $2\pi\chi_{c1}$ in blue, the Monte Carlo matched signal in red and all the background channels combined in green for this analysis. For the other 2 cases, the reconstructed signal is plotted and the MC matched ones are shown in black.

2.8 Beamtime estimation and conclusion

The cross section of the process $p\bar{p} \to \eta\tilde{\eta_{c1}}$ is currently unknown. The only measured cross section until now of this type ($p\bar{p} \to m\Psi$, where m is a light meson and Ψ is a charmonium state) is $p\bar{p} \to \pi^0 J/\Psi$. It was measured by E760 to be (130±25) pb at $\sqrt{s} = 3.5 - 3.6$ GeV. [31] A method to estimate cross sections of this type was suggested in [11]. They estimate the production cross section $\sigma_{p\bar{p}\to m\Psi}$ from measured charmonium partial width into a three-body state, $\Gamma_{\Psi\to mp\bar{p}}$, assuming a constant decay amplitude. They used results from BESII [32] and CLEO-c [33] from J/Ψ and Ψ' initial states at e^+e^- colliders. The process with a known partial width most similar to our process of interest is the $\Psi' \to \eta p\bar{p}$: $\Gamma_{BES} = (16.3 \pm 3.8)$ eV and $\Gamma_{CLEO} = (22 \pm 12)$ eV. Based on this, the cross section $\sigma_{p\bar{p}\to\eta\Psi'}$ is estimated to be (33 ± 8) pb at $E_{CM} = 5.38$ GeV. Based on these values an optimistic 100 pb was used for the estimation of the beam time in this work. If we multiply this by $\overline{\text{P}}$ANDAs design luminosity of $L = 2 \cdot 10^{32} \frac{1}{cm^2 s}$ in high luminosity mode, then we get 1728 charmonium hybrid produced per day.
Table 2.12 summarizes the branching ratios of the involved processes in the decay trees based on the PDG database [34] and the total branching ratio of the decay trees assuming 100% branching ratio for the given $\tilde{\eta_{c1}}$ decay. If we put this together with the reconstruction efficiency and the number of hybrids produced per day, the number of events reconstructed per day can be calculated.

Process	Branching ratio
$\pi^0 \to \gamma\gamma$	98.82%
$\eta \to \gamma\gamma$	39.41%
$\chi_{c1} \to \gamma J/\Psi$	34.3%
$J/\Psi \to \mu^+\mu^-$	6%
$\Phi \to K^+K^-$	48.9%
$\omega \to \pi^+\pi^-\pi^0$	89.2%

Decay tree	Branching ratio	MC truth reconstructed per day	Total signal reconstructed per day
$2\pi\chi_{c1}$	0.8%	0.22	0.48
$\eta\chi_{c1}$	0.3%	0.22	0.49
$\Phi J/\Psi$	0.4%	0.98	1.04
$\omega J/\Psi$	0.7%	0.27	0.64

Table 2.12: Branching ratios of the involved processes in the decay trees based on the PDG database [34] on the left and the total branching ratio of the decay trees assuming 100% branching ratio for the given $\tilde{\eta_{c1}}$ decay on the right. The number of MC and total signal events reconstructed per day with the optimized cuts is also shown on the right.

To estimate the DPM background, the total $p\bar{p}$ cross section was estimated to be 50 mb at 15 GeV beam energy based on data from the PDG database [34] shown in Figure 2.27. This means that 127.5 s of beamtime was simulated with the DPM background if we also take the rejection rate of the event filter into consideration. As only 3 events were reconstructed in the mass region of interest and only 1 of them is directly under the charmonium peak, this translates to 667-2000 reconstructed background events per day, even if all 4 channels are combined.

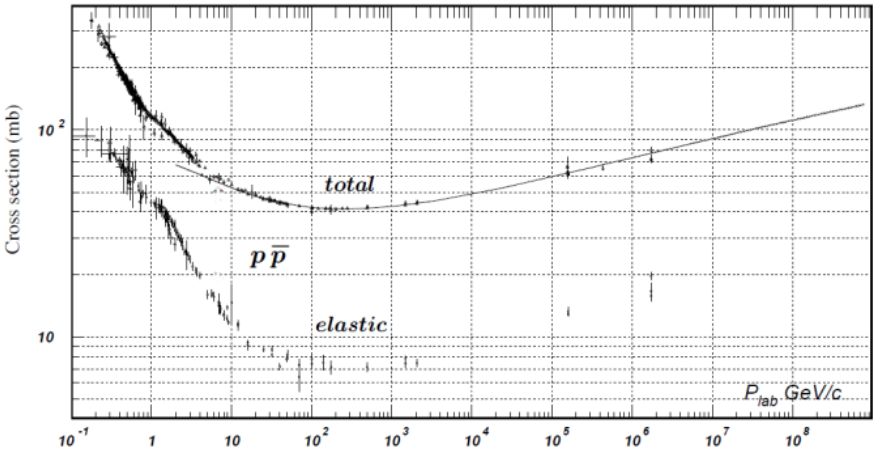


Figure 2.27: Total and elastic cross sections for $p\bar{p}$ collisions as a function of beam momentum [34]

Using equation 2.1 the number of days of measurement required to reach 1 σ significance can be calculated with different assumptions. The results are shown in Table 2.13. This translates to approximately 4 years of beamtime to reach 5 σ, with the worst case scenario being 9.7 years and the best case scenario being 1.3 years. This means that according to this study, the measurement of this hybrid charmonium state will be feasible with $\overline{\text{P}}$ANDA.

	MCM	All
3 DPM	700 days	285 days
1 DPM	300 days	95 days

Table 2.13: Beamtime required to reach 1 σ significance combining all 4 decay channels, assuming 1 or 3 DPM background events and calculating with MC matched signal events or all reconstructed signal events.

As a comparison, the same calculation was performed, but using the hand optimized cuts and the old PandaRoot version using the old MC matching, the old cluster finding algorithm and other not mentioned, small bugs in the code discovered while studying these channels. The result was, that at least 83 years were needed to achieve 1 σ significance, meaning that the measurement is not possible with $\overline{\text{P}}$ANDA. This striking difference shows the importance of the software improvements and the efficiency of the neural network optimized cuts. As all methods are implemented in the newest software version, future feasibility studies can make use of the improvements of the simulation and neutral particles will be reconstructed with a much better efficiency during the reconstruction of measured data.

Chapter 3

A multidimensional study of charged kaon SIDIS beam spin asymmetry over a wide range of kinematics

3.1 Introduction and physics motivation

The moment $A_{LU}^{\sin\phi}$ and the directly related ratio of the structure functions $F_{LU}^{\sin\phi}/F_{UU}$ have been extracted from the polarized electron beam-spin asymmetry in semi-inclusive deep inelastic scattering for the reactions eK^+X and eK^-X. A multidimensional analysis in Q^2, x_B, z and P_T has been performed. This document will present the obtained results and discuss the systematic uncertainties of the measurement. The following sections will closely follow the analysis on single π^+ SIDIS [35], since the physics and the analysis methods are very similar to what is presented there.

3.1.1 Definition of kinematic variables

In the single kaon SIDIS process, the information about the four-vectors of the incoming and outgoing electron (l, l') with the energies (E and E') the four-vector of the target proton P and the four-vector of the produced kaon P_K are known and/or measured. Based on these properties, the required kinematic variables can be defined as:

- $Q^2 = -q^2$ with $q = l - l'$ lepton momentum transfer squared
- $x_B = \frac{Q^2}{2Pq}$ fraction of the proton momentum that is carried by the struck quark
- $\nu = E - E'$ energy of the virtual photon
- $W^2 = m_p^2 + 2m_p\nu - Q^2$ center-of-mass energy squared
- $y = \frac{Pq}{Pl}$ energy fraction of the incoming lepton carried by the virtual photon
- $z = \frac{PP_K}{Pq}$ energy fraction of the virtual photon carried by the outgoing hadron
- P_T transverse momentum of the final state hadron
- $\epsilon = \frac{1-y-\frac{1}{4}\gamma^2y^2}{1-y+\frac{1}{2}y^2+\frac{1}{4}\gamma 2y^2}$ with $\gamma = \frac{2m_px_B}{\sqrt{Q^2}}$ ratio of the longitudinal and transverse photon flux
- ϕ angle between the lepton scattering plane and the meson production plane, according to the "Trento-convention" [45]

3.1.2 Deeply inelastic scattering, structure functions and the parton model

The electron-proton scattering experiment by Hofstadter and McAllister [36] showed that there is a large contribution from inelastic scattering and that the cross section of this process is therefore inconsistent with a point-like proton, meaning that the nucleons have internal structure. Based on the theoretical works of Gell-Mann [37] and Zweig [38], the hadrons can be described as different combinations of quarks.
By increasing the energy of the scattered electron the wavelength of the virtual photon becomes smaller. If it interacts with a constituent quark causing the proton to break up, the scattering process is called deeply inelastic scattering (DIS). The most commonly used kinematic variables are defined in section 3.1.1. The cross section in the unpolarized case can be described using QED in terms of 2 structure functions ($F_{1,2}(x, Q^2)$):

$$\frac{d\sigma}{dxdQ^2} = \frac{4\pi\alpha^2}{Q^4}\left[\left(1 - y - \frac{m_p^2 y^2}{Q^2}\right)\frac{F_2(x,Q^2)}{x} + y^2 F_1(x,Q^2)\right] \tag{3.1}$$

where x_B is the fraction of the protons momentum carried by the struck quark, y is the energy fraction of the incoming lepton carried by the virtual photon, Q^2 is the virtuality of the collision, m_p is the mass of the proton and α is the fine-structure constant. Measurements at Stanford [39] showed that these structure functions are approximately independent of Q^2, this behavior is called Bjorken-scaling, and that they are related through the Callan-Gross relation:

$$F_2(x) = 2xF_1(x) \tag{3.2}$$

These behaviors are well described by the parton model proposed by Feynman et al. [40]. The idea is that since the electron-proton collisions occur on a very short timescale, the process can be described as an electromagnetic interaction of the electron and a single quark in the proton and with a parton distribution function (PDF) $q_i(x)$, which describes the probability of finding a quark of flavor i carrying a momentum fraction of x of the hadron:

$$\frac{d\sigma}{dxdQ^2} = \frac{4\pi\alpha^2}{Q^4}\left(1 - y + \frac{y^2}{2}\right)\sum_i e_i^2 q_i^p(x) \tag{3.3}$$

But measurements of the H1 and ZEUS experiments at HERA showed that the results actually have a weak Q^2-dependence. This feature can be explained with the fact that more than half of the proton momentum is carried by the sea quarks and not by just the valence quarks. This can be calculated by the DGLAP (Dokshizer Gribov Lipatov Altarelli Parisi) equations. When the cross section is calculated, the integral is regularized by factorizing the protons non-perturbative dynamics from the perturbative hard cross section. The choice of this factorization scale is arbitrary between 1 GeV^2 and Q^2. By varying it, one can get renormalization group equations for the PDFs. These are the DGLAP equations. They are the following in leading order:

$$Q^2\frac{\partial}{\partial Q^2}\begin{pmatrix} q_i(x,Q^2) \\ \bar{q}_i(x,Q^2) \\ g(x,Q^2)\end{pmatrix} = \frac{\alpha_s(Q^2)}{2\pi}\sum_j\int_x^1\frac{\mathrm{d}\xi}{\xi}\begin{pmatrix} P_{q_iq_j}(x/\xi) & 0 & P_{q_ig}(x/\xi) \\ 0 & P_{\bar{q}_i\bar{q}_j}(x/\xi) & P_{\bar{q}_ig}(x/\xi) \\ P_{gq_j}(x/\xi) & P_{g\bar{q}_j}(x/\xi) & P_{gg}(x/\xi)\end{pmatrix}\begin{pmatrix} q_j(\xi,Q^2) \\ \bar{q}_j(\xi,Q^2) \\ g(\xi,Q^2)\end{pmatrix} \tag{3.4}$$

where P are called splitting functions and for example P_{qg} describes the process where a gluon (g) radiates a quark (q) and $\alpha_s(Q^2)$ is the strong coupling constant.

3.1.3 Semi-inclusive deeply inelastic scattering and transverse momentum dependent functions

To obtain a deeper understanding of the origin of the proton spin, the focus of the hadron physics community has moved beyond the collinear PDFs towards the study of the partons motion and its spatial distribution in the transverse plane, perpendicular to the momentum of the parent hadron. To describe these distributions, two sets of non-perturbative functions have been introduced. On the one side, Transverse-momentum-dependent parton distributions (TMDs) are used to describe the longitudinal and transverse momentum distributions. On the other side, Generalized Parton Distributions (GPDs) are used to describe the longitudinal momentum distributions and the transverse positions of the partons.
Both TMDs and GPDs can be derived from the most general objects describing the nucleon structure, called

Wigner functions [41]. It was first introduced to study quantum corrections to classical statistical mechanics by linking the wavefunction to a probability distribution in phase space. Later it was recognized as the quantum moment-generating functional encoding all quantum expectation values. [42] It's Fourier-transform is called the Generalized Transverse Momentum Distribution (GTMD). All other distributions can be derived from these ones by different combinations of the following operations: taking the forward limit in hadron momentum ($\Delta = 0$), integrating over x and integrating over k_T. This is illustrated by the so-called GTMD cube on Figure 3.1.

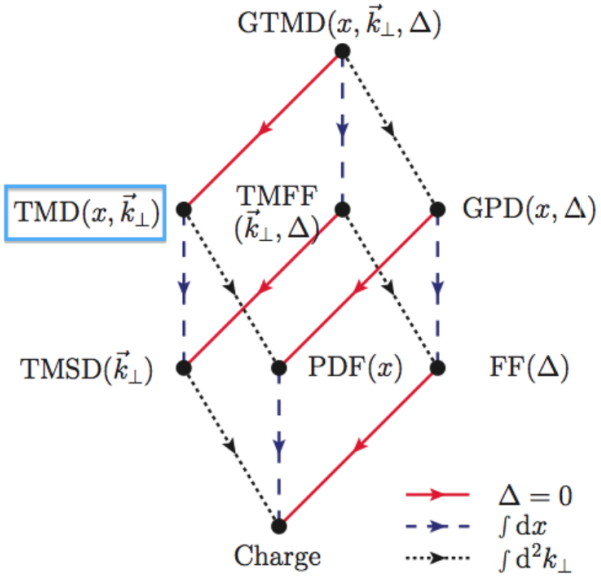


Figure 3.1: GTMD cube illustrating the relations between different distribution and fragmentation functions [43]

The main goal of this thesis is to provide means to better constrain the TMDs, so they will be discussed in more details. Table 3.1 lists the TMDs for leading and sub-leading twist. The twist is 2 plus the power of M/Q. For the twist-3 terms it is impossible to define a quark polarization because each term also contain a gluon in addition to the quark. In case of an unpolarized nucleon, the only twist-2 terms are f_1 which describes the unpolarized quarks in an unpolarized nucleon and the Boer-Mulders function ($h_1^\perp$) which describes the transversely polarized quarks inside an unpolarized proton. While these terms describe parton densities ($< \bar{q}q >$), the twist-3 terms are correlations of quark and gluon fields ($< \bar{q}gq >$). In case of an unpolarized nucleon, only 4 of them can be measured. They include the twist-3 T-odd TMD $g^\perp$, which is the higher-twist analog to the Sievers function and the chiral-odd TMD e. Its x^2 moment describes the transverse force acting on transversely polarized quarks in the unpolarized proton. [43]

	twist-2			twist-3
N \ q	U	L	T	
U	f_1		$h_1^\perp$	$f^\perp$, $g^\perp$, h, e
L		g_{1L}	$h_{1L}^\perp$	$f_L^\perp$, $g_L^\perp$, h_L, e_L
T	$f_{1T}^\perp$	g_{1T}	h_1, $h_{1T}^\perp$	f_T, $f_T^\perp$, g_T, $g_T^\perp$, $h^\perp$, $h_T^\perp$, e_T, $e_T^\perp$

Table 3.1: TMDs for leading and sub-leading twist. The labels in the top row denote the polarization of the quark and the labels in the first column denote the polarization of the nucleon.

By detecting a final state hadron in DIS, information about the quark momenta in the plane transverse to the hard scattering direction can be obtained. This process is called semi-inclusive deeply inelastic scattering (SIDIS). The diagram in Fig. 3.2 shows the SIDIS scattering process including the involved transverse momentum dependent parton distribution (TMD PDF) and fragmentation function (FF). In a single photon exchange model, the differential cross section of this process can be written as a product of leptonic ($L_{\mu\nu}$) and hadronic ($2MW_{\mu\nu}$) tensors [44]:

$$\frac{d\sigma}{dxdQ^2dzdp_T} = \frac{\pi\alpha^2 yz}{2Q^4} L_{\mu\nu} 2MW^{\mu\nu} \tag{3.5}$$

where z is the energy fraction of the virtual photon carried by the outgoing hadron and P_T is the transverse momentum of the final state hadron. The leptonic tensor describes the emission of the virtual photon and can be solved exactly [43]:

$$L_{\mu\nu} = Tr\left(\gamma_\mu(\not{k}' + m)\gamma_\nu(\not{k} + m)\frac{1 \pm \gamma_5 \not{S}}{2}\right) \tag{3.6}$$

The hadronic tensor in the Bjorken limit in factorization approach can be expressed as a convolution of the correlator describing the quarks inside the proton ($\Phi(x, p_T)$) and the correlator describing the fragmentation process or in other words the final state ($\Delta(z, k_T)$) [43]:

$$2MW_{\mu\nu} = 2z \int d^2p_T d^2k_T \delta^{(2)}\left(p_T - k_T - \frac{P_{h\perp}}{z}\right) Tr(\Phi(x, p_T)\gamma_\mu \Delta(z, k_T)\gamma_\nu) \tag{3.7}$$

The correlators are the following [43]:

$$\Phi(x, p_T) = \int \frac{d\xi^- d^2\xi_T}{(2\pi)^3} e^{ip\xi} \langle P| \bar{\Psi}_j(0) U^{n_-}_{(0,\text{inf})} U^{n_-}_{(\text{inf},\xi)} \Psi_i(\xi) |P\rangle |_{\xi^+=0} \tag{3.8}$$

$$\Delta(z, k_T) = \frac{1}{2z} \sum_X \int \frac{d\xi^- d^2\xi_T}{(2\pi)^3} e^{ik\xi} \langle 0| U^{n_+}_{(\text{inf},\xi)} \Psi_i(\xi) |P_h, X\rangle \langle P_h, X| \bar{\Psi}_j(0) U^{n_+}_{(0,\text{inf})} |0\rangle |_{\xi^-=0} \tag{3.9}$$

Here the U^n are the Wilson-lines. They allow the comparison of two quark fields at different space-time points. This is analogous to the parallel transport in general relativity which compares tangent vectors that live in the tangent spaces at different points. This way the equations above are guarantied to be gauge invariant. [43]
The correlators can be expressed as functions of TMDs and FFs for the special case of a polarized electron beam fragmenting an un-polarized target fragmenting into an unpolarized hadron in leading twist ($t2$) [43]:

$$\Phi^{\pm}_{t2}(x, p_T) = \frac{1}{2}\left(f_1(x, p_T)\not{n}_+ \pm i h_1^\perp(x, p_T)\frac{[p_T, \not{n}_+]}{2M}\right) \tag{3.10}$$

$$\Delta_{t2}(z, k_T) = \frac{1}{2}\left(D_1 \not{n}_- + i H_1^\perp \frac{[k_T, \not{n}_-]}{2M_h}\right) \tag{3.11}$$

With $\not{n}_\pm = \frac{\gamma^0 \mp \gamma^3}{\sqrt{2}}$. Here the TMDs can be interpreted as the probability of measuring a parton with a certain momentum fraction x, and fragmentation functions give the probability of measuring a hadron with momentum fraction z in the decay products. The twist-3 ($t3$) terms are the following [43]:

$$\Phi_{t3}(x, p_T) = \frac{M}{P^+}\left(\frac{k_T^i}{M} f^\perp(x, k_T^2) - \frac{\varepsilon_T^{ij} k_{Tj}}{M} g^\perp(x, k_T^2)\right) \tag{3.12}$$

$$\Delta_{t3}(z, k_T) = \frac{M_h}{2P_h^-}\left(E + D^\perp \frac{\not{k}_T}{M_h} + iH\frac{[\not{n}_-, \not{n}_+]}{2} + G^\perp \gamma_5 \frac{\varepsilon_T^{\rho\sigma}\gamma_\rho k_{T\sigma}}{M_h}\right) \tag{3.13}$$

Here P^+ is the target momentum in the light cone coordinates, defined by: $P^\mu = P^+ n_+^\mu + \frac{M^2}{2P^+} n_-^\mu$
In case of a polarized electron beam, interacting with an un-polarized target, the SIDIS cross section is given by [7]:

$$\frac{d\sigma}{dxdQ^2dzdp_Td\phi} = K(x,y,Q^2)\left[F_{UU,T} + \epsilon F_{UU,L} + \epsilon F_{UU}^{\cos 2\phi}\cos 2\phi + \sqrt{2\epsilon(1+\epsilon)}F_{UU}^{\cos\phi}\cos\phi + \lambda_e\sqrt{2\epsilon(1-\epsilon)}F_{LU}^{\sin\phi}\sin\phi\right] \tag{3.14}$$

where $\lambda_e = \pm 1$ is the electron helicity and $K(x,y,Q^2)$ is a kinematic factor. The cross-section consists of several model independent structure functions (F) and depends on the azimuthal angle ϕ between the electron scattering plane and the hadron production plane as shown in Fig. 3.2, according to the definition of the trento convention [45]. The subscripts on the structure functions indicate the beam and target polarization (U: unpolarized, L: longitudinally polarized).

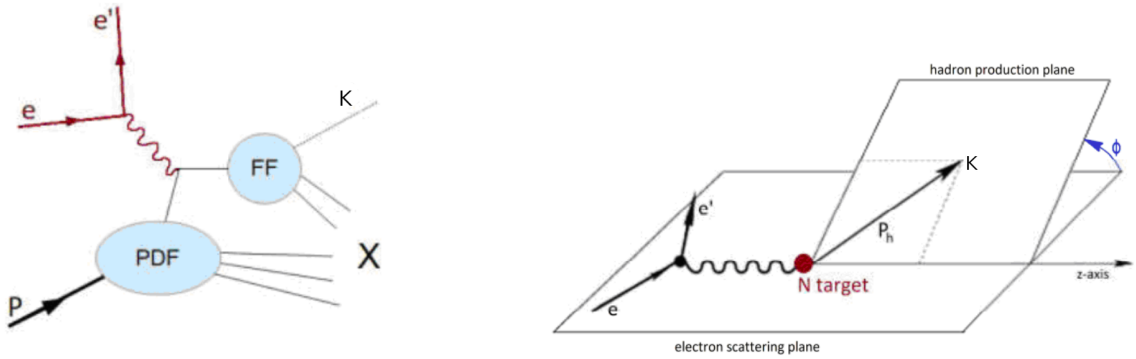


Figure 3.2: Process (left) and kinematics (right) of single kaon SIDIS [46]

The structure functions can be expressed theoretically as a convolution of TMDs ($f^a(x,k_T^2)$) and FFs ($D^a(z,p_T^2)$) as [7, 47] up to order $1/Q^2$ based on the equations above:

$$F_{UU,T} = \zeta(f_1 D_1) \tag{3.15}$$

$$F_{UU,L} = 0 \tag{3.16}$$

$$F_{UU}^{\cos\phi} = \frac{2M}{Q}\zeta\left(-\frac{\hat{h}k_T}{M_h}\left(xhH_1^\perp + \frac{M_h}{M}f_1\frac{\tilde{D}^\perp}{z}\right) - \frac{\hat{h}p_T}{M}\left(xf^\perp D_1 + \frac{M_h}{M}h_1^\perp\frac{\tilde{H}}{z}\right)\right) \tag{3.17}$$

$$F_{UU}^{\cos 2\phi} = \zeta\left(-\frac{2(\hat{h}k_T)(\hat{h}p_T) - k_Tp_T}{MM_h}h_1^\perp H_1^\perp\right) \tag{3.18}$$

$$F_{LU}^{\sin\phi} = \frac{2M}{Q}\zeta\left(-\frac{\hat{h}k_T}{M_h}\left(xeH_1^\perp + \frac{M_h}{M}f_1\frac{\tilde{G}^\perp}{z}\right) + \frac{\hat{h}p_T}{M}\left(xg^\perp D_1 + \frac{M_h}{M}h_1^\perp\frac{\tilde{E}}{z}\right)\right) \tag{3.19}$$

Here the convolution $\zeta(\omega f D)$ is defined as:

$$\zeta(\omega f D) = x\sum_a e_a^2\int d^2p_Td^2k_T\delta^{(2)}\left(p_T - k_T - \frac{P_{h\perp}}{z}\right)\omega(p_T,k_T)f^a(x,p_T^2)D^a(z,k_T^2) \tag{3.20}$$

where $\omega(p_T,k_T)$ is the kinematic weighting function and $\hat{h}$ is a unit vector pointing in the direction of the outgoing hadron.
The $\tilde{E}$, $\tilde{G}^\perp$, $\tilde{D}^\perp$ and $\tilde{H}$ come from the quark-gluon-quark fragmentation correlator and and they are related to the corresponding FFs and quark-quark correlators as the following [43]:

$$\frac{E}{z} = \frac{\tilde{E}}{z} + \frac{m}{M_h}D_1 \tag{3.21}$$

$$\frac{G^{\perp}}{z} = \frac{\tilde{G}^{\perp}}{z} + \frac{m}{M_h} H_1^{\perp} \tag{3.22}$$

$$\frac{D^{\perp}}{z} = \frac{\tilde{D}^{\perp}}{z} + D_1 \tag{3.23}$$

$$\frac{H}{z} = \frac{\tilde{H}}{z} + \frac{k_T^2}{M_h^2} H_1^{\perp} \tag{3.24}$$

The fragmentation functions appearing above describe the fragmentation process by containing the information on how partons produced in hard-scattering processes are turned into an observed colorless hadronic bound final-state. D_1 describes the probability of a (spin-averaged) quark to produce a hadron, G_1 and H_1 describe the difference of probabilities that a polarized quark produces a polarized hadron with the same or the opposite polarization in the longitudinal and transverse case respectively. The FFs with the $\perp$-denotement stand for a transverse parton or hadron polarization, while the hadron or parton is spin-averaged. For example the Collins function $H_1^{\perp}$ describes the fragmentation of a transversely polarized quark into an unpolarized hadron. [48]

3.1.4 Beam spin asymmetry and SIDIS structure functions

The main focus of this study will be the structure function $F_{LU}^{\sin\phi}$, which is related to the quark-gluon correlations in the proton. The $F_{LU}^{\sin\phi}$ has been studied in several theoretical and phenomenological works. [49–57] To sum up the previous section: On the TMD side the most important contributors are the Boer-Mulders function $h_1^{\perp}$, which is a leading-twist time-reversal-odd TMD [58] as well as the higher twist counter part of the Sivers function $g^{\perp}$, a twist-3 time-reversal-odd TMD [53,59] and the chiral-odd twist-3 PDF $e(x)$ which can be related to the transverse force acting on the transversely polarized quarks in an un-polarized nucleon [60,61]. In each term, a TMD is convoluted with a twist-2 or 3 fragmentation function like $\tilde{G}^{\perp}$, $\tilde{E}$ and the Collins function $H_1^{\perp}$ [62]. At leading order, each convolution represents a pure twist-3 term.
Previous experiments like CLAS [63] and HERMES [64] showed that kaon signals generally follow the pion signals, but they have larger values. With these measurements the s-quark can be accessed, but due to the low statistics the uncertainties are large and the kinematic bins are wide. The high statistics on an extended kinematic range, which is available with the new CLAS12 data, enables a fully differential analysis for the first time and therefore provides an excellent basis for the extraction of TMDs and fragmentation functions in global fits as well as an extended understanding of the effects which take place in the different kinematic regions. To access the s-quarks, a flavor separation (i.e. measuring and comparing results for both the pions and kaons) is required. In the valence quark picture the charged pions are composed of $u\bar{d}$ and $d\bar{u}$ and the charged kaons are composed of $u\bar{s}$ and $s\bar{u}$. By measuring the SIDIS process for these particles, the TMD of the strange quark can be accessed.
By introducing the moments $A_{LU}^{\sin\phi}$, $A_{UU}^{\cos\phi}$ and $A_{UU}^{\cos 2\phi}$ the cross-section from Eq. 3.14 can be written as:

$$d\sigma = d\sigma_0(1 + A_{UU}^{\cos\phi}\cos\phi + A_{UU}^{\cos 2\phi}\cos 2\phi + \lambda_e A_{LU}^{\sin\phi}\sin\phi) \tag{3.25}$$

Based on this definition, the beam spin asymmetry (BSA) can be written as:

$$BSA = \frac{\sigma^+ - \sigma^-}{\sigma^+ + \sigma^-} = \frac{A_{LU}^{\sin\phi}\sin\phi}{1 + A_{UU}^{\cos\phi}\cos\phi + A_{UU}^{\cos 2\phi}\cos 2\phi} \tag{3.26}$$

Be defining the total ϕ-integrated cross section $F_{UU} = F_{UU,T} + \epsilon F_{UU,L}$, these three moments can be related to the structure functions:

$$\frac{F_{LU}^{\sin\phi}}{F_{UU}} = \frac{A_{LU}^{\sin\phi}}{\sqrt{2\epsilon(1-\epsilon)}} \tag{3.27}$$

$$\frac{F_{UU}^{\cos\phi}}{F_{UU}} = \frac{A_{UU}^{\cos\phi}}{\sqrt{2\epsilon(1+\epsilon)}} \tag{3.28}$$

$$\frac{F_{UU}^{\cos 2\phi}}{F_{UU}} = \frac{A_{UU}^{\cos 2\phi}}{\epsilon} \tag{3.29}$$

3.2 Experimental setup

The data-set used in this work was taken at Jefferson Laboratory, Newport New, Virginia. The electron beam was provided by the Continuous Electron Beam Accelerator Facility (CEBAF) [65–67]. The electron helicity was flipped periodically with a frequency of 33 Hz which provided a practically equal luminosity for both helicity states. To further minimize systematic effects, a half wave plate was used to switch the definition of positive and negative helicity periodically for certain run periods. The beam polarization was measured regularly with a Møller polarimeter in the beam-line upstream of the detector. The measurements were performed with the upgraded CEBAF Large Acceptance Spectrometer (CLAS12) [67, 68]. The CLAS12 detector setup consists of a central detector (CD), surrounding the liquid hydrogen target cell and a dedicated forward detector (FD). The forward detector starts with a high threshold Cherenkov counter (HTCC), which is used as a trigger for electrons. The tracking is provided by a torus magnet and a drift chamber system (DC) with 3 regions before, inside and after the magnet coils. The torus coils split the detector in 6 independent sectors. A low threshold Cherenkov counter (LTCC) available in 2 sectors and a Ring-imaging Cherenkov detector (RICH) available in 1 sector can be used for particle identification in addition to the forward time of flight system (FTOF), which provides the standard particle ID. Finally an electromagnetic calorimeter (ECAL) is used for the detection and identification of electrons and photons. The central detector is equipped with a solenoid magnet, a micromegas (MM) and a barrel silicon tracker (SVT) as well as a central time-of-flight system (CTOF) and a neutron detector (CND). To extend the detector coverage to small Q^2, a forward tagger (FT), consisting of a tracker, a hodoscope and an electromagnetic calorimeter was installed in the most forward region around the beam-line. This section will give a detailed overview of the experimental setup.

3.2.1 Accelerators

The CEBAF accelerator at Jefferson Lab is composed of 2 linear accelerators (linac) with 25 cryomodules and recirculating arcs arranged in a racetrack shape. It provides an extremely high duty factor and an average beam current of up to 200 μA to the 4 experimental halls. This means that the beam is continuous on the scale of the integration time of the detectors. This high duty factor is achieved by the use of superconducting radio-frequency cavities (SRF). The electron beam is a continuous train of short 100 ps pulses recurring at 1497 MHz. The delivery of electron pulses to to the Halls occur at 249.5 MHz during 4 hall operation and 499 MHz during 3 hall operations. The electrons are created from a GaAs photo-cathode gun which is illuminated with 4 lasers. The 850 nm lasers have a phase difference with respect to each other operating at 250 MHz. This offset is used to separate the beam delivered to the halls after the acceleration. The electrons, injected to CEBAF, gain an energy of 1.1 GeV per pass in each linac. The bunches can make up to 5 passes, which means that the beam energy received by Halls A, B and C can be as high as 11 GeV. The beam energy in Hall D can reach 12 GeV, since the beam makes another half pass to reach that hall. [69] Figure 3.3 shows the schematic overview of CEBAF.

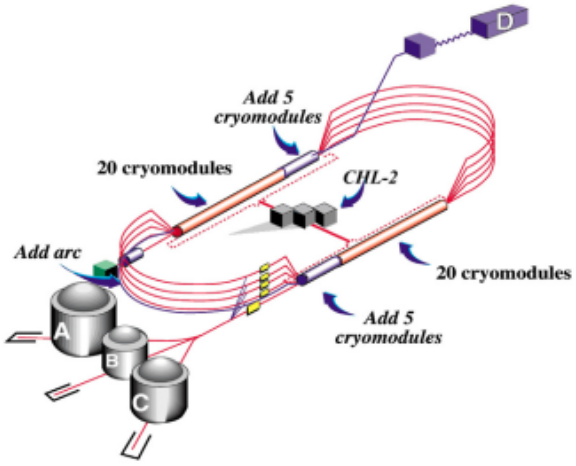


Figure 3.3: Schematic overview of CEBAF at Jefferson Lab [67]

3.2.2 The CLAS12 detector

CLAS12 is a fixed target experiment and this is reflected in the detector design shown in Figure 3.4. The RG-A target system consists of a liquid hydrogen target cell and an associated cryogenic system. The target is located from -5.5 to -0.5 cm in z-axis of the CLAS12 coordinate system. [70]

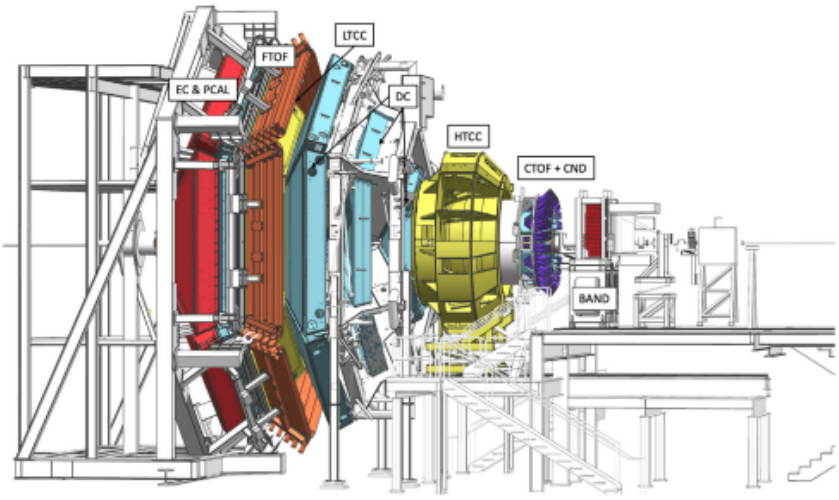


Figure 3.4: Schematic overview of the CLAS12 setup [67]

Since this analysis only uses the Forward Detector, the Central Detector and the Forward tagger will not be discussed in detail. The FD covers the polar angles (θ) from 5° to 35°, except Panel-2 of the FTOF system, which covers up to 45°. The detectors are split by the torus magnet to 6 independent sectors, each covering 60° azimuth angle (Φ). The numbering of the sectors follow a counterclockwise convention starting with sector 1, which has its middle at 0°. The peak magnetic fields in the solenoid and in the torus superconducting magnets surrounding the detectors for momentum measurement are 5 T and 3.58 T, respectively. The integrated magnetic field strengths are 7 Tm and 0.54-2.78 Tm. The later values correspond to 40° and 5° in polar angle. [70]
The HTCC is the detector closest to the target. It's main purpose is to distinguish electrons and charged pions below 4.9 GeV and to serve as a trigger for the scattered electrons as part of the standard PID. It can also be used to separate kaons and pions above ∼4.5 GeV as a supplementary PID method. [71]
The DC system measures the momentum of charged particles in the magnetic field. It consists of 3 regions, each encompassing 2 superlayers, in every 6 sectors. Each superlayer has 6 layers with 112 wires per layer. The regions are approximately 2, 3 and 4 m away from the target. The DC has a spacial resolution of 250-350 μm and a momentum resolution of dp/p> 1%. The angular resolution is 1 mrad in θ and 1 mrad/sin θ in Φ. [71]
The time information of the DC can also be utilized in additional PID refinements, since it has a very good time resolution of a few ns [72].
The main purpose of the LTCC and of the RICH detectors is to separate kaons and pions above 3 GeV, where the standard TOF PID is not sufficient. During the runs used in this work, only 2 sectors of the LTCC were filled with gas and used for data taking. The RICH which is planned to replace the LTCC was installed in one sector and its reconstruction is limited to a certain kinematic range [73]. The reconstruction is currently only reliable for forward scattered particles which have polar angles less than 15°. This is detailed in section 3.4.9. Although these detectors are not used during this analysis to extract any physics, they are used to validate the pion contamination in the kaon sample of the Monte Carlo and also to estimate the capability of the machine learning method which improves the standard PID and is based on TOF information from multiple detectors. Both detectors have a reliable kaon-pion separation above 3-4 GeV up to 9 GeV for LTCC and 6 GeV for RICH. The pion/kaon rejection factor of the RICH is claimed to be less than 500. [71]
The FTOF system provides the standard PID for charged pions, kaons and protons. It achieves a 4σ separation for kaon-pion up to 2.8 GeV and for kaon-proton up to 4.8 GeV. It consists of 3 arrays of counters in each sector. The first array, panel-1a, is made of 23 counters and has a time resolution of 90-160 ps, while the second one, panel-1b, is made of 62 counters providing 60-110 ps time resolution. Panel-2, which covers the larger polar angles, with a time resolution of 140-165 ns consists of 5 counters. [71]

The outermost detector package is the electromagnetic calorimeter, located approximately 7 m from the target. It consists of the legacy ECAL of CLAS6, which has 2 layers (ECin and ECout), complemented with the pre-shower calorimeter (PCAL). Its main purpose is to identify electrons, photons, neutral pions and neutrons. All components are sampling calorimeters consisting of six modules. Each module has a triangular shape and is made of 54 layers (15/15/24-PCAL/ECin/ECout) of 1 cm thick scintillators segmented into 4.5 cm/10 cm (PCAL/ECAL) wide strips and sandwiched between 2.2 mm thick lead sheets. This gives a total thickness of around 20.5 radiation lengths. The scintillator layers are grouped into 3 readout views (u,v,w) with 5/5/8 (PCAL/ECin/ECout) layers per view. The whole system gives an energy resolution of $10\%/\sqrt{E}$, a position resolution of 0.5 cm and a time resolution of 500 ps.

3.2.3 Data acquisition

The CLAS12 Data Acquisition system (DAQ) collects the data from all of the CLAS12 detector components. This network-based system collects data from more than 100 front-end components including the trigger system signals, assembles them into events, monitors the data quality and records the data to tapes. The typical data rates are 15-30 kHz and 500-1000 MByte/s with a livetime above 90%. The CLAS12 DAQ software is a modular and multi-threaded C/C++ library based on the CEBAF Online Data Acquisition (CODA) system. The front-end components, also known as Readout Controllers (ROC), are commercial VME/VXS crates with Intel-based controllers and custom boards, commercial Linux servers and JLab-designed VXS Trigger Processor (VTP) boards. All of these components are connected via TCP sockets over Ethernet to the Event Builder (EB) component. The events built are then transferred through the Event Transfer (ET) system to the Event Recorder (ER). The CLAS12 Event Display (ced) is also part of the DAQ system. It is a ROOT [74] based customizable GUI, that allows online data monitoring during the data taking. [75]

3.3 Preparation of the data set

This section summarizes the treatment of the data-set used during this analysis. The analysis is based on the common analysis procedure of the CLAS12 Collaboration [76].

3.3.1 Analyzed data-set

The data-sets used in this work was taken using a polarized electron beam, interacting with an un-polarized liquid hydrogen target during the RG-A run period in fall 2018 and in spring 2019. 3 different data-sets were analyzed. The first contains 174 runs from the run range between 5032-5419. It had an in-bending torus field with 100% field strength. The beam current varied between 40 nA and 55 nA. The beam energy was 10.6041 GeV. The second data-set contains 176 runs from the run range between 5422-5666. It had an out-bending torus field with 100% field strength and the beam energy was also 10.6041 GeV. The beam current varied between 40 nA and 59 nA. The third data-set contains 121 runs from the run range between 6616-6783. It had an in-bending torus field with 100% field strength and the beam energy was 10.1998 GeV. The beam current varied between 40 nA and 50 nA.

3.3.2 Quality checks and cuts

A series of quality checks have been applied to the data set on a file by file basis, following the common quality criteria. The common QA is primarily based on the FC-charge normalized electron yield (N/F), but also considers sector losses as well as irregularities of the live time and miscellaneous defects.
First a trigger electron is identified with a $|\chi^2_{PID}| < 3$ per event. The number of trigger electrons N is expected to scale proportionally with F, the accumulated FC charge. Normalizing N by F accounts for fluctuations in beam conditions, allowing N/F to be a metric for detecting changes in the data acquisition conditions. For each file the difference between the minimum and maximum accumulated FC charge gated by the DAQ is taken as F. For every sector and epoch (time period of runs where N/F was relatively constant) the Interquartile Range (IQR), which is the difference between the 75th and 25th percentiles, is calculated. Files are identified as outliers if their N/F falls outside the 75th or the 25th percentiles by more than 4 IQR. This identification is enhanced by considering the statistical uncertainty of N/F, by assuming a Poisson distribution of both F and N. If a certain N/F is outside of the above described cut within its uncertainty, then it is marked as a marginal outlier, instead of a total outlier. If N/F is a total outlier, but it is the first or last file of a run, then it is instead marked as a terminal outlier, because this behavior is caused by the DAQ. If there is a sudden, visible and consecutive drop in the N/F in one or more sectors, then the files are marked as sector loss.
The setting "OkForAsymmetry" combines multiple criteria to remove files which could influence the asymmetry extraction: Total outliers represent an issue effecting all sectors, like an empty target test, and thus they are removed. Although terminal outliers should have no impact on the extracted symmetries, they are still removed for safety. Also marginal outliers are rejected for the same reason. As sector losses can introduce azimuthal modulations they are removed as well. Also any additional files were removed if they were marked during the data taking as having special issues, such as a DAQ test, high beam trip rate, empty target, no helicity info or anything similar.
To apply the data quality cut, the setting "OkForAsymmetry" has been selected from the common quality database. Since the quality assessment checks the data quality on a file by file basis, in most cases no complete runs are rejected; only single files which did not pass the defined quality criteria were rejected from the analysis. [76]

3.3.3 Fiducial cuts

To ensure that charged particles and photons are only detected in regions of the fiducial volume in which their momentum, energy and angular information can be reconstructed in a reliable way, a series of fiducial cuts has been applied. This way the particles reconstructed by unreliable parts (usually the edges) of the detectors are removed. An example would be the uncertain energy reconstruction in the periphery of the calorimeter due to shower leakage or the inaccurate drift-time reconstruction in the edge of the drift chamber due to the irregular electric field.
For electrons, a homogeneous, loose PCAL fiducial cut is used, requiring that v and w are larger than 9 cm. The effect of the PCAL fiducial cuts on the hit distribution of electrons in the PCAL is shown in Fig. 3.5. For

electrons and hadrons, drift chamber fiducial cuts on the 3 regions are used. The effect of the DC fiducial cuts on the hit distribution in the single DC regions is shown in Fig. 3.6 for electrons, in Fig. 3.7 for positive kaons and in Fig. 3.8 for negative kaons for the inbending case.

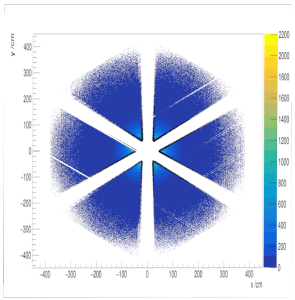


Figure 3.5: Hit distributions of electrons in the PCAL, the distribution without the PCAL fiducial cut is shown as black points, while the distribution after the cut is overlaid as colored points

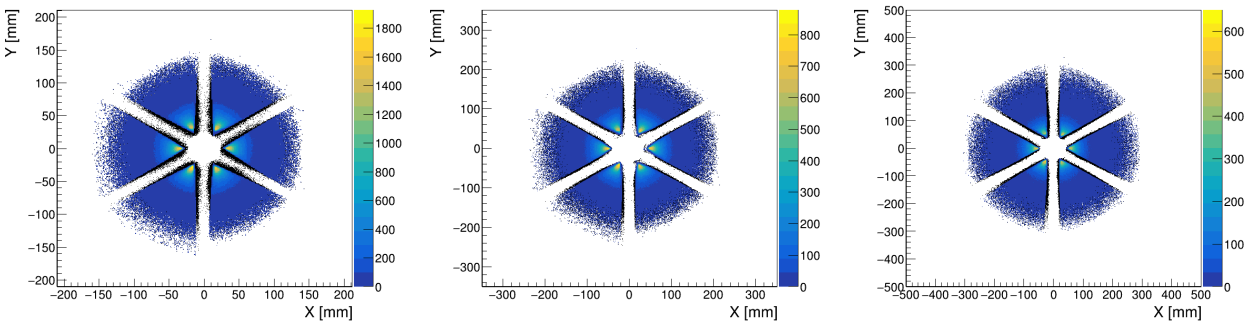


Figure 3.6: Hit distributions of electrons in the 3 regions of the drift chambers with inbending setting (region 1: left, region 2: center, region 3: right), the distribution without the DC fiducial cut is shown as black points, while the distribution after the cut is overlaid as colored points

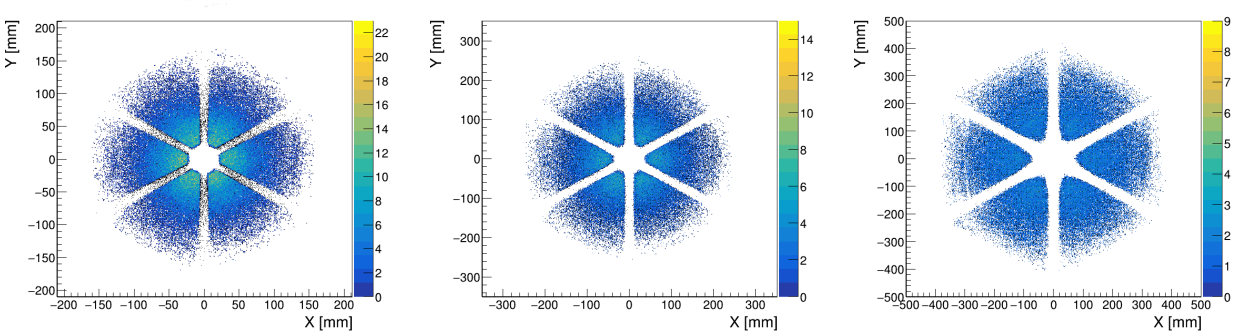


Figure 3.7: Hit distributions of positive kaons in the 3 regions of the drift chambers with inbending setting (region 1: left, region 2: center, region 3: right), the distribution without the DC fiducial cut is shown as black points, while the distribution after the cut is overlaid as colored points

The CLAS12 common analysis procedure defines only the inbending fiducial cuts. [76] The outbending fiducial cuts were determined for electrons and hadrons with exactly the same method used for inbending electrons: In bins of x and y the average tracking χ^2/NDF is calculated. The x distributions of averaged χ^2/NDF in slices of y are analyzed. 15 slices are created between 25.7 cm and 151.2 cm, 28.8 cm and 262.8 cm and 31.8 cm and

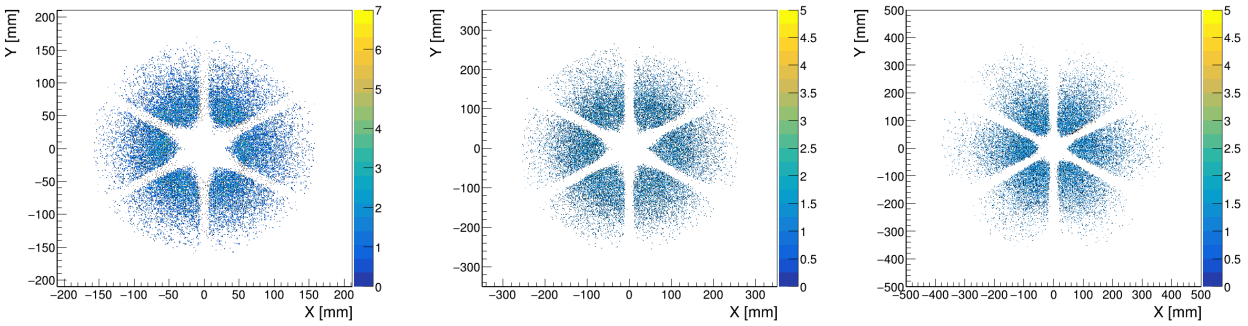


Figure 3.8: Hit distributions of negative kaons in the 3 regions of the drift chambers with inbending setting (region 1: left, region 2: center, region 3: right), the distribution without the DC fiducial cut is shown as black points, while the distribution after the cut is overlaid as colored points

355.8 cm in region 1, 2, and 3, respectively. An example for the resulting distributions can be seen in Fig. 3.9. The center of the distributions are fitted with a constant around $x = 0$ cm. Then the x values are determined where a 20% increase or a 50% decrease in the averaged χ^2/NDF level is reached compared to the fitted value, whichever occurs first. Similar to the inbending case, in case of electrons in region 1 a 5% increase or a 10% decrease determines the x values.
After the removal of outliers the points are fitted with $x = a + b * y$. An example can be seen in Fig. 3.10.
Fig. 3.11 shows the resulting cuts in $x-y$ coordinates in region 3 for all sectors. The cuts are sector independent, which is illustrated by Fig. 3.12.
The result of the cuts for electrons can be seen in Fig. 3.13, for positive kaons in Fig. 3.14 and for negative kaons in Fig. 3.15. The black points are the hits without the fiducial cuts and the colored ones are those that remained after applying the cuts. The survival rates can be found in Table 3.2.
In case of inbending hadrons, the local Φ and θ was used instead of X and Y. 15 slices are created between 5.2°and 30.4°in θ. To the θ-Φ values obtained this way an additional fix point is added at $\theta = 40°$and $\Phi = \pm$ 27.5°to reduce the influence of the fluctuations present at high θ values to the asymptotic behavior of the fit function, which was chosen to be $\Phi = a + b \cdot \log\theta + c \cdot \theta + d \cdot \theta^2$. This method allowed a much better description of the observed borders of the distributions, but it was not stable enough to be used for the other cases. The method is illustrated on figure 3.16.

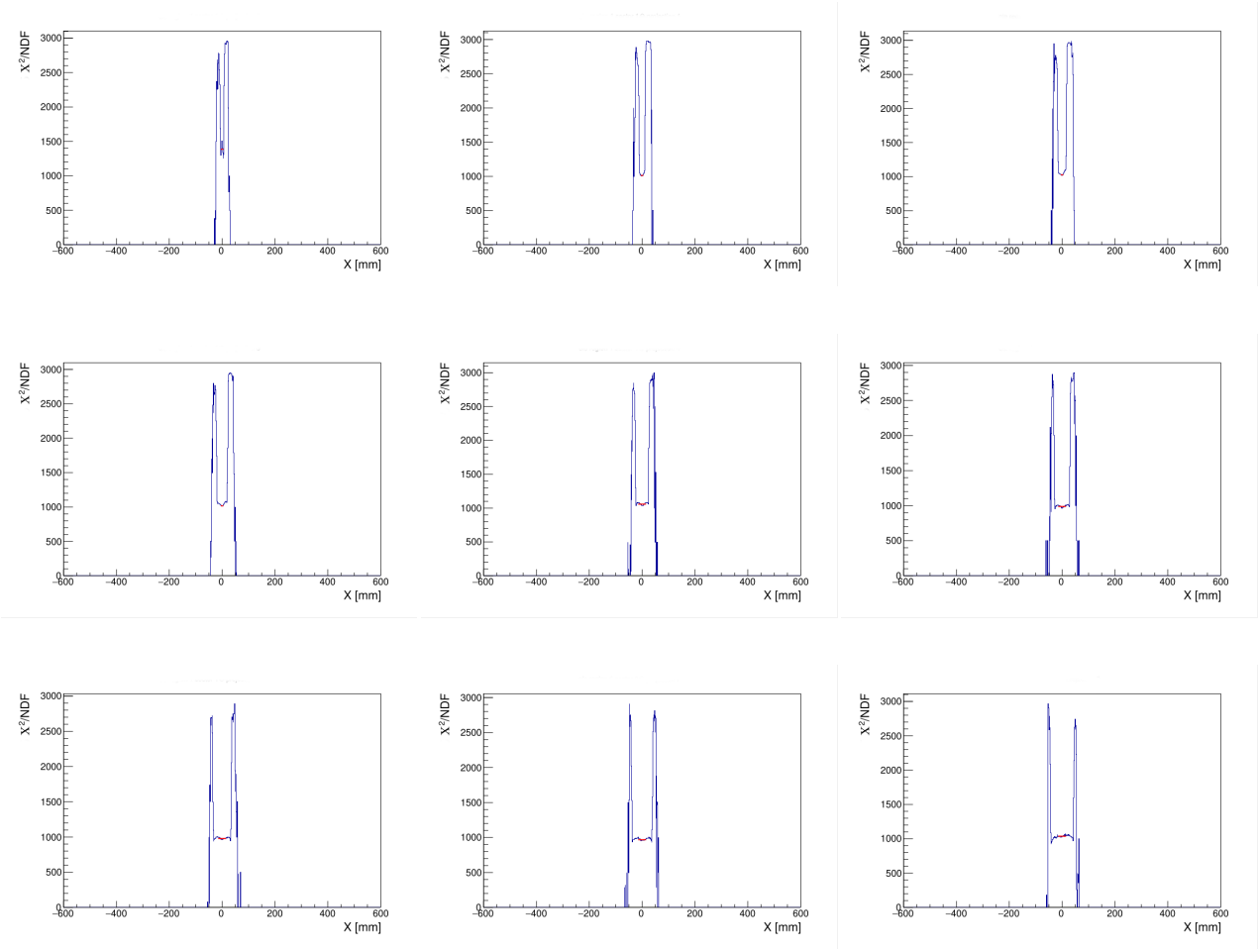


Figure 3.9: Tracking χ^2/NDF distributions as a function of local X in Y bins in region 1 sector 1 for outbending electrons

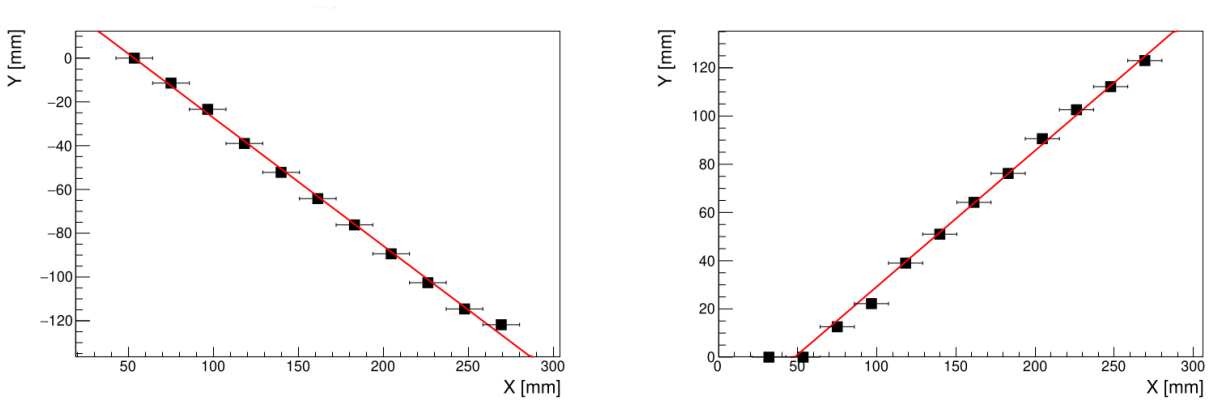


Figure 3.10: An example for the fitting of the cut boundary in region 3 sector 3 for outbending electrons

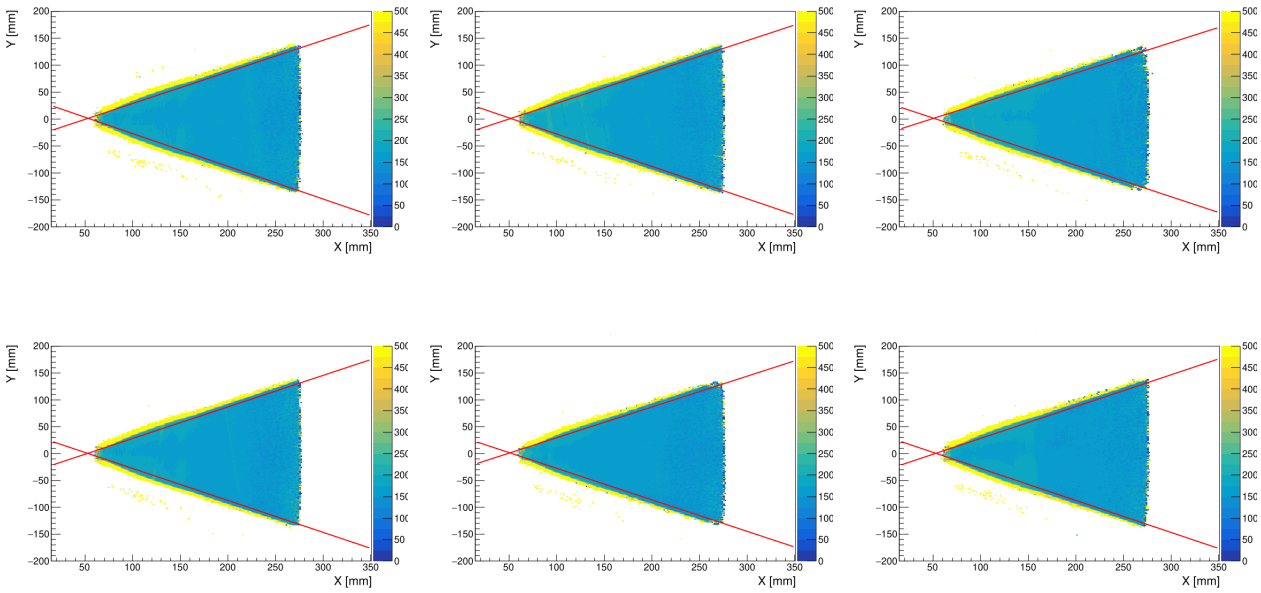


Figure 3.11: Average tracking χ^2/NDF distributions in local X-Y bins and the fiducial cuts sector by sector in region 3 for outbending electrons

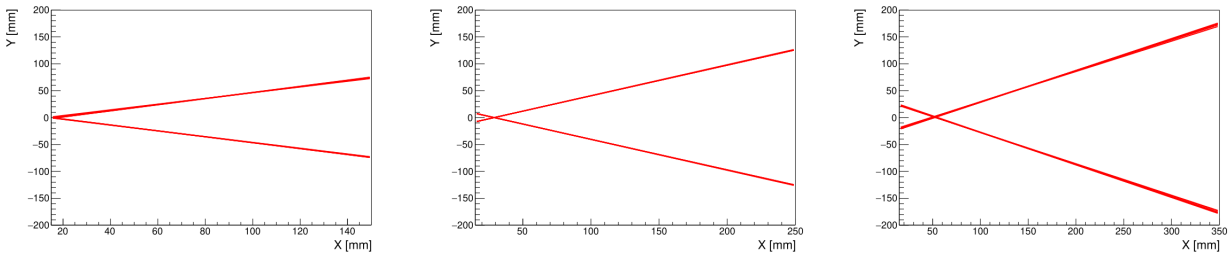


Figure 3.12: Sector comparison of fiducial cuts in region 1 (left), 2 (middle) and 3 (right) for outbending electrons

	Electron	π^+	π^-	K^+	K^-
Region 1	87.6 %	97.4 %	96.8 %	96.3 %	93.7 %
Region 2	87.4 %	95.9 %	95.9 %	95.9 %	96.0 %
Region 3	96.4 %	97.4 %	98.6 %	96.3 %	97.6 %

Table 3.2: Survival rates of the outbending DC fiducial cuts

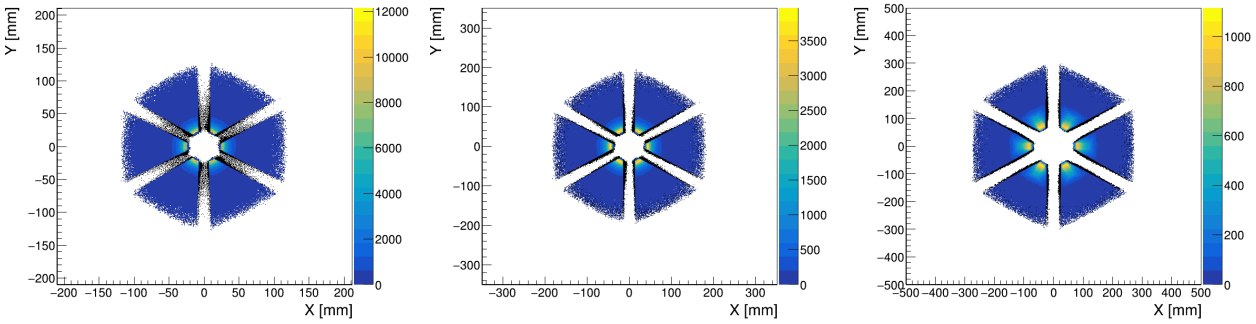


Figure 3.13: Hit distributions of electrons in the 3 regions of the drift chambers with outbending setting (region 1: left, region 2: center, region 3: right). The distribution without the DC fiducial cut is shown as black points, while the distribution after the cut is overlaid as colored points.

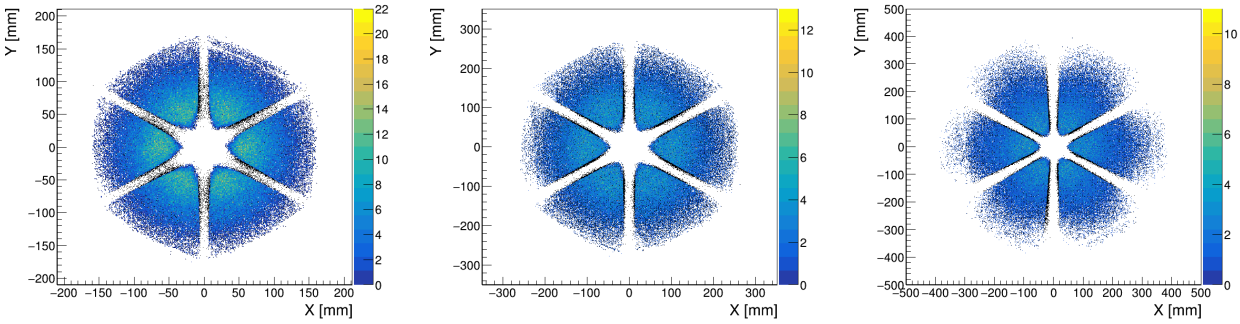


Figure 3.14: Hit distributions of positive kaons in the 3 regions of the drift chambers with outbending setting (region 1: left, region 2: center, region 3: right). The distribution without the DC fiducial cut is shown as black points, while the distribution after the cut is overlaid as colored points.

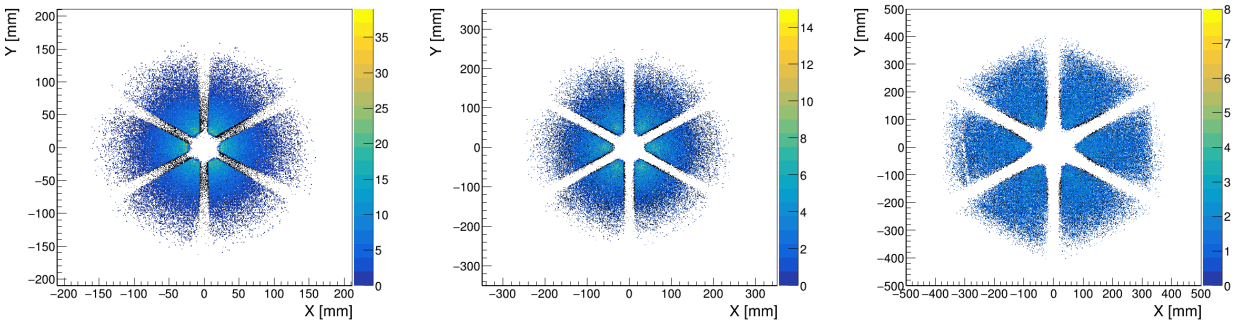


Figure 3.15: Hit distributions of negative kaons in the 3 regions of the drift chambers with outbending setting (region 1: left, region 2: center, region 3: right). The distribution without the DC fiducial cut is shown as black points, while the distribution after the cut is overlaid as colored points.

3.3.4 Event Builder PID

As a basis for the particle identification, the PID information provided by the event-builder (EB) [77] is used. The EB service is part of the COATJAVA framework written in Java using the CLARA framework [78].
The reconstruction of the particles starts with the track reconstruction in the DC. For each reconstructed track a charged particle is created. For each track an association with the other non-tracking detectors is performed.

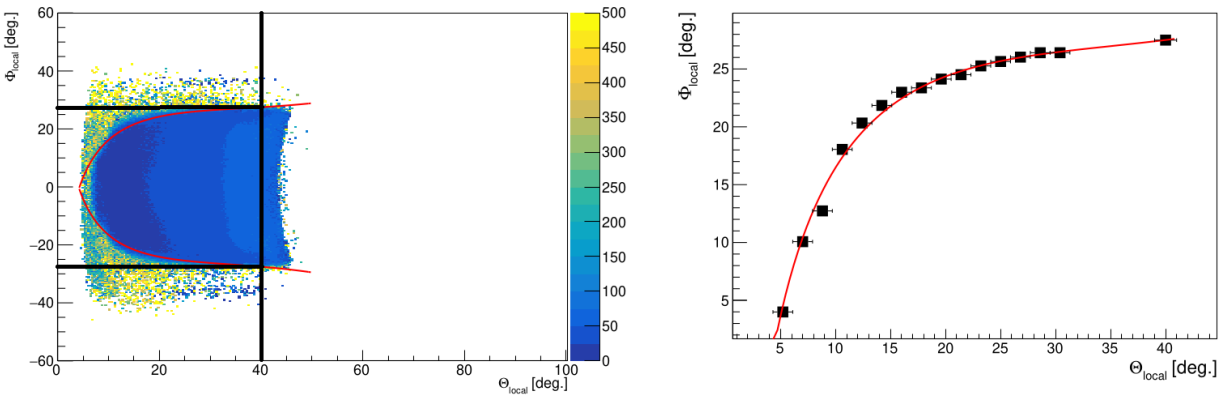


Figure 3.16: The fixed points determined in black and the fitted cut boundaries in red on the left. An example for the fitting of the cut boundary in region 3 sector 1 for inbending π^+ on the right. The local angles are defined as polar and azimuth angles of the local X, Y, and Z coordinates.

This association is based on loose geometric matching on the distance of closest approach with fixed windows. After all charged particles are matched, the neutral candidates are created from the remaining unassociated calorimeter responses. The event start time is determined from the RF signal and the best detected particle. This particle must have a track in the DC with an associated FTOF hit in panel 1B or 1A and it must be identified as an electron or if there is none, then a charged pion. [77]

Electron identification

The electrons are the first particles to be assigned. They must have an associated FTOF hit, create at least 2 photoelectrons in the HTCC, deposit at least 60 MeV energy in the PCAL and the measured sampling fraction (ratio of the measured momentum and the measured total energy in all calorimeter layers) is consistent with the parameterized momentum-dependent sampling fraction within 5 sigmas. Both the sampling function and the resolution is parametrized with the same functional form: $SF = p_1(p_2 + p_3/E + p_4/E^2)$. The fitted parameters are stored in the CLAS Calibration Database (CCDB) and are regularly updated. [77]

Hadron identification

All other charged particles which have an FTOF hit and do not satisfy the conditions for electrons are identified as a pion, a kaon, a proton or a deuteron. The identification is purely based on the difference between the event start time and the vertex time of the given particle. Only one FTOF timing response is used with the following prioritization: 1B, 1A, 2. A PID χ^2 value is assigned as the number of sigmas from the expected vertex time for the best hypothesis ($\Delta t/\sigma$). [77]

3.3.5 PID refinements

Since the EB PID is relatively loose, additional PID refinements have been added to improve the purity of the SIDIS sample. The refinements are used commonly for most CLAS12 analyses [76]. This section will provide a short summary.

Electron PID refinement

Since an electron scattering experiment has been performed, the detection of the scattered electron is a key requirement for this analysis. In addition to the event-builder, the following cuts have been applied:

- PCAL minimum energy deposition cut
- Electromagnetic calorimeter sampling fraction cut
- z-vertex position cut

PCAL minimum energy deposition cut

The minimal energy deposition of the PCAL, which is set to 0.06 GeV in the event-builder has been tightened to 0.07 GeV to improve the pion rejection from the electron sample. The cut is shown on figure 3.17. [76]

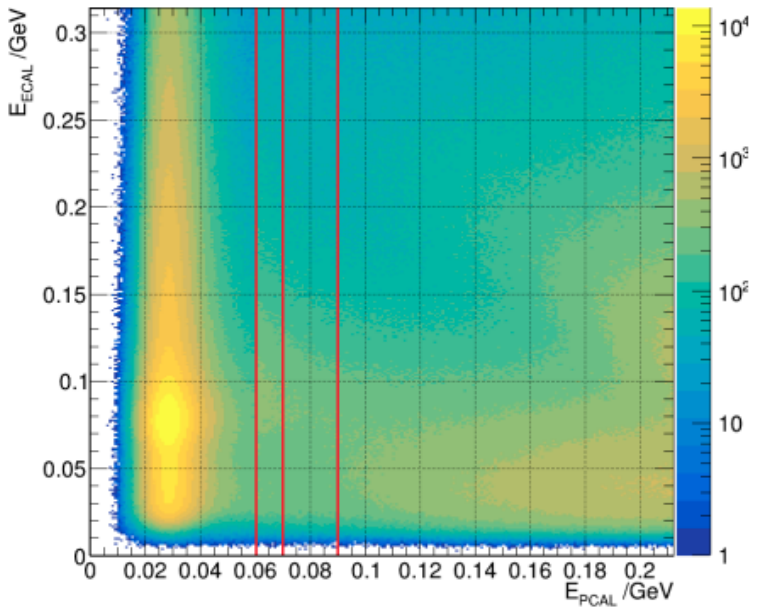


Figure 3.17: Energy deposition in ECin + ECout vs the energy deposition in the PCAL [76]

Electromagnetic calorimeter sampling fraction cut

A cut on the sampling fraction is important to control the pion contamination in the electron sample. Based on a fit of the total sampling fraction $(E_{PCAL} + E_{inner} + E_{outer})/p$ in momentum slices, a momentum dependent 3.5 σ band cut around the mean value is applied. This cut already helps to reduce the pion contamination. However detailed contamination studies showed, that above the HTCC threshold at 4.5 GeV a significant pion contamination is present. Therefore another cut has been placed on the correlation between the sampling fractions of the inner calorimeter and the PCAL. Detailed studies showed, that a diagonal cut on $\frac{E_{inner}}{p} < 0.2 - \frac{E_{PCAL}}{p}$ can remove most of the remaining pion contamination above the HTCC threshold. Therefore this additional cut has been applied for p > 4.5 GeV. The effect of the cut is illustrated by figure 3.18 [76]

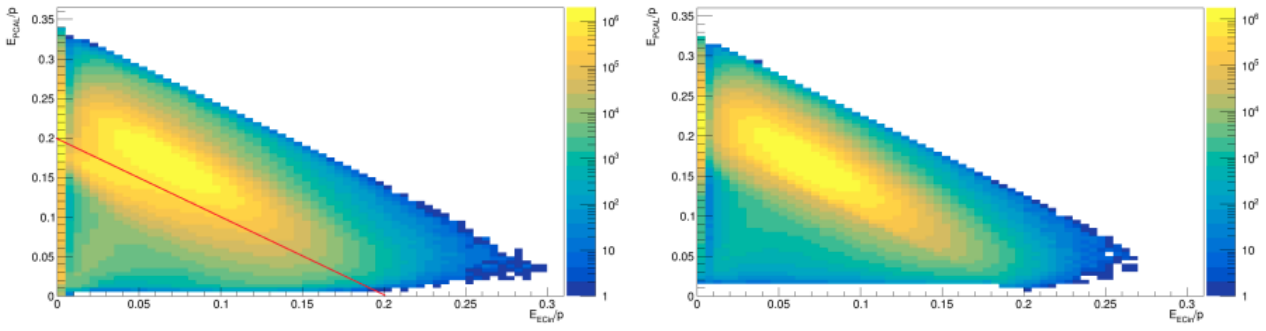


Figure 3.18: Correlation between the sampling fraction of the inner calorimeter and the PCAL before any cut (left) and after a the 3.5 σ sampling fraction cut and the triangular cut on this correlation for p > 4.5 GeV (right). [76]

z-vertex position cut

For the in/outbending data set, only electron candidates which have a z-vertex $v_z \in$ [-13 cm, +12 cm] or [-18 cm, +10 cm] are accepted respectively to reject electrons originating from interaction within the beam pipe or with the exit window of the target cell. The distributions are shown on figure 3.19. [76]

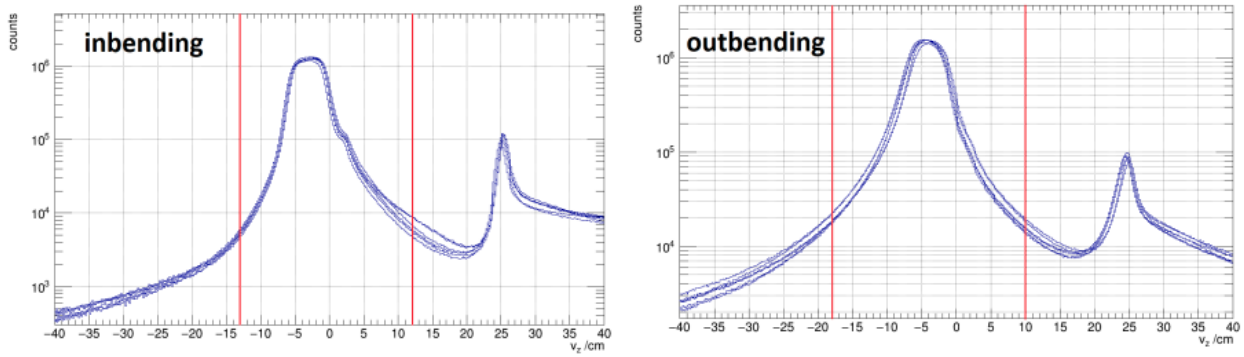


Figure 3.19: Distribution of the z-vertex position for electrons with an inbending torus field (left) and with an outbending torus field (right) with the cuts shown in red. [76]

Hadron PID refinement

The hadron identification of CLAS is mainly based on the time of flight information from the time of flight counters (TOF). In the following paragraphs the refinement cuts are shortly summarized. In addition to the event-builder, the following cuts have been applied:

- A cut on the vertex difference between electrons and hadrons
- $|\chi^2_{PID}| < 3$ from the event-builder

A cut on the vertex difference between electrons and hadrons

To reject hadrons which are produced outside the target region, a cut on the vertex difference between electrons and hadrons is applied. A cut is placed at $|v_z^{(electron)} - v_z^{(hadron)}| < 20$ cm. [76]

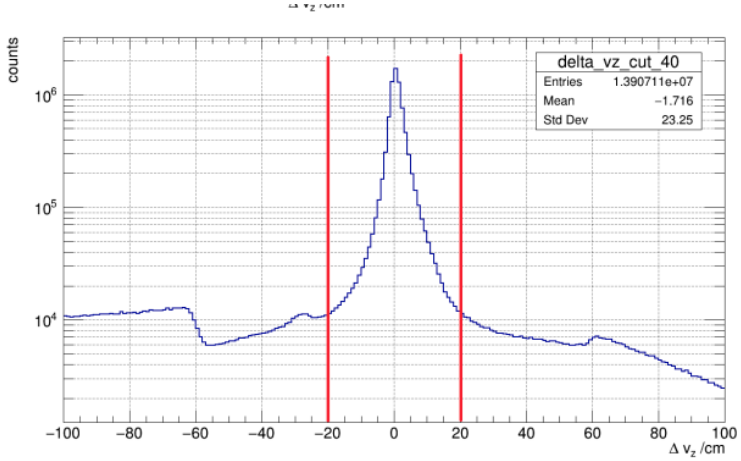


Figure 3.20: Vertex difference between the positive kaons and the electron. The cut positions are marked as vertical red lines. [76]

$|\chi^2_{PID}| < 3$ from the event-builder

As mentioned previously, the event-builder assigns a χ^2_{PID} value as the number of sigmas from the expected vertex time for the best hypothesis ($\Delta t/\sigma$) to identified hadrons. To reject contamination from other hadrons a $|\chi^2_{PID}| < 3$ was placed, which is especially important at high momenta. A more detailed description and analysis can be found in section 3.4.6.

3.3.6 Total survival rates of the fiducial cuts and PID refinements

The total survival rates of the PID refinements and fiducial cuts for the experimental data and the MC simulation are summarized in Table 3.3 for electrons and in Tables 3.4 and 3.5 for kaons. The output of the event-builder PID is considered as 100 % since it is used as a basis for the cuts and refinements. A comparison between experimental data and MC simulation is given.

	inbending data	inbending MC	outbending data	outbending MC
PCAL fiducial	89.8%	83.1%	99.9%	99.6%
DC region 1 fiducial	98.3%	85.0%	91.1%	87.4%
DC region 2 fiducial	96.2%	92.9%	93.8%	88.2%
DC region 3 fiducial	99.3%	98.0%	99.3%	97.5%
All applied cuts	79.1%	65.6%	82.8%	76.3%

Table 3.3: Survival rates for electrons. Except for the last row, the survival rates are for the stated cut only.

	inbending data	inbending MC	outbending data	outbending MC
DC region 1 fiducial	97.0%	97.5%	96.7%	88.3%
DC region 2 fiducial	99.8%	99.5%	95.5%	97.7%
DC region 3 fiducial	99.8%	98.7%	96.1%	96.7%
All applied cuts	82.0%	83.2%	71.4%	71.2%

Table 3.4: Survival rates for positive kaons. Except for the last row, the survival rates are for the stated cut only.

	inbending data	inbending MC	outbending data	outbending MC
DC region 1 fiducial	94.3%	94.1%	91.3%	91.7%
DC region 2 fiducial	96.5%	97.6%	93.7%	94.4%
DC region 3 fiducial	96.3%	98.1%	94.9%	86.2%
All applied cuts	63.7%	61.0%	79.5%	60.2%

Table 3.5: Survival rates for negative kaons. Except for the last row, the survival rates are for the stated cut only.

3.3.7 Selection of the topology

After the fiducial cuts and particle ID refinements are applied, only events are considered which have at least one good (survived the fiducial and refinement cuts) electron. In addition it is required, that the event contains at least one good kaon. If an event contains more than one kaon, all of them are used for the extraction of the BSA. In case of multiple electrons, the one with the highest energy is used in an event.

3.4 Improving kaon-pion identification with Machine Learning Techniques

Above 3 GeV momentum the kaons in CLAS12 are almost indistinguishable from pions with a pure time of flight based particle identification as shown in Figure 3.21. Machine learning methods which could complement the other traditional particle identification methods were investigated. In this section these methods are described and compared, followed by detailed checks of the best method.

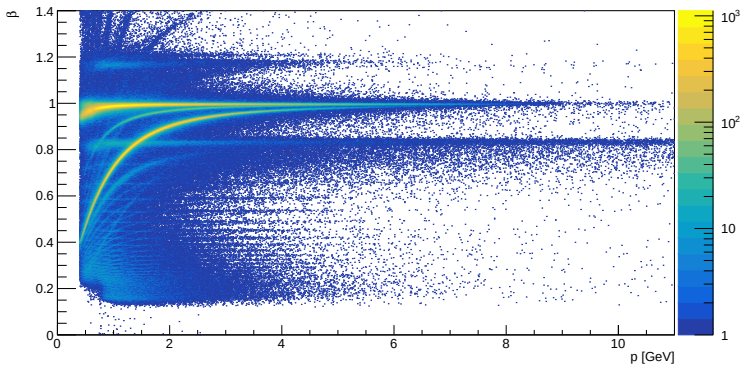


Figure 3.21: β versus momentum for positively charged particles

3.4.1 Monte Carlo Matching

For the training of the machine learning algorithms, a common CLAS12 large scale MC SIDIS production was used [79]. The same amount of kaons and pions were used for the training. To match the detected particles with the generated ones the following procedure was applied:
The first 60 reconstructed particles and the first 60 generated particles of each event were considered during the matching. Those reconstructed particles which had a momentum of 0 and those generated particles which had an energy of 0 were removed from the sample. A 60x60 distance matrix was initialized with a very high value. The distance in 3 momenta was calculated for each generated and reconstructed particle pair: $d = \sqrt{(p_x^r - p_x^g)^2 + (p_y^r - p_y^g)^2 + (p_z^r - p_z^g)^2}$. Also the ϕ and the θ difference was calculated for each pair. If the ϕ difference was less than 2° and the θ difference was less than 1° than the difference in momenta was put into the distance matrix at the corresponding position. After all pairs were processed the smallest value in the matrix was found and the corresponding particles were matched. All matrix elements in that row and column were set to a very high value. This procedure was repeated while the found minimum value was less than 5. Since most of the events usually contain just a few high energetic electrons, kaons or pions, this matching procedure is reliable and applicable for this sample.

3.4.2 Event selection

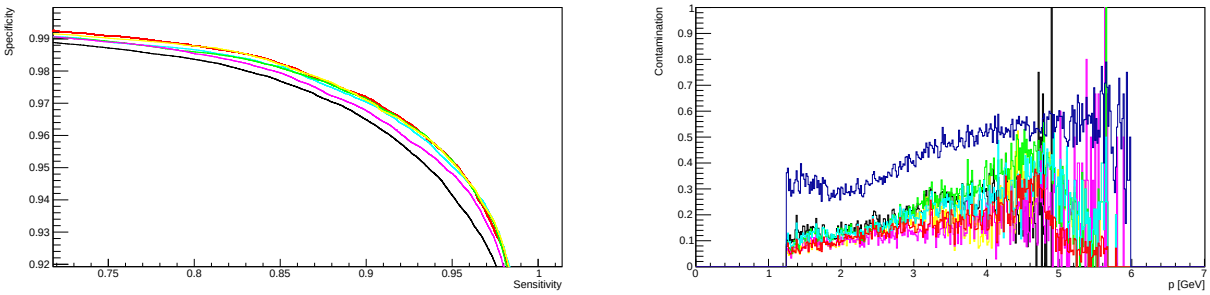
Those events were selected where at least one electron with an energy of at least 2 GeV was reconstructed. The matched positive kaons and pions were used during the training, if they satisfied the following criteria: azimuth angle is $5° < \theta < 35°$, $0 < \beta < 1.1$, energy depositions, time information and cluster momenta from the ECal and the number of photoelectrons in the HTCC are non negative and none of the values are NaN. This way it is assured that the machine learning algorithms are not mislead by wrongly reconstructed values.
The same number of kaons and pions were selected randomly for training and for validation. All kaons were used, 150000 for validation and all the rest in the input file for training. By using the same number of signal and background, the machine learning method will really try to label the input based on correlations in the input variables and not just label everything with the label that has the majority in the sample.

3.4.3 Used variables

The following variables were used as input during the training:

- EventBuilder PID
- Momentum and β
- Deposited energies in the 3 calorimeter layers (ECin, ECout, PCAL)
- Calorimeter time information: path/time
- Cluster moments: m2u, m2v, m2w, m3u, m3v, m3w
- HTCC number of photoelectrons and time information: path/time
- Energy depositions and time information (path/time) in the 3 FTOF layers

The simulated and measured distributions of these variables were compared and the only difference found was a consequent shift in the times.
To check the effect of the variables, the best method (described in section 3.4.5) was trained with removing the variables one by one. The resulting ROC curves and the obtained contaminations from simulations (described in section 3.4.6) can be seen in Figure 3.22. The resulting curve is red, when all variables are used during the training, yellow, when the EventBuilder PID and the χ^2_{PID} is not used, green, when the FTOF is not used, blue, when the calorimeter is not used, purple, when the HTCC is not used and black, when only the momentum and β is used. In the contamination plot the results with the EventBuilder PID are plotted in dark blue for comparison.



ROC curves

Contamination obtained from simulated data

Figure 3.22: Comparison of the effect of the different variables used for training (The resulting curve is red, when all variables are used during the training, yellow, when the EventBuilder PID and the χ^2_{PID} is not used, green, when the FTOF is not used, blue, when the calorimeter is not used, purple, when the HTCC is not used and black, when only the momentum and β is used. In the contamination plot the results with the EventBuilder PID are plotted in dark blue for comparison.)

Here it can be observed, that the inclusion of the Ecal and the FTOF variables improves the separation above 3 GeV. Below that they have little effect. The HTCC variables decrease the contamination significantly above 4.5 GeV and have absolutely no effect below that momentum, as expected. Using only the momentum and β gives good results below 4 GeV. Above that the neural network removes all statistic, since it is unable to reliably distinguish kaons and pions. The inclusion of the EventBuilder PID and χ^2_{PID} doesn't influence the results, but they reduce the training speed.
The separation power of the variables measured by TMVA during the training of the best method can be seen in Table 3.6. The variables with the most separation power are the ones which contain time of flight information from different sources. Also the energy depositions are important for separation kaons from pions.

Rank	Variable	Separation power
1	pid	7.983e-01
2	beta	3.394e-01
3	p	1.364e-01
4	FTOF path layer1/FTOF time layer1	7.478e-02
5	FTOF path/FTOF time	6.414e-02
6	Cal PCAL path/Cal PCAL time	5.105e-02
7	FTOF path layer3/FTOF time layer3	2.984e-02
8	Cal ECin path/Cal ECin time	2.370e-02
9	FTOF energy layer3	2.333e-02
10	Cal energy total	2.147e-02
11	Cal ECout path/Cal ECout time	1.977e-02
12	Cal PCAL energy	1.738e-02
13	FTOF energy	7.069e-03
14	Cal PCAL m2u	4.811e-03
15	Cal PCAL m2w	4.509e-03
16	Cal PCAL m2v	4.103e-03
17	CC HTCC nphe	3.654e-03
18	Cal ECin energy	3.529e-03
19	FTOF energy layer1	2.954e-03
20	Cal ECin m2u	2.505e-03
21	CC HTCC path/CC HTCC time	2.301e-03
22	Cal ECin m2w	1.990e-03
23	Cal ECin m2v	1.922e-03
24	Cal ECout energy	1.869e-03
25	Cal PCAL m3w	1.646e-03
26	Cal PCAL m3u	1.621e-03
27	Cal PCAL m3v	1.582e-03
28	chi2pid	8.683e-04
29	Cal ECout m2u	6.321e-04
30	Cal ECout m2v	4.933e-04
31	Cal ECout m2w	4.594e-04
32	Cal ECout m3v	3.670e-05
33	Cal ECout m3w	3.665e-05
34	Cal ECout m3u	2.952e-05
35	Cal ECin m3w	2.747e-05
36	Cal ECin m3u	2.472e-05

Table 3.6: Separation power of the variables used during the training

3.4.4 Root TMVA

The root TMVA (Toolkit for Multivariate Data Analysis with ROOT [80]) was chosen as a machine learning environment for building, training and evaluating multivariate classification algorithms. It uses root trees as input and output format and it is easy to include the application of the methods into a C++ analysis code. For these reasons TMVA was an ideal choice in our case.

3.4.5 Method selection and description

Many methods available in TMVA were tested. The settings can be found in Table 3.7. The methods were tested on a smaller dataset. According to Figure 3.23, the best method is the custom built neural network across the whole momentum range.

Type	Preprocessing	Settings
BDT - Boosted Decision Trees		NTrees=200:MinNodeSize=2.5%:MaxDepth=2: BoostType=AdaBoost:AdaBoostBeta=0.5: UseBaggedBoost:BaggedSampleFraction=0.5: SeparationType=GiniIndex:nCuts=20
MLP - Multi-Layer Perceptron	Normalization	NCycles=100:HiddenLayers=N+5:TestRate=5: !UseRegulator
Fisher - Fisher determinant	Gaussianisation +Decorrelation	
RuleFit		
DNN - Deep Neural Network	Normalisation	Layout=TANH|128,TANH|128,TANH|128

Table 3.7: Machine learning methods tested

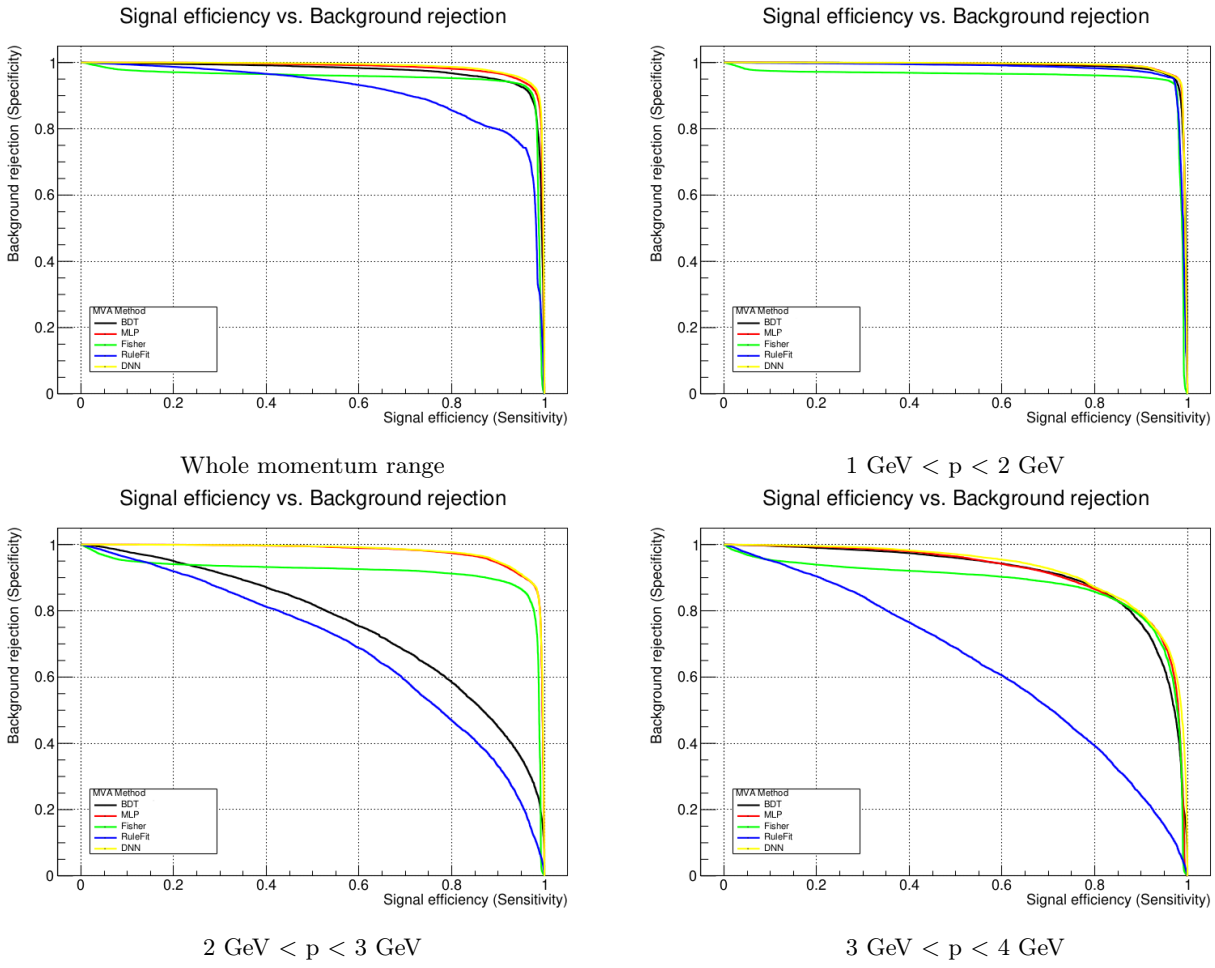
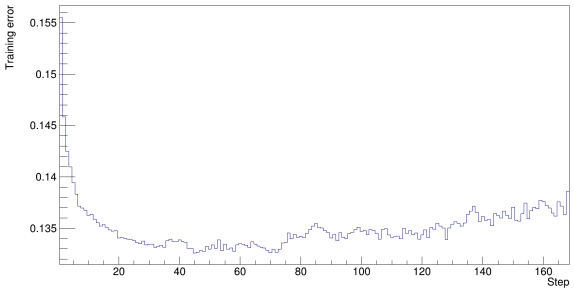


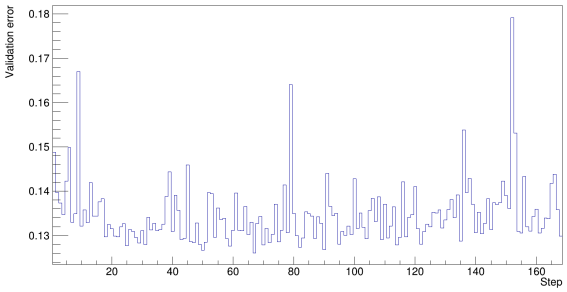
Figure 3.23: ROC curves comparing different methods and momentum ranges. The detailed description of the methods can be found above in Table 3.7.

This fully connected neural network consist of 3 hidden layers with 128 neurons per layer. The neurons have tanh activation function, except the last one, which is linear. The weights were initialized using the Xavier initialization method. They were optimized using the ADAM optimizer with cross-entropy loss function, with a batch size of 30 and with a learning rate of 10^{-5}. The input variables were normalized to 0-1 interval. To prevent over-fitting, the training was stopped after 50 steps, after both the training and the validation error started to rise (Figure 3.24). The resulting ROC curve can be seen in Figure 3.25 and the DNN response in

Figure 3.26. To achieve a clean sample, but to still keep reasonable statistic, the optimal cut on the response was determined to be 0.96. The efficiency and contamination as a function of the DNN response based on MC for K^+ is shown in Figure 3.27.

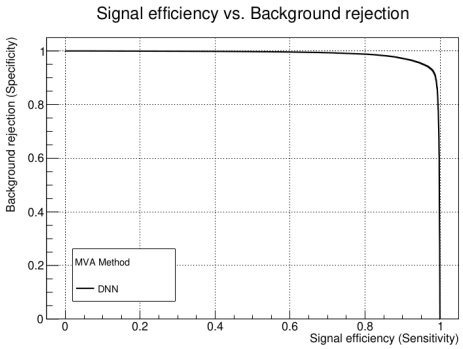


Training error

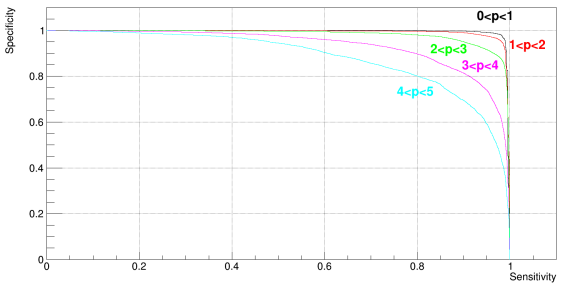


Validation error

Figure 3.24: Errors during the training

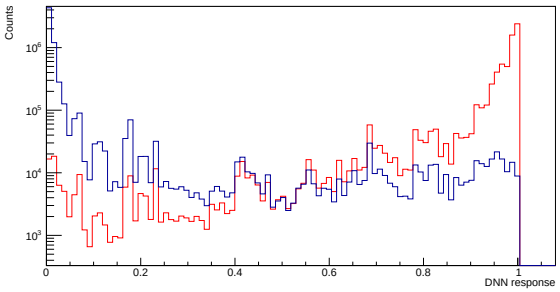


ROC curve for the whole momentum range

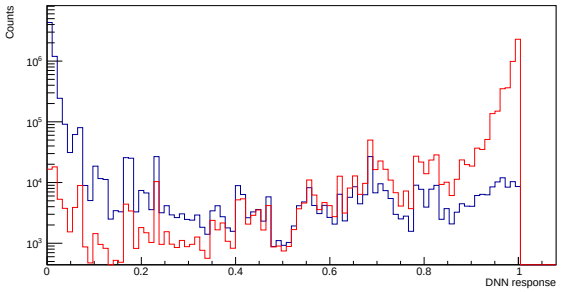


ROC curves at different momenta ranges

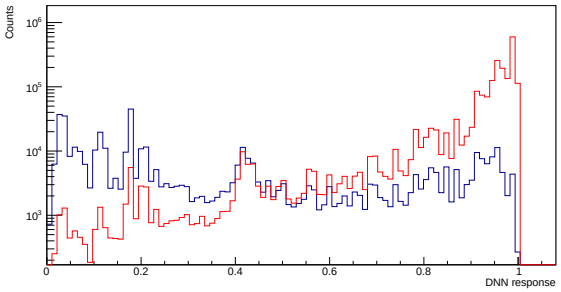
Figure 3.25: ROC curves of the final selected method



DNN response for the whole momentum range

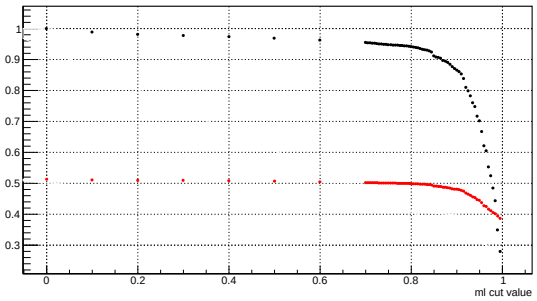


DNN response below 3 GeV momentum

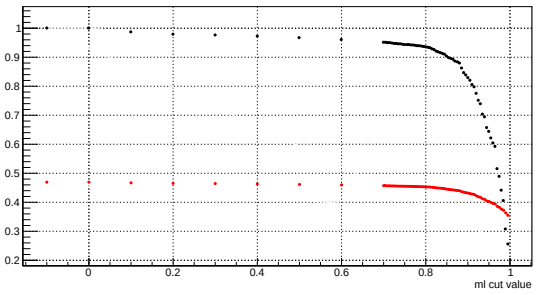


DNN response above 3 GeV momentum

Figure 3.26: DNN response for kaons (red) and for pions (blue)



Inbending



Outbending

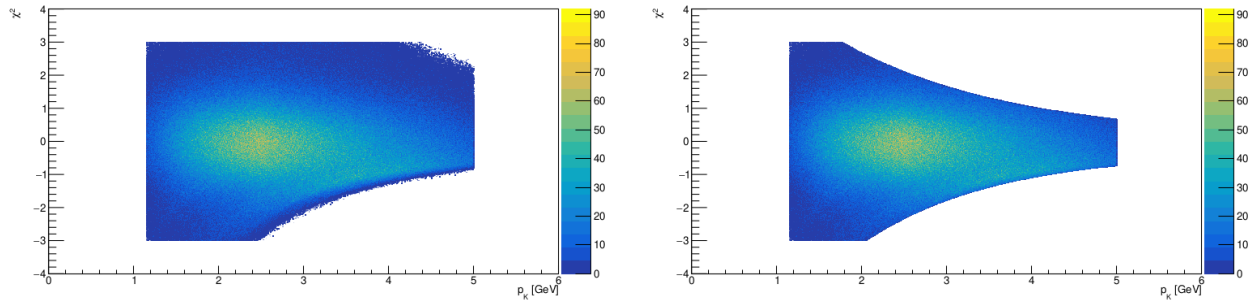
Figure 3.27: Efficiency (black) and contamination (red) as a function of the DNN response based on MC for K^+.

3.4.6 Checks with simulations

The performance of the neural network was tested on another Monte Carlo sample. The newest large scale SIDIS production was used for this purpose. The same matching and event selection was applied as previously, described in sections 3.4.1 and 3.4.2. The achieved efficiencies and the contaminations were measured for different pid criteria and methods:

- EventBuilder PID
- EventBuilder PID with $|\chi^2_{PID}| < 3$
- EventBuilder PID with momentum dependent χ^2_{PID} cut
- EventBuilder PID with $|\chi^2_{PID}| < 3$ and with neural network PID

A momentum dependent structure can be observed in the χ^2_{PID} distribution of the kaons. These χ^2_{PID} values were determined at different momentum values for both the pion and proton sides of the distribution. These points were fitted and the resulting functions are the borders of the momentum dependent χ^2_{PID} cut: $\chi^2_{PID} < 0.122 + 7.1 \cdot e^{-0.51 \cdot p}$ and $\chi^2_{PID} > -0.35 - 10.08 \cdot e^{-0.65 \cdot p}$. The cut borders can be seen in Figure 3.28.
The reconstruction efficiency is defined as number of reconstructed mc matched kaons / number of generated kaons, while the contamination is the number of reconstructed mc matched pions reconstructed as kaons / number of all reconstructed kaons. The measured values are shown in Figure 3.29. The red is the EB PID, the green is the EB PID with $|\chi^2_{PID}| < 3$, the black is with momentum dependent χ^2_{PID} cut and the blue is the EB PID with neural network.

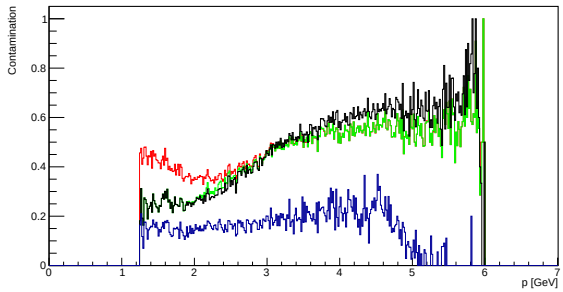


χ^2_{PID} distribution for kaons

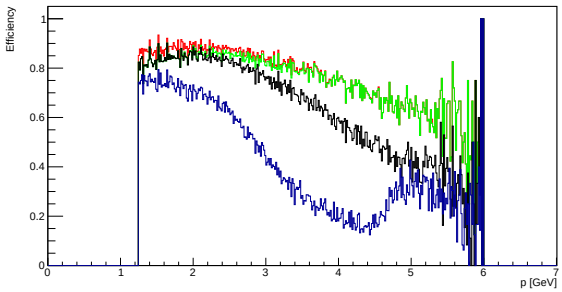
χ^2_{PID} distribution with the cut

Figure 3.28: Momentum dependent χ^2_{PID} cut

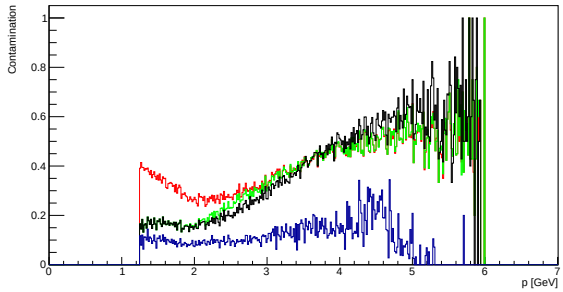
The $|\chi^2_{PID}| < 3$ criteria lowers the contamination at low momentum significantly and it only reduces the efficiency slightly. The momentum dependent $|\chi^2_{PID}|$ cut improves things only below 3 GeV and it shouldn't be used above that. With the neural network, the contamination can be kept below 20%. The efficiency drops continuously to 20% at 4 GeV, but it rises above 4.5 thanks to the HTCC. The results are very similar for both in- and outbending data.



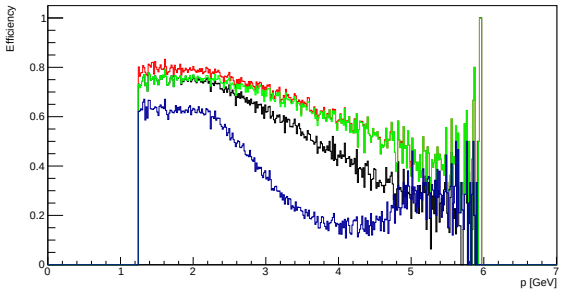
Inbending kaon contamination



Inbending kaon efficiency



Outbending kaon contamination



Outbending kaon efficiency

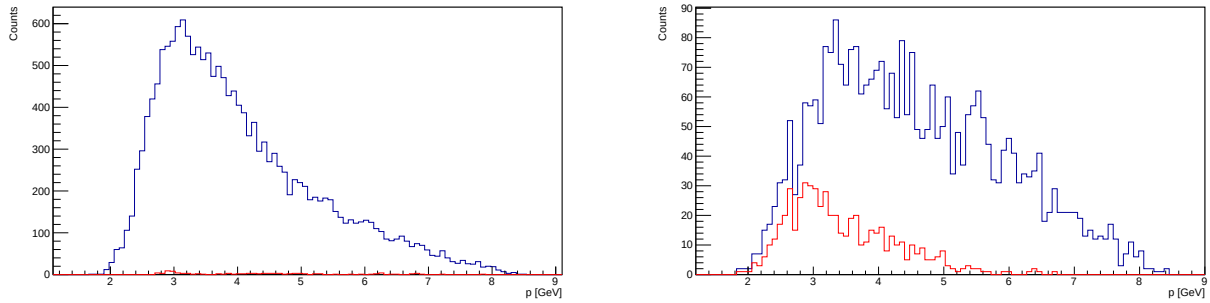
Figure 3.29: Kaon contamination and efficiency with different PID methods (see the description of the colors in the text above)

3.4.7 Checks with an exclusive channel

By using exclusive events, the pion rejection efficiency of the neural network can be checked. For that purpose events where at least one electron and at least one pion and at least one neutron was detected were selected. The following exclusivity cuts were used:

- 0.8 GeV $< M_{missing\ n^0} <$ 1.2 GeV
- -0.4 GeV $^2 < M^2_{missing\ \pi} <$ 0.5 GeV^2
- -0.01 GeV $^2 < M^2_{missing\ tot} <$ 0.005 GeV^2
- $E_{missing} <$ 0.5 GeV
- $p_{T,\ missing} <$ 0.15 GeV
- cone angle difference $< 3^\circ$

The same procedure was repeated with kaons treated as pions. After the tight exclusivity cuts it is very likely that all these EB kaons were actually misidentified pions. Figure 3.30 shows that the neural network removes almost all pions that were correctly identified by the EventBuilder and it also removes most of the misidentified pions as well below 5 GeV momentum. The ratio of the misidentified pions removed by the neural network is 77.7% in the whole momentum range.



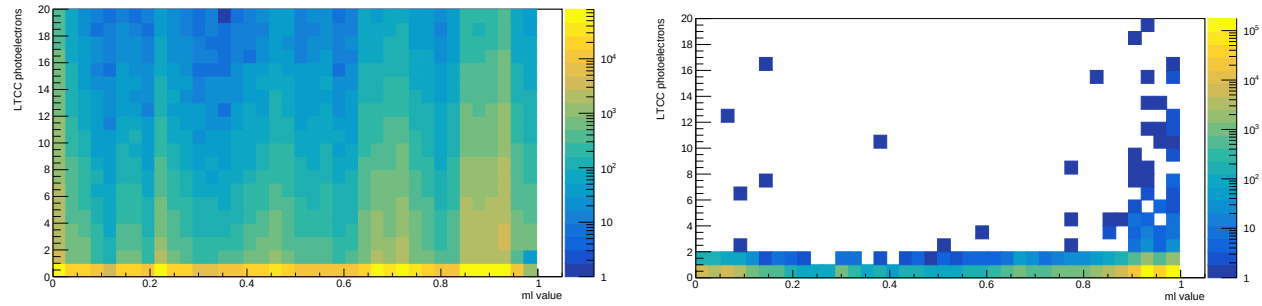
Correctly identified EB pions

Misidentified EB pions

Figure 3.30: Exclusive pions without (blue) and with (red) kaon neural network ID cut

3.4.8 Checks with the LTCC

The LTCC which is installed in sectors 3 and 5 can separate kaons and pions above 3 GeV. The number of photoelectrons in the LTCC was plotted against the output of the neural network in Figure 3.31. Here we can observe that above 0.9 return value the number of EB pions that didn't have an LTCC signal is much lower, then below 0.9. While in case of the EB kaons some of the particles with a signal remain after a cut at 0.96, a large portion of them can is correctly removed. Table 3.8 shows the correlation matrix of EB and LTCC PID (number of photoelectrons > 1) for particles in sectors 3 and 5, with momentum between 3 and 5 GeV with ML return value > 0.96. These numbers show that the NN effectively removes the LTCC pions regardless of the EB PID.



EB pions

EB kaons

Figure 3.31: Correlation of the number of LTCC photoelectrons and the neural network output

EB\LTCC	pion	kaon
pion	92	71
kaon	1	5479

Table 3.8: Correlation matrix of EB and LTCC PID (number of photoelectrons > 1) for particles in sectors 3 and 5, with momentum between 3 and 5 GeV with ML return value > 0.96

3.4.9 Checks with the RICH

For a more accurate check with the RICH detector, a rudimentary RICH reconstruction algorithm was developed. Skim files made for the RICH group were used. These files contained events, which had one trigger electron and exactly one hadron was detected in the RICH. They contain banks with information about the reconstructed photons in the RICH, besides all the standard banks.
For every EB pions, kaons and protons in sector 4 with $\theta < 16°$ the $cos\theta_C = \frac{1}{n\beta}$ Cherenkov angle was calculated for all 3 mass hypothesis with $n = 1.05$. The number of reconstructed photons with Cherenkov angles within $3\sigma_e$ were counted for all 3 Cherenkov angles. Here σ_e is the single photon resolution for electrons. The σ_e value was calculated based on a fit from the RICH technical design report [81], shown on Figure 3.32.

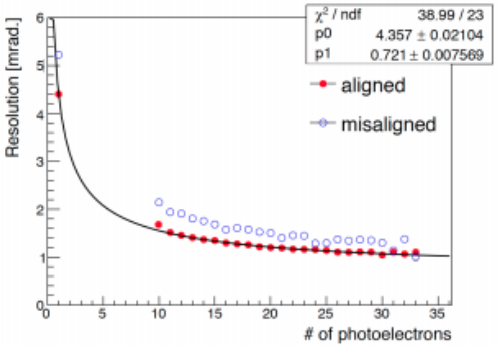


Figure 3.32: Single photon resolution for electrons from the RICH TDR [81]

The quality of the identification is defined with the ratio of the number of photons for the best and second best particle hypothesis as $Q_{ID} = 1 - \frac{N_{phot,2^{nd}}}{N_{phot,1^{st}}}$. If the best hypothesis has at least 4 photons and $Q_{ID} > 0.05$, then the particle is successfully identified.

The momentum distribution of the EB and the RICH identified kaons (Figure 3.33.) look very similar and their ratio is in good agreement with the contaminations measured in the simulations in section 3.4.6, showing that this simplified RICH reconstruction works.

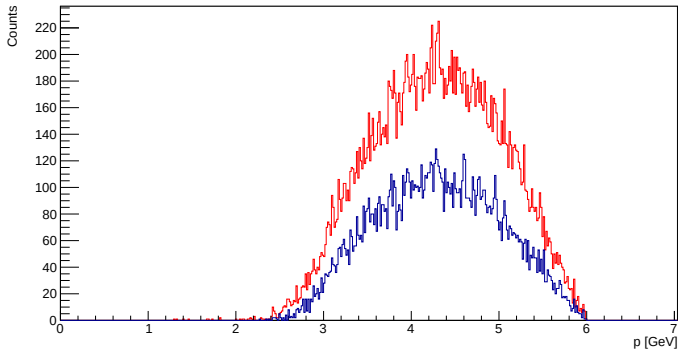
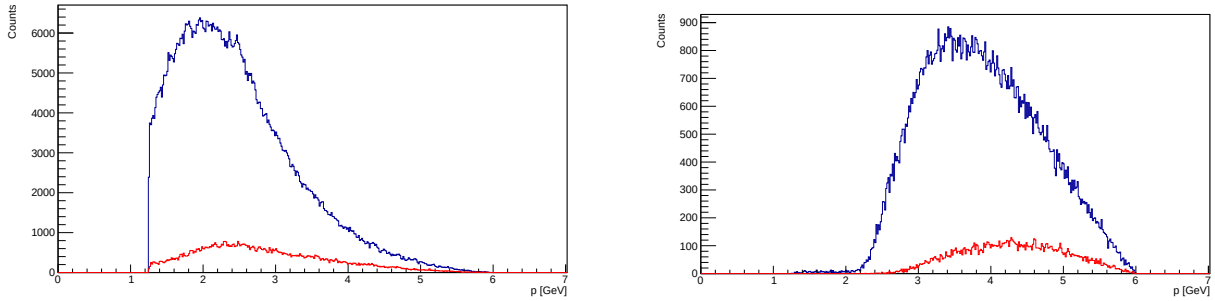


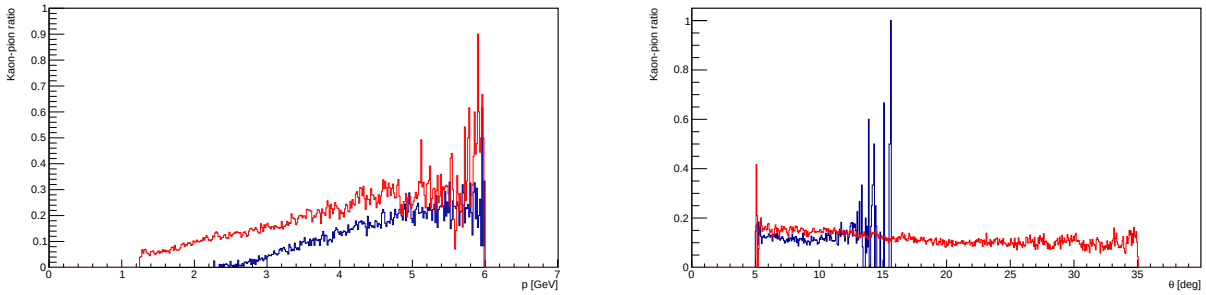
Figure 3.33: Momentum distribution of the EB (red) and the RICH (blue) identified kaons

Figure 3.34 compares the kaons and pions momentum distribution generated in the simulation and detected in the RICH, while Figure 3.35. compares the ratios directly. At low θ and high momentum, where the RICH is more reliable, the distributions and the ratios are very similar to the generated ones. Below 3 GeV the distributions look different, since the RICH files were skimmed to have exactly one hadron in the RICH per event. This means that the contamination and efficiency values from the RICH should only be trusted above 4 GeV momentum.



Monte Carlo RICH ID

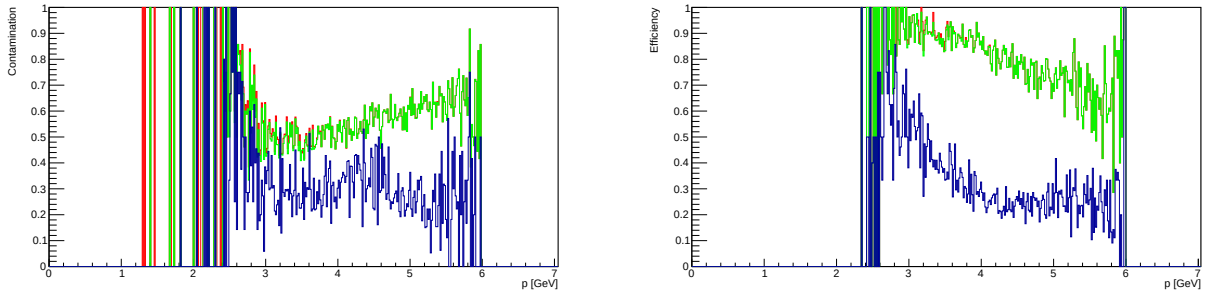
Figure 3.34: Momentum distribution of pions (blue) and kaons (red)



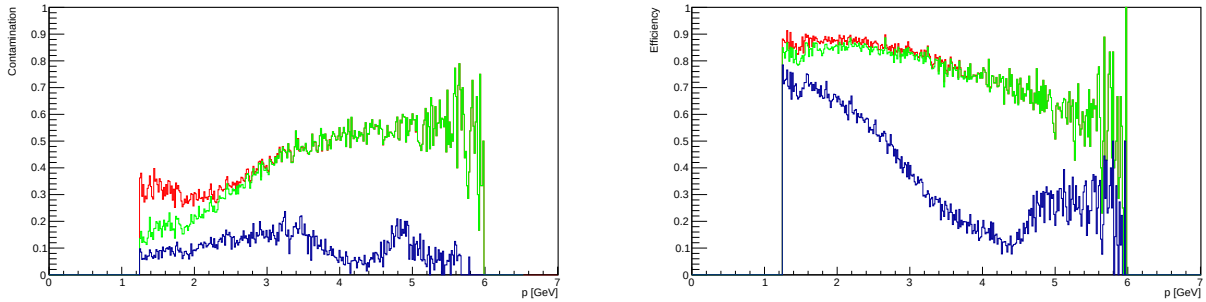
Momentum dependence with $5^\circ < \theta < 12^\circ$ θ dependence with 3 GeV<p<6 GeV

Figure 3.35: Ratio of pions and kaons generated in the simulations (red) and measured by the RICH (blue)

The reconstruction efficiency for RICH is defined as the number of reconstructed RICH kaons which were identified as kaons by the specific PID / number of RICH kaons, while the contamination is the number of reconstructed RICH pions identified as kaons by the specific PID / number of reconstructed particles identified as kaons by the specific PID. In Figure 3.36 the contamination and efficiency are compared for different PID methods using the RICH ID or the Monte Carlo information. The same methods are compared as in section 3.4.6, except the momentum dependent $|\chi^2_{PID}|$ cut. As before the red is the EB PID, the green is the EB PID with $|\chi^2_{PID}| < 3$ and the blue is the EB PID with neural network.
Based on the RICH ID we get slightly higher contamination and better efficiency for all methods. The improvement achieved with machine learning is not as significant according to the RICH ID as with the MC matching, but it is clearly visible.
Table 3.9 shows the correlation between different PID methods. The ML is usually in agreement with the EB, but it is much more accurate when compared with the RICH. The contamination of the kaon sample is the lowest, when the EB and the RICH is used together and the kaon efficiency is just slightly lower than using only the ML.



RICH ID contamination RICH ID efficiency



Monte Carlo contamination Monte Carlo efficiency

Figure 3.36: Comparison of different PID methods using the RICH ID or MC information (see the description of the colors in the text above)

EB\RICH	pion	kaon
pion	104860	2372
kaon	14454	11514

EB\ML	pion	kaon
pion	106832	441
kaon	19738	6279

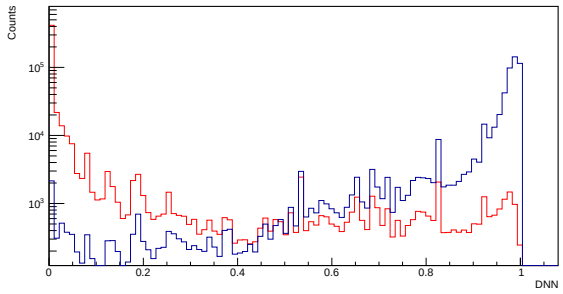
RICH\ML	pion	kaon
pion	118892	2401
kaon	9922	4371

RICH\EB+ML	pion	kaon
pion	119253	2040
kaon	10001	4292

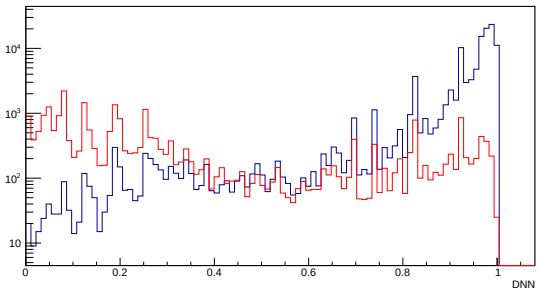
Table 3.9: Correlation matrices of PID methods

3.4.10 Negative kaons

Since the application of machine learning proved to be useful in separating positive pions and kaons, the applicability of the same technique for negative particles was investigated. The same procedure was repeated as described for positive kaons. The output of the neural network can be seen in Figure 3.37. Here the the cut value was chosen to be 0.95 to match the structure appearing in the distribution and to achieve a good compromise between efficiency and contamination. They are shown in Figure 3.38. The same studies for negative kaons were repeated with MC simulations and with RICH (described for positive kaons in sections 3.4.6 and 3.4.9). The efficiencies and contaminations can be found in Figure 3.39. In case of negative kaons the level of improvement achieved by the neural network is somewhat lower than for positive kaons. This is due to the lower kaon to pion ratio for negative particles, shown in Figure 3.40.
The contamination is much lower below 3 GeV momentum and the cut at 0.95 is almost identical to a hard cut at 3 GeV, so the pion contamination can also be handled as described in section 3.6. This results in a much lower statistics, than for positive kaons, but the asymmetries can still be extracted reliably.

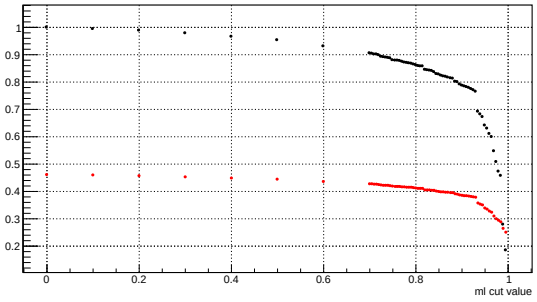


DNN response on the whole momentum range

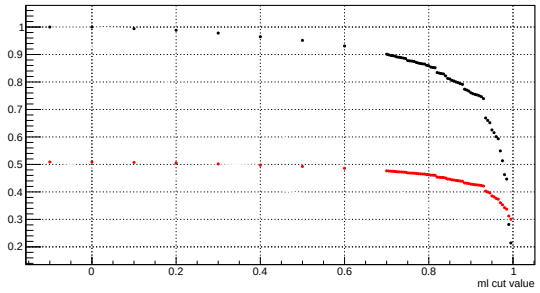


DNN response above 3 GeV momentum

Figure 3.37: DNN response for negative kaons (blue) and for pions (red)

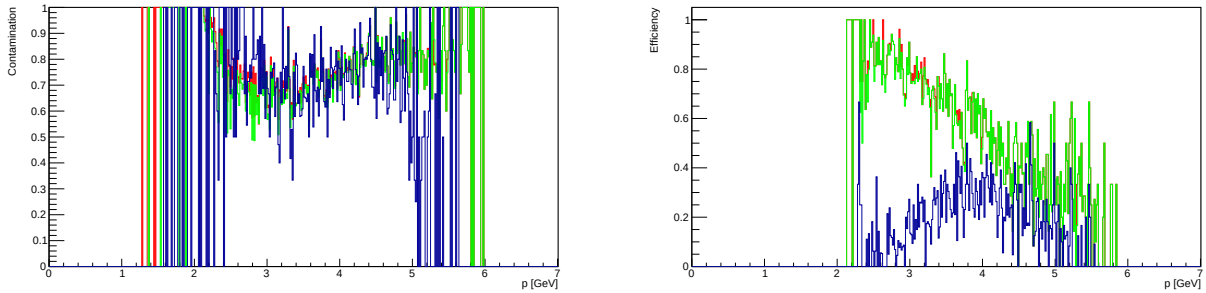


Inbending

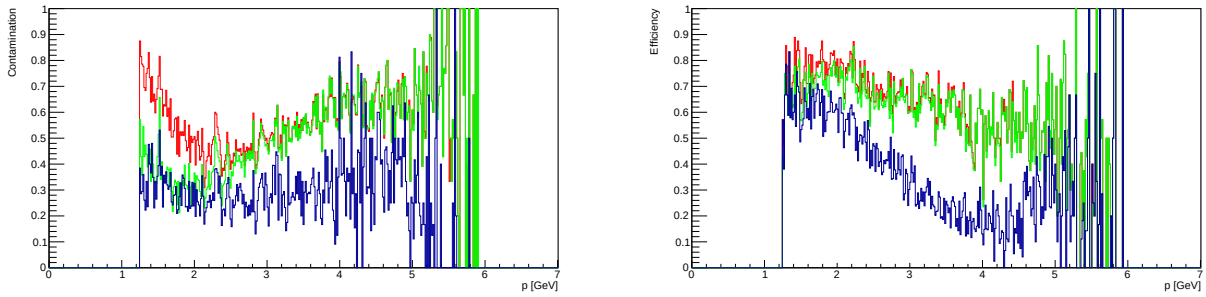


Outbending

Figure 3.38: Efficiency (black) and contamination (red) as a function of the DNN response based on MC for K^-.



RICH ID contamination RICH ID efficiency



Monte Carlo contamination Monte Carlo efficiency

Figure 3.39: Comparison of different PID methods using the RICH ID or MC information for negative kaons (see the description of the colors in the text above)

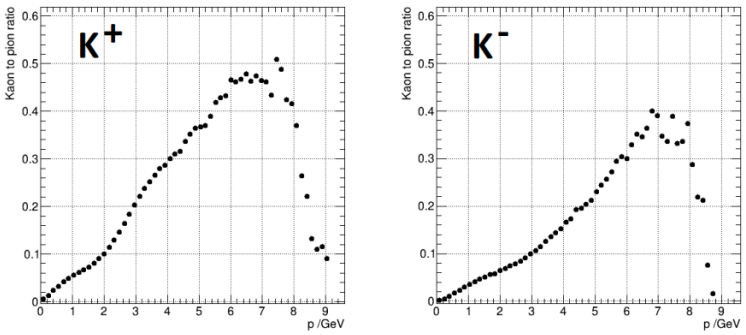


Figure 3.40: Kaon to pion ratio as a function of the particle momentum for positive kaons (left) and negative kaons (right). The ratio has been estimated from the generated MC events. [76]

3.4.11 Conclusion on the machine learning PID

A new particle identification method was developed using neural networks. Based on multiple independent tests it can efficiently reduce the pion contamination in the kaon sample in the whole momentum range, by still keeping the statistics on a reasonable level especially for positive kaons.

3.5 Analysis of the Beam Spin Asymmetry

The procedure used during the extraction of the Beam Spin Asymmetry and the obtained results are described in this chapter. Most of the methods and cuts follow the pion SIDIS analysis [35, 79] and the common analysis of the CLAS collaboration.

3.5.1 Kinematic cuts

The following kinematic cuts were applied on the single electrons and kaons in order to get a clean data sample:

- y < 0.75 (this corresponds to a minimal electron momentum of $\sim$2.65 GeV)
- kaon momentum: 1.25 GeV $< p_K <$ 3 or 5 GeV
- kaon polar angle: $5° < \theta_K < 35°$
- Electron polar angle: $5° < \theta_e < 35°$

To select the deep inelastic scattering region:

- W > 2 GeV
- $Q^2 > 1$ GeV^2

To reject the kaons from the fragmentation region a cut is applied on x_F Feynman x. A cut is also set on z, if the data is not binned in z:

- $x_F > 0$
- $0.3 <$ z

To reduce the contamination from exclusive processes a cut is also applied on the eKX missing mass:

- z < 0.7
- $M_X > 1.6$ GeV

The cut on the missing mass was determined to remove the structures that are clearly visible on the missing mass distributions shown in Fig. 3.41. To see the effect of this contamination on the results the $F_{LU}^{\sin\phi}/F_{UU}$ was extracted in bins of the missing mass. The results are shown in Fig. 3.42. At lower missing masses a falling (K^+)/rising (K^-) tendency can be observed, compared to the smoother behavior at higher values. There is an outlier value at 1.8 GeV missing mass, which was not removed, because cutting those events out would remove too much statistics.

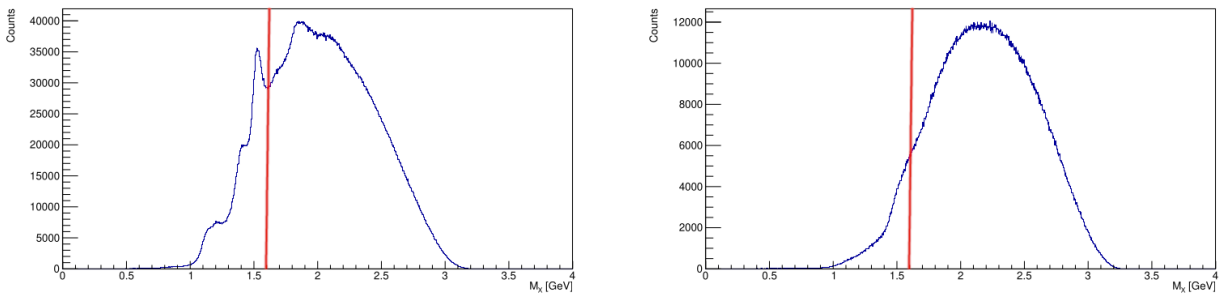


Figure 3.41: Missing mass distribution for e^-K^+X (left) and e^-K^-X (right). The cut borders are shown in red.

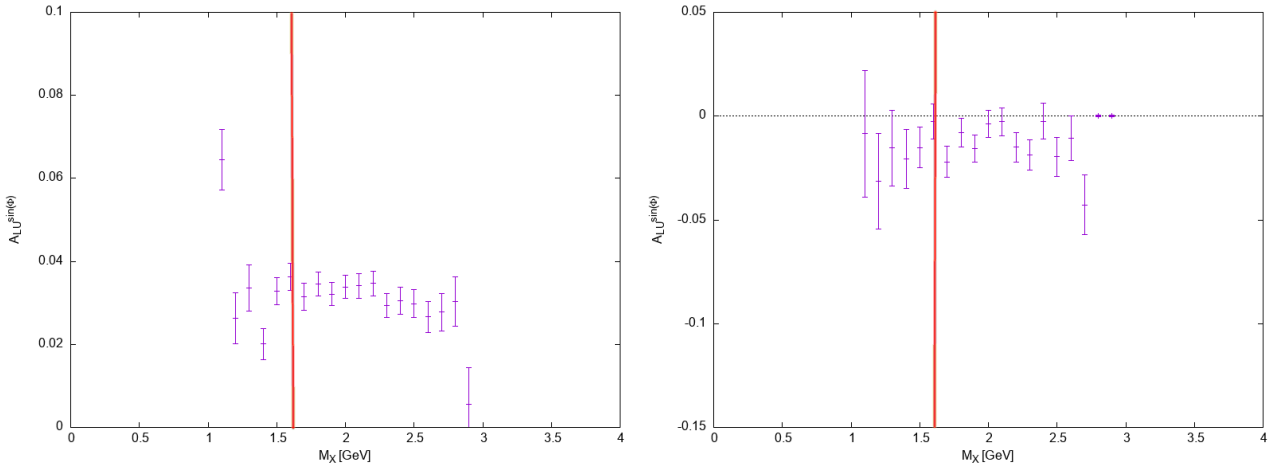


Figure 3.42: $F_{LU}^{\sin\phi}/F_{UU}$ as a function of missing mass for e^-K^+X (left) and e^-K^-X (right). The cut borders are shown in red.

3.5.2 Available statistics

The available statistics after the application of the cuts described above are shown in Table 3.10 for maximal kaon momentum of 3 GeV and in Table 3.11 using machine learning PID cut in addition and in Table 3.12 for maximal kaon momentum of 5 GeV and in Table 3.13 using machine learning PID cut in addition.

	inbending	outbending	spring2019	total
eK^+X	3188971	3861403	3672641	10723015
eK^-X	536002	2748619	631138	3915759

Table 3.10: Available statistics for maximal kaon momentum of 3 GeV

	inbending	outbending	spring2019	total
eK^+X	1858933	2289596	1942882	6091411
eK^-X	67075	223507	159743	450325

Table 3.11: Available statistics for maximal kaon momentum of 3 GeV using machine learning PID

	inbending	outbending	spring2019	total
eK^+X	4775497	5255214	5342699	15373410
eK^-X	765874	3761633	862771	5390278

Table 3.12: Available statistics for maximal kaon momentum of 5 GeV

	inbending	outbending	spring2019	total
eK^+X	2391101	2932680	2805885	8129666
eK^-X	143636	434968	256576	835180

Table 3.13: Available statistics for maximal kaon momentum of 5 GeV using machine learning PID

3.5.3 Particle distributions

The momentum, θ and ϕ distribution of electrons and kaons with all the applied cuts are shown in Fig. 3.43 for the e^-K^+X and in Fig. 3.44 for the e^-K^-X sample. The angles are in radian.

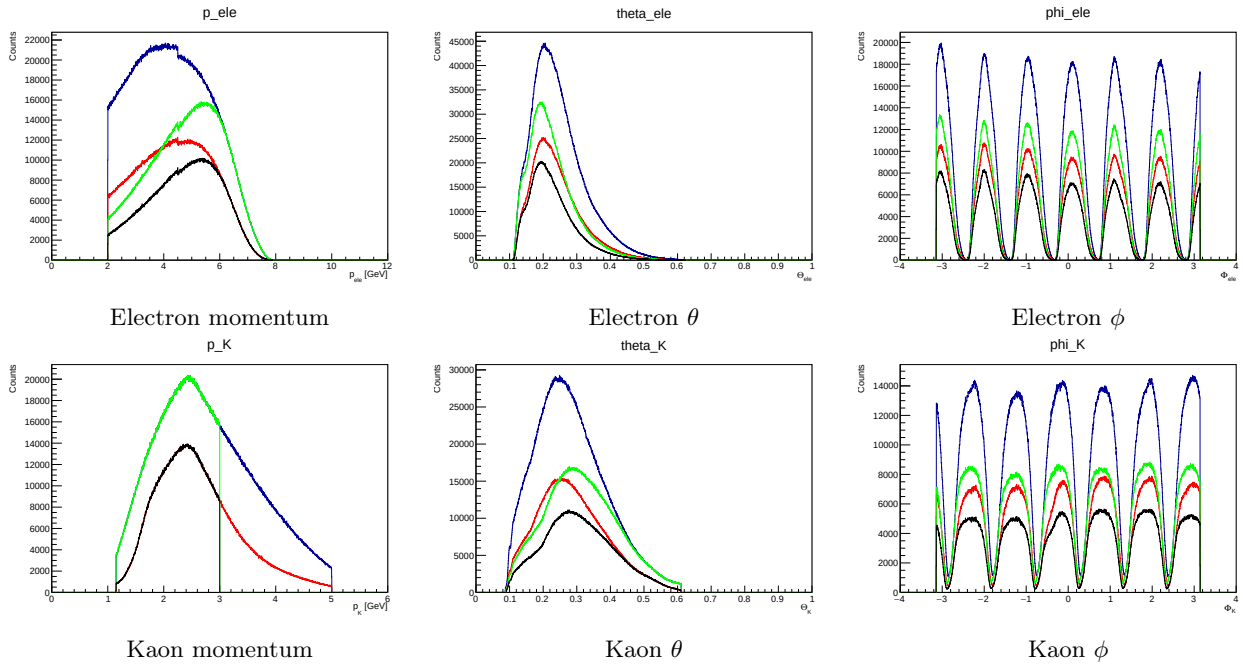


Figure 3.43: Particle distributions for the e^-K^+X sample with different PID (blue: $p_K < 5$ GeV, red: $p_K < 5$ GeV with ml, green: $p_K < 3$ GeV, black: $p_K < 3$ GeV with ml)

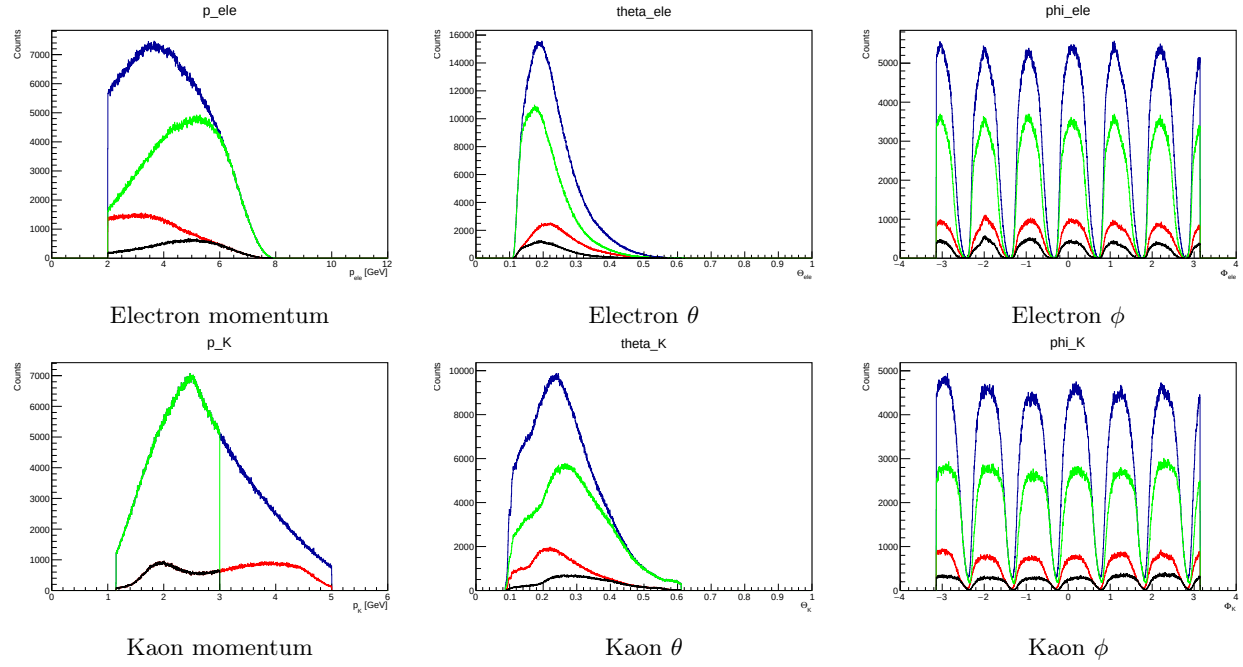


Figure 3.44: Particle distributions for the e^-K^-X sample with different PID (blue: $p_K < 5$ GeV, red: $p_K < 5$ GeV with ml, green: $p_K < 3$ GeV, black: $p_K < 3$ GeV with ml)

3.5.4 Kinematic distributions

The kinematic distributions in Q^2, x_B, z and P_T with all the applied cuts are shown in Fig. 3.45 for the e^-K^+X and in Fig. 3.46 for the e^-K^-X sample. For both samples the dataset covers the Q^2 range up to 8 GeV^2, the x_B region between 0.1 and 0.6, the z region between 0.2 and 0.8 and the P_T region up to 1.2 GeV. The correlations of Q^2-x_B and P_T-z are shown in Fig. 3.47 and 3.48 for the e^-K^+X and in Fig. 3.49 and 3.50 for the e^-K^-X sample with the loosest and tightest PID cut respectively.

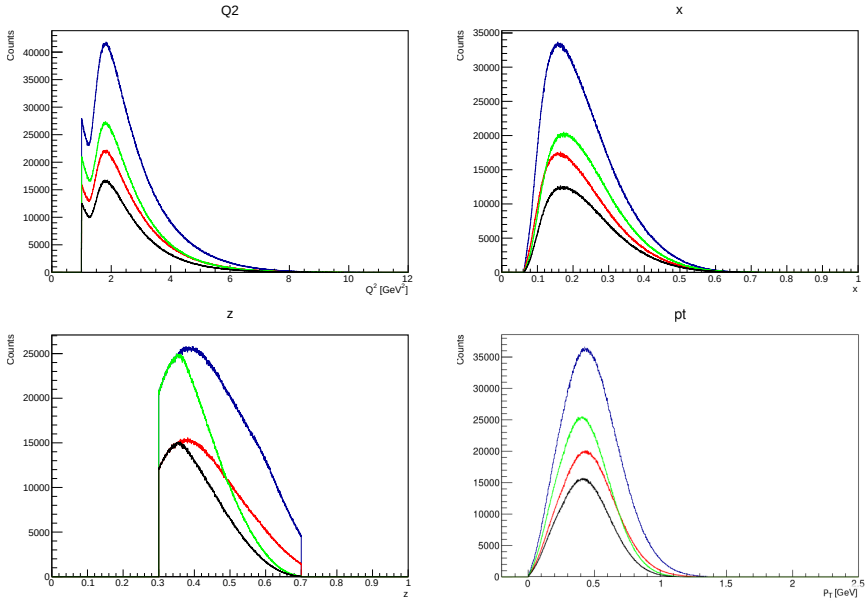


Figure 3.45: Kinematic distributions for the e^-K^+X sample with different PID (blue: $p_K < 5$ GeV, red: $p_K < 5$ GeV with ml, green: $p_K < 3$ GeV, black: $p_K < 3$ GeV with ml)

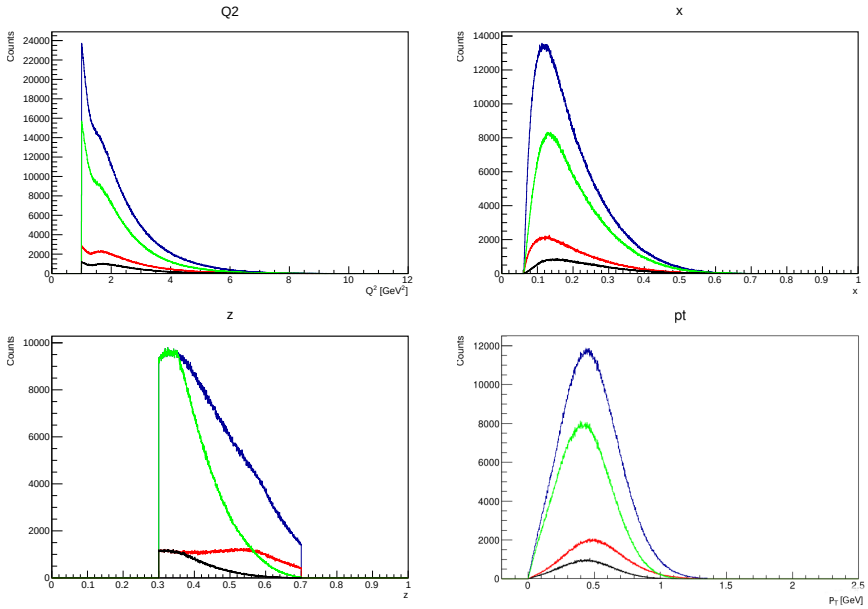


Figure 3.46: Kinematic distributions for the e^-K^-X sample with different PID (blue: $p_K < 5$ GeV, red: $p_K < 5$ GeV with ml, green: $p_K < 3$ GeV, black: $p_K < 3$ GeV with ml)

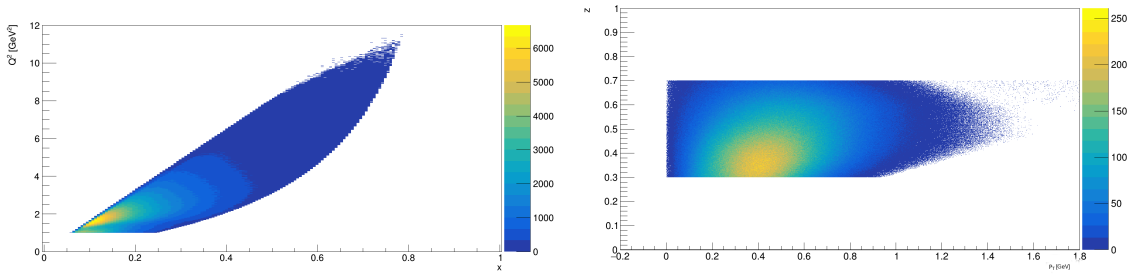


Figure 3.47: The correlations of Q^2-x_B (left) and P_T-z (right) for the e^-K^+X sample with $p_K < 5$ GeV cut

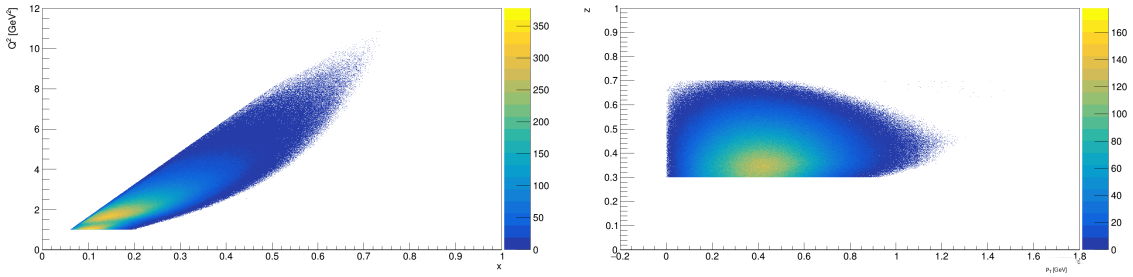


Figure 3.48: The correlations of Q^2-x_B (left) and P_T-z (right) for the e^-K^+X sample with $p_K < 3$ GeV and ml cut

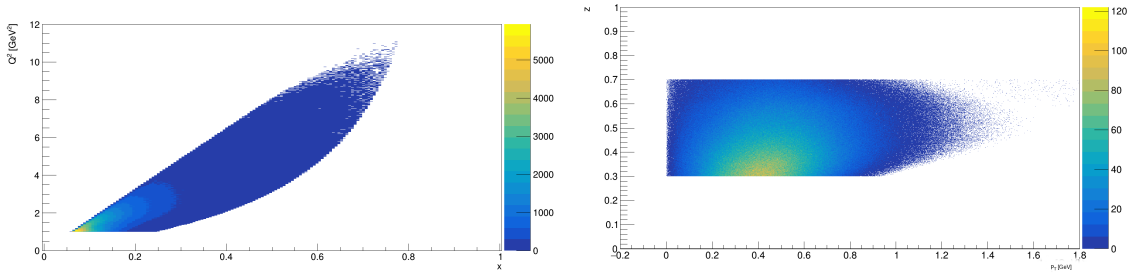


Figure 3.49: The correlations of Q^2-x_B (left) and P_T-z (right) for the e^-K^-X sample with $p_K < 5$ GeV cut

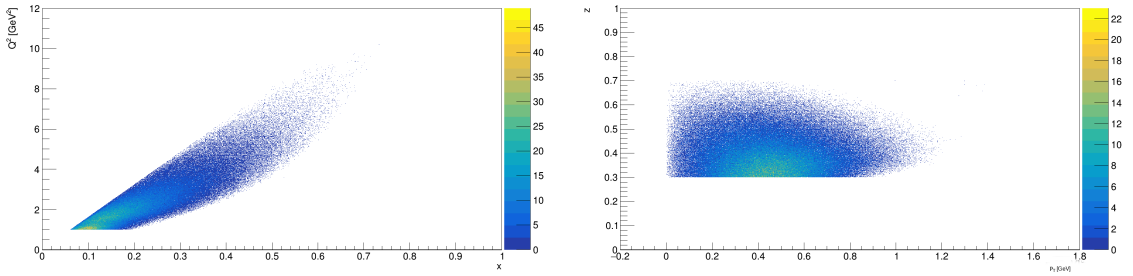


Figure 3.50: The correlations of Q^2-x_B (left) and P_T-z (right) for the e^-K^-X sample with $p_K < 3$ GeV and ml cut

Figure 3.51 shows the correlation of missing mass and momentum. Here we can observe that all structures seen in the missing mass spectra span over the whole momentum range.

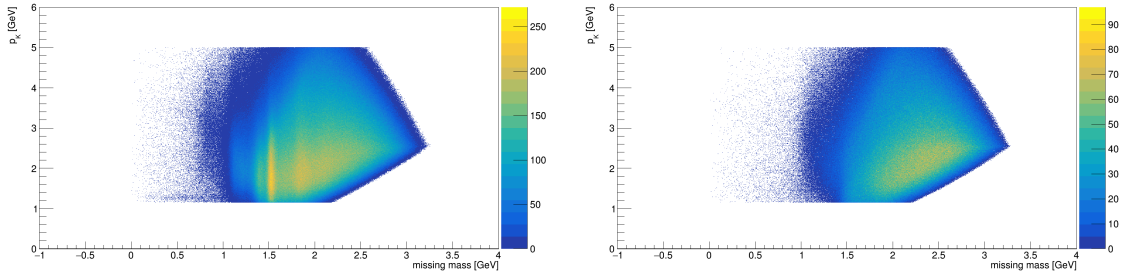


Figure 3.51: The correlation of missing mass and momentum for the e^-K^+X (left) and e^-K^-X (right) samples

3.5.5 Comparison of the used datasets

Figures 3.52 and 3.53 compare the distributions of the kinematic variables for the in- and the outbending and the 2019 datasets for the e^-K^+X and e^-K^-X samples respectively.

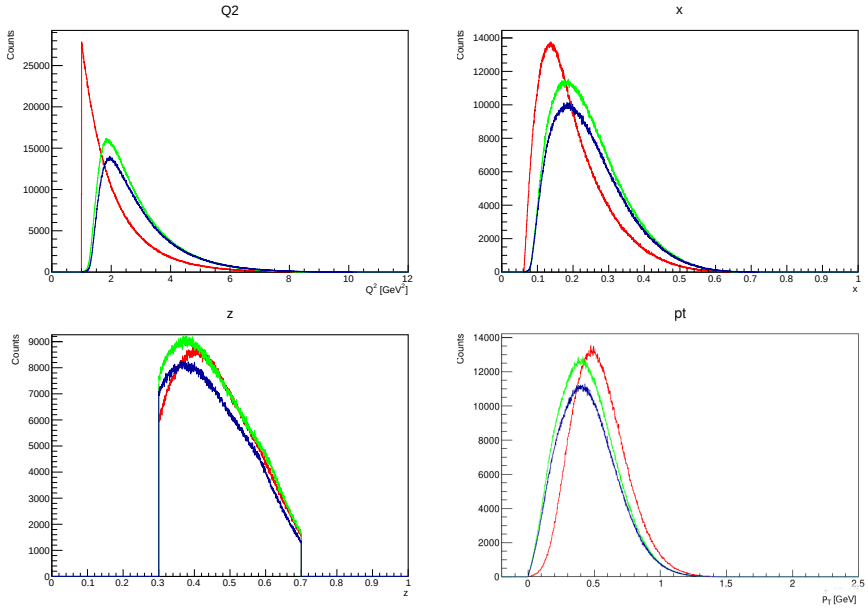


Figure 3.52: Comparison of the kinematic distributions from different datasets (blue-inbending fall 2018, red-outbending fall 2018, green-inbending spring 2019) for the e^-K^+X sample

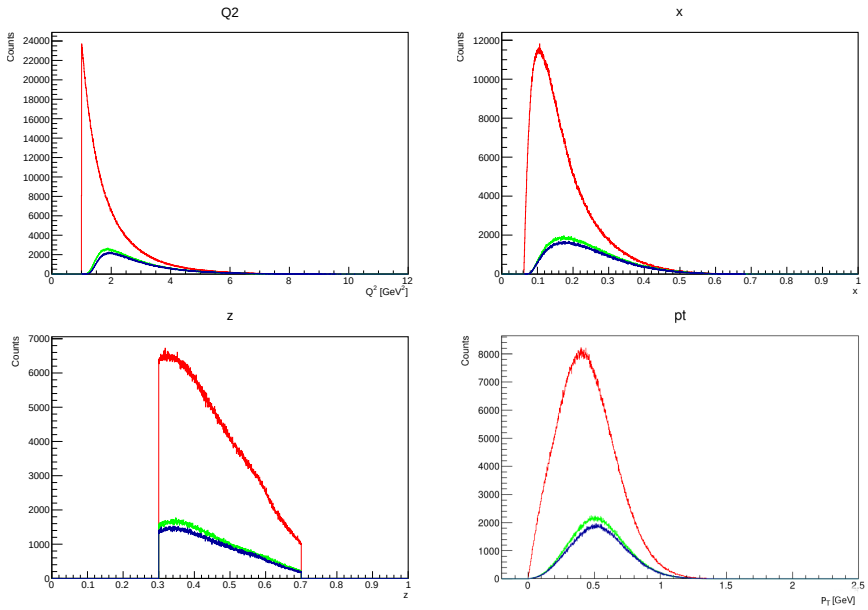


Figure 3.53: Comparison of the kinematic distributions from different datasets (blue-inbending fall 2018, red-outbending fall 2018, green-inbending spring 2019) for the e^-K^-X sample

Figures 3.54 and 3.55 show the extracted $F_{LU}^{\sin\phi}/F_{UU}$ values from the different datasets. In all cases the loosest PID setting was used: $p < 5$ GeV.

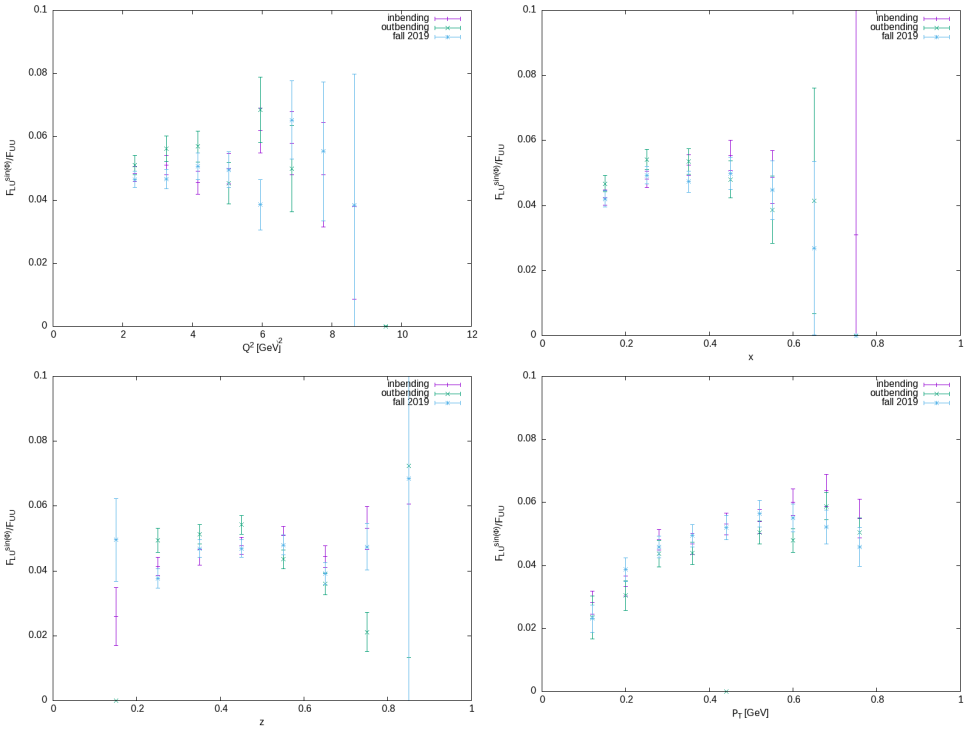


Figure 3.54: $F_{LU}^{\sin\phi}/F_{UU}$ as a function of Q^2 (top left), x_B (top right), z (bottom left) and P_T (bottom right) for K^+ using different datasets (inbending fall 2018, outbending fall 2018 and inbending spring 2019)

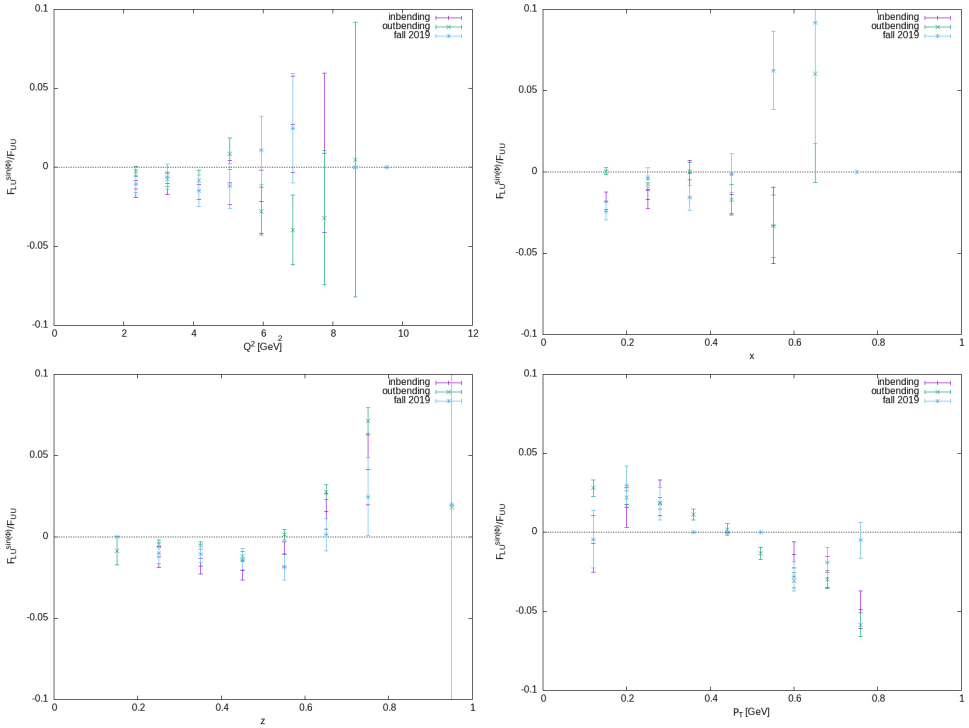


Figure 3.55: $F_{LU}^{\sin\phi}/F_{UU}$ as a function of Q^2 (top left), x_B (top right), z (bottom left) and P_T (bottom right) for K^- using different datasets (inbending fall 2018, outbending fall 2018 and inbending spring 2019)

In most bins the extracted asymmetries are in very good agreement for the different datasets. In some cases the extracted results are different for outbending dataset and for the other two inbending datasets. This difference can come from 2 sources. The first one is that in a given kinematic bin the distributions of the other 3 variables are different for the in- and outbending datasets as the figures below demonstrate. The other one is the fact that the kaon-pion contamination is different for the in- and for the outbending datasets. This pushes the dataset with the smaller contamination further away from the pion asymmetries. This is discussed in more details in section 3.6. Based on these checks, the different datasets can be used combined for the extraction of the asymmetries. This way the uncertainties of the extraction can be highly reduced.

3.5.6 Binning scheme

In every bin 12 ϕ-bins with a width of 30° each are used to extract the ϕ-dependence. Multiple binning schemes were used during the analysis. They offer small uncertainties, show the different a in the different kinematic regions or offer a compromise between these two.

One dimensional binning

To obtain a one dimensional binning each variable is binned independently while the dataset is integrated over the complete distribution of all the other 3 variables, except for z, on which a cut is applied as described in section 3.5.1. For P_T and Q^2 10 bins, for z 9 bins and for x_B 8 bins are used with equal widths. This width is 0.1 for z. The first bins mean is 0.15 and the last ones is 0.95. For Q^2 the width is 0.9 GeV^2, with bin means ranging from 1.45 GeV^2 to 9.55 GeV^2. For x the width is 0.1, with bin means ranging from 0.05 to 0.75. For P_T the width is 0.08 GeV, with bin means ranging from 0.04 GeV to 0.76 GeV.

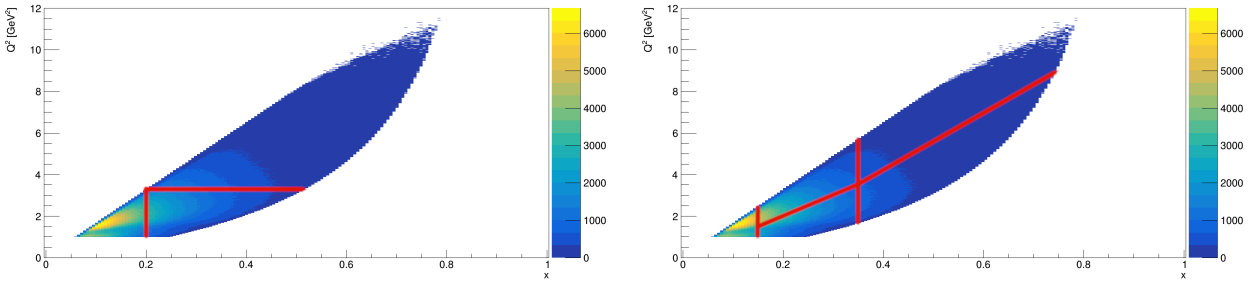
Three dimensional binning

In this case a binning is applied to 3 kinematic variables in parallel and the dataset is integrated in the fourth one. To consider the correlation between Q^2 and x_B, the first set of bin borders were determined in their 2

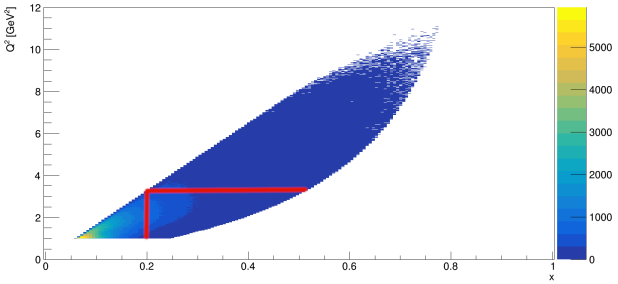
dimensional plane. In case of the $e^- K^+ X$ dataset the results where extracted with 3 and 5 bins on the Q^2-x_B-plane. The cut borders can be found in Table 3.14 and they are shown in Fig. 3.56a. For the $e^- K^- X$ dataset the same 3 bins were used. They are shown in Fig. 3.56b.

Bin number	3 bins	5 bins
1	$x < 0.2$	$x < 0.15$
2	$x > 0.2$ and $Q^2 < 3.3$	$0.15 < x < 0.35$ and $Q^2 < 1.5 + 10(x - 0.15)$ GeV^2
3	$x > 0.2$ and $Q^2 > 3.3$	$0.15 < x < 0.35$ and $Q^2 > 1.5 + 10(x - 0.15)$ GeV^2
4	-	$x > 0.35$ and $Q^2 < 3.5 + 13.75(x - 0.35)$ GeV^2
5	-	$x > 0.35$ and $Q^2 > 3.5 + 13.75(x - 0.35)$ GeV^2

Table 3.14: Bin borders and numbering of the bins on the Q^2-x_B-plane



(a) $e^- K^+ X$ dataset with 3 (left) and 5 (right) bins



(b) $e^- K^- X$ dataset with 3 bins

Figure 3.56: Bin borders on the Q^2-x_B-plane

In a third variable (z or P_T) the same bins are used as in the one dimensional case. The z and P_T distributions in different Q^2-x_B bins are shown in Fig. 3.57 and 3.58 for the e^-K^+X dataset and in Fig. 3.59 and 3.60 for the e^-K^-X dataset with loosest and strictest PID cut respectively.

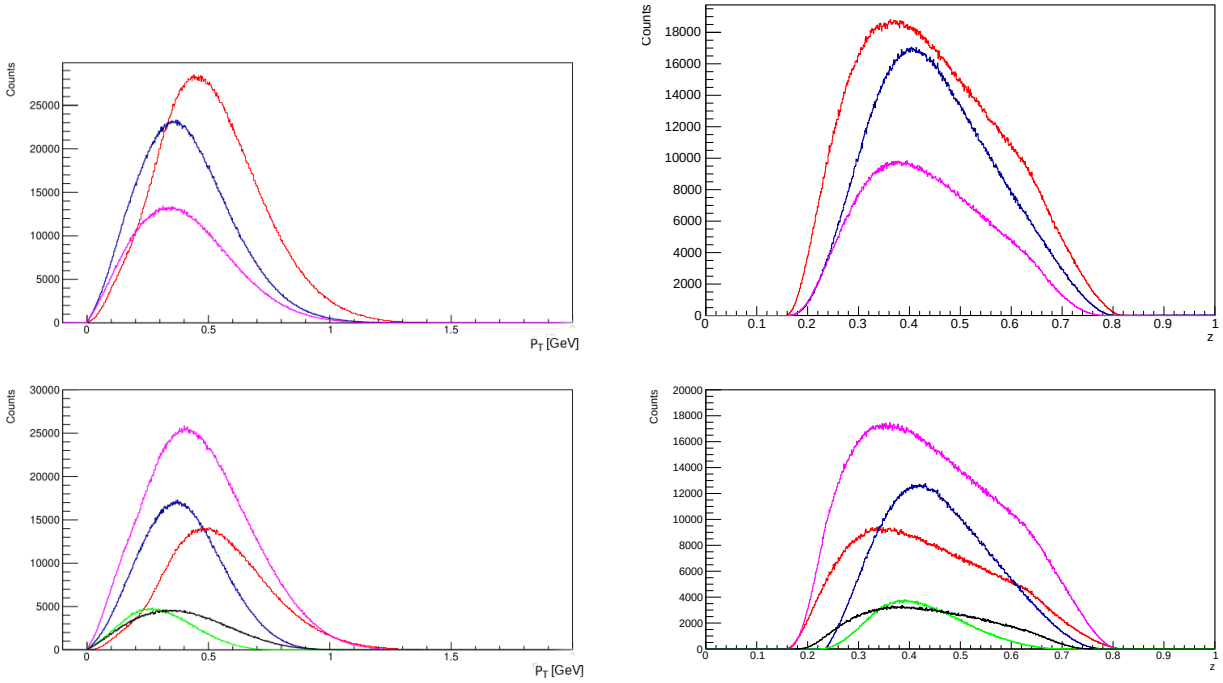


Figure 3.57: P_T (left) and z (right) distributions in different Q^2-x_B bins (red - bin 1, blue - bin 2, purple - bin 3, green - bin 4, black - bin 5) with 3 (top) and 5 bins (bottom) for the e^-K^+X dataset with $p_K < 5$ GeV cut

Fully multidimensional binning

For a fully multidimensional binning, a binning is applied to all 4 kinematic variables. The same 3 Q^2-x_B bins and the same binning in z or P_T are used as described in the previous sections. The difference is that the dataset is not integrated over in the last variable, but binned in it. 3 bins are used for the e^-K^+X dataset and 2 for the e^-K^-X. The bin borders can be found in Table 3.15.

Bin number	P_T bin	x bin
1	$P_T < 0.33$ GeV	$z < 0.34$
2	0.33 GeV$< P_T < 0.66$	$0.34 < z < 0.52$
3	$P_T > 0.66$ GeV	$z > 0.52$

Bin number	P_T bin	x bin
1	$P_T < 0.5$ GeV	$z < 0.5$
2	$P_T > 0.5$ GeV	$z > 0.5$

Table 3.15: The binning of the last remaining kinematic variable for fully multidimensional binning for the e^-K^+X dataset (left) and for the e^-K^-X dataset (right)

3.5.7 Extraction of the BSA

The extraction of the asymmetries is identical to what was used for pions [35, 79]. In each kinematic bin, the BSA is extracted for each ϕ bin separately. In bin i it is defined as:

$$A_i = \frac{1}{P_e} \frac{N_i^+ - N_i^-}{N_i^+ + N_i^-} \tag{3.30}$$

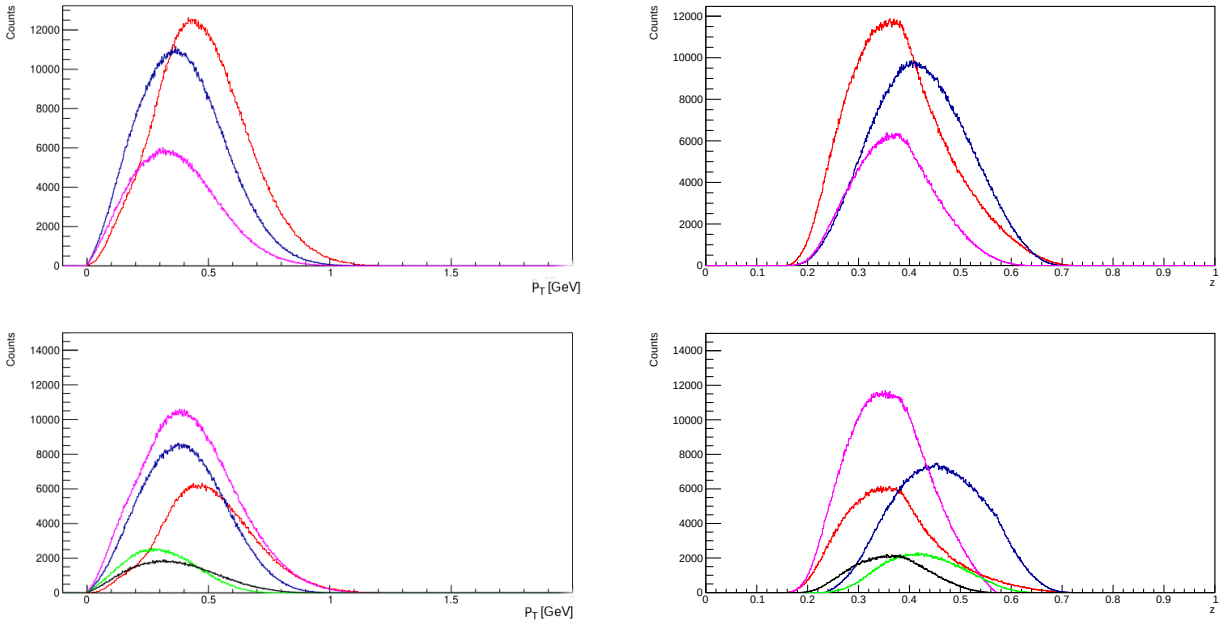


Figure 3.58: P_T (left) and z (right) distributions in different Q^2-x_B bins (red - bin 1, blue - bin 2, purple - bin 3, green - bin 4, black - bin 5) with 3 (top) and 5 bins (bottom) for the e^-K^+X dataset with $p_K < 3$ GeV and ml cut

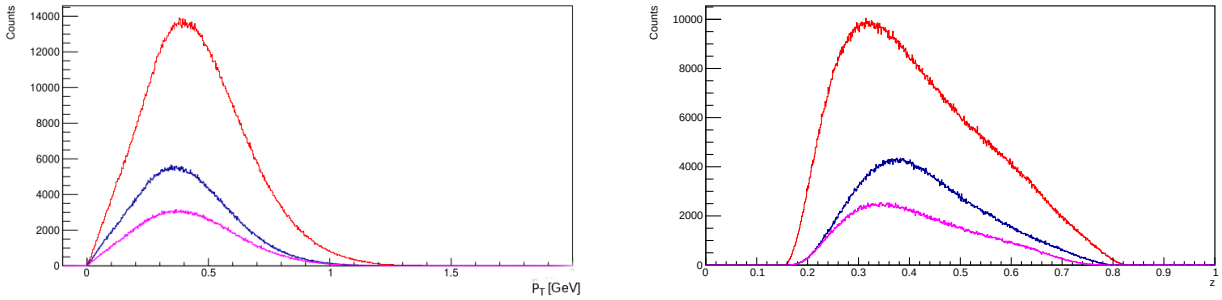


Figure 3.59: P_T (left) and z (right) distributions in different Q^2-x_B bins (red - bin 1, blue - bin 2, purple - bin 3) for the e^-K^-X dataset with $p_K < 5$ GeV cut

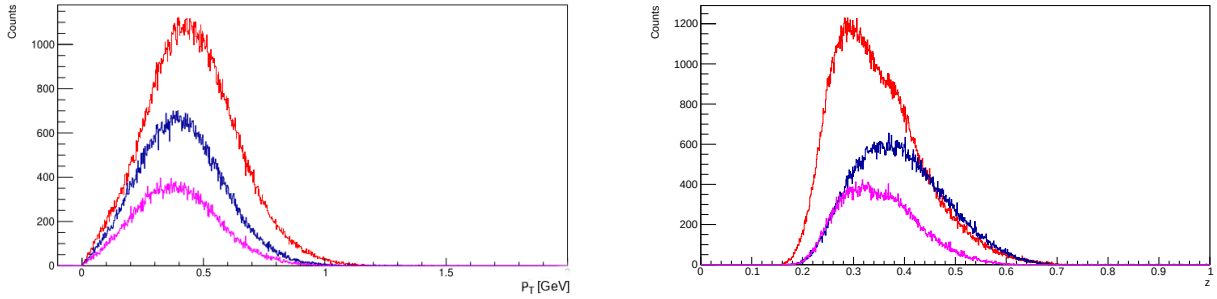


Figure 3.60: P_T (left) and z (right) distributions in different Q^2-x_B bins (red - bin 1, blue - bin 2, purple - bin 3) for the e^-K^-X dataset with $p_K < 3$ GeV and ml cut

Here P_e is the average beam polarization measured to be $(86.3 \pm 3.7)\%$ and $N_i^{\pm}$ are the number of events in bin i with $\pm$ helicity.
The uncertainty of the measured A_i values comes from 2 sources. One is the statistical uncertainty of the counts $N_i^{\pm}$, the other is the uncertainty of the measurement of P_e. The latter is included in the systematic uncertainties. So the statistical uncertainty can be given as:

$$\delta A_i = \frac{2}{P_e}\sqrt{\frac{N_i^+ N_i^-}{(N_i^+ + N_i^-)^3}} \tag{3.31}$$

The extracted values are then fitted with the theoretical function $A_{LU}^{\sin\phi}\sin\phi$ in ϕ to determine $A_{LU}^{\sin\phi}$. The effect of the two additional moments $A_{UU}^{\cos\phi}$ and $A_{UU}^{\cos 2\phi}$ from Eq. 3.26 on the extraction of $A_{LU}^{\sin\phi}$ was found to be small and discussed in section 3.7.6 in more details. As a final step, $F_{LU}^{\sin\phi}/F_{UU}$ is calculated according to Eq. 3.27. For this purpose the mean value of ϵ is calculated in each kinematic bin.
All figures in this section will show the "Raw" $F_{LU}^{\sin\phi}/F_{UU}$ values, which are the directly extracted values, containing also the pion asymmetries due to the contamination. The final extraction method is described in section 3.6.4 and the final results are presented and discussed in section 3.8.
Figure 3.61 shows the integrated BSA values in bins of Φ together with the fitted $\sin\Phi$ functions for different PID methods (turquoise-p<5 GeV, purple-p<3 GeV, blue-p<5 GeV and ml cut, red-p<3 GeV and ml cut, black-strict(0.99) ml cut, green-RICH). Here we can observe that the amplitude of the extracted asymmetries increase as the pion contamination decrease, which means that the kaons have a higher asymmetry than the pions. An other explanation of the variation of the values is that the different methods change the underlying kinematic distributions. It can also be observed that the positive kaons have a positive asymmetry, while the negative ones have a slightly negative one, close to zero. A similar behavior was also observed for pions [64, 79].

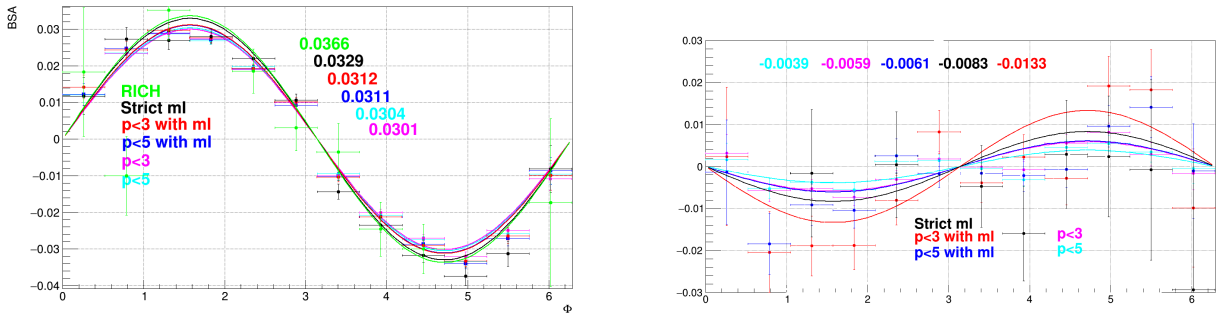


Figure 3.61: Integrated BSA values in bins of Φ together with the fitted $\sin\Phi$ functions for K^+ (left) and K^- (right) using different PID methods: turquoise-p<5 GeV, purple-p<3 GeV, blue-p<5 GeV and ml cut, red-p<3 GeV and ml cut, black-strict(0.99) ml cut, green-RICH (in order of decreasing pion contamination)

Results of the one dimensional binning

For the one dimensional binning each variable is binned independently while the dataset is integrated over the complete distribution of all the other 3 variables, except for z, on which a cut is applied, when the binning is not done in z. The ϕ dependence of the extracted BSA values with the fits can be seen in Fig. 3.62 for positive kaons using the $p < 5$ GeV PID in bins of P_T. All the fits were inspected visually and if the fit cannot be done sufficiently, the kinematic bin were removed from the final plots and tables. These fits are replaced or marked with a red cross. Also some ϕ bins with very high statistical uncertainty were excluded from the fits. All kinematic bins, which have sufficient statistics, can be very well described with a sinusoidal shape.

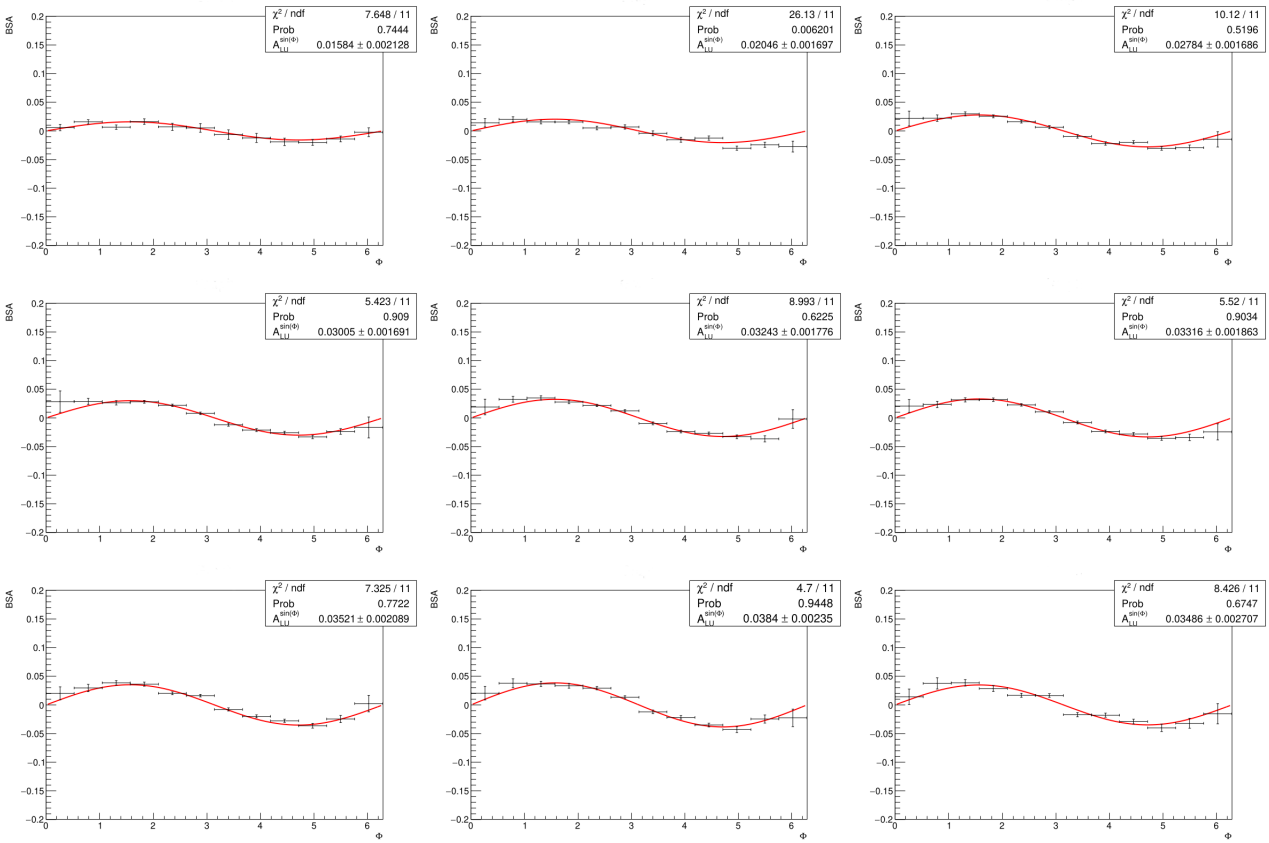


Figure 3.62: BSA as a function of Trento ϕ in radians in P_T bins in increasing order from top left to bottom right for the $e^- K^+ X$ dataset with $p_K < 5$ GeV cut

Based on the fits and Eq. 3.27 the $F_{LU}^{\sin\phi}/F_{UU}$ ratios are calculated. The results are shown for positive kaons in Fig. 3.63. and for negative kaons in Fig. 3.64. These figures compare the results obtained with the different PIDs. They also contain the ratios obtained from pions with the same binning and same cuts from the same data-sample. The results obtained with different PIDs follow the same trends and are very close to each other. These trends are similar to the pions one, but they are clearly different especially for positive kaons. Here we can also observe that the measured values move away from the pion values when the machine learning PID is used. These behaviors show that the machine learning works as intended and that the positive kaon sample doesn't only consist of misidentified pions. For negative kaons the obtained ratios are in agreement with the negative pion ratios within errors for most PIDs. But using the strictest PID ($p < 3$ GeV with ml) the behavior and magnitude of the ratios start to differ for pions and kaons.

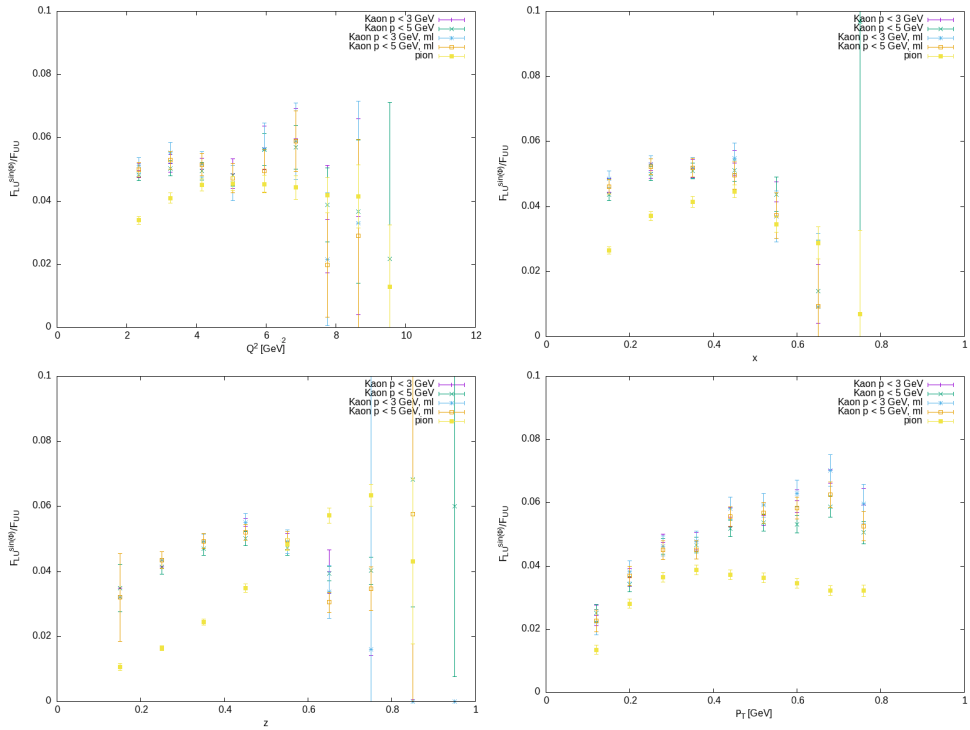


Figure 3.63: "Raw" $F_{LU}^{\sin\phi}/F_{UU}$ as a function of Q^2 (top left), x_B (top right), z (bottom left) and P_T (bottom right) for K^+ with different PIDs compared with π^+ results

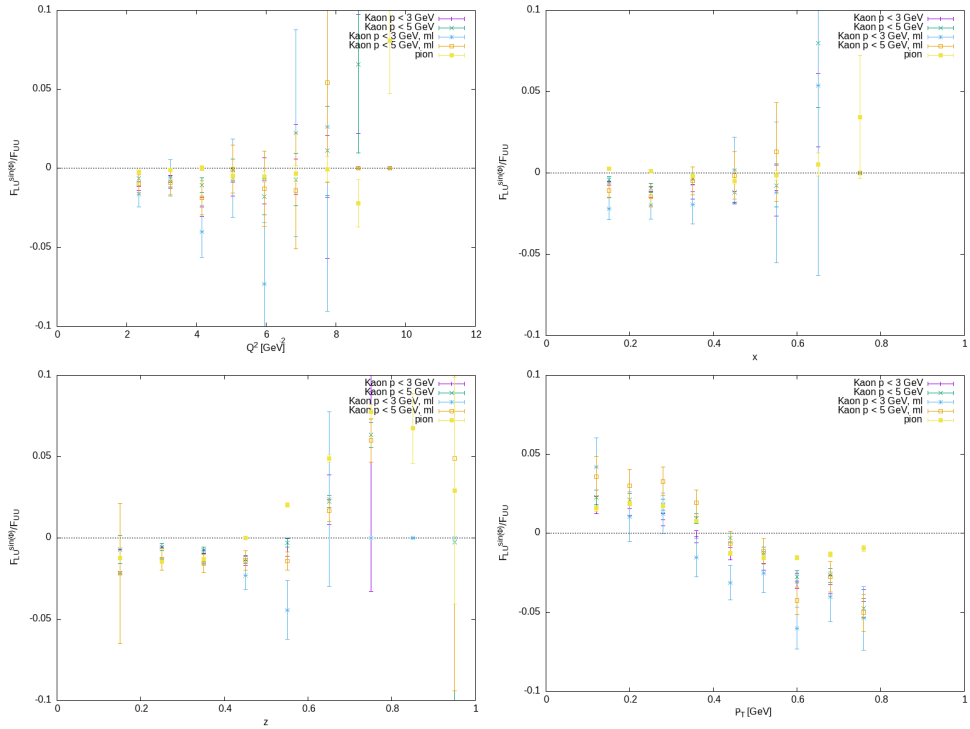


Figure 3.64: "Raw" $F_{LU}^{\sin\phi}/F_{UU}$ as a function of Q^2 (top left), x_B (top right), z (bottom left) and P_T (bottom right) for K^- with different PIDs compared with π^- results

Results of the three dimensional binning

In this case a binning is applied to 3 kinematic variables in parallel and the dataset is integrated in the fourth one. To consider the correlation between Q^2 and x_B, the first set of bin borders were determined in their 2 dimensional plane. In case of the e^-K^+X dataset the results where extracted with 3 and 5 bins on the Q^2-x_B-plane. For the e^-K^-X dataset the same 3 bins were used. All the fits were inspected visually in this case as well and ϕ bins and kinematic bins were removed with the same criteria as in the one dimensional case. Figures 3.65 and 3.66 show the extracted asymmetries with the 3 and 5 bin binning scheme respectively. We can observe that the asymmetries rise with the increasing Q^2 and x_B values, but the asymmetries show a similar tendency in P_T and z in all Q^2-x_B bin.

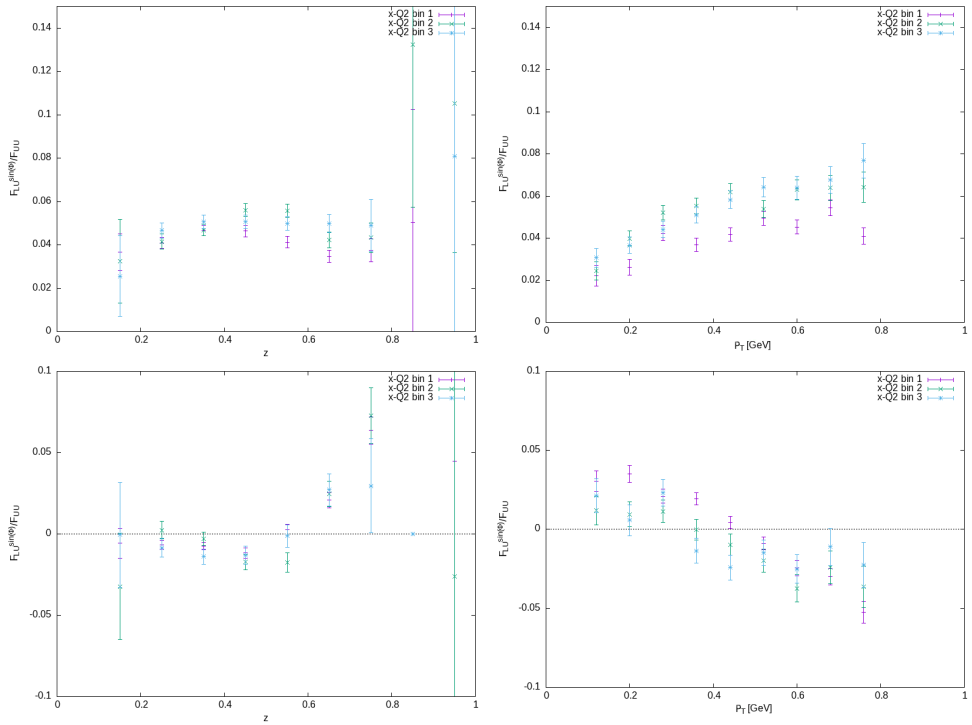


Figure 3.65: "Raw" $F_{LU}^{\sin\phi}/F_{UU}$ as a function of z (left) and P_T (right) for K^+ (top) and K^- (bottom) in different Q^2-x_B bins with p<5 GeV PID

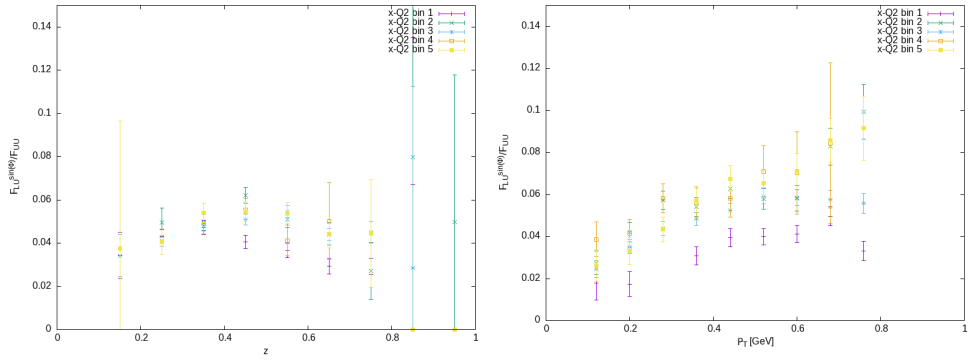
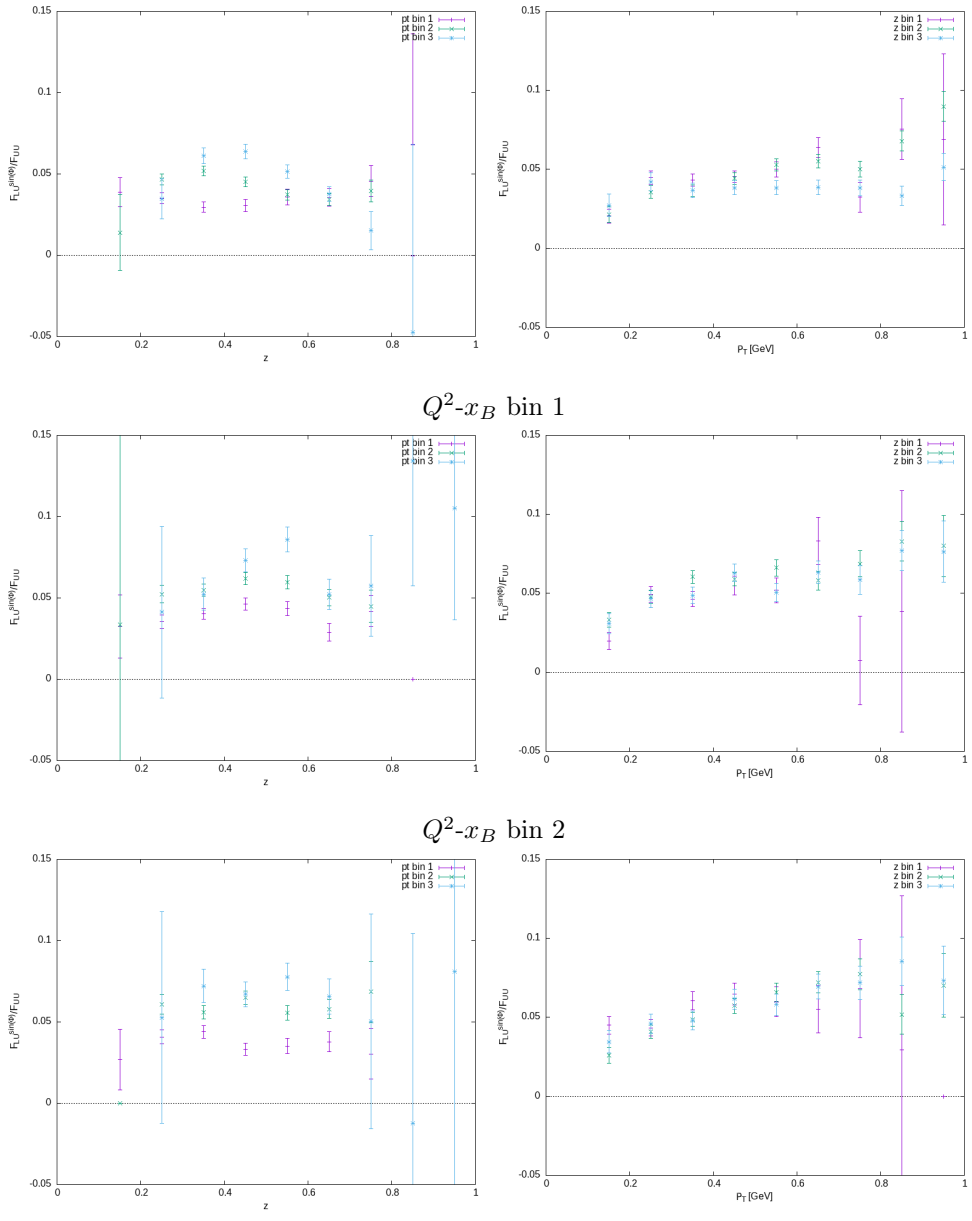


Figure 3.66: "Raw" $F_{LU}^{\sin\phi}/F_{UU}$ as a function of z (left) and P_T (right) for K^+ in different Q^2-x_B bins with p<5 GeV PID

Results of the complete four dimensional binning

In this case a binning is applied to all 4 kinematic variables in parallel. The same procedure is applied as in the three dimensional binning scheme with 3 Q^2-x_B, but this time also both 2 remaining dimensions (P_T, z) are binned using 2 (K^-) or 3 (K^+) bins in one dimension and 10 in the other. Figures 3.67 and 3.68 show the extracted asymmetries.



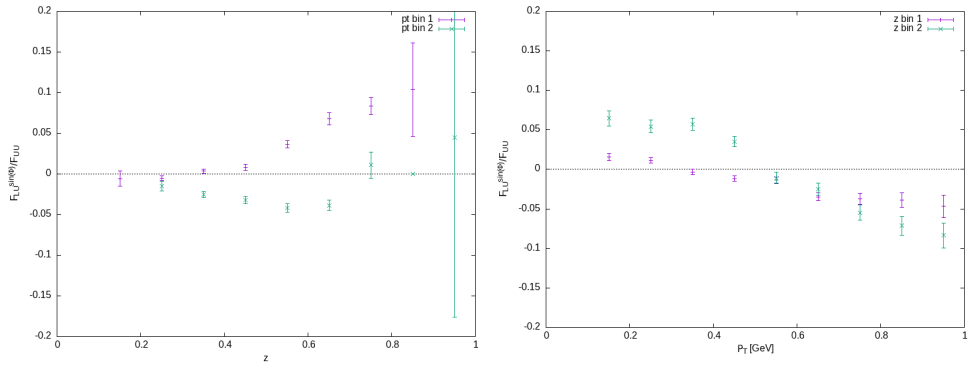
Q^2-x_B bin 1

Q^2-x_B bin 2

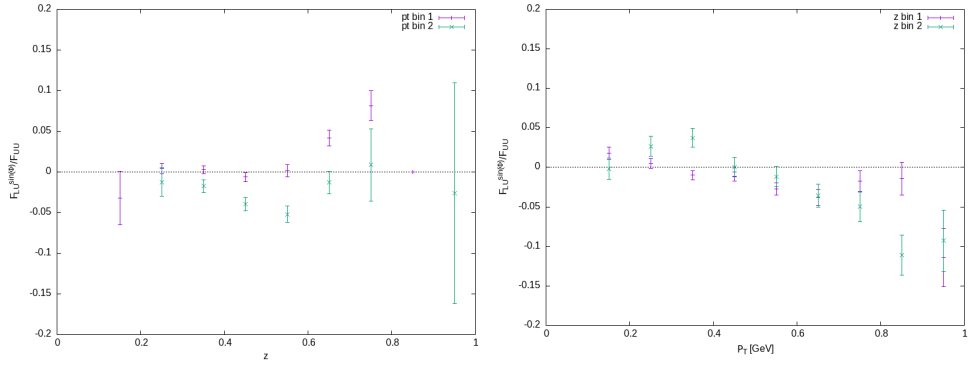
Q^2-x_B bin 3

Figure 3.67: "Raw" $F_{LU}^{\sin\phi}/F_{UU}$ as a function of z (left) and P_T (right) for K^+ in different Q^2-x_B and z or P_T bins with p<5 GeV PID

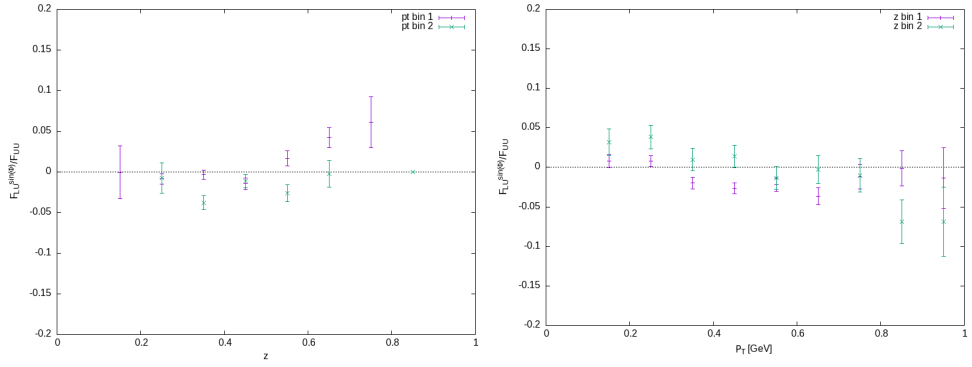
It can be observed that the overall behavior of the asymmetries are independent of Q^2 and x_B. In case of K^+ the P_T-dependence looks the same for all z bins. For K^- we can observe a rising and then a falling tendency in P_T for both z bins, although this effect is more prominent at high z values.
If we look at K^+-s the z-dependence in different P_T bins we can observe a falling trend at low z values and a rising trend at high z values at low P_T. This trend changes to a rising and then falling one at high P_T values.



Q^2-x_B bin 1



Q^2-x_B bin 2



Q^2-x_B bin 3

Figure 3.68: "Raw" $F_{LU}^{\sin\phi}/F_{UU}$ as a function of z (left) and P_T (right) for K^- in different Q^2-x_B and z or P_T bins with p<5 GeV PID

For K^-, a similar difference can be observed in the z-dependence between the 2 P_T bins. It shows a rising trend at low P_T and a falling trend followed by a sharp increase at high P_T.

3.6 Subtraction of the pion asymmetries

As the pion contamination in the kaon sample is significant, the pion asymmetries have to be subtracted from the "raw" kaon asymmetries. This is done based on contamination from MC validated by the RICH and by β-p fits of the data. This section describes the validation of the MC samples and the RICH reconstruction and then the pion subtraction method in details.

3.6.1 Monte Carlo simulations

The CLASDIS [82] generator, which is based on the PEPSI [83] generator was used for the SIDIS Monte Carlo production. The parton level interaction is based on the electroweak cross sections, implemented in leading order for any lepton of arbitrary polarization. First-order QCD matrix elements for the boson-gluon fusion and the gluon radiation are implemented and higher order QCD radiation is treated using parton showers. The hadronization process follows the Lund string model, implemented in Pythia (Jetset) [84]. Several parameters enter in the tuning of the hadronization process, and those have been tuned to reproduce the electron, photon and pion distributions obtained in the CLAS12 RG-A data for the semi-inclusive regime.
For the present work, the new large scale SIDIS MC production for RG-A from Orlando Soto with the more accurate kaon-pion ratio has been used which contains approx. 2 billion (10^9) inbending and 0.7 billion outbending events and is used within the SIDIS community of the CLAS12 collaboration. [85] The generated particles were transmitted through the CLAS12 detectors using the Geant4 Monte-Carlo (GEMC) package. [86,87] The MC has been analyzed exactly the same way as the data and the same cuts are applied, except the quality assessment (QA) cut which is not relevant for simulations since the MC has no bad files or runs which need to be rejected due to detector problems. Most cut parameters are identical between MC and simulation, only for the sampling fraction cut the mean and sigma value have been adjusted to provide a proper 3.5 σ region for the contamination studies of the electron sample. This is described and discussed in detail in the updated version of the common analysis note.
To reproduce the realistic resolutions the reconstructed momentum, θ and ϕ were smeared with the following resolution functions provided by F. X. Girod [88]:

$$p_{S1} = 0.0184291 - 0.0110083 \cdot \theta + 0.00227667 \cdot \theta^2 - 0.000140152 \cdot \theta^3 + 3.07424 \cdot 10^{-6} \cdot \theta^4$$

$$p_R = 0.02\sqrt{(p_{S1} \cdot p)^2 + (0.02 \cdot \theta)^2}$$

$$\theta_R = 2.5\sqrt{((0.004 \cdot \theta + 0.1)(p^2 + 0.13957^2)/(p^2))^2}$$

$$\phi_{S1} = 0.85 - 0.015 \cdot \theta$$

$$\phi_{S2} = 0.17 - 0.003 \cdot \theta$$

$$\phi_R = 3.5\sqrt{(\phi_{S1}\sqrt{p^2 + 0.13957^2}/p^2)^2 + \phi_{S2}^2}$$

$$\phi_{new} = \phi + \phi_R \cdot \text{gRandom}->\text{Gaus}(0,1)$$

$$\theta_{new} = \theta + \theta_R \cdot \text{gRandom}->\text{Gaus}(0,1)$$

$$p_{new} = p + p_R \cdot \text{gRandom}->\text{Gaus}(0,1) \cdot p$$

All angles are in radians in this section, unless stated otherwise.
The MC very well reproduces the kinematic distributions of the experimental data. Figures 3.69, 3.70, 3.71, 3.72, 3.73, 3.74, 3.75, 3.76 and 3.77 show the comparison of the experimental and simulated kinematic distributions. In all cases the loosest PID setting was used: $p < 5$ GeV. This means that the highest possible pion contamination was present in both the data and mc.
Since the MC reproduces most of the distributions very well, it can be used to validate the extraction method, to check the acceptances and bin-migration effects and to extract the kaon-pion contamination in kinematic bins for subtracting the pion asymmetries.
Figure 3.78 compares the missing mass distribution obtained from the simulation with and without the smearing with the one obtained from data without any scaling. Although the structure seen in the data is not reproduced completely even with the smearing, by utilizing it the resolutions are better reproduced.

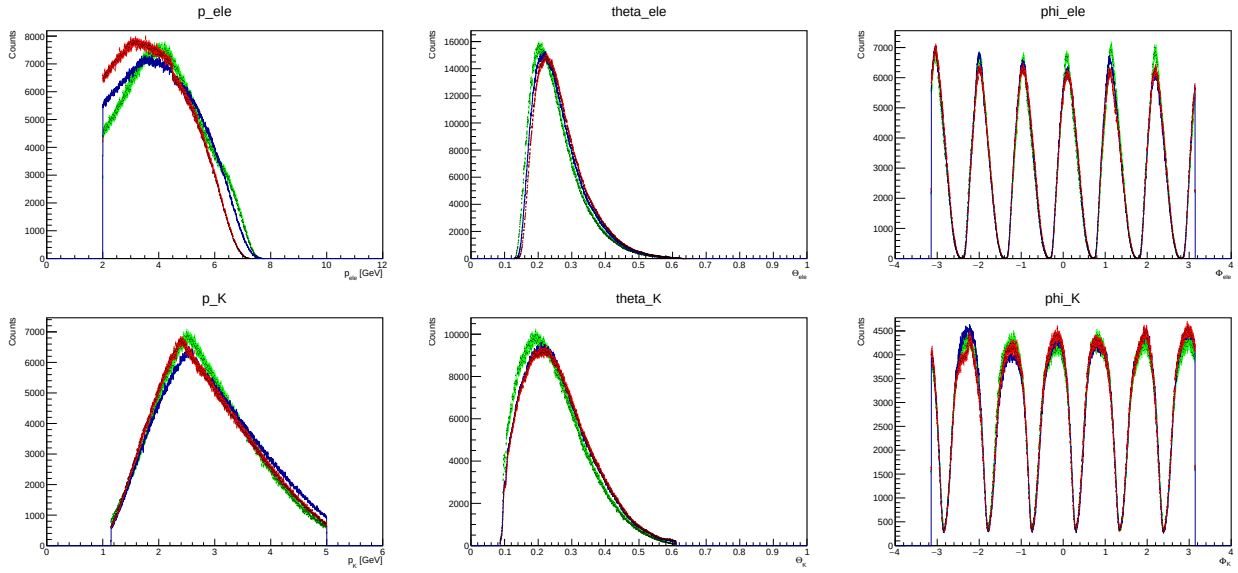


Figure 3.69: Comparison of the inbending simulated (green) and inbending experimental (blue-2018, red-2019) distributions for the electron (up) and kaon (down) for the e^-K^+X sample

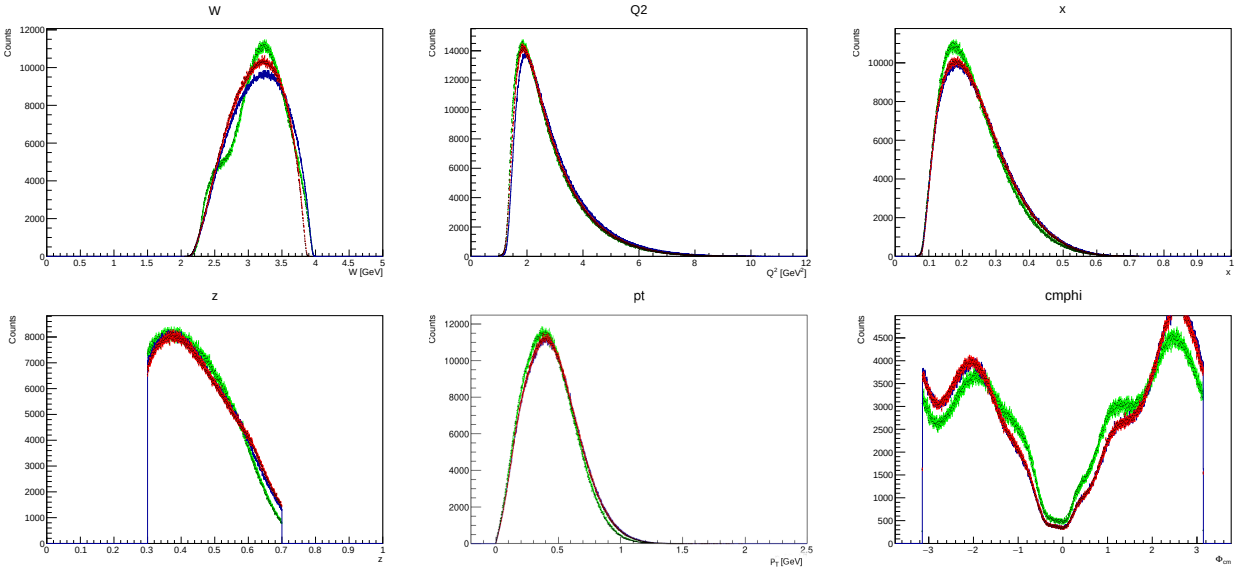


Figure 3.70: Comparison of the inbending simulated (green) and inbending experimental (blue-2018, red-2019) kinematic distributions of W, Q^2, x_B, z, P_T and ϕ for the e^-K^+X sample

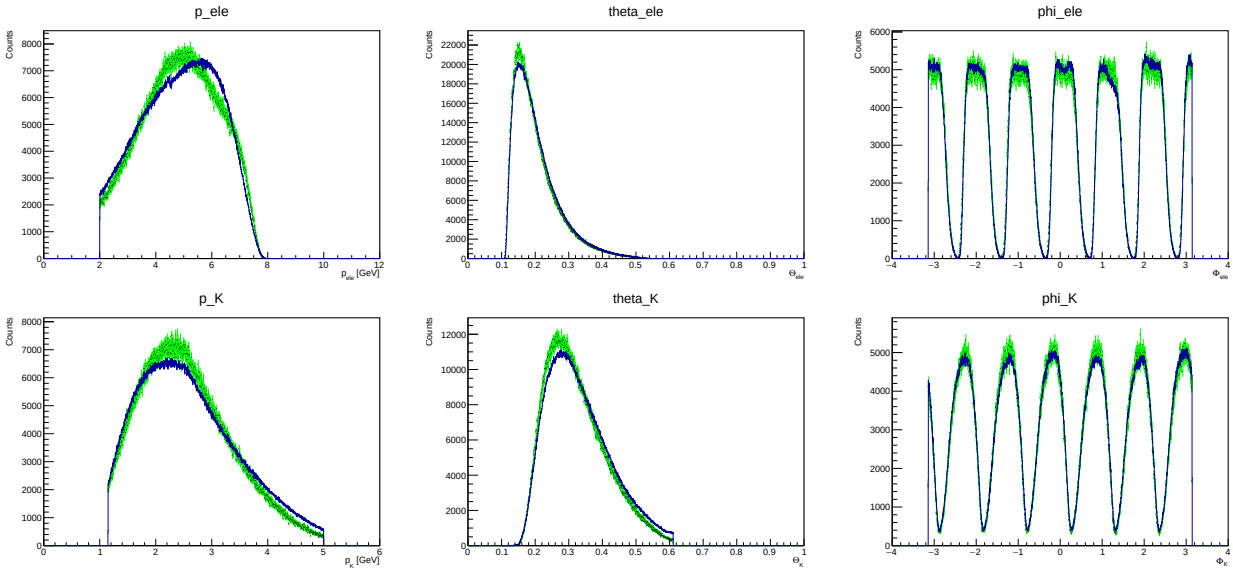


Figure 3.71: Comparison of the outbending simulated (green) and outbending experimental (blue) distributions for the electron (up) and kaon (down) for the $e^- K^+ X$ sample

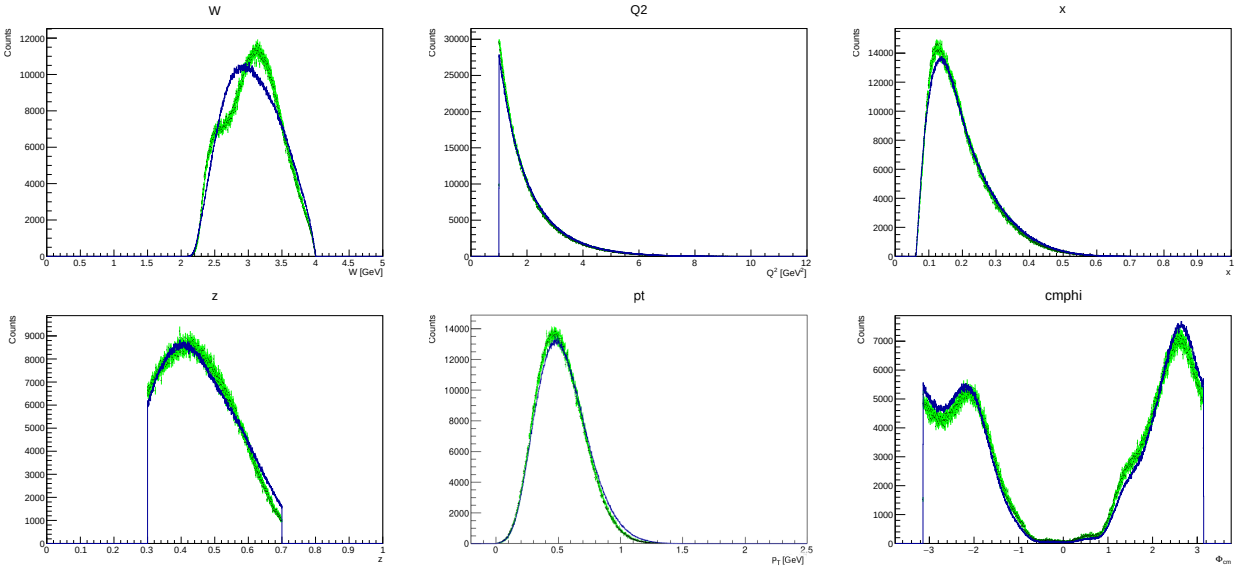


Figure 3.72: Comparison of the outbending simulated (green) and outbending experimental (blue) kinematic distributions of W, Q^2, x_B, z, P_T and ϕ for the $e^- K^+ X$ sample

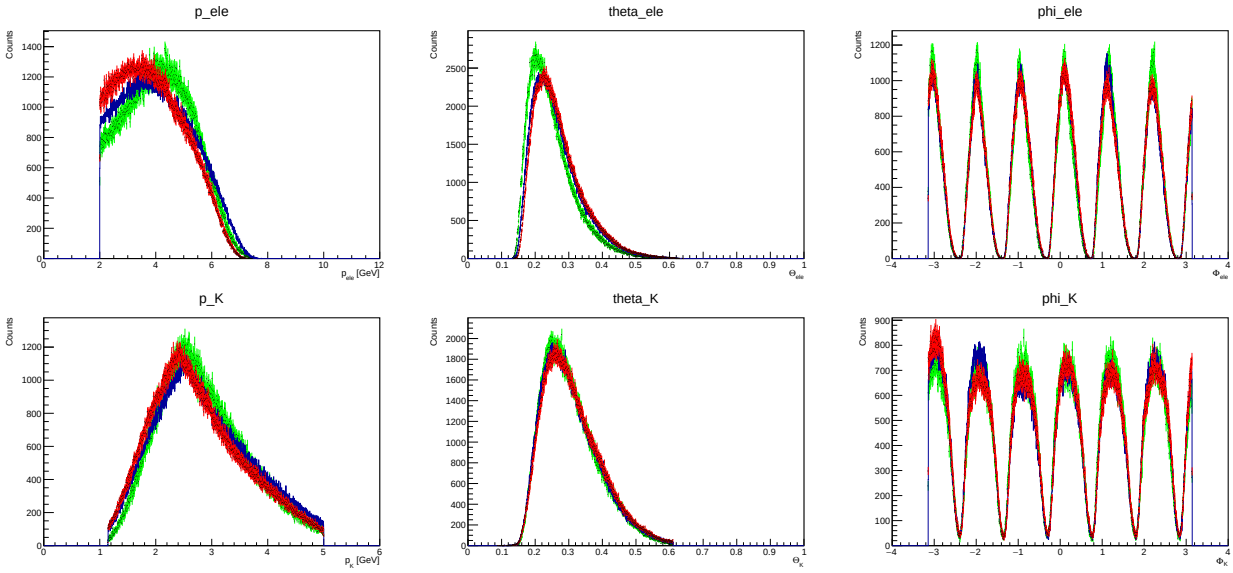


Figure 3.73: Comparison of the inbending simulated (green) and inbending experimental (blue-2018, red-2019) distributions for the electron (up) and kaon (down) for the e^-K^-X sample

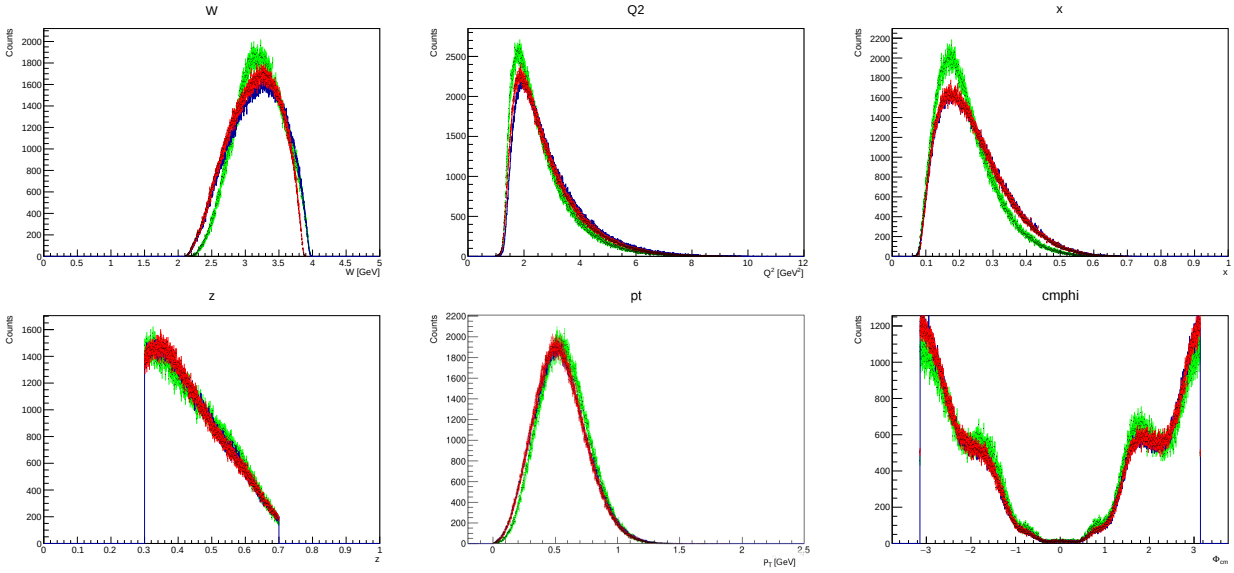


Figure 3.74: Comparison of the inbending simulated (green) and inbending experimental (blue-2018, red-2019) kinematic distributions of W, Q^2, x_B, z, P_T and ϕ for the e^-K^-X sample

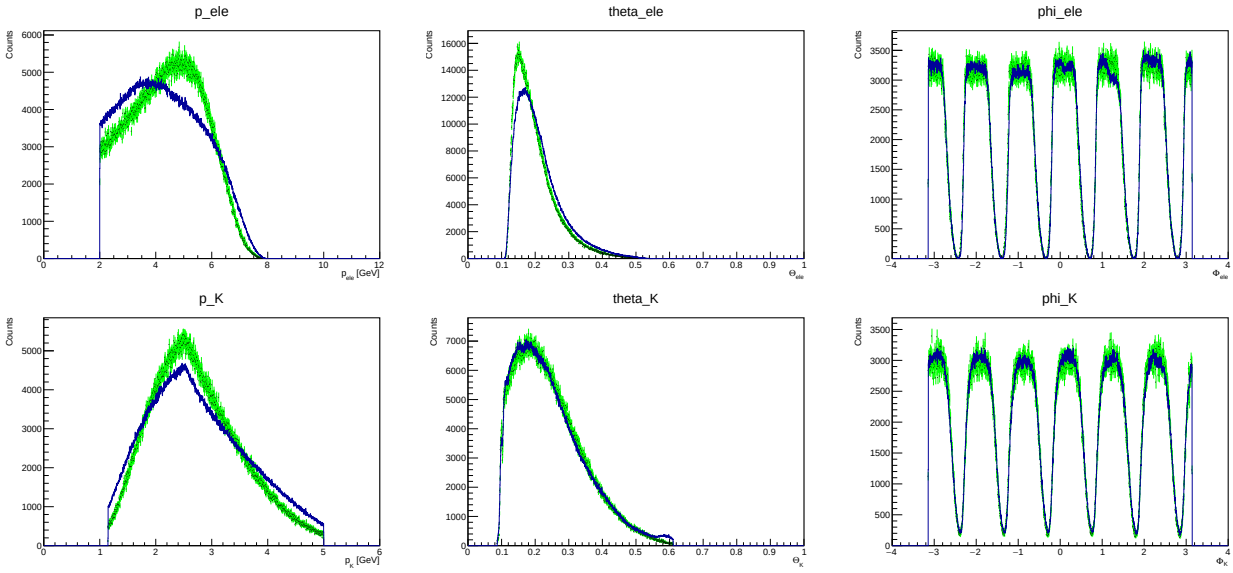


Figure 3.75: Comparison of the outbending simulated (green) and outbending experimental (blue) distributions for the electron (up) and kaon (down) for the e^-K^-X sample

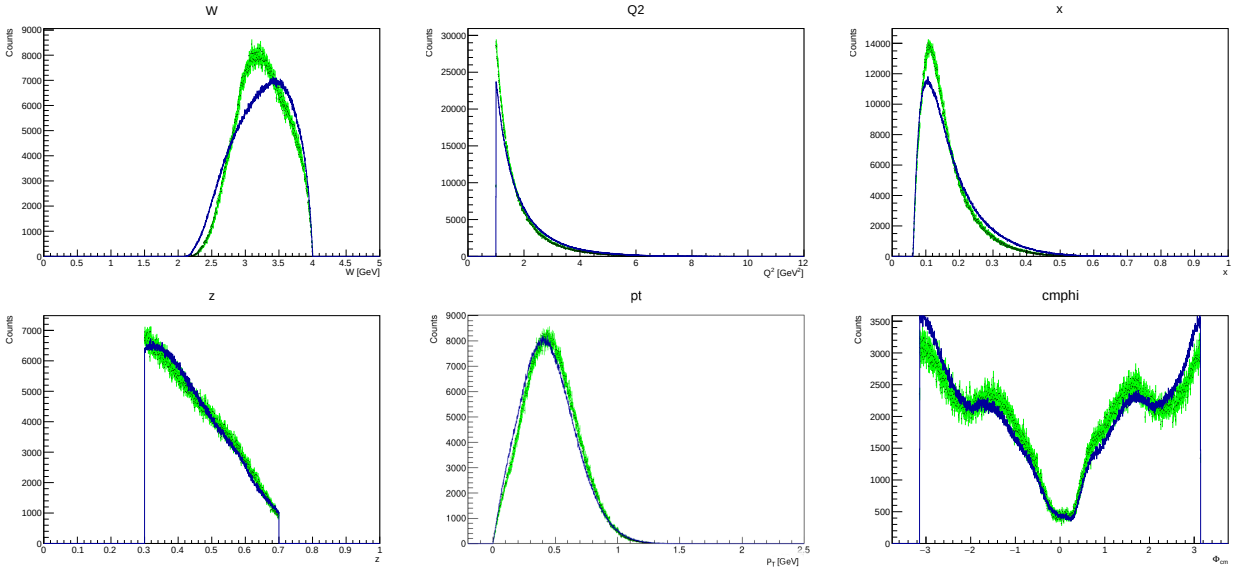


Figure 3.76: Comparison of the outbending simulated (green) and outbending experimental (blue) kinematic distributions of W, Q^2, x_B, z, P_T and ϕ for the e^-K^-X sample

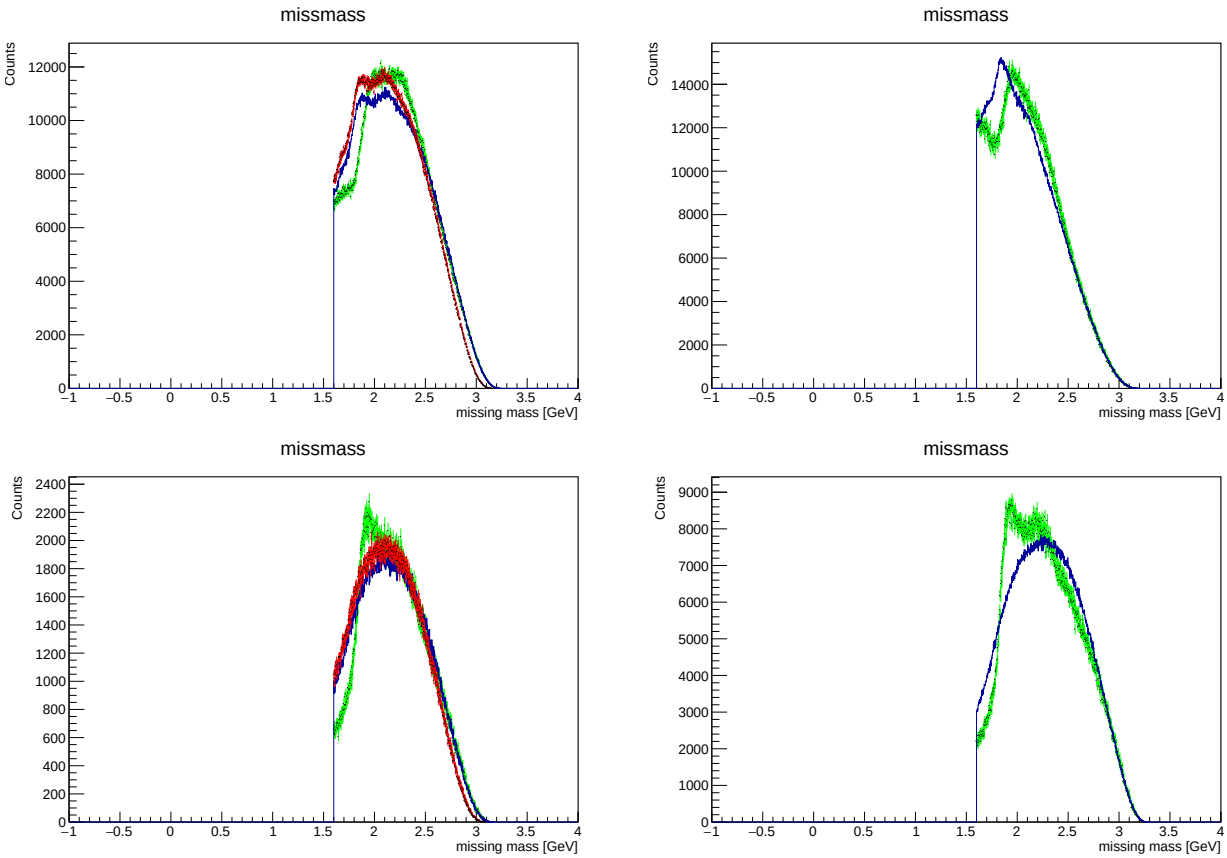


Figure 3.77: Comparison of the simulated (green) and experimental (blue-in- or outbending 2018, red-2019) missing mass distributions in the in- (left column) and outbending (right column) setting for the e^-K^+X (upper row) and e^-K^-X (lower row) sample

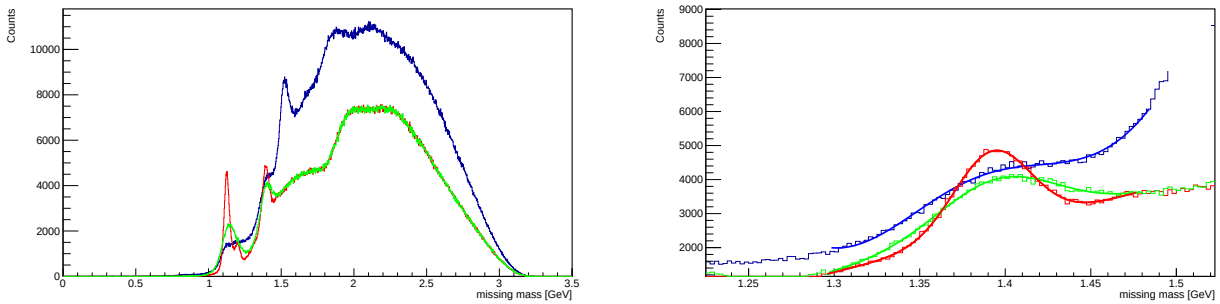


Figure 3.78: Comparison of the simulated with (green) and without the smearing (red) and experimental (blue) missing mass distributions in the inbending setting for the e^-K^+X sample on the whole missing mass range (left) and near the peak at 1.4 GeV (right)

3.6.2 Results using a more sophisticated RICH reconstruction

Studies were performed using the currently available files with the official RICH reconstruction (pass 1 version 11) made by Marco Contalbrigo [73, 89, 90]. First the correlation matrix between the RICH and the machine learning PID was calculated for positive kaons, that have a momentum larger than 3 GeV, are part of the SIDIS sample and have a RICH PID (Figure 3.79). This figure shows that the machine learning PID removes the pions from the kaon sample effectively, especially in the interesting kinematic regions.

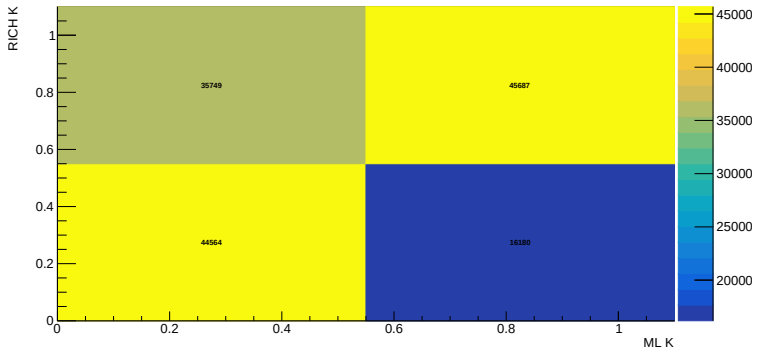


Figure 3.79: Correlation matrix between the RICH and the machine learning PID was calculated for positive kaons, that have a momentum larger than 3 GeV, are part of the SIDIS sample and have a RICH PID

One of the most important part of this analysis is the subtraction of the pion asymmetries. This procedure, which is described in the following section, highly relies on the kaon-pion contamination determined using the Monte Carlo sample. To check this estimation, the kaon-pion contamination was compared with the contamination measured by the RICH. Figure 3.80 shows the MC contamination (blue) and RICH contamination (red) as a function of momentum using positive Eventbuilder kaons. We can observe that above 3.5 GeV kaon momentum, where the RICH works reliably, the contaminations are very close.

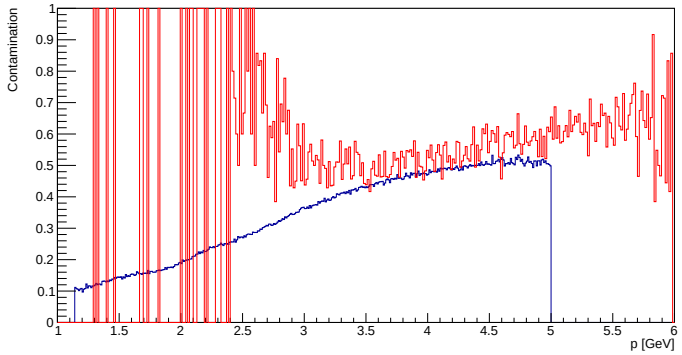


Figure 3.80: Comparison of the kaon-pion contamination obtained from MC (blue) and from RICH (red) as a function of momentum using positive Eventbuilder kaons

Figure 3.81 compares the extracted K^+ asymmetries using EB, ML and RICH PIDs to the pion asymmetries. The RICH results, which have the lowest pion contamination, follow the results obtained with the other PID methods. Most of the RICH asymmetries are smaller, but this is due to the fact that the RICH sample contains mostly events with high z, where the asymmetry is smaller than at low z.

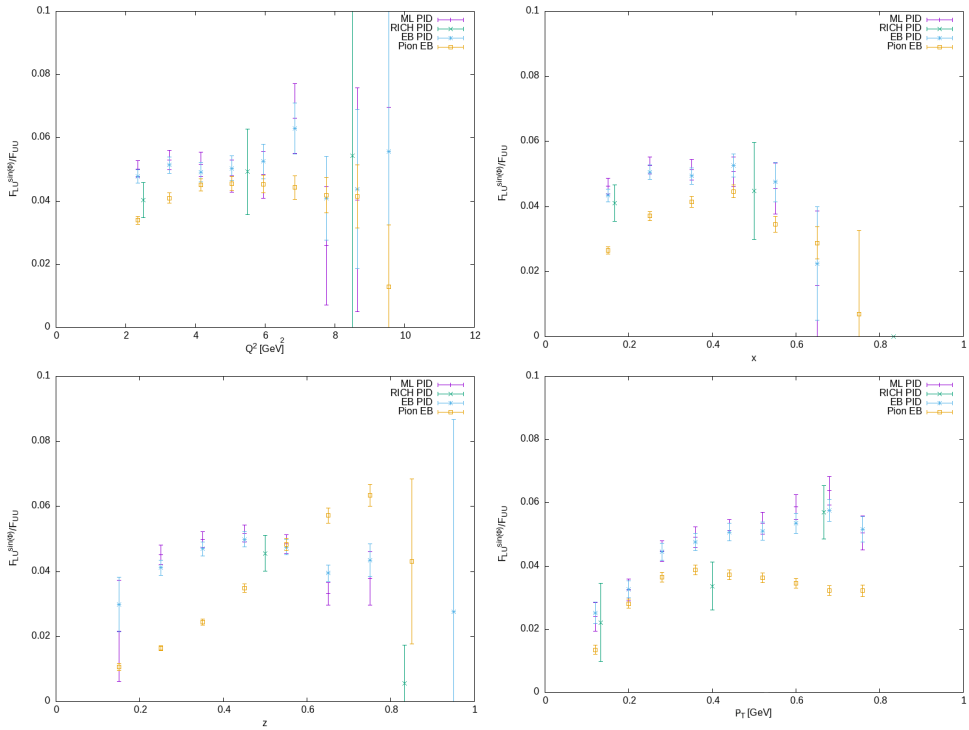
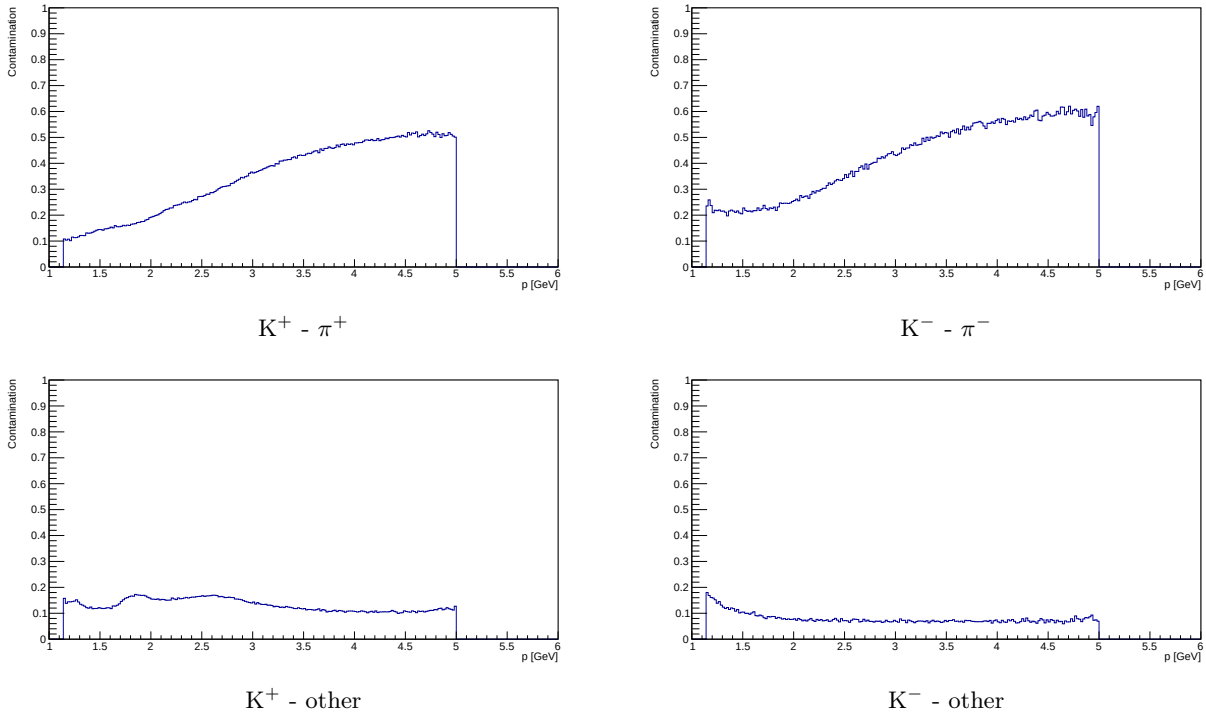


Figure 3.81: "Raw" $F_{LU}^{\sin\phi}/F_{UU}$ as a function of Q^2 (top left), x_B (top right), z (bottom left) and P_T (bottom right) for K^+ with different PIDs compared with π^+ results

Based on these studies we can conclude that the current RICH reconstruction is reliable to be used to validate the contamination estimates of the MC but it is not yet ready to be used to extract physics.

3.6.3 Pion subtraction

To subtract the pion asymmetries an accurate estimation of the pion contamination is necessary. For this detailed contamination studies were performed. The reconstructed kaons were matched with the generated kaons and pions based on there momenta. The same matching method was used as for the machine learning studies. More details can be found in section 3.4.1. The contamination values determined this way can be seen in Figure 3.82 as a function of momentum for the final selected kaon samples using the loosest PID in the inbending case. Here we can observe that the contamination steadily rises with the momentum as expected. At low momentum the contamination is around 10%, which is caused by the long reaching tail of the pions β-distribution. This is demonstrated on Figure 3.83. The average contamination is around 30% and 40% for positively and negatively charged kaons, respectively. The contamination from non-matched particles is mainly constant at 10%. These particles are mainly secondary kaons. This is supported by the fact that the missing mass spectrum of these non-matched particles is very similar to the matched kaons distribution, which can be seen in Figure 3.84. In case of positive kaons a proton band with incorrectly reconstructed time information crosses the kaon band on the β-momentum plane around 2 GeV. This creates a peak in the non-matched contamination there.



K^+ - π^+ K^- - π^-

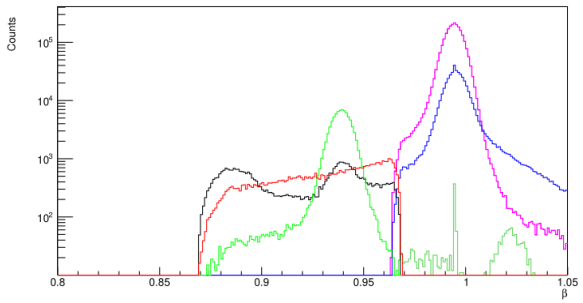
K^+ - other K^- - other

Figure 3.82: Contamination as a function of momentum for the final selected kaon samples using the loosest PID in the inbending case

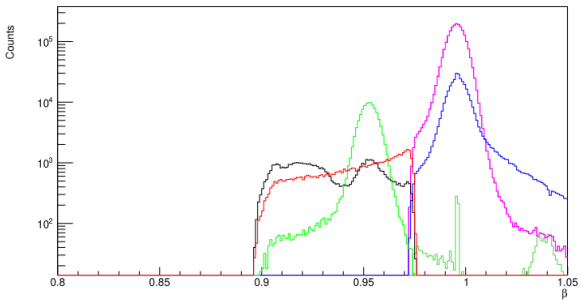
In every kinematic bin the pion contamination is calculated with the following formula for the combination of the 3 datasets:

$$C = \frac{1}{N_{in}^{rec} + N_{out}^{rec} + N_{2019}^{rec}} \left(\frac{N_{inmc}^{mcm\pi}}{N_{inmc}^{rec}} (N_{in}^{rec} + N_{2019}^{rec}) + \frac{N_{outmc}^{mcm\pi}}{N_{outmc}^{rec}} N_{out}^{rec} \right) \tag{3.32}$$

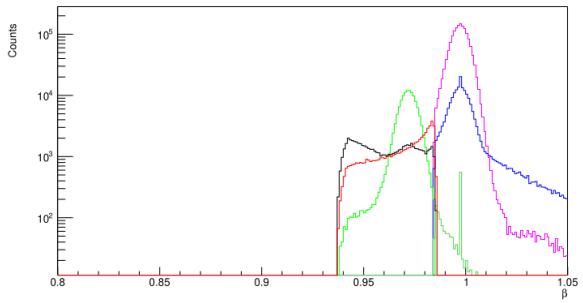
This way the overall contamination in a bin is the combination of the contaminations measured in the inbending and in the outbending MC samples in the given bins weighted with the ratio of the amount of events coming from the different settings ((inbending+2019):outbending) in the given bin.
In every kinematic bin the pion asymmetry is extracted with exactly the same cuts and methods used for kaons. The only exception is that the machine learning cut is not applied and that the highest momentum is fixed at 5 GeV in all cases. The four momenta of the pions were recalculated as they were kaons so that they would fall into the same bins as the pions in the kaon sample. These measured asymmetries are then subtracted from the



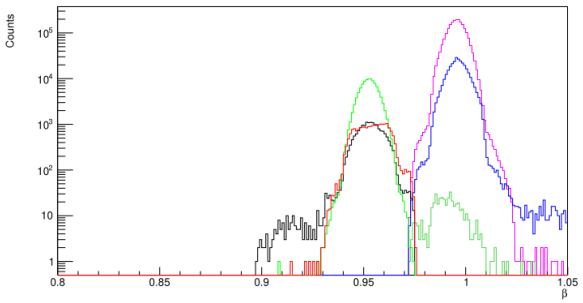
1.3 GeV $< p <$ 1.4 GeV



1.5 GeV $< p <$ 1.6 GeV

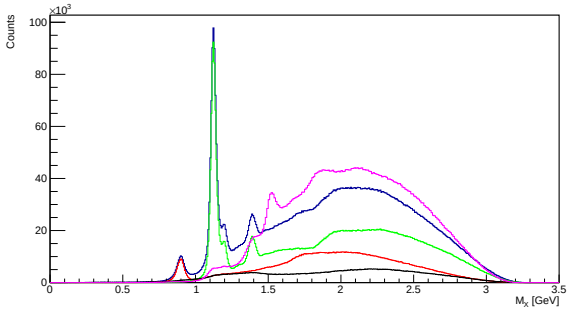


2.0 GeV $< p <$ 2.1 GeV



1.5 GeV $< p <$ 1.6 GeV, $|\chi^2| < 3$

Figure 3.83: Beta distributions in momentum bins without $|\chi^2| < 3$ (except right bottom) for the final selected positive kaon (red-generated as pion, green-generated as kaon and black-non-matched) and pion (purple - generated as pion, dark green-generated as kaon and dark blue-non-matched) sample using the loosest PID in the inbending case



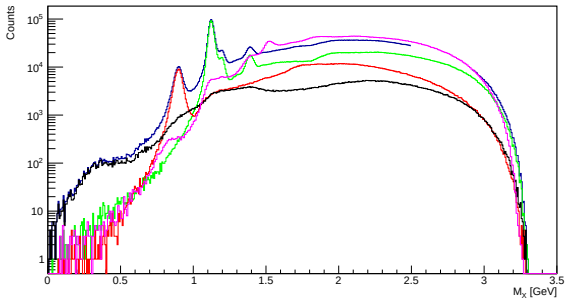
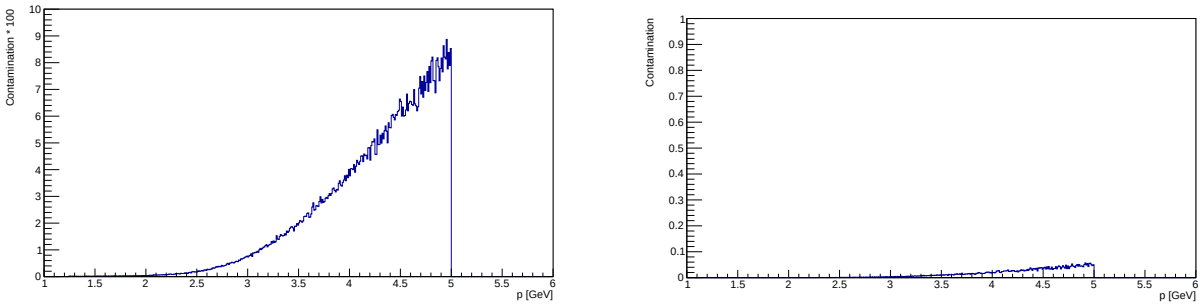


Figure 3.84: Missing mass distribution from the inbending simulation in blue for the e^-K^+X sample. The misidentified pions are shown in red and the correctly identified kaons are shown green. The non-matched particles are in black. The data is plotted in purple.

measured kaon asymmetries with the following formula in every bin:

$$A^K_{corr} = \frac{A^K_{meas} - CA^\pi_{meas}}{1 - C} \tag{3.33}$$

Here we assumed that the kaon contamination in the pion sample is negligible: $A^\pi_{meas} \approx A^\pi_{corr}$. This is indeed the case according to Figure 3.85. The effect of the bin migration of the pions on the subtraction was also investigated. Since it was found to be small, it was treated as a systematic uncertainty and described in the next section.



π^+ - K^+, In percentage π^- - K^-

Figure 3.85: Contamination as a function of momentum for the final selected pion samples using the loosest PID in the inbending case

The effect of the subtraction is demonstrated in Figures 3.86 and 3.87 for the different datasets. It can be observed, that the agreement between them is much better with the correction.

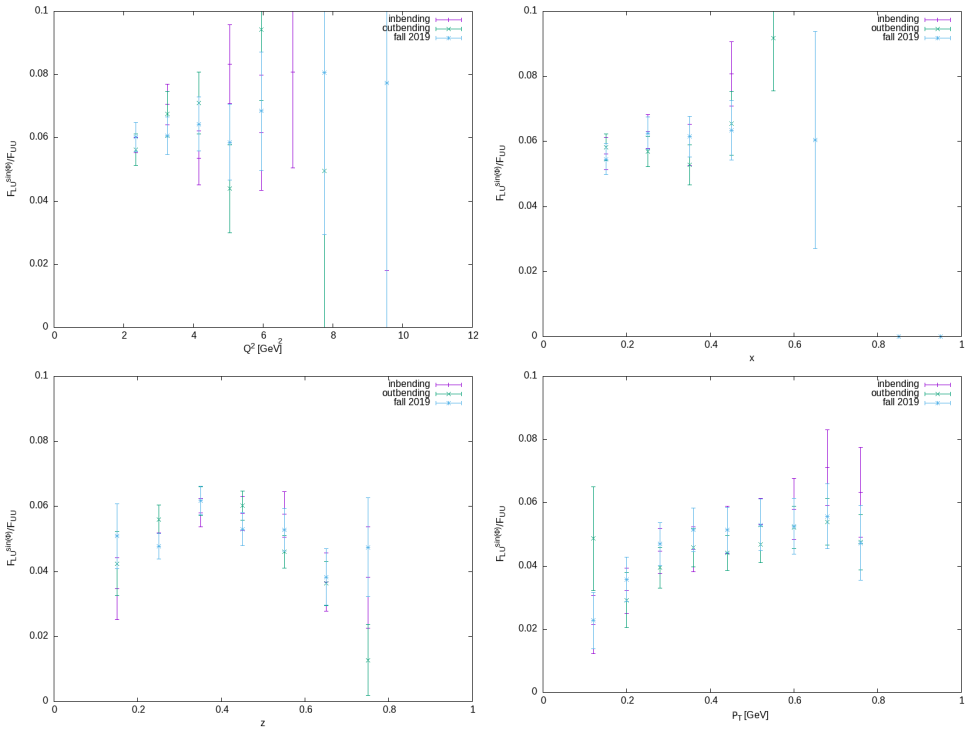


Figure 3.86: $F_{LU}^{\sin\phi}/F_{UU}$ as a function of Q^2 (top left), x_B (top right), z (bottom left) and P_T (bottom right) for K^+ using different datasets with subtracted π^+ asymmetries

For comparison the asymmetries were also extracted using a machine learning cut of 0.99. This way the statistics is very limited, but the contamination is estimated to be a few percent. The corrected asymmetries with different PIDs and the uncorrected asymmetries extracted with the very strict ml cut are compared in Figures 3.88 and 3.89. Here we can observe that the results obtained with an extremely strict PID and the corrected results with the loosest PID are consistent, which means that the subtraction works as intended.

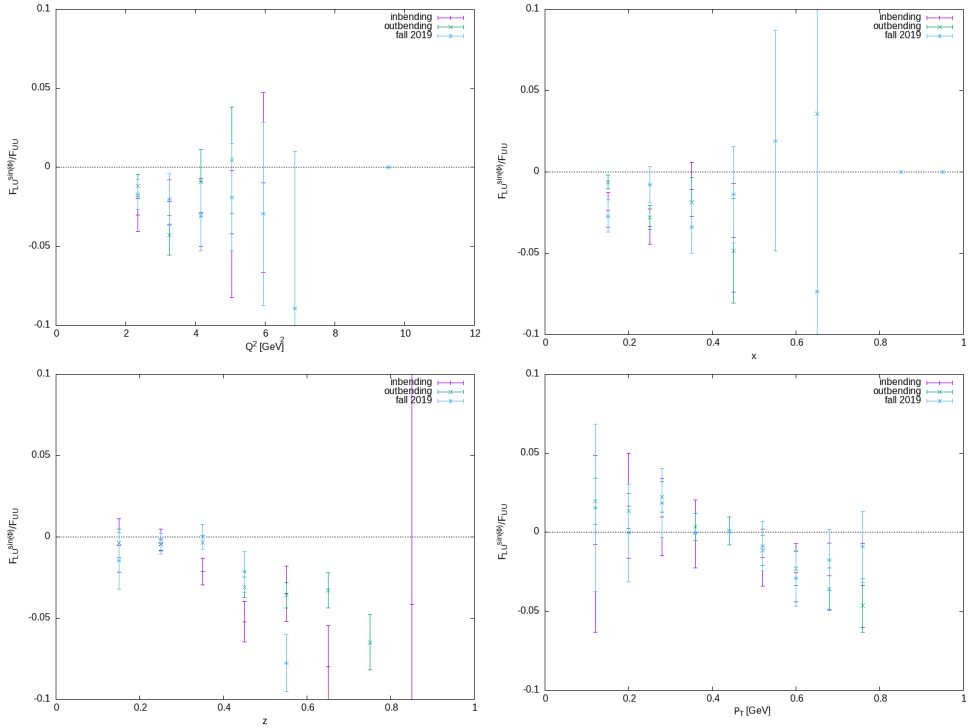


Figure 3.87: $F_{LU}^{\sin\phi}/F_{UU}$ as a function of Q^2 (top left), x_B (top right), z (bottom left) and P_T (bottom right) for K^- using different datasets with subtracted π^- asymmetries

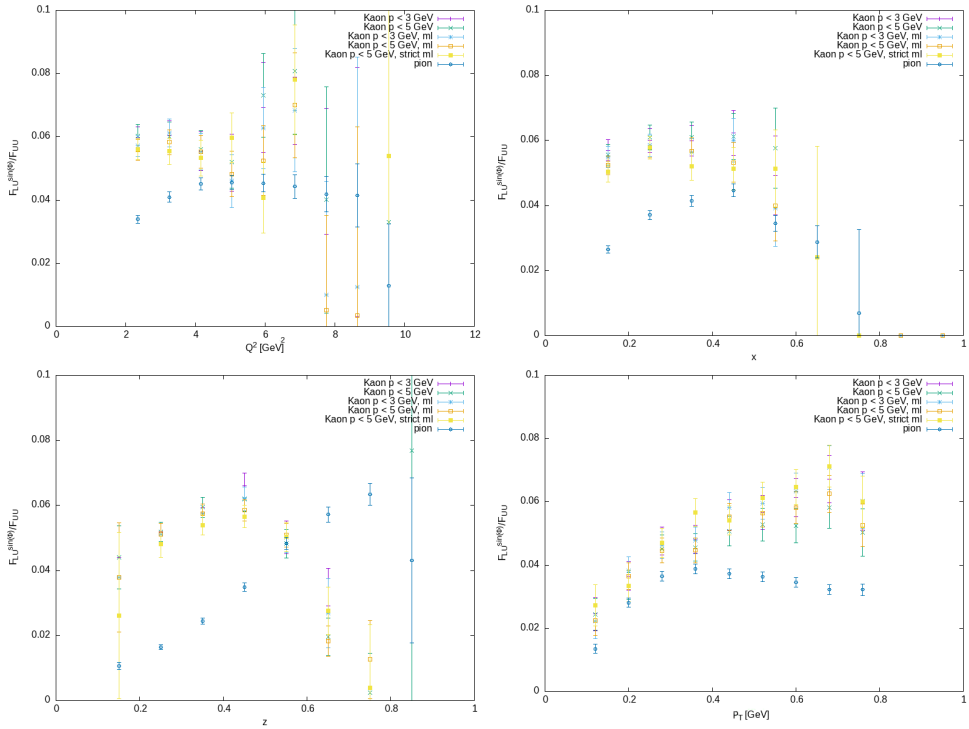


Figure 3.88: $F_{LU}^{\sin\phi}/F_{UU}$ as a function of Q^2 (top left), x_B (top right), z (bottom left) and P_T (bottom right) for K^+ using different PID with subtracted π^+ asymmetries

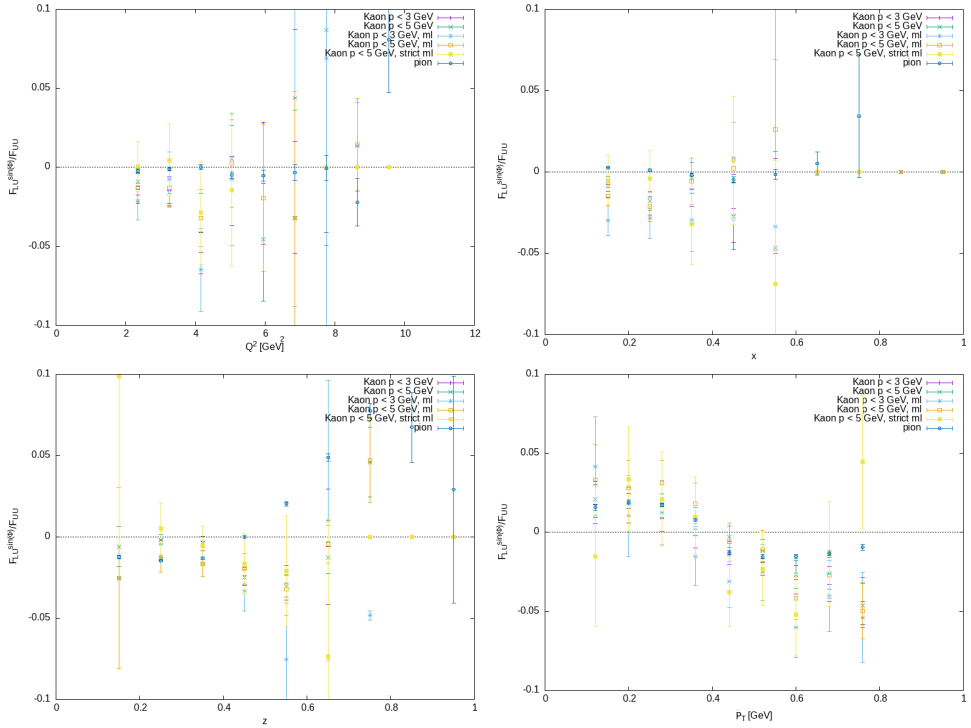


Figure 3.89: $F_{LU}^{\sin\phi}/F_{UU}$ as a function of Q^2 (top left), x_B (top right), z (bottom left) and P_T (bottom right) for K^- using different PID with subtracted π^- asymmetries

To verify the contamination estimated from the MC, one can look at the contamination determined by the RICH above 4.5 GeV momentum, where it works reliably. To verify the contamination at low momenta, the β distributions of EB kaons were fitted in momentum bins. As shown in Figure 3.83, there is significant background under the kaon peak coming from the the tail of the pion peak. As Figure 3.90 demonstrates, the β-distribution could be well described with a Gaussian and a second order polynomial background. To estimate the contamination the integrals of the fitted functions were calculated within 3 sigma.

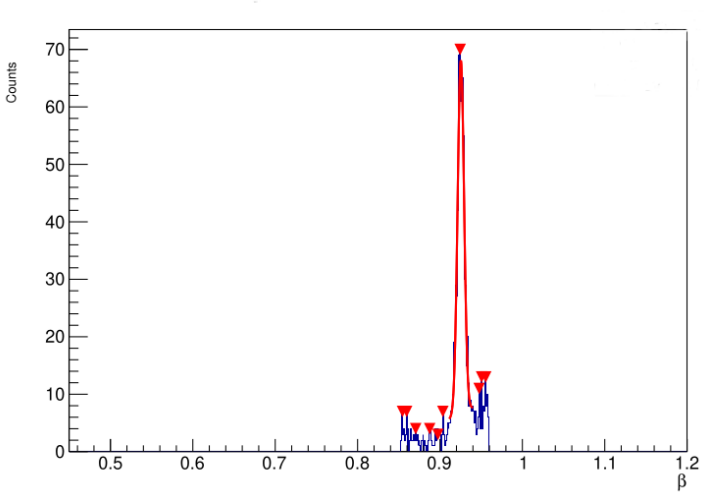
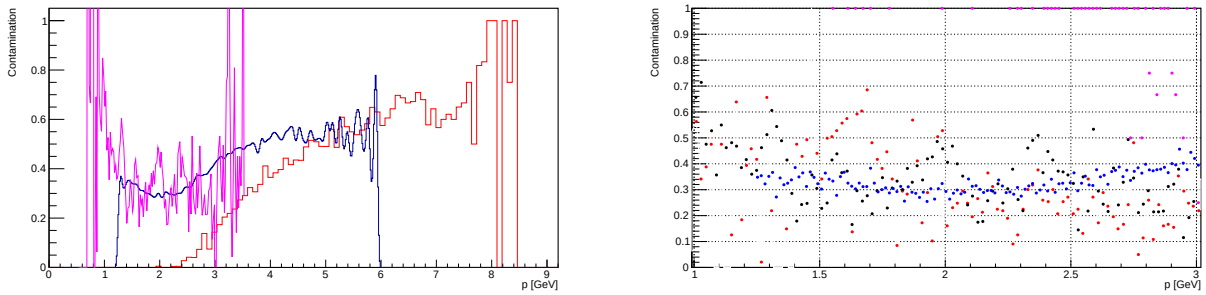


Figure 3.90: β-distribution of EB kaons fitted with a Gaussian and a polynomial background in one of the momentum bins

Figure 3.91 compares the pion contamination in the kaon sample determined by different methods. The RICH

contamination is in good agreement above 4.5 GeV with the MC contamination. That is also in agreement with the fits made to beta-p slices of data below 2.7 GeV, where the pion and kaon peaks start to overlap and the distributions are not described well by the fits. We can also observe an increase in the contamination just below 2 GeV, which is not present in the simulations. This is caused by the crossing bands from other beam bunches and it is only present in data. Its effect on the final results is evaluated in section 3.7.3. Due to the limited statistics, there are a lot of outliers in the β-p contaminations, but the tendencies clearly agree with the contamination determined from the MC. The fluctuation is a result of the uncertainties, which are not shown on the plots to allow a better visualization for the comparison.



(a) purple-β-p fits on data, blue-MC, red-RICH

(b) red-β-p fits on data, black-β-p fits on MC, blue-MC

Figure 3.91: Comparison of contaminations as a function of momentum determined by different methods over the whole momentum range on the left and at low momentum on the right.

3.6.4 Extraction method of the final results

As Figures 3.88 and 3.89 demonstrate all PID methods combined with the pion subtraction give the same results within statistical uncertainties. To extract the final results the $p < 5$ GeV with the standard machine learning PID was used together with the pions subtraction, since this PID setting requires the lowest correction due to the pion contamination, while keeping enough statistics to extract reasonable results in all multidimensional binning schemes. And in addition, this PID provides a smooth distribution in the kinematic variables which have large bins or are not binned. This way the distributions are well described with the averages and helps in the future interpretation of the results. The behavior of the variables are shown in Figures 3.57-3.60. The difference between the results obtained with different methods were assigned as a systematic uncertainty, which is almost exclusively coming from the uncertainty of the MC contamination as shown in section 3.7.9.
The final detailed results can be seen in Appendices I, II and III.

3.7 Systematic uncertainties

To determine the statistical uncertainty the following sources were considered in this analysis:

- Uncertainty of the beam polarization
- Effect of the fiducial cuts
- Contamination of the electron and kaon samples
- Contamination of the SIDIS sample
- Acceptance effects
- Effect of the extraction method and higher order moments
- Bin migration and resolution effects
- Radiative effects
- Effects of the pion subtraction

In the one dimensional binning case the systematic uncertainties are usually in the order of the statistical uncertainty. This means that the extraction of these uncertainties is possible without problems for most of the bins. The few problematic bins were not used for the final result anyway due to improper fits caused by the low statistics. For the multidimensional binning the systematic uncertainty is dominated by the statistical uncertainty. To overcome this problem an averaging was introduced in the last dimension, where the binning is the finest. The value of the systematic uncertainty in a given bin was averaged with its direct neighbors to reduce the impact of the bin to bin fluctuations. This procedure could be performed since we expect a smooth behavior of the systematic uncertainty from bin to bin.
Most of the studies presented in this section follows the precedent set by the pion SIDIS analysis [35, 79].

3.7.1 Uncertainty of the beam polarization

The beam polarization was measured upstream of CLAS12 several times during the run periods with a Møller polarimeter. The results of these Møller measurements are shown in Fig. 3.92 for the fall 2018 and spring 2019 data taking periods.
Th average beam polarizations for the 3 different run ranges are the following:

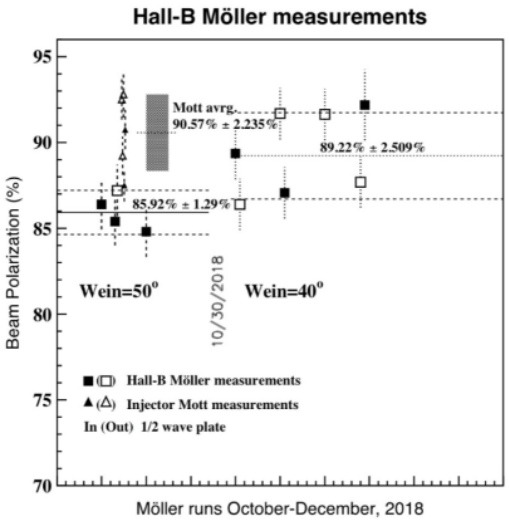
- Run 5032-5328 - (85.92±1.29)%
- Run 5331-5666 - (89.22±2.51)%
- Run 6616-6783 - (84.53±1.47)%

The average beam polarization for the combined dataset was calculated as a weighted average of the 3 values. The weights were determined as the proportion of the electron-kaon pairs in the given run range. The weights are the following for the e^-K^+X dataset:

- Run 5032-5328 - 22.5%
- Run 5331-5666 - 51.2%
- Run 6616-6783 - 26.3%

And for the e^-K^-X dataset:

- Run 5032-5328 - 9.5%
- Run 5331-5666 - 76.8%
- Run 6616-6783 - 13.7%



Polarization (%)	-84.529 +/- 1.474
Beam Charge Asymmetry (%)	-0.050 +/- 0.016
Half Wave Plate	IN
V Wien Angle (deg)	60.0
Target Position	Right
Beam Energy (MeV)	10199.8
Quad Current (A)	2924.5/2932.4
Helmholtz Current (A)	3.5
SLM Voltage (V)	1400.0
CODA Run Number	6725
Moller Run Number	78
Run Start Time	04/08/2019 20:37:30
Run End Time	04/08/2019 21:12:45

Figure 3.92: Møller measurements during the fall 2018 run period (left) and spring 2019 (right). The beam polarization was changed after October 28, 2018 and so two separate polarizations were considered for the fall data set. [91]

With the calibration uncertainty of the Møller polarimeter estimated to be 2%, this gives an average polarization (P) of (87.24±1.37±2)%=(87.24±2.42)% for positive and (88.26±1.94±2)%=(88.26±2.79)% for negative kaons. The sum of the statistical and systematic uncertainty of the beam polarization is considered to contribute to the systematic uncertainty of the BSA measurement:

$$\frac{\delta F_{LU}^{\sin\phi}/F_{UU}}{F_{LU}^{\sin\phi}/F_{UU}} = \frac{\delta A_{LU}^{\sin\phi}}{A_{LU}^{\sin\phi}} = \frac{\delta BSA}{BSA} = \frac{\delta P}{P} = \begin{cases} 2.8\%, & \text{for } K^+ \\ 3.2\%, & \text{for } K^- \end{cases} \tag{3.34}$$

3.7.2 Effect of the fiducial cuts

Since the same fiducial cuts have been applied to both helicity states, the systematic effect on the beam-spin asymmetry due to fiducial cuts is expected to be small, especially in a multidimensional binning. Nevertheless, to study the effect of the fiducial cuts on the result, the data was extracted with 3 different combinations of the DC and PCAL fiducial cuts. The DC fiducial cuts were loosened and tightened with 10% and the PCAL fiducial cuts were changed from normal to tight setting. Figures 3.93 and 3.94 show the absolute uncertainties in case of the one-dimensional binning. While Figure 3.95 shows the absolute and relative uncertainties in the fully multidimensional case.

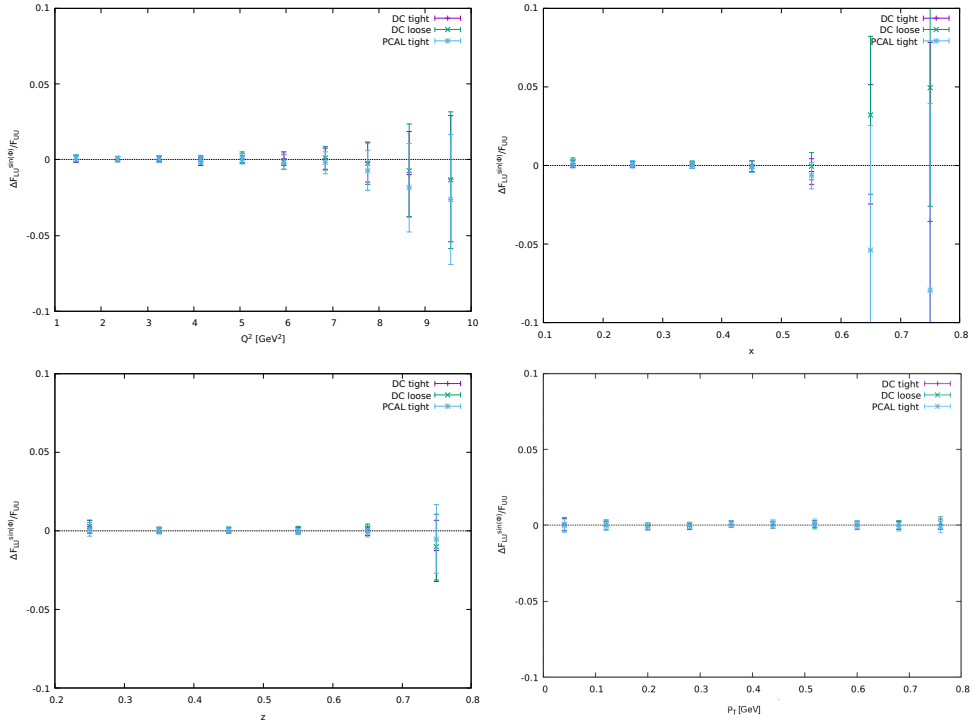


Figure 3.93: Absolute uncertainty of $F_{LU}^{\sin\phi}/F_{UU}$ as a function of Q^2 (top left), x_B (top right), z (bottom left) and P_T (bottom right) for K^+

In bins with sufficient statistics, the relative uncertainty is below 10% for K^+ and below 30% for K^-. The higher value in case of K^- is due to the fact that the asymmetries are very close to 0.

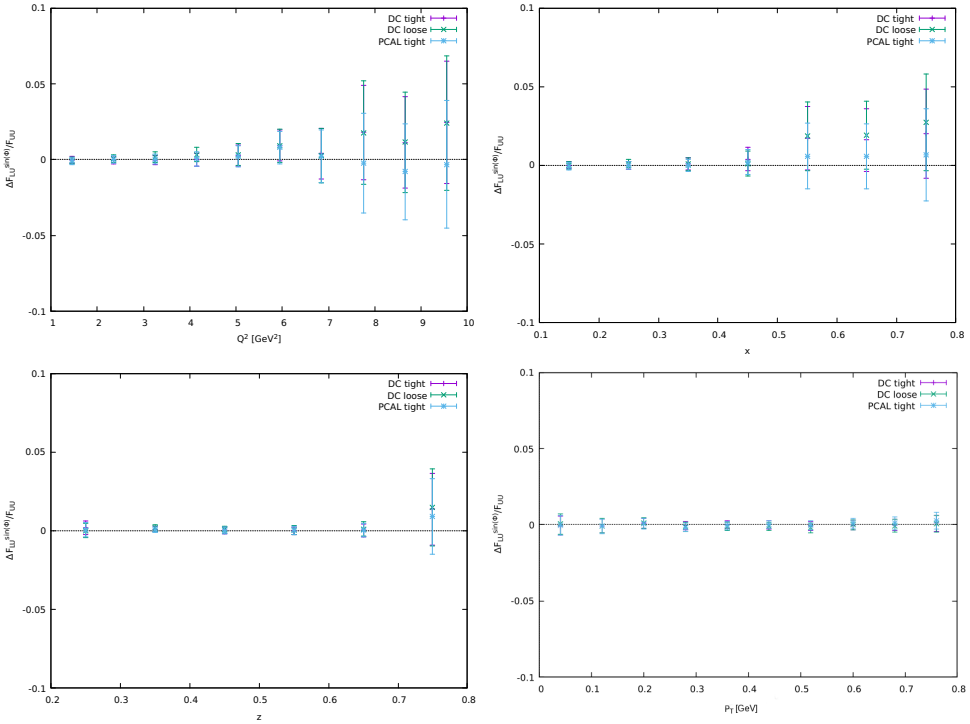


Figure 3.94: Absolute uncertainty of $F_{LU}^{\sin\phi}/F_{UU}$ as a function of Q^2 (top left), x_B (top right), z (bottom left) and P_T (bottom right) for K^-

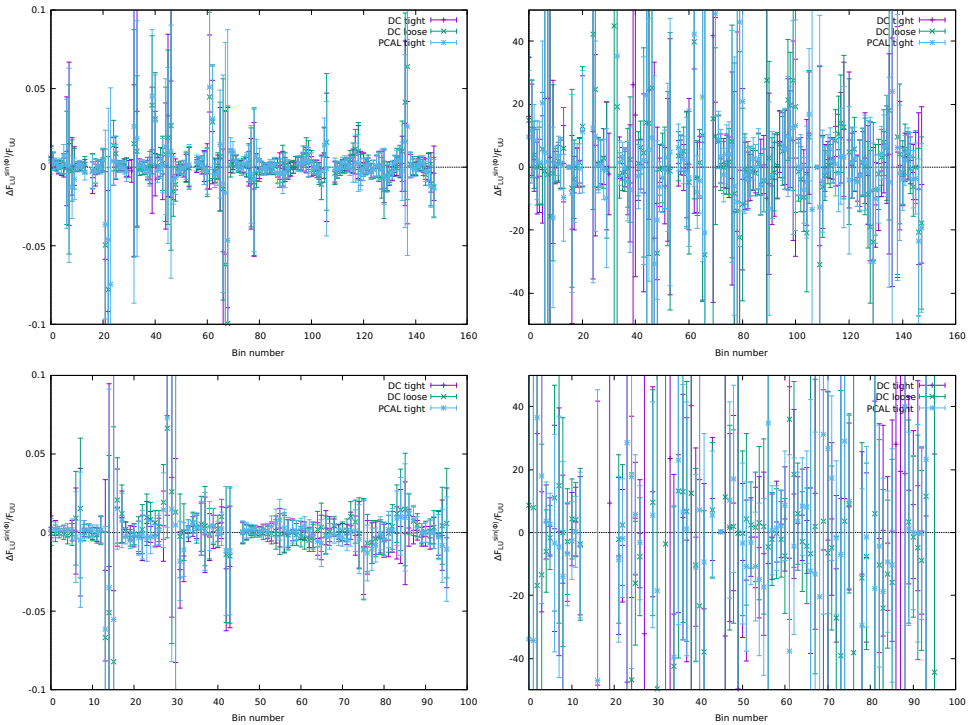


Figure 3.95: Absolute (left) and relative (right) uncertainty of $F_{LU}^{\sin\phi}/F_{UU}$ as a function of the fully multidimensional bin number for K^+ (top) and K^- (bottom)

3.7.3 Contamination of the electron and kaon samples

A very good single particle identification is essential for a SIDIS analysis, since no exclusivity cuts are applied in addition. Therefore many possible sources of contaminations were thoroughly investigated.

Contamination of the kaon sample with pions

Since the contamination of the kaon sample with π is very high, around 50% in some kinematic bins, it cannot be treated as a systematic uncertainty. Instead the pion asymmetries were subtracted bin by bin. See section 3.6 for more details.
The effect of the bin migration of the pions on the subtraction introduces a systematic uncertainty in the corrected asymmetries. This uncertainty is very small in most cases, but can be significant when the extracted asymmetries differ significantly in two neighboring bins. To estimate this the measured values were also corrected with contamination values determined with the consideration of bin migration. The bin purity and the migration rates for the neighboring 3 bins were determined using the MC sample with the same method used for kaons (section 3.7.7). The contamination was calculated in each bin according to Equation 3.32. Using these contamination values the corrected kaon asymmetry can be calculated with the following formula:

$$A^i_{corr} = (A^i_K - C^i(p^i \cdot A^i_\pi + r^i_{+1} \cdot A^{i+1}_\pi + r^i_{+2} \cdot A^i_\pi + r^i_{+3} \cdot A^{i+3}_\pi + r^i_{-1} \cdot A^{i-1}_\pi + r^i_{-2} \cdot A^{i-2}_\pi + r^i_{-3} \cdot A^{i-3}_\pi))/(1 - C^i) \quad (3.35)$$

,where C^i is the contamination from Equation 3.32, p^i is the bin purity and $r^i_{\pm 1,2,3}$ is the $\pm 1, 2, 3$ migration rate in the ith bin.
The difference between the corrected asymmetries with and without considering the pions bin migration can be treated as a systematic uncertainty. Figures 3.96 and 3.97 shows these absolute uncertainties in the 1 dimensional binning case. The relative uncertainties are typically around a few percent for K^+ and below 20% for K^-.

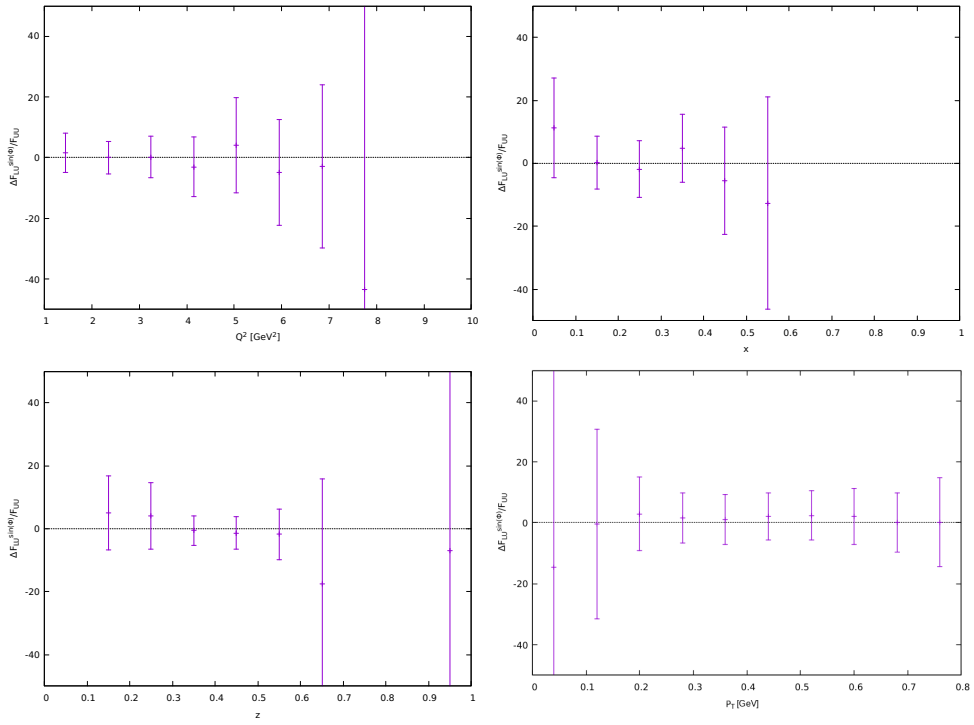


Figure 3.96: Relative uncertainty of $F^{\sin\phi}_{LU}/F_{UU}$ as a function of Q^2 (top left), x_B (top right), z (bottom left) and P_T (bottom right) for K^+

In the multidimensional cases the bin migration was only considered in the last binning dimension, similar to the kaon bin migration studies.

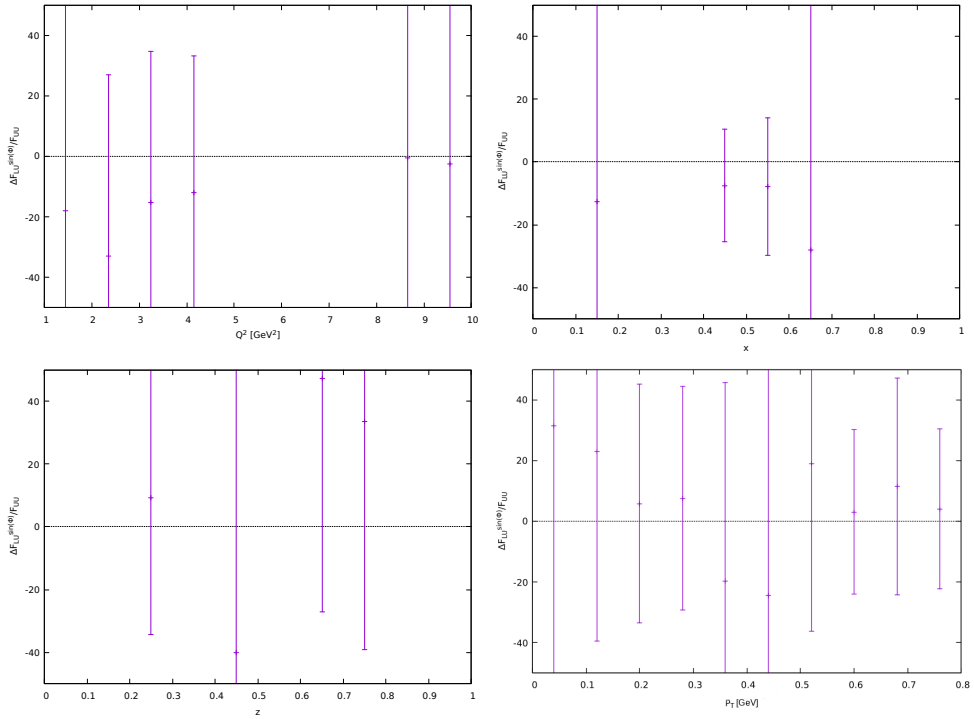


Figure 3.97: Relative uncertainty of $F_{LU}^{\sin\phi}/F_{UU}$ as a function of Q^2 (top left), x_B (top right), z (bottom left) and P_T (bottom right) for K^-

Contamination of the electron sample with π^-

To suppress the contamination of the electron sample by π^- different cuts on the energy deposition and sampling fraction of the different calorimeter components were used. Simulations showed that with these cuts the contamination above the HTCC threshold of 4.5 GeV is at a level of 2%, as stated previously. To estimate the effect of this contamination the cut on the sampling fraction was changed from 3.5σ to 3σ. This change was included in the investigations of the effects of the PCAL fiducial cuts (section 3.7.2).

Contamination of the kaon sample with accidentals from other beam bunches

The accidentals from the previous or from the following beam bunches appear as shifted bands with reduced intensity in the β-p-plane. For the kaon samples these shifted bands can introduce contamination if they cross the kaon band in the relevant momentum range. But as Figure 3.98 shows only the proton band crosses the main kaon band slightly below 1.25 GeV, which is below the lowest used momentum. So there is no systematic uncertainty caused by this effect.

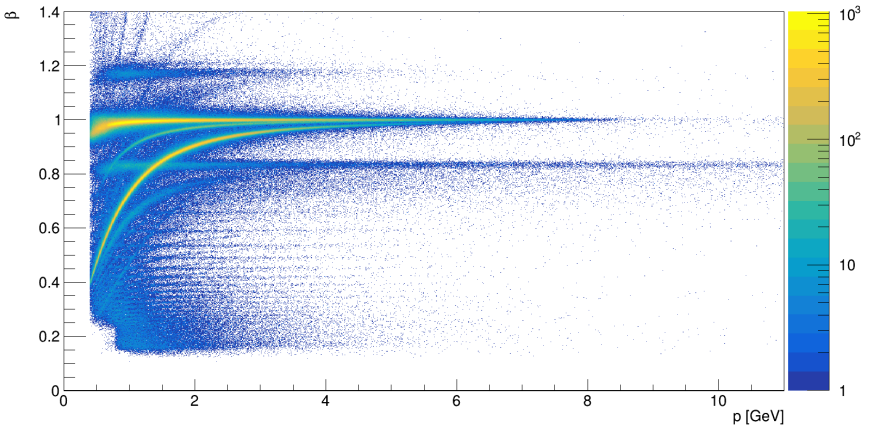


Figure 3.98: β versus momentum for hadrons in run 5197

Contamination of the kaon sample with accidentals from the same beam bunch

The other source of possible accidentals are the secondary interactions within the same beam bunch: secondary electro-production for which the electron is not detected and photo-productions. The rate of these accidentals can be estimated from the pion accidentals from the neighboring beam bunch. It is assumed that this rate is the same for pions and kaons. The ratio (f) of the pions in the band at $\beta \approx 1.1$ and in the main band at $\beta \approx 1$ was determined in momentum bins and shown in Figure 3.99.

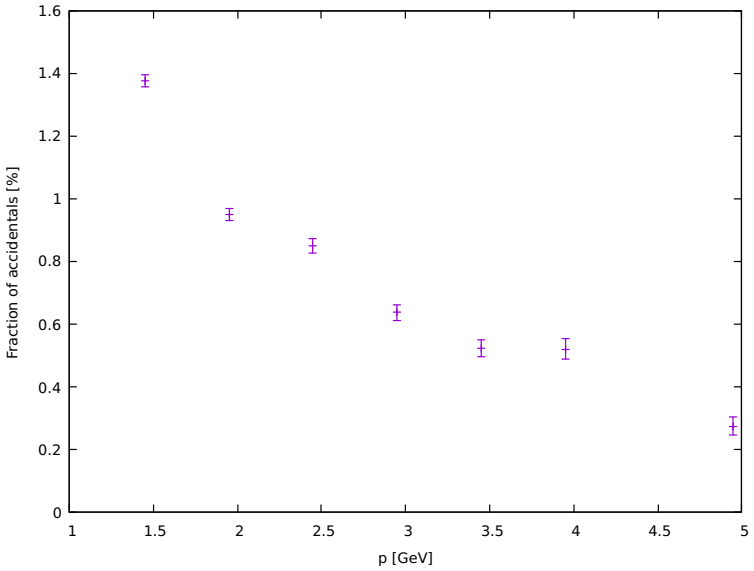


Figure 3.99: Fraction of accidentals as a function of momentum

It can be seen that the ratio is decreasing with the momentum and in the following we will estimate it to be 1%. Since the accidental kaons are not correlated with the electron, it can be assumed that the asymmetry of these events is 0. Thus the systematic uncertainty coming from this effect is the following:

$$\frac{\Delta A}{A} = \frac{f}{1-f} = 0.01 \tag{3.36}$$

3.7.4 Contamination of the SIDIS sample

In the following, the sources of the contamination of the selected $e^- K^{\pm} X$ samples besides the contamination of the single particle species are discussed.

Contamination by hadrons from photo-production

Electron-produced decays of photo-produced hadrons, low-Q^2 electro-production, when the original DIS electron goes down the beam pipe, and pair production are important sources of non-DIS contamination of the trigger electron sample. The fraction of these electrons can be estimated from measurements of positrons, since most of them are produced with a corresponding positron. Given that the inbending electrons show similar acceptance characteristics as outbending positrons and vice versa, the background contribution and asymmetry was estimated by analyzing the $e^+ K^{\pm} X$ (skim6) samples. It has been checked that the kaon acceptance has no significant impact on the obtained results. The in- and outbending skim6 dataset was analyzed with the same methods and cuts used for the main analysis with the only difference that the electrons were replaced with positrons.
Figures 3.100 and 3.101 show the distributions of these skim6 inbending $e^+ K^+ X$ and outbending $e^+ K^- X$ events respectively. Here we can observe that the background is focused on the upper Q^2 border of the low x_B region.

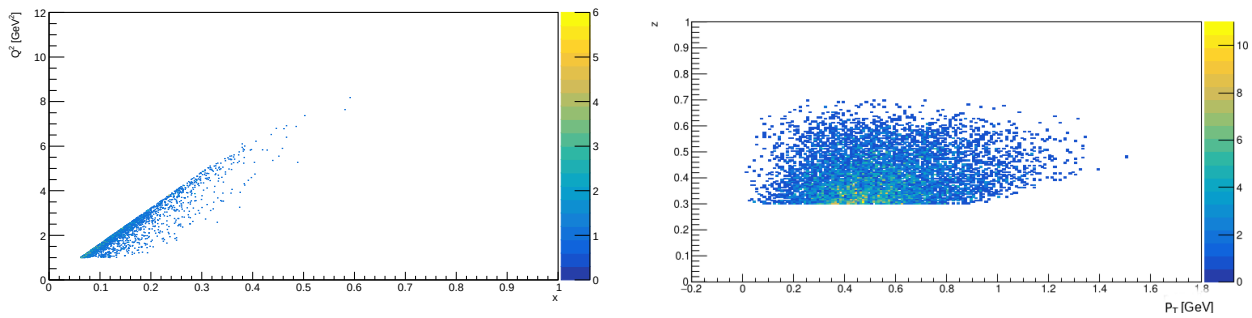


Figure 3.100: The correlations of Q^2-x_B (left) and P_T-z (right) for the $e^+ K^+ X$ inbending sample

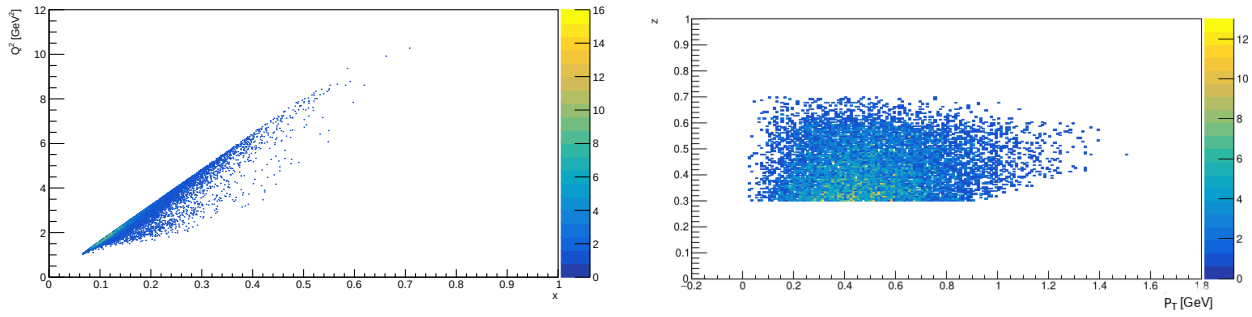


Figure 3.101: The correlations of Q^2-x_B (left) and P_T-z (right) for the $e^+ K^- X$ outbending sample

To calculate the contamination fraction (c), the corresponding electron and positron datasets have been normalized based on the region of the electron/positron angle θ between 15° and 25°, in which both show a comparable acceptance. Figure 3.102 shows the contamination as a function of θ and y. The contamination is below 0.5% if we are below 0.7 in y, which is why this cut was applied.
The measured asymmetry can be expressed as:

$$F_M = (1 - c)F_T + cF_C \tag{3.37}$$

, where F_T is the true asymmetry and F_C is the asymmetry of the charge symmetric background, which is

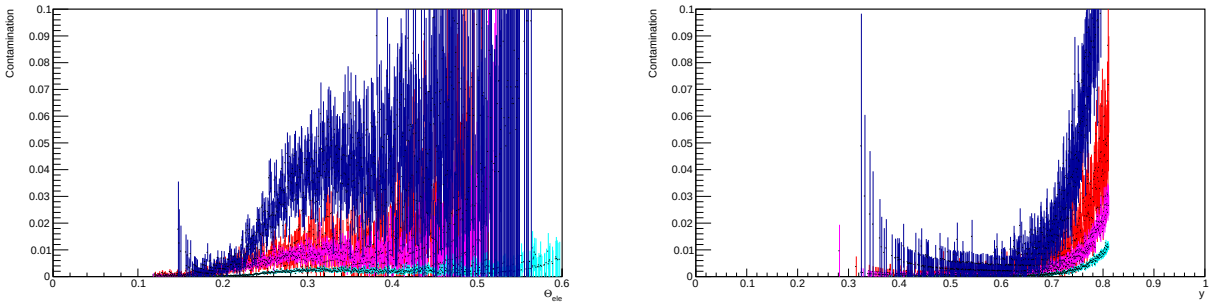


Figure 3.102: Charge-symmetric background contamination as a function of θ (left) and y (right). Turquoise - inbending K^+, Red - outbending K^+, Blue - inbending K^-, Purple - outbending K^-

assumed to be 0, like for pions [35]. This assumption could not be verified since the ϕ distribution cannot be fitted due to very low statistics. Then the relative systematic uncertainty of $F_{LU}^{\sin\phi}/F_{UU}$ due to the charge symmetric background is the following:

$$\frac{\Delta F}{F_M} = \frac{F_T - F_M}{F_M} = \frac{c}{1-c} \tag{3.38}$$

Figures 3.103 and 3.104 show the relative systematic uncertainty in case of the 1 dimensional binning. For K^+ it is below 0.5% and for K^- below 2%.

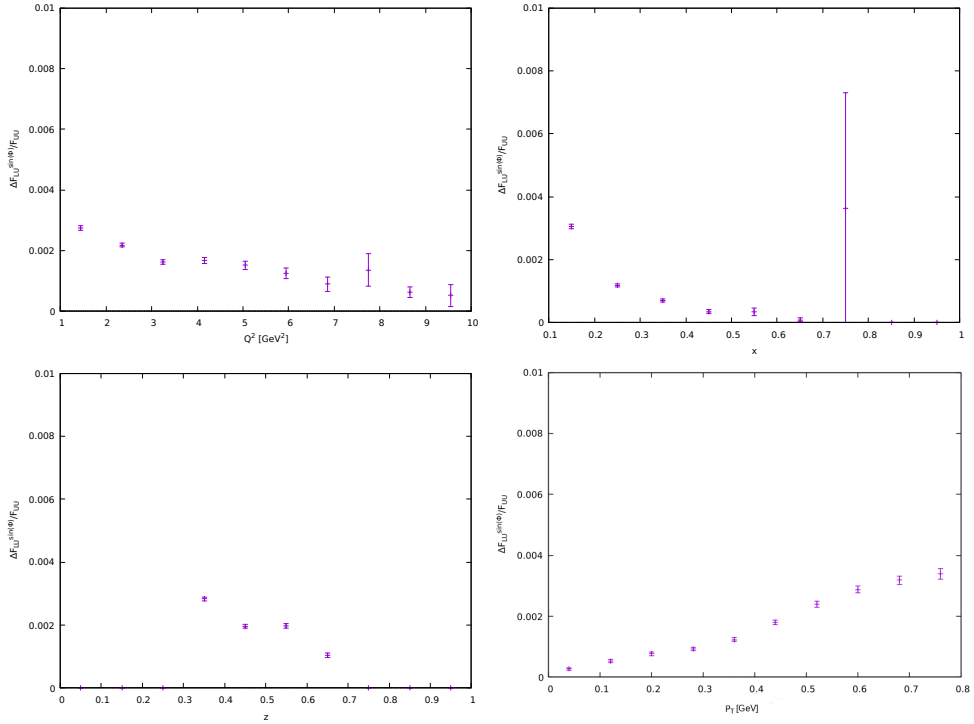


Figure 3.103: Relative systematic uncertainty of $F_{LU}^{\sin\phi}/F_{UU}$ due to the charge symmetric background as a function of Q^2 (top left), x_B (top right), z (bottom left) and P_T (bottom right) for K^+

The statistics of the skim6 datasets were not enough to repeat the studies for the multidimensional case, so the uncertainties were taken from the 1-dimensional bin corresponding to the bin of the last-dimension in the multidimensional cases. This is reasonable, since the uncertainty values show a very smooth behavior in all variables and the x-Q^2 and primary p_T/z bins are very wide.

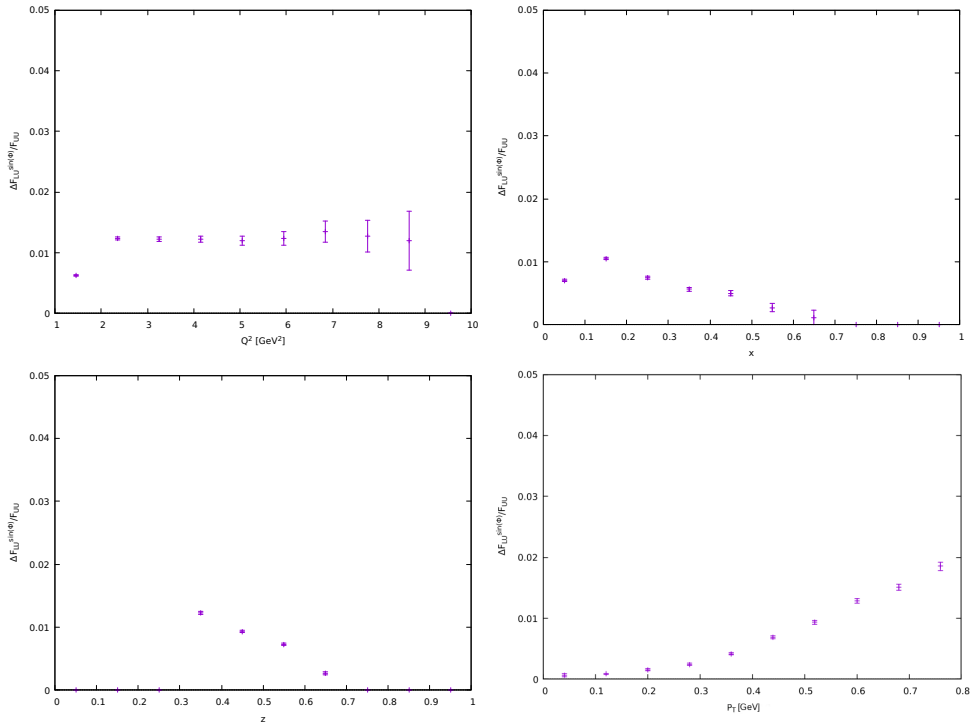


Figure 3.104: Relative systematic uncertainty of $F_{LU}^{\sin\phi}/F_{UU}$ due to the charge symmetric background as a function of Q^2 (top left), x_B (top right), z (bottom left) and P_T (bottom right) for K^-

Contamination by exclusive processes

An other possible source of contamination of the SIDIS sample are the events from exclusive processes. Most of these events were removed by the cut on the missing mass at 1.6 GeV. The leakage of these events into the SIDIS samples via radiative effects are evaluated in the radiative effects section later. There is a peaks in the missing mass spectra, which is not completely removed by this cut for K^+ at $\sim$2 GeV. But since this peak has almost no effect on the extracted asymmetries according to Figure 3.42, the related uncertainty can be neglected. The leakage of these events via radiative effects are evaluated in section 3.7.8.

Contributions from baryonic resonances

Factorization of the cross section into distribution and current fragmentation functions does not apply to to particles coming from the target fragmentation region. Any event corresponding to the production and decay from a baryon resonance is not part of SIDIS in the current fragmentation region and if incorrectly included in the event sample will alter the true asymmetry. The PYTHIA MC, optimized for CLAS12 was used to estimate the contamination coming from the decay of baryonic resonances. Figure 3.105 shows the PID of the mothers of all generated positive (blue) and negative (red) kaons. All kaons above 2000 are coming from baryonic resonances.
The reconstructed and generated kaons were matched using the method describes in section 3.4.1. After placing the kinematic cuts on the reconstructed particles described in section 3.5.1, all events above 2000 disappear for K^- and only a few survives for K^+. In the second case they are 0.06% of all events. This shows that these cuts are working very well and the uncertainty coming from the baryonic resonances are negligible compared to the other sources.

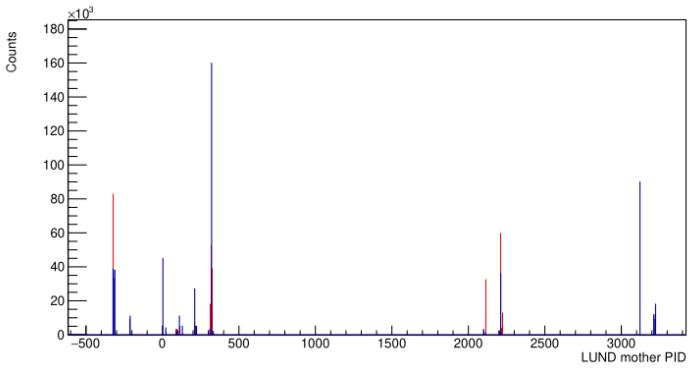


Figure 3.105: PID of the mothers of all generated psoitive (blue) and negative (red) kaons from the PYTHIA MC, optimized for CLAS12

3.7.5 Acceptance effects

Although the contribution of this source is expected to be small, since helicity independent acceptance terms cancel out in the cancellation of the BSA, it was investigated using the large scale MC sample. After matching the reconstructed and generated particles with the same method used previously (section 3.4.1), all kinematic variables are calculated for the reconstructed and for the generated particles. Using the generated values a weight is calculated according to the following:

$$\text{asymm} = P_e A_{inj} \sin\phi_{gen} \tag{3.39}$$

$$\text{weight} = \begin{cases} \frac{1+\text{asymm}}{1-\text{asymm}}, & \text{if helicity} = +1 \\ 1, & \text{if helicity} = -1 \end{cases} \tag{3.40}$$

, where P_e is the beam polarization, A_{inj} is the injected asymmetry and helicity is the helicity state which is randomly assigned to MC events. Different cases of A_{inj} have been investigated for the one dimensional binning:

$$A_{inj} = \begin{cases} 0.031 \\ 0.025 + 0.012 \cdot z \\ 0.015 + 0.04 \cdot p_T \end{cases} \qquad \text{for } K^+ \tag{3.41}$$

$$A_{inj} = \begin{cases} -0.013 \\ 0.04 - 0.28 \cdot z + 0.35 \cdot z^2 \end{cases} \qquad \text{for } K^- \tag{3.42}$$

A few examples of the comparison of the injected and extracted asymmetries are shown in Figure 3.106.

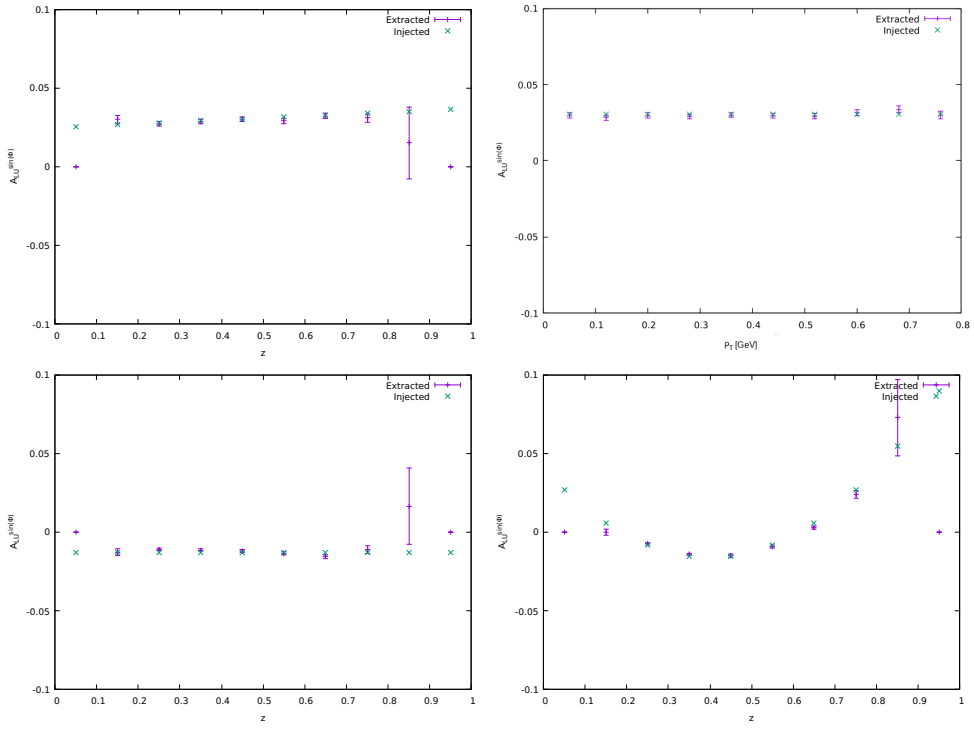


Figure 3.106: Injected and extracted $A_{LU}^{\sin\phi}$ using realistic z- (bottom left) and realistic P_T-dependence (bottom right) for K^+ and using constant (bottom left) and realistic z-dependence (bottom right) for K^-

The difference between the injected and extracted asymmetry is assigned as a systematic uncertainty bin by bin. Figures 3.107 and 3.108 show the relative uncertainties for the 1 dimensional case.
For the multidimensional binning cases the same study was performed, but only using the constant asymmetry.

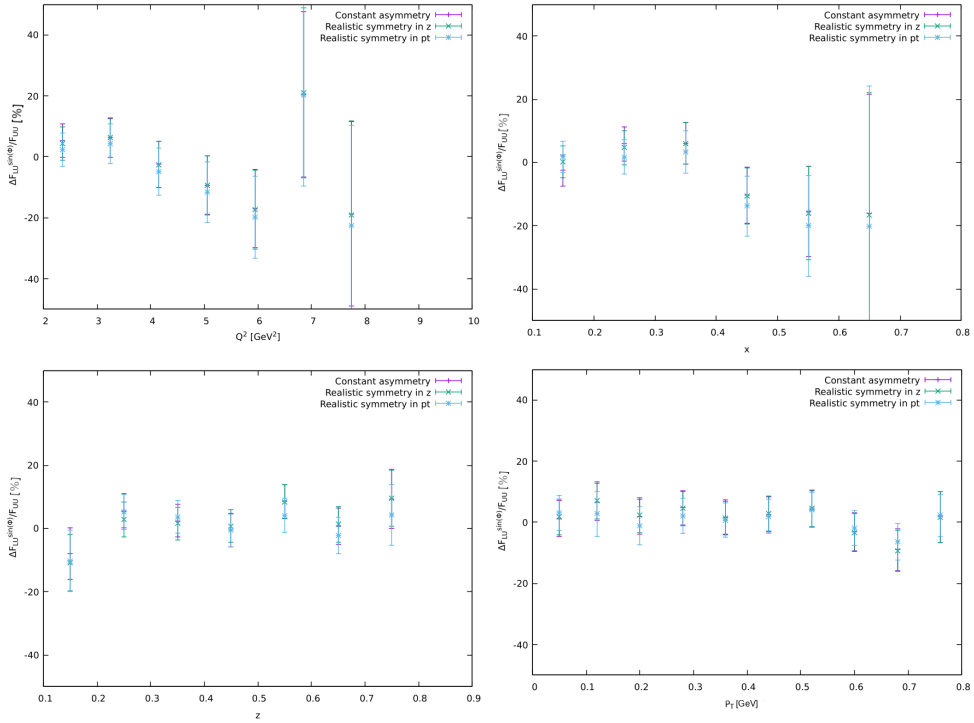


Figure 3.107: Relative systematic uncertainty of $F_{LU}^{\sin\phi}/F_{UU}$ due to acceptance effects as a function of Q^2 (top left), x_B (top right), z (bottom left) and P_T (bottom right) for K^+

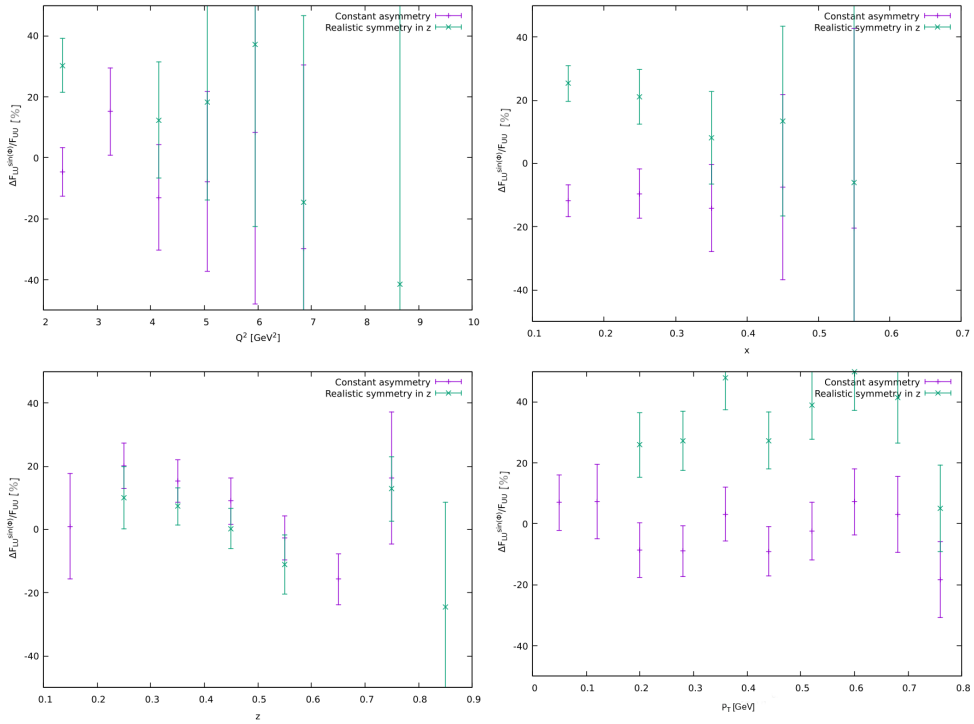


Figure 3.108: Relative systematic uncertainty of $F_{LU}^{\sin\phi}/F_{UU}$ due to acceptance effects as a function of Q^2 (top left), x_B (top right), z (bottom left) and P_T (bottom right) for K^-

3.7.6 Effect of the extraction method and higher order moments

Due to the reduced statistics, the BSA values were fit with the following function:

$$BSA = A_{LU}^{\sin\phi} \sin\phi \tag{3.43}$$

This assumes that the correlation is low between the $\sin\phi$ moment and the other moments of the BSA. To check the validity and the effect of this assumption on the extracted moments, the fits were also done with the full dependence:

$$BSA = \frac{A_{LU}^{\sin\phi} \sin\phi}{1 + A_{UU}^{\cos\phi} \cos\phi + A_{UU}^{\cos 2\phi} \cos 2\phi} \tag{3.44}$$

Since the fits could not be performed reliably due to the limited statistics, the $\cos\phi$ and $\cos 2\phi$ moments were limited. The limits are based on the values obtained from fitting the integrated BSA and from fits of one dimensional bins with sufficient statistics. In these cases the limits had a small effect on the results. The limits are the following:

$$|A_{UU}^{\cos\phi}| < \begin{cases} 0.2 \text{ or } 0.5, & \text{for } K^+ \\ 1, & \text{for } K^- \end{cases} \tag{3.45}$$

$$|A_{UU}^{\cos 2\phi}| < \begin{cases} 0.05 \text{ or } 0.3, & \text{for } K^+ \\ 0.05, & \text{for } K^- \end{cases} \tag{3.46}$$

To get a reliable result, in case of K^- the $\sin\phi$ term also had to be limited:

$$|A_{LU}^{\sin\phi}| < 0.5 \tag{3.47}$$

To check to validity of the limits, the K^+ sample was fit with 2 different limits and without a limit as well in the 1 dimensional case. The fits are reasonable in most bins. The extracted moments shown in Figures 3.109, 3.110 and 3.111 are in agreement within uncertainties in bins with sufficient statistics.
The difference of the obtained results with the two different fit functions is assigned as a systematic uncertainty. Figures 3.112 and 3.113 show the absolute uncertainties in case of the one-dimensional binning. While Figure 3.114 shows the absolute and relative uncertainties in the fully multidimensional case.

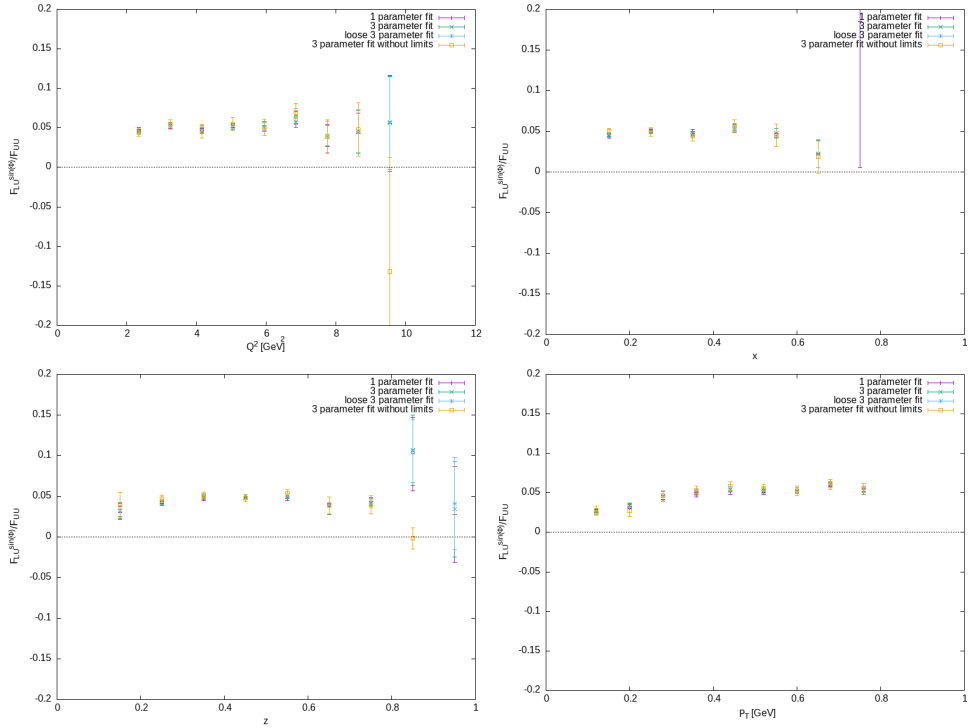


Figure 3.109: Extracted $F_{LU}^{\sin\phi}/F_{UU}$ with different limits and fit functions as a function of Q^2 (top left), x_B (top right), z (bottom left) and P_T (bottom right) for K^+

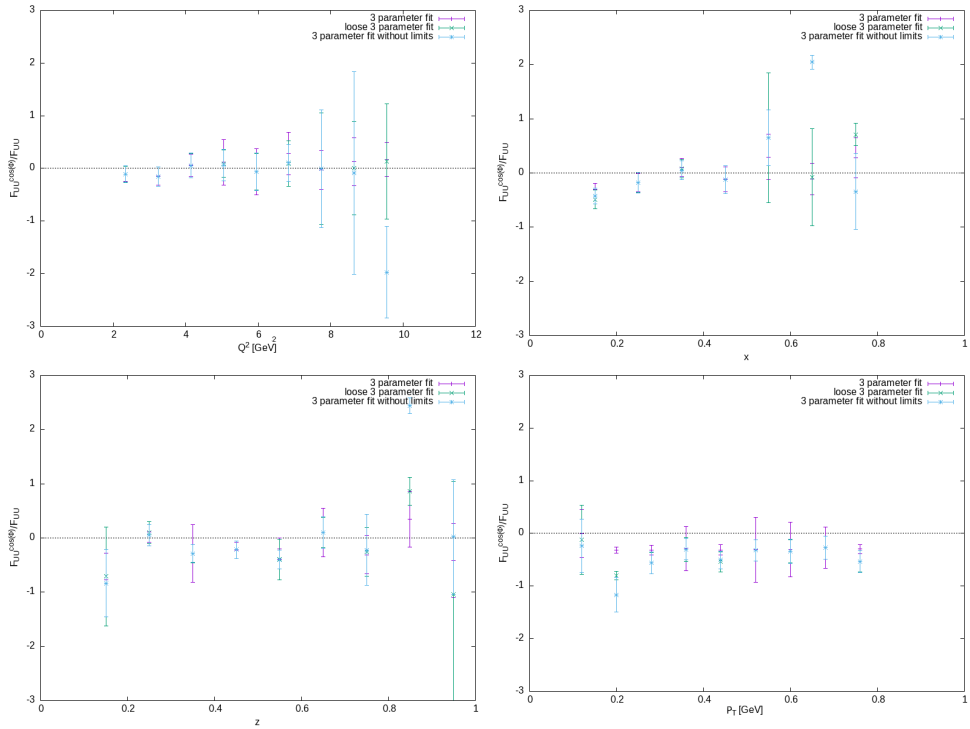


Figure 3.110: Extracted $F_{UU}^{\cos\phi}/F_{UU}$ with different limits and fit functions as a function of Q^2 (top left), x_B (top right), z (bottom left) and P_T (bottom right) for K^+

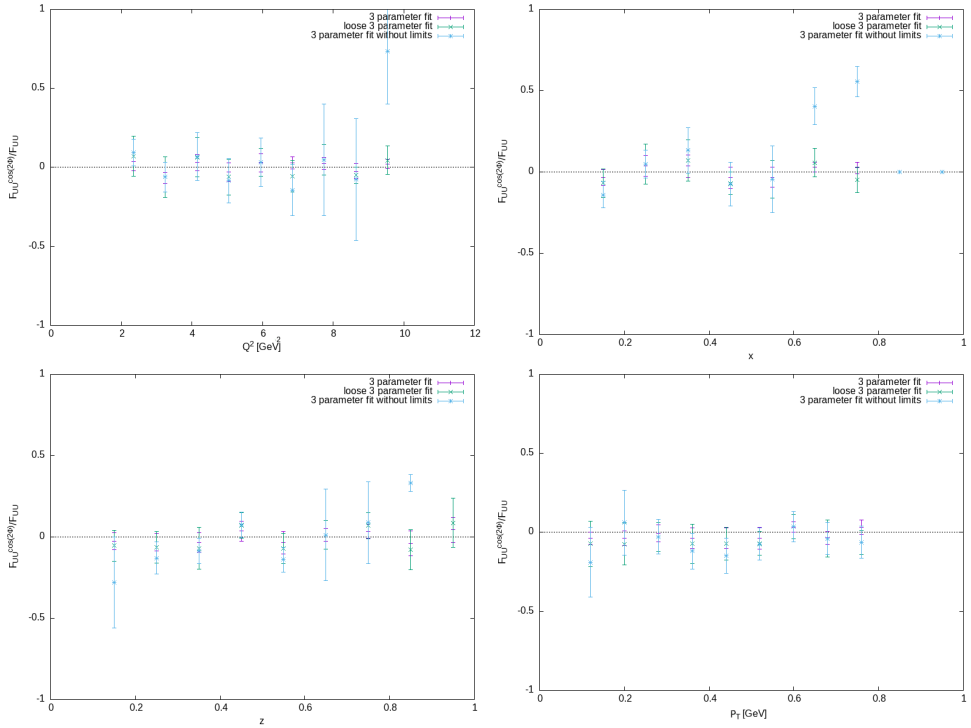


Figure 3.111: Extracted $F_{UU}^{\cos 2\phi}/F_{UU}$ with different limits and fit functions as a function of Q^2 (top left), x_B (top right), z (bottom left) and P_T (bottom right) for K^+

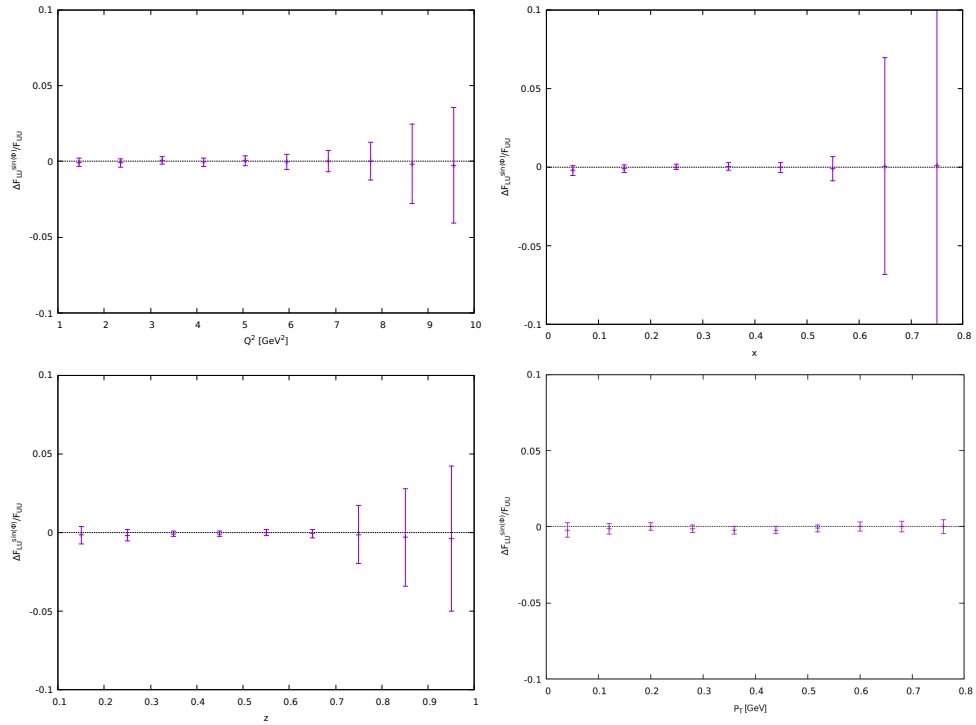


Figure 3.112: Absolute uncertainty of $F_{LU}^{\sin\phi}/F_{UU}$ as a function of Q^2 (top left), x_B (top right), z (bottom left) and P_T (bottom right) for K^+

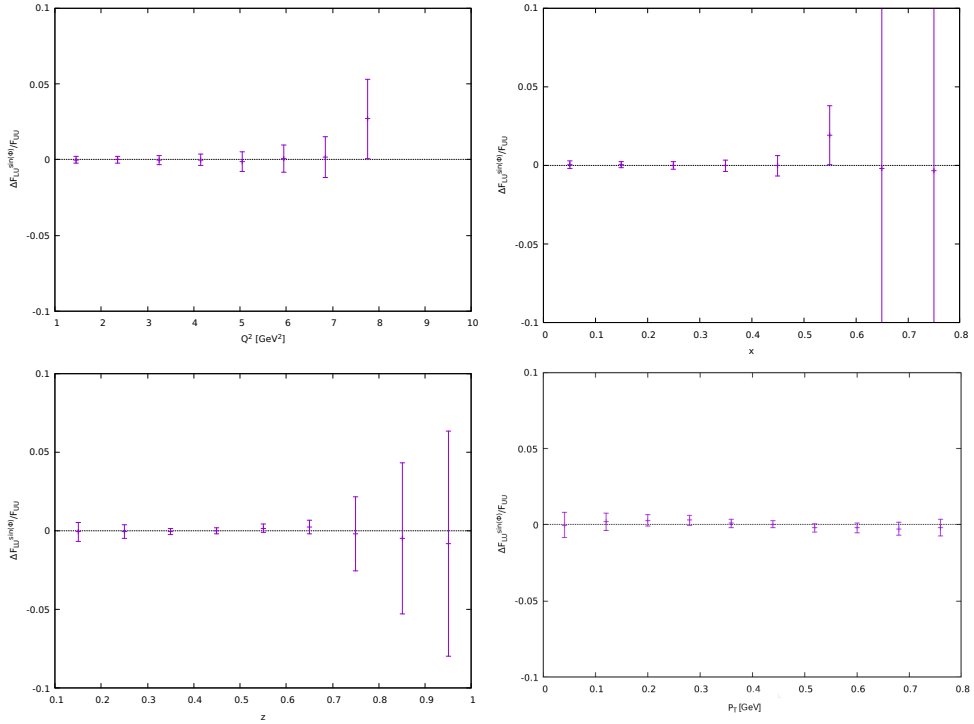


Figure 3.113: Absolute uncertainty of $F_{LU}^{\sin\phi}/F_{UU}$ as a function of Q^2 (top left), x_B (top right), z (bottom left) and P_T (bottom right) for K^-

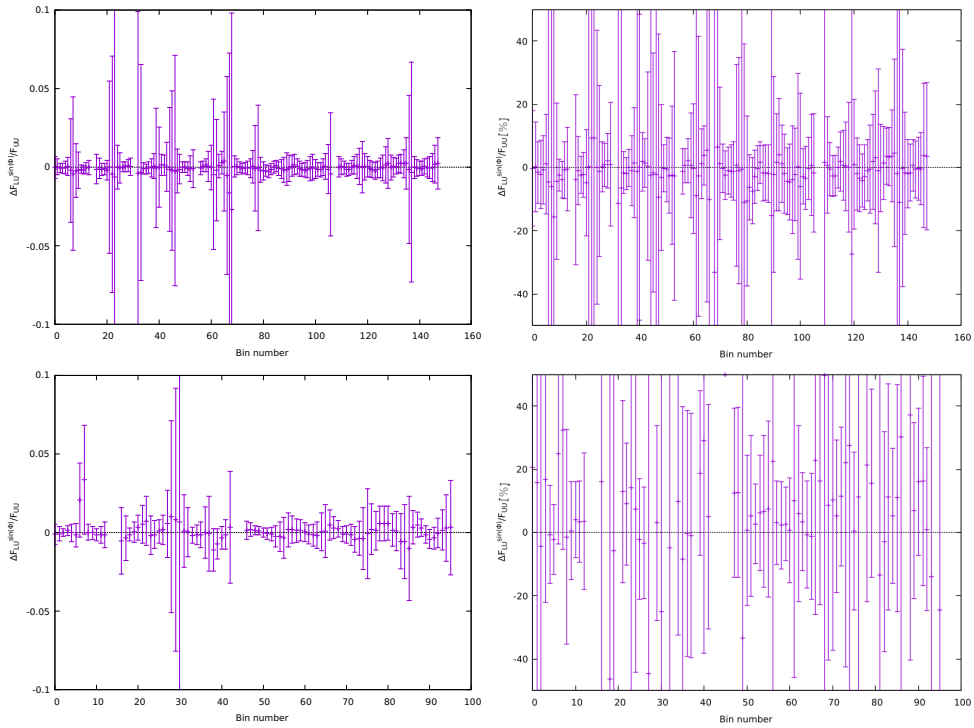


Figure 3.114: Absolute (left) and relative (right) uncertainty of $F_{LU}^{\sin\phi}/F_{UU}$ as a function of the fully multidimensional bin number for K^+ (top) and K^- (bottom)

3.7.7 Bin-migration effects

Due to the finite resolution of the measured kinematic variables, the reconstructed values used to study the beam spin asymmetry could be assigned to a different bin than they should be assigned to. This source of uncertainty is small if the measured asymmetries in neighboring bins are only slightly different, which is the case for the majority of the bins.
The large scale MC production was used to study this systematic effect. The reconstructed and generated particles were matched with the same method used previously (section 3.4.1) and the realistic resolution function was used from section 3.6.1. For each event and each kinematic variable, the bin in which the kinematic variable was generated and the bin in which it was reconstructed have been determined. Based on these values the bin purities and the contamination rates due to bin migration have been calculated. The result is shown in Fig. 3.115 for the one dimensional binning in case of positive kaons.

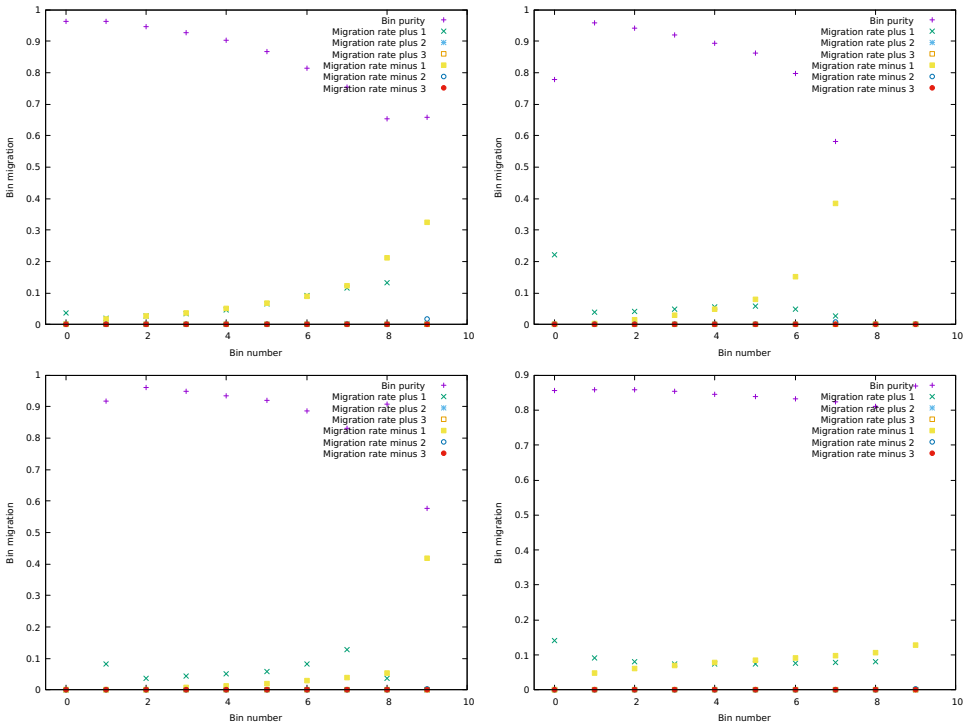


Figure 3.115: Bin purity and bin migration rates for K^+ in case of the 1 dimensional binning in bins of Q^2 (top left), x_B (top right), z (bottom left) and P_T (bottom right)

To estimate the effect of the bin migration on the result, the contamination rate has been separated based on the generated bin where it originates from. It has been found that the major effect originates from the 3 neighboring bins in each direction. For larger distances the rate can be neglected. The absolute systematic uncertainty due to the bin migration in bin i can be calculated with the following formula:

$$\Delta A_i = \sum_{n=1}^{3} f_{i-n} A_{i-n} \sum_{n=1}^{3} f_{i+n} A_{i+n} - \left(\sum_{n=1}^{3} f_{i+n} + \sum_{n=1}^{3} f_{i-n} \right) A_i \tag{3.48}$$

,where A_i is the asymmetry measured in bin i and f_{i+n} is the contamination from bin i+n to bin i. The systematic uncertainties estimated this way in the 1 dimensional binning case are shown in Figures 3.116 and 3.117. It can be observed that the effect stays below a few percents, since the asymmetry shows a very smooth behavior and the kinematic bins are significantly larger than the resolution.

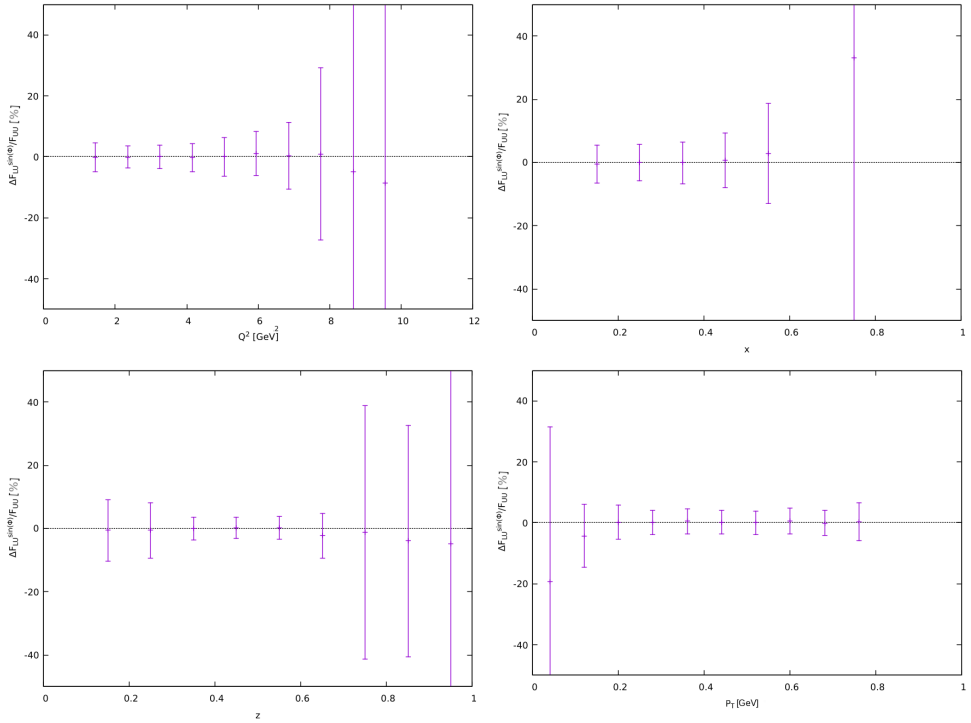


Figure 3.116: Relative uncertainty of the asymmetry due to bin migration for K^+ in case of the 1 dimensional binning

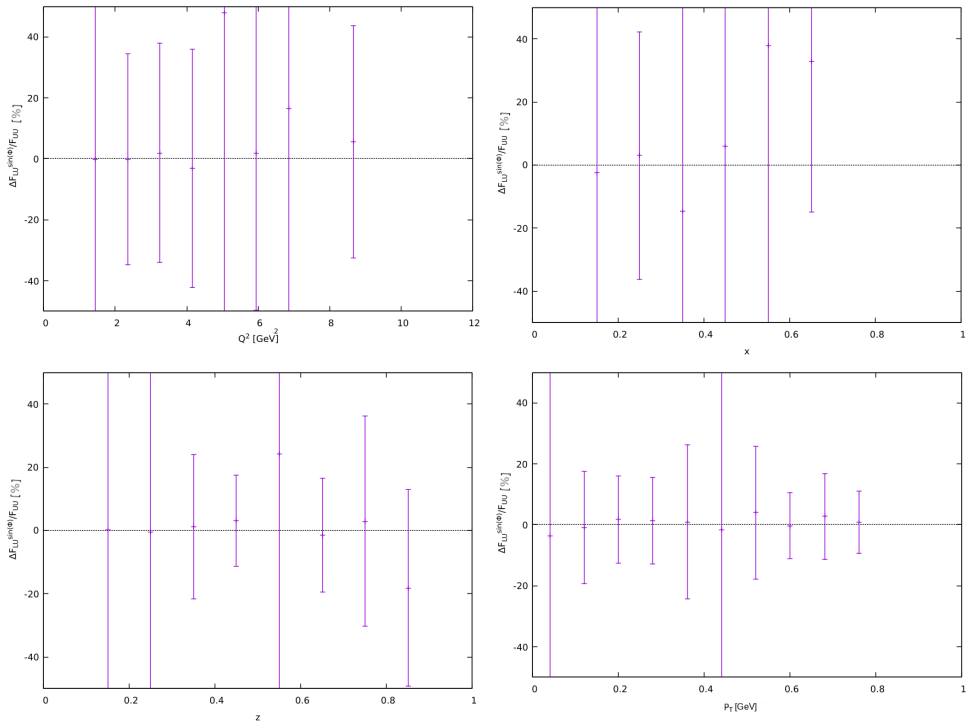


Figure 3.117: Relative uncertainty of the asymmetry due to bin migration for K^- in case of the 1 dimensional binning

In the multidimensional binning cases it is expected that most of the bin migration happens in the last binning dimension, where the bins are much smaller than in the other dimensions, so for the uncertainty estimation the bin migration in the first few dimensions were neglected. To check this assumption the bin purities and migration rates were determined in the first few dimensions. The lowest bin purity values were measured in case of the 5 x-Q^2 bins, but they were also above 93%. Figure 3.118 shows the bin purities and migration rates in that case.

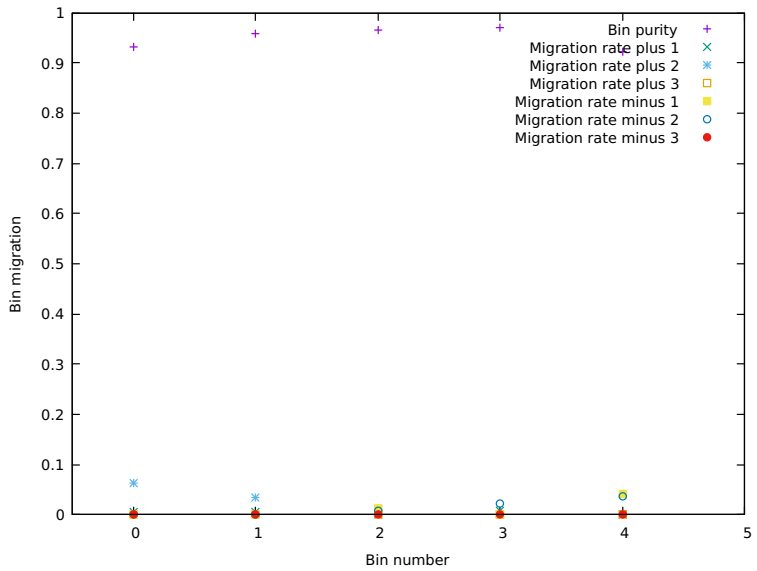


Figure 3.118: Bin purity and bin migration rates for K^+ in case of the multidimensional binning in the x-Q^2 binning plane

ϕ can also show bin migration, influencing the extracted asymmetry. To investigate this effect, the asymmetry in ϕ was calculated based on an ideal sin curve with a given asymmetry. Then the asymmetry in each bin was calculated with the effect of the bin migration taken into consideration. As a final step these points were fitted with a sin function. The results are shown in Fig 3.119. The effect is found to be in the order of one percent. This was added as an additional systematic uncertainty for each bin.

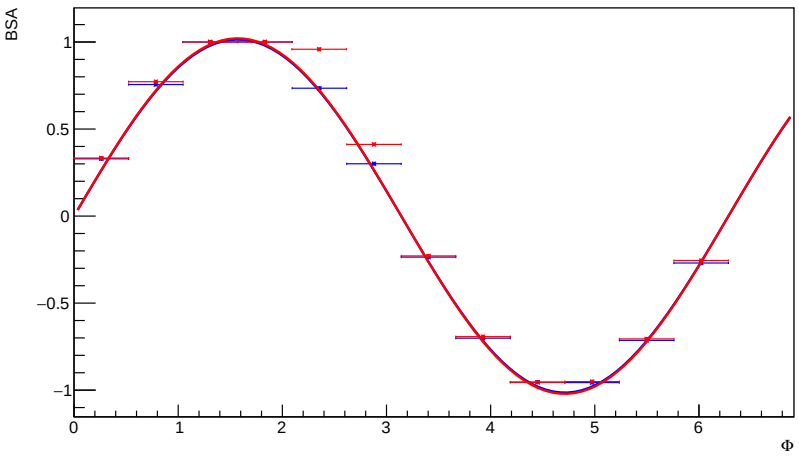


Figure 3.119: Injected (blue) and obtained (red) asymmetries

3.7.8 Radiative effects

Radiative photons emitted in the scattering process modify the reconstructed virtual photon's 4-momentum. This introduces a bias in the SIDIS event kinematics that needs to be corrected. These radiative corrections on the measured asymmetries are expected to be small for SIDIS beam spin asymmetries, because of the requirement of at least one hadron in the final state. This limits the radiative corrections to those for the inelastic part of the cross-section, unlike for inclusive deep inelastic scattering. In addition to this argument, we can also have a direct look on the BSA, if we would consider acceptance ($A^{\pm}$) and radiative ($R^{\pm}$) corrections for both helicity states ($\pm$). In this case, the BSA can be expressed as:

$$BSA' = \frac{N'^{+} - N'^{-}}{N'^{+} + N'^{-}} = \frac{\frac{N^+}{A^+R^+} - \frac{N^-}{A^-R^-}}{\frac{N^+}{A^+R^+} + \frac{N^-}{A^-R^-}} \tag{3.49}$$

In the performed multidimensional analysis, the bin sizes are small. Therefore it can be assumed that the acceptance as well as the radiative corrections are approximately the same for both helicity states ($A^+ \approx A^- \approx A$ and $R^+ \approx R^- \approx R$). This leads to

$$BSA' \approx \frac{\frac{N^+}{AR} - \frac{N^-}{AR}}{\frac{N^+}{AR} + \frac{N^-}{AR}} = \frac{N^+ - N^-}{N^+ + N^-} \tag{3.50}$$

, which shows that the effect of radiative corrections on the BSA is expected to be small. Nevertheless, the radiative effects have been evaluated by means of Monte Carlo simulations using the dedicated software (RADGEN) in combination with LEPTO. The estimated radiative corrections do not exceed a few percent of the value of the beam spin asymmetries. The radiative corrections have not been applied to the measured asymmetries, but they are included in the systematic uncertainties. [35]
To study the radiative effects and to estimate their contribution to the systematic uncertainties, RADGEN and LEPTO was used to simulate the distribution shown on Figure 3.120. The full analysis was then repeated, but instead of a fixed beam energy, the energy was chosen according to this distribution.

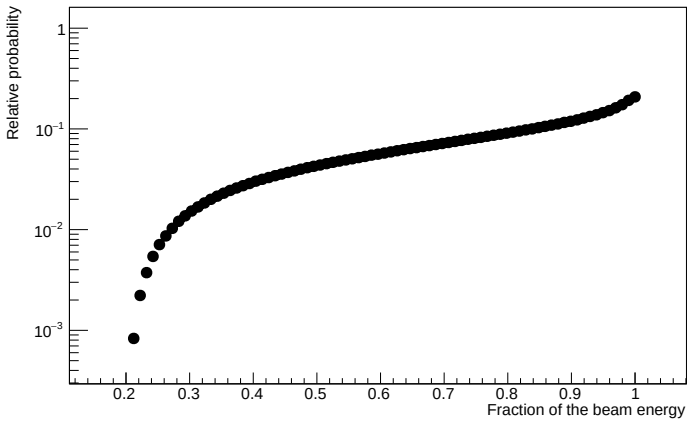


Figure 3.120: Distribution of the relative beam energy if radiative effects are considered, based on a "radgen" simulation

The resulting missing mass distributions are shown in Figure 3.121. The distributions calculated with a fixed beam energy are shown in blue and the ones with the simulated beam energy are shown in red. A negative shift of the missing mass spectra due to the lower overall beam momentum can be observed. In addition, the statistics is reduced, since the applied cuts are rejecting radiated events. These radiative effects can limit the effect of the missing mass cut placed at 1.6 GeV, since radiated events can be reconstructed at a higher missing mass than they should be. This way exclusive channels can contaminate the SIDIS sample. As Figure 3.42, the effect on the BSA of these channels are expected to be small.
By using the samples obtained with a fixed beam energy and with the beam energy chosen according to the distribution, the contribution of events below the missing mass cut was simulated on an event-by-event basis.

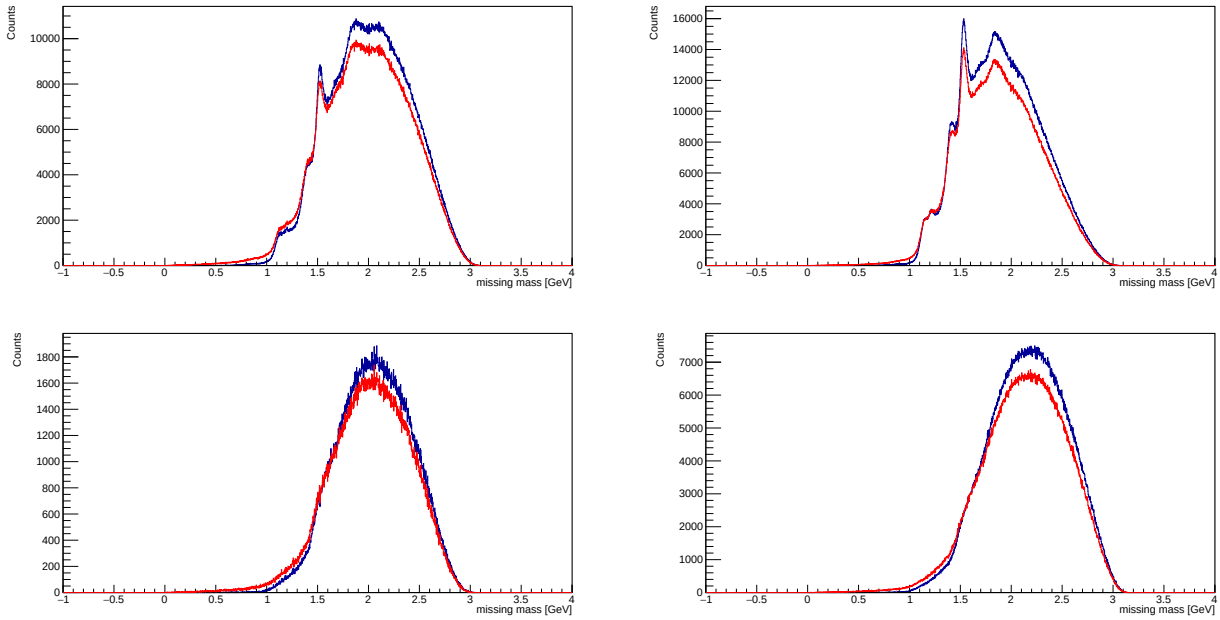


Figure 3.121: Missing mass of e^-K^+X (top) and e^-K^-X (bottom) and without (blue) and with (red) simulated radiative effect with inbending (left) and with outbending (right) setting.

For each kinematic bin the proportion of the events that should not be part of the SIDIS sample because they actually came from below the missing mass cut and are only accepted in the SIDIS sample due to radiative effects was determined:
f=Number of events in bin i (determined with recalculated variables) with standard cuts on both normal and recalculated variables, except the cut on recalculated M_X is reversed/Number of events in bin i (determined with normal variables) with standard cuts on normal variables.
The measured asymmetry can then be expressed as:

$$F_M = (1 - f)F_T + fF_\gamma \tag{3.51}$$

, where F_T is the true asymmetry and F_γ is the asymmetry with the $M_X < 1.6$ GeV cut. The F_γ values were extracted for each kinematic bin. The values are slightly higher than the F_M values and are shown in Figures 3.122 and 3.123. The difference between the true and the measured value is assigned as the systematic uncertainty.
The resulting relative systematic uncertainties are shown in Figures 3.124 and 3.125. It can be observed that the relative uncertainties stay below 5% within the whole kinematic range. Since also within the higher missing mass region and within other kinematic variables, bin migrations can be introduced by radiative effects, which are not considered by this study, the final systematic uncertainty due to radiative effects has been set to 5% for all bins in all binning schemes.

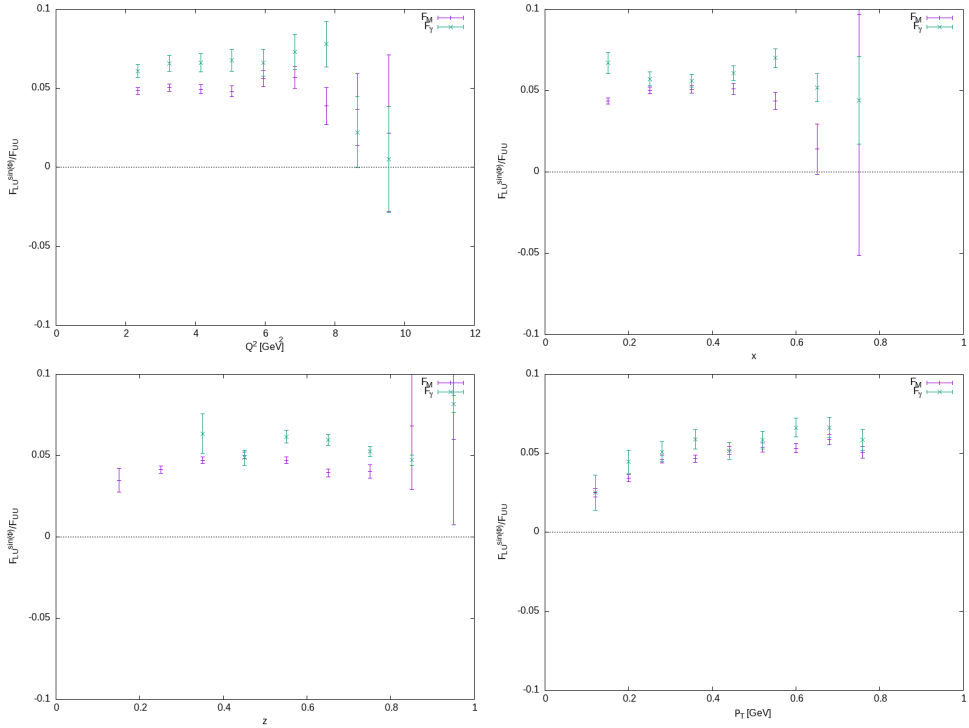


Figure 3.122: Comparison of the extracted $F_{LUM}^{\sin\phi}/F_{UU}$ and $F_{LU\gamma}^{\sin\phi}/F_{UU}$ as a function of Q^2 (top left), x_B (top right), z (bottom left) and P_T (bottom right) for K^+

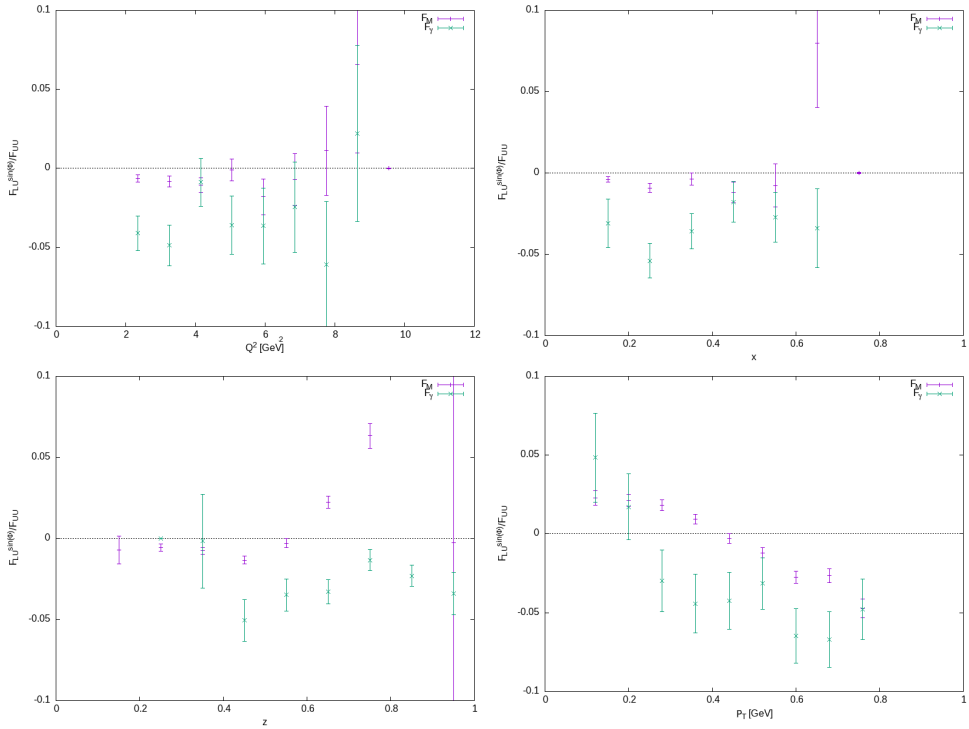


Figure 3.123: Comparison of the extracted $F_{LUM}^{\sin\phi}/F_{UU}$ and $F_{LU\gamma}^{\sin\phi}/F_{UU}$ as a function of Q^2 (top left), x_B (top right), z (bottom left) and P_T (bottom right) for K^-

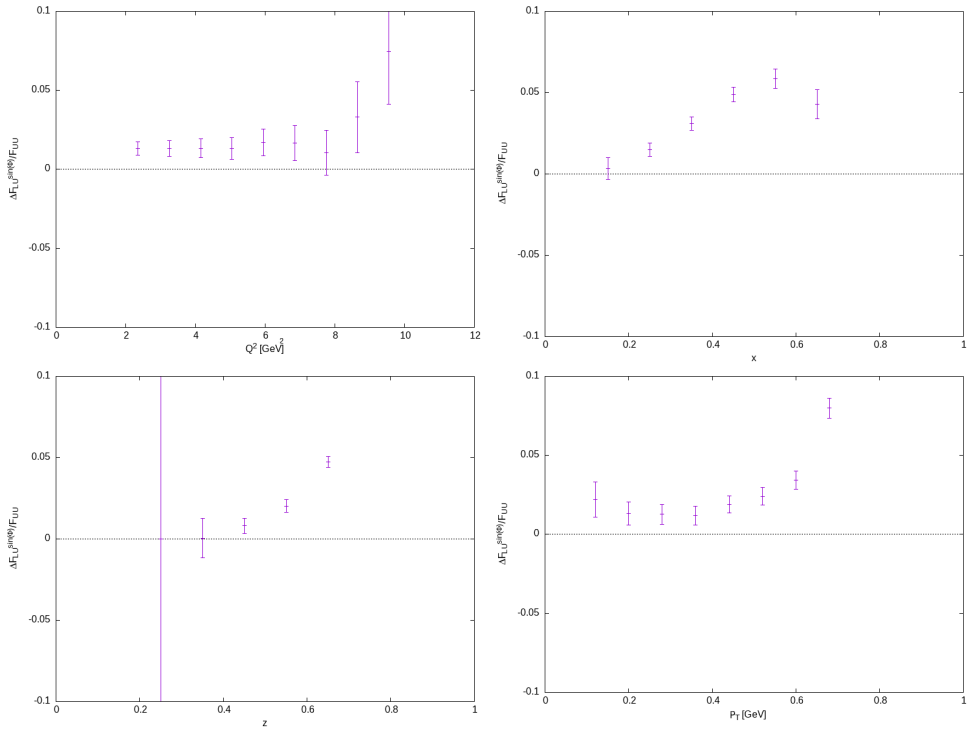


Figure 3.124: Relative uncertainty of the extracted $F_{LU}^{\sin\phi}/F_{UU}$ as a function of Q^2 (top left), x_B (top right), z (bottom left) and P_T (bottom right) for K^+ due to radiative uncertainties

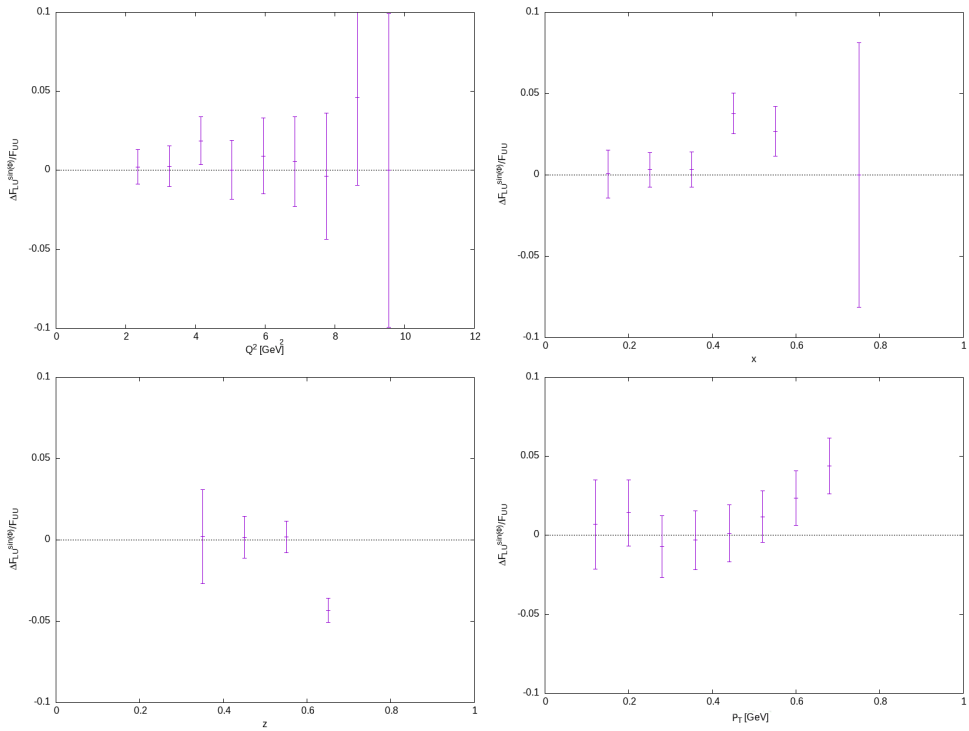


Figure 3.125: Relative uncertainty of the extracted $F_{LU}^{\sin\phi}/F_{UU}$ as a function of Q^2 (top left), x_B (top right), z (bottom left) and P_T (bottom right) for K^- due to radiative uncertainties

3.7.9 Effect of the pion subtraction and uncertainty of the contamination

Using different PID methods together with the pion subtraction give very similar final results. The small differences between the results can be used to estimate the systematic uncertainties introduced by these methods. Since a certain amount of the difference of the extracted value with different methods in a given bin can also be caused by the slightly different distributions of the other kinematic variables, the standard deviation of all of the differences was assigned as a systematic uncertainty. Figures 3.126 and 3.127 show the relative difference of the extracted asymmetries between p <5 GeV with ML PID and other PIDs after the pion subtraction.

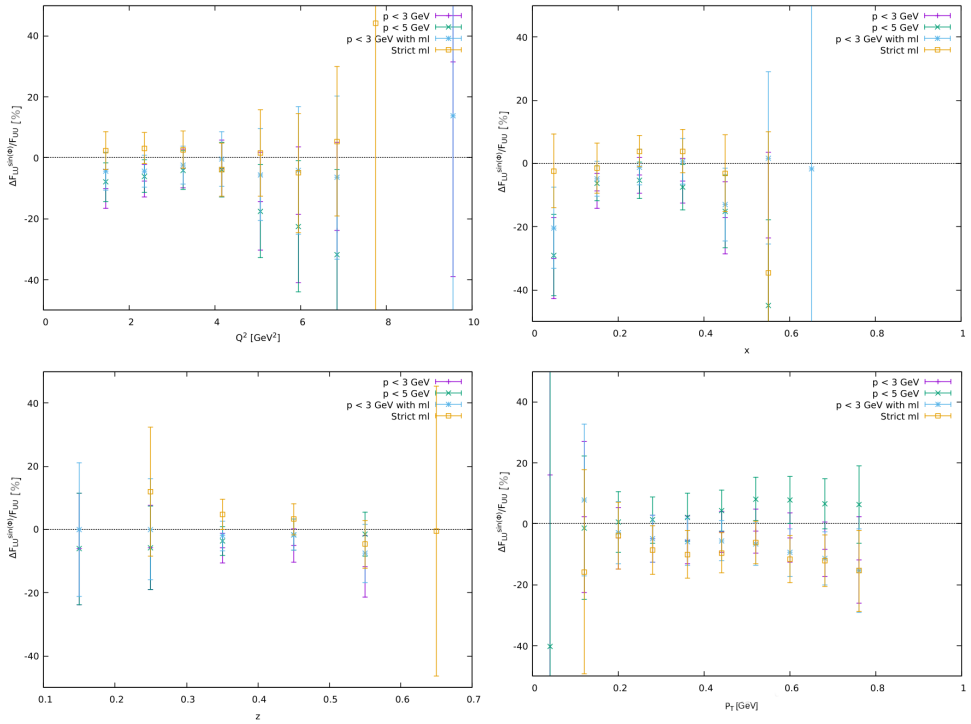


Figure 3.126: Relative difference of the extracted $F_{LU}^{\sin\phi}/F_{UU}$ values between p <5 GeV with ML PID and other PIDs as a function of Q^2 (top left), x_B (top right), z (bottom left) and P_T (bottom right) for K^+

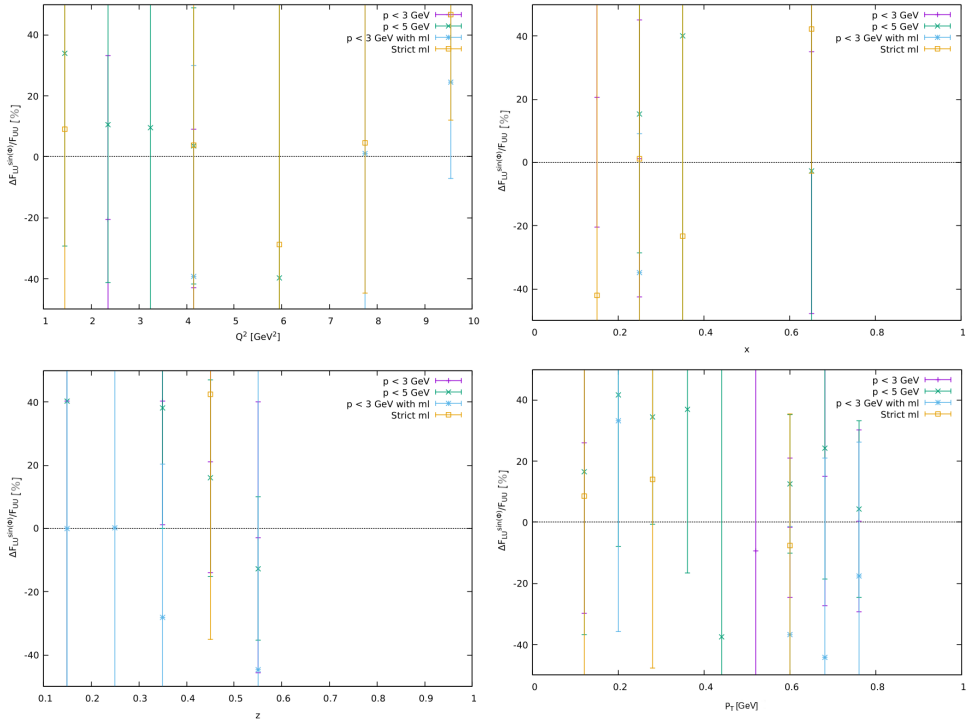


Figure 3.127: Relative difference of the extracted $F_{LU}^{\sin\phi}/F_{UU}$ values between p <5 GeV with ML PID and other PIDs as a function of Q^2 (top left), x_B (top right), z (bottom left) and P_T (bottom right) for K^-

To check that this uncertainty is mostly coming from the uncertainty of the contamination, further studies were performed. First the uncertainty of the contamination was estimated by calculating the average contamination between 1.2 and 2.9 GeV momentum obtained by different methods. The data-points used are shown on Figure 3.91b. The determined average contaminations are the following:

- β-p fits from data - 35.07%
- β-p fits from MC - 36.99%
- MC-matching from MC - 35.1%

The contamination value around 4 GeV is 53% by MC-matching and 56% by the RICH. Based on these values, the uncertainty of the contamination from MC can be estimated to be 10% across the whole momentum range. The uncertainty of the corrected asymmetries coming from the uncertainty of the contamination was calculated in every bin with the following formula:

$$\Delta A_{corr}^{K} = \frac{A_{meas}^{K} + A_{meas}^{\pi}}{(1-C)^2}\Delta C \tag{3.52}$$

These estimated uncertainties are plotted together with the difference of the corrected values extracted with different PID for K^+ in the one dimensional kinematic bins in Figure 3.128. The plots show that the uncertainty coming from the contamination is slightly smaller than the one determined from the differences of the different PID methods.

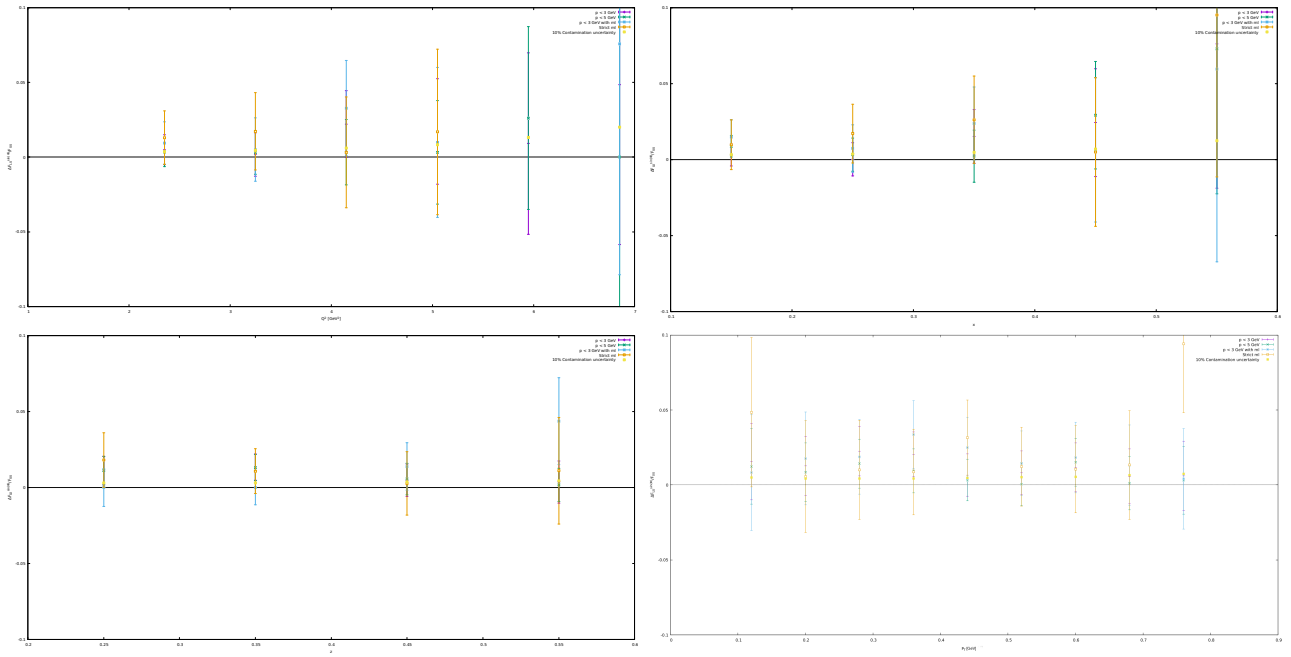


Figure 3.128: Uncertainty of the extracted $F_{LU}^{\sin\phi}/F_{UU}$ values using p <5 GeV with ML PID coming from the uncertainty of the contamination and relative difference of the extracted $F_{LU}^{\sin\phi}/F_{UU}$ values between p <5 GeV with ML PID and other PIDs in bins of Q^2 (top left), x_B (top right), z (bottom left) and P_T (bottom right) for K^+

It was also checked that the results with different PID methods are only different because the distribution of the other kinematic variables are different. The raw asymmetries were extracted using different PID and a $z > 0.55$ cut, to achieve a similar underlying kinematic distribution for K^+ in 3 one dimensional bins. The values are shown in Figure 3.129. As a cross check the pion asymmetries were also extracted using only the EB PID and also requiring both EB and RICH pion PID to show that they agree. The EB values are approximately 10% closer to the kaon values, showing that there is a non-negligible kaon contamination in the pion sample in the high z region, which corresponds to high momenta. The next group of PIDs is the kaon EB PID and the kaon EB PID combined with a RICH PID which is either pion or kaon. The fact that these values coincide shows that that the kinematic region of RICH availability doesn't add any other bias to the asymmetries. Since these points have a high contamination, we expect these to be between the pion values and the "zero" contamination kaon values, which is the case in all bins. The "zero" contamination PIDs are the EB and RICH kaon PID and the EB and strict ML PID, which means that the cut on the DNN response is increased to 0.99, as before. In the last case we expect that the contamination is just a few percent based on the MC. Since the strict ML PID is in good agreement with the RICH and it is also in good agreement with the corrected asymmetries using the other 4 PIDs, we can conclude that the contamination estimates are accurate within 10% uncertainty and that the different PID methods do not have any effect on the extraction of the asymmetries.
Since z is highly correlated with momentum, the contamination can be estimated around 1.8 GeV ($z = 0.17$) and 4.2´ GeV ($z = 0.83$) momenta using the following formula:

$$C = \frac{F^K_{meas} - F^K_{\mathrm{MC/RICH}}}{F^\pi_{meas} - F^K_{\mathrm{MC/RICH}}} \tag{3.53}$$

Here we assumed that the contamination is zero using the strict ML at low momenta and using the RICH PID at high momenta, so $F^K_{\mathrm{MC/RICH}} = F^K_{real}$. The contamination determined this way around 4 GeV is 0.58, which is very close to the values from MC ($C = 0.53$) and RICH ($C = 0.56$). The low z contamination is 0.16, which is also within 10% difference from the MC contamination using the $|\chi^2_{PID}| < 3$ cut ($C = 0.18$). The agreement is illustrated by Figure 3.130.

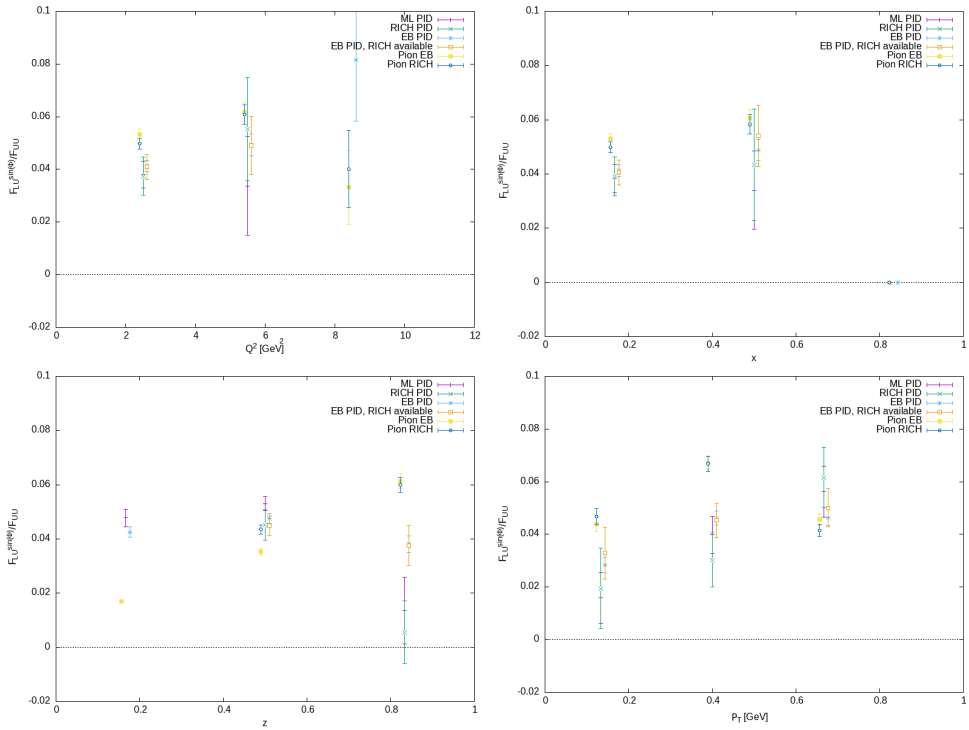


Figure 3.129: Comparison of the "raw" extracted $F_{LU}^{\sin\phi}/F_{UU}$ values using different PIDs in bins of Q^2 (top left), x_B (top right), z (bottom left) and P_T (bottom right) for K^+ and π^+

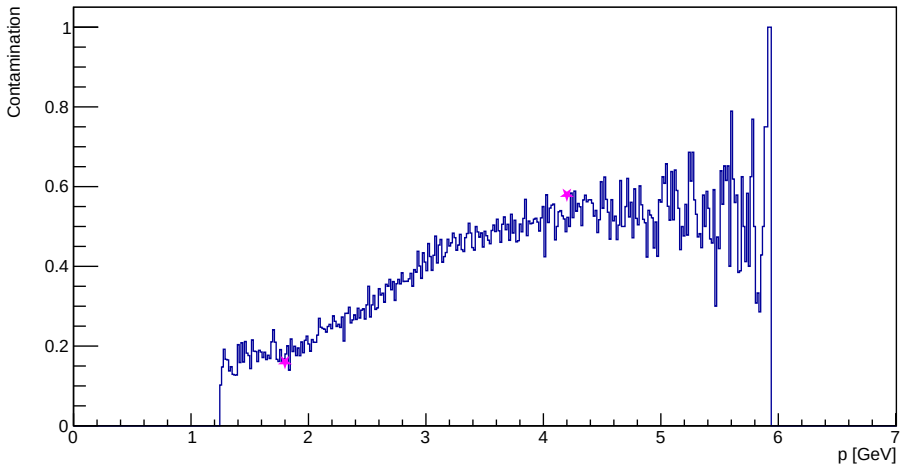


Figure 3.130: Contamination as a function of momentum from the MC in blue and contamination calculated from the difference of the extracted asymmetries using different PID (purple stars).

3.7.10 Total systematic uncertainty

To estimate the total systematic uncertainty the different contributions are calculated for each bin and they are added quadratically. Figures 3.131 and 3.132 show the contributions from different sources to the total uncertainty in each bins in the one dimensional binning case for positive and negative kaons respectively.

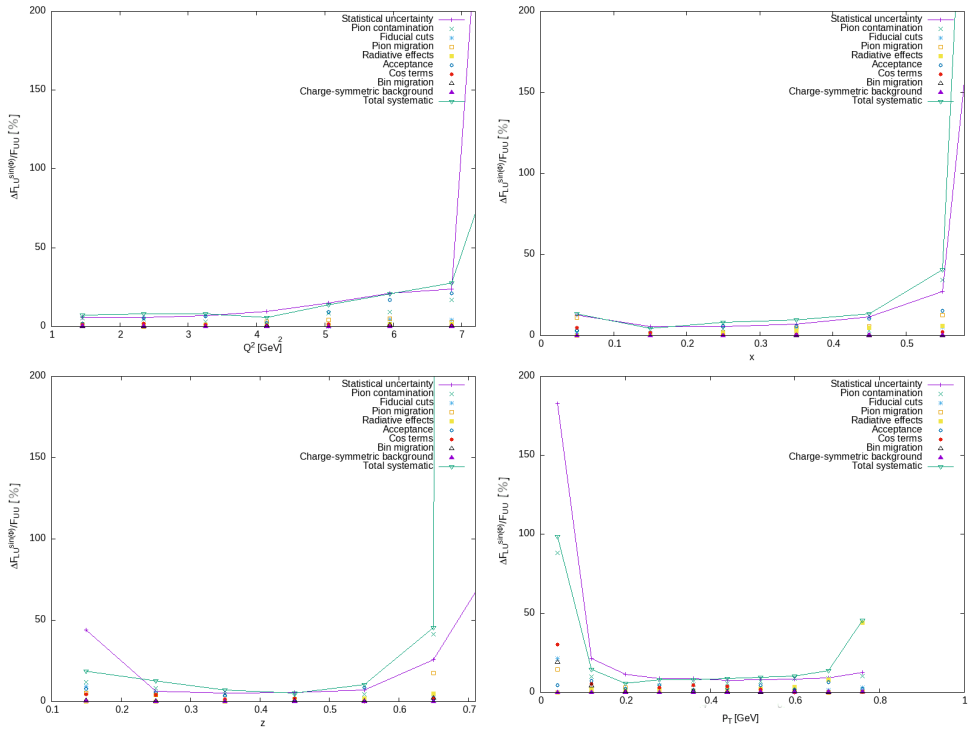


Figure 3.131: Contributions from different sources to the total relative uncertainty in bins of Q^2 (top left), x_B (top right), z (bottom left) and P_T (bottom right) for K^+

We can observe that the total systematic uncertainty is at the same level as the statistical uncertainty. The acceptance effects and the uncertainty of the pion value subtraction have the most dominant contribution in most bins. The uncertainty from radiative effects also have a significant contribution in most kinematic bins, especially at high P_T.
Table 3.16 shows the relative uncertainty from the different sources averaged over all bins of the fully multidimensional binning. Here we can observe that the pion contamination causes the most uncertainty in the multidimensional binning as well. The average systematic uncertainty is below the average statistical uncertainty. The absolute uncertainties are slightly bigger for K^- than for K^+, while the relative uncertainties are much higher. This is due to the fact that the asymmetries for K^- are very close to 0. The uncertainties are shown in every bin in the fully multidimensional case on Figure 3.133.

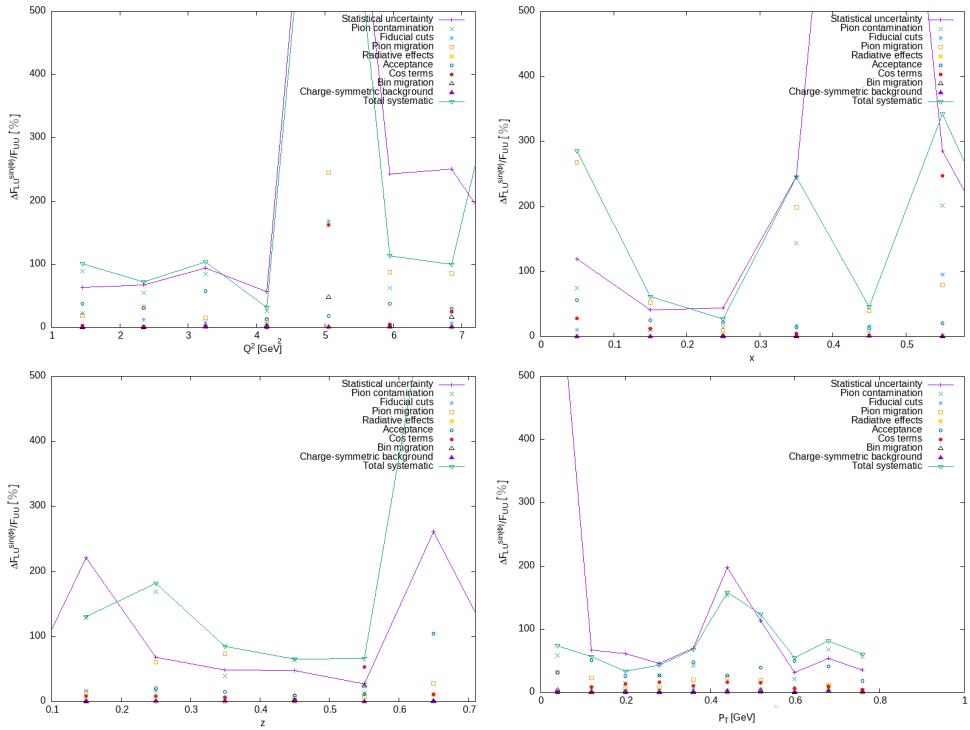


Figure 3.132: Contributions from different sources to the total relative uncertainty in bins of Q^2 (top left), x_B (top right), z (bottom left) and P_T (bottom right) for K^-

	Relative uncertainty		Absolute uncertainty	
Source	K^+	K^-	K^+	K^-
Statistical uncertainty	28.8%	123.7%	0.012	0.023
Pion contamination	11.5%	79.9%	0.005	0.015
Beam polarization	2.8%	3.2%	0.001	0.001
Fiducial cuts	8.5%	29.3%	0.003	0.006
Pion bin migration	5.3%	50.3%	0.002	0.009
Radiative effects	5%	5%	0.002	0.0005
Acceptance	3.5%	31.4%	0.001	0.006
Cos terms	5.4%	32%	0.002	0.006
Bin migration	1.8%	3.2%	0.001	0.001
Phi bin migration	1%	1%	0.0004	0.0001
Charge-symmetric background	0.2%	0.5%	0.0001	0.0002
Accidentals	1%	1%	0.0004	0.0001
Total systematic uncertainty	21.6%	89.9%	0.009	0.017

Table 3.16: Relative and absolute uncertainty from the different sources averaged over all bins of the fully multidimensional binning

Figure 3.133: Contributions from different sources to the total relative uncertainty in the fully multidimensional binning for K^+ (top) and for K^- (bottom)

3.8 Final results and discussion

The $F_{LU}^{\sin\phi}/F_{UU}$ values extracted using the $p < 5$ GeV cut and the machine learning PID, together with the mean values of the most important kinematic variables can be found in appendices I to III for all binning schemes described in section 3.5.6. Figures 3.134 and 3.135 show the final values together with the systematic and statistical uncertainties in every kinematic bin and compare it with previously measured values, when available. The comparison is made using results obtained by CLAS6 [63] and HERMES [64]. The new results are in good agreement with the previous measurements. They have smaller uncertainties and much narrower kinematic bins due to the much higher statistics. The average Q^2 for CLAS6 is much lower than in case of this study, where the $F_{LU}^{\sin\phi}/F_{UU}$ values are lower. This results in a lower asymmetry in every other kinematic bin. The HERMES dataset is mostly concentrated at low x_B values, but since the $F_{LU}^{\sin\phi}/F_{UU}$ is flat as a function of x_B in comparison with the uncertainty of the HERMES asymmetries, so this does not affect the bins in the other kinematic variables significantly.

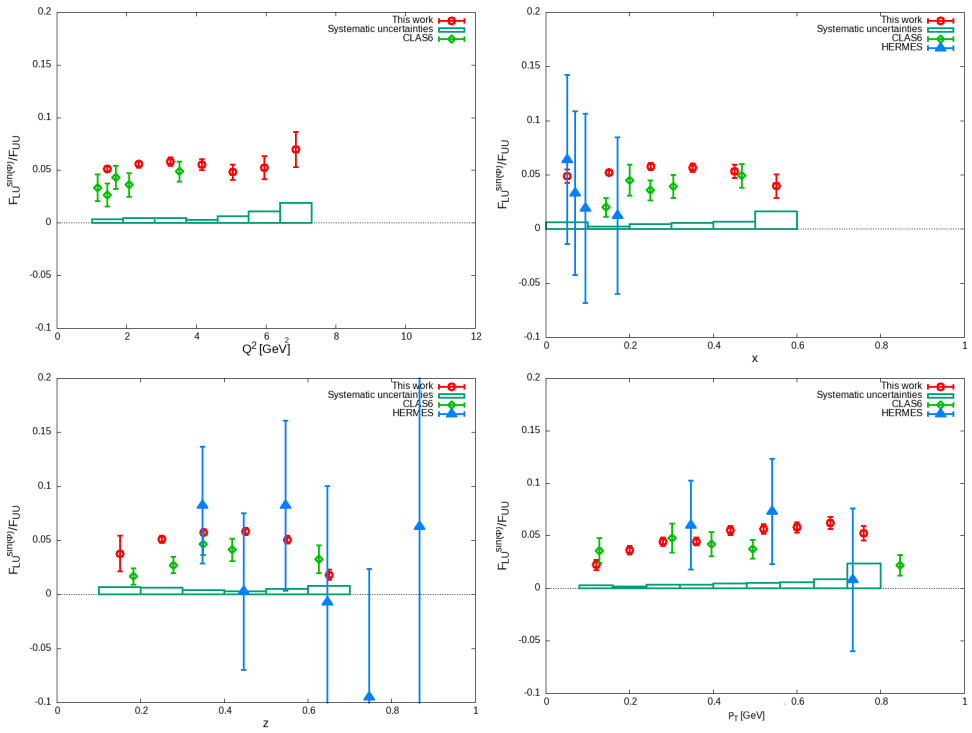


Figure 3.134: Final results after subtracting the pion asymmetries compared with previous results in bins of Q^2 (top left), x_B (top right), z (bottom left) and P_T (bottom right) for K^+. The results are integrated over all the other kinematic variables. ($\langle x_B \rangle = 0.226$, $\langle Q^2 \rangle = 2.477$ GeV^2, $\langle z \rangle = 0.433$, $\langle P_T \rangle = 0.419$ GeV)

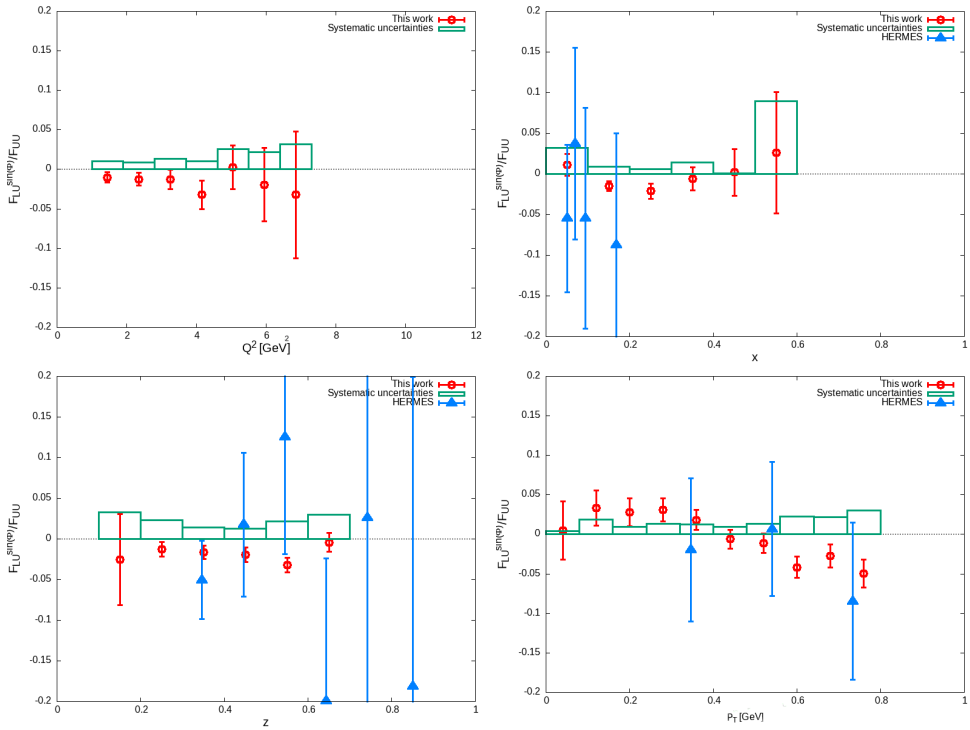


Figure 3.135: Final results after subtracting the pion asymmetries compared with previous results in bins of Q^2 (top left), x_B (top right), z (bottom left) and P_T (bottom right) for K^-. The results are integrated over all the other kinematic variables. ($\langle x_B \rangle = 0.201$, $\langle Q^2 \rangle = 2.349$ GeV^2, $\langle z \rangle = 0.466$, $\langle P_T \rangle = 0.446$ GeV)

Figures 3.136 to 3.149 show the final values together with the systematic and statistical uncertainties in every kinematic bin and compare it with the pion values extracted from the same dataset with the same binning. The binning in the $x_B - Q^2$-plane is shown in Figures 3.137 and 3.146.

Here only the statistical uncertainty is shown for pions, but according to [35] the systematic uncertainties are usually significantly smaller than the statistical ones and thus can be neglected compared to the much higher ones in the kaon case. Especially that the 2 highest sources of the pion systematic uncertainties are the bin migration and the radiative effects. Since much wider kinematic bins are used in our case we expect much lower bin migration and most of the kinematic bins, in which the radiative uncertainties are the highest for pions (high x_B, P_T and z and low Q^2), were removed from the kaon analysis due to insufficient statistics.
It can be observed that the overall behavior of the asymmetries are almost independent of Q^2 and x_B. In case of K^+ the P_T-dependence looks the same for all z bins. For K^- we can observe a rising and then a falling tendency in P_T for both z bins, although this effect is more prominent at high z values.
If we look at K^+, the z-dependence in different P_T bins shows a falling trend at low z values and a rising trend at high z values at low P_T. This trend changes to a rising and then falling one at high P_T values. For K^-, a similar difference can be observed in the z-dependence between the two P_T bins. It shows a rising trend at low P_T and a falling trend followed by a sharp increase at high P_T.
In case of a few kinematic bins the asymmetry could not have been extracted with the stricter PID methods, due to the limited statistics. This results in a systematic uncertainty that is much higher in that bin, than in the surrounding ones, since the uncertainty of the pion asymmetry subtraction is higher.
Comparing the kaon and pion asymmetries could reveal the differences in the contributing TMDs and FFs of the sea s quarks. The valence quark content of the charged pseudoscalar mesons is shown in Table 3.17. This comparison of the quark content shows, that the substitution of d quarks by s quarks can cause a difference in the observed asymmetries for pions and kaons. The details on the theoretical interpretation are still under discussion, but a few ideas will be discussed in the following.

Meson	π^+	π^-	K^+	K^-
Quark content	$u\bar{d}$	$d\bar{u}$	$u\bar{s}$	$s\bar{u}$

Table 3.17: Valence quark content of the charged pseudoscalar mesons

3.8.1 Positive Kaons

The fact that every other data point follow the pion values, but have higher absolute values, show that at least one of the 4 contributions to $\frac{F_{LU}^{\sin\Phi}}{F_{UU}}$ is different for the sea $\bar{s}$-quarks than for the sea $\bar{d}$-quarks in the proton. As these mesons valence quark content only differs in these quarks. The following terms contribute to $\frac{F_{LU}^{\sin\Phi}}{F_{UU}}$:

$$\frac{F_{LU}^{\sin\Phi}}{F_{UU}} \approx \frac{\zeta(eH_1^{\perp} + f_1\tilde{G}^{\perp} + g^{\perp}D_1 + h_1^{\perp}\tilde{E})}{\zeta(f_1 D_1)} \tag{3.54}$$

In [35], the π^+ asymmetries are compared to theoretical models from [92] and [93]. Both of these models describe the proton as an active quark and spectator scalar and axial-vector diquarks. Both models include the $eH_1^{\perp}$ and the $g^{\perp}D_1$ terms as the other terms are assumed to be small. The first model uses a complicated propagator for the axial-vector diquarks and a ratio for axial-vector and scalar strengths fitted to data. In contrast to this, the second model uses a simple propagator for the axial-vector diquarks and the ratio of axial-vector and scalar is fixed by SU(4) spin-flavor symmetry. The FFs used in both models are described in [94]. These models are also compared to a third model [95], which only includes the $eH_1^{\perp}$, with an $e(x)$ based on a chiral quark soliton model. In this case the P_T dependence of the TMD is narrow, thus that the FF's transverse momentum dependence is more important than that of the target TMD when describing BSA. They found that model 2 describes best the π^+ datapoints, which also provides very similar $eH_1^{\perp}$ terms to model 3. This also provides additional support to this model. Based on the comparison with model 2, they found that the $eH_1^{\perp}$ is dominant at small Q^2 and x_B, but the $g^{\perp}D_1$ term, describing quark-gluon correlations, become increasingly important at larger Q^2 and x_B.
As the kaon asymmetries are closer to the pion ones at high Q^2 and x_B, than at lower values, one could assume that this shows that the quark-gluon correlations pay a bigger role for the sea $\bar{s}$ than for $\bar{d}$ quarks at small Q^2 and x_B. But the difference could also come from the other unknown terms $f_1\tilde{G}^{\perp}$ and $h_1^{\perp}\tilde{E}$.

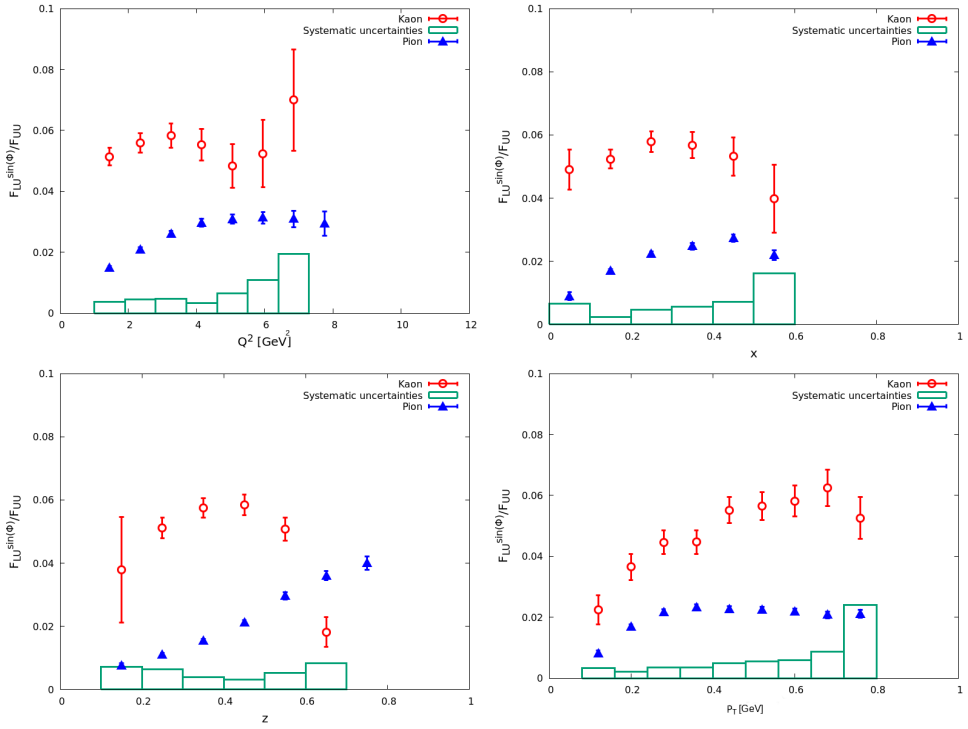


Figure 3.136: Final results after subtracting the pion asymmetries compared with pions in bins of Q^2 (top left), x_B (top right), z (bottom left) and P_T (bottom right) for K^+. The results are integrated over all the other kinematic variables. ($\langle x_B\rangle = 0.226$, $\langle Q^2\rangle = 2.477$ GeV^2, $\langle z\rangle = 0.433$, $\langle P_T\rangle = 0.419$ GeV)

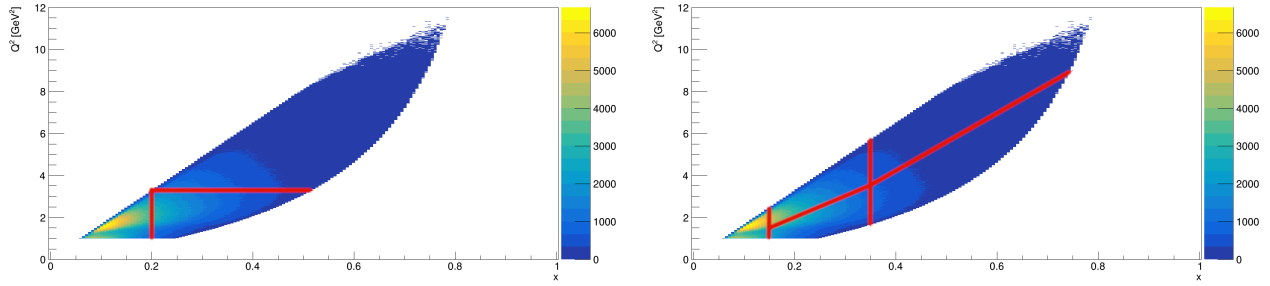


Figure 3.137: Bin borders on the Q^2-x_B-plane for the e^-K^+X dataset with 3 (left) and 5 (right) bins

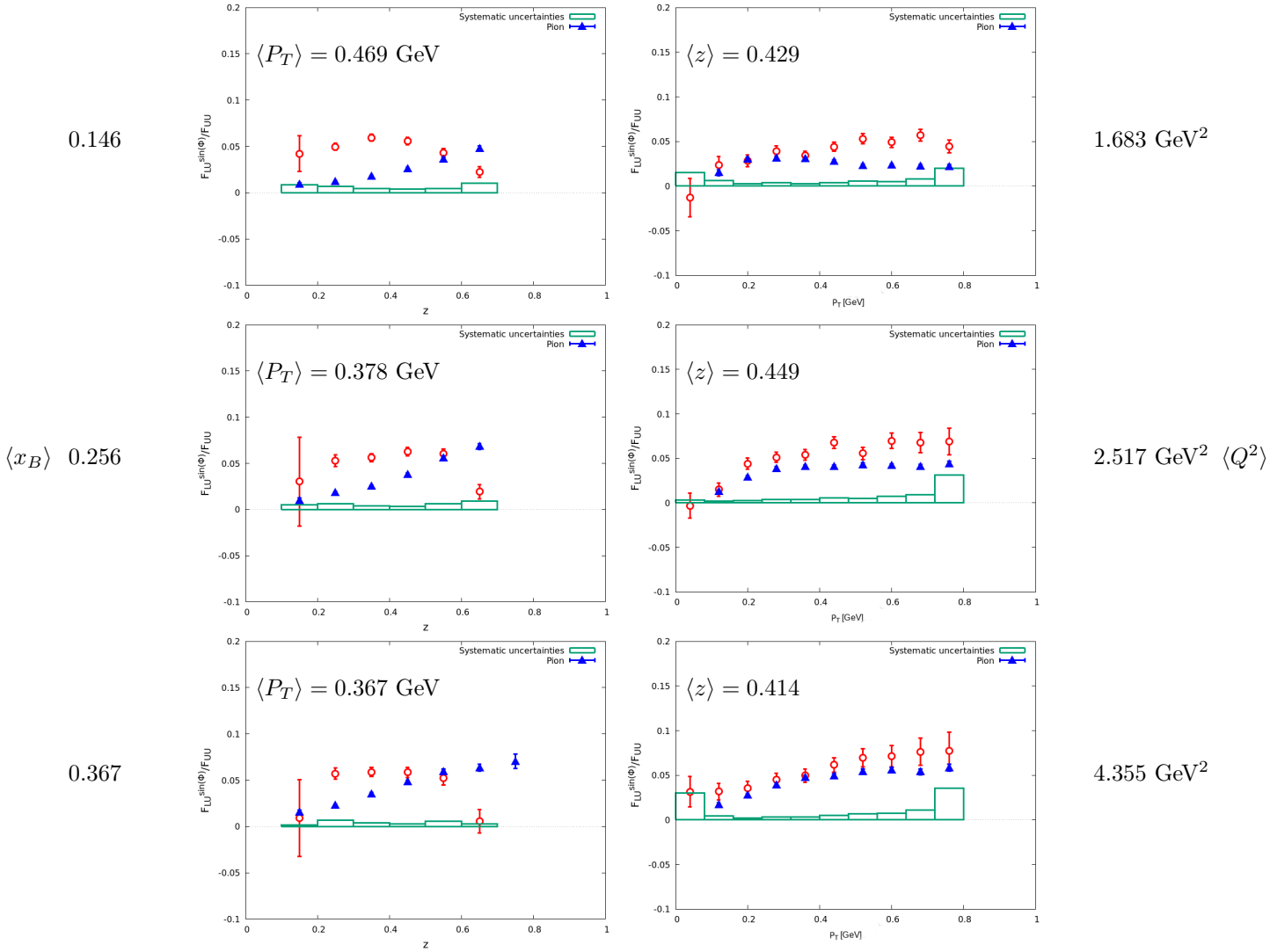


Figure 3.138: Final results after subtracting the pion asymmetries in bins of z (left) and P_T (right) for low x_B and low Q^2 (top), high x_B and low Q^2 (middle) and high x_B and high Q^2 (bottom) for K^+ (x_B-Q^2-binning: Figure 3.137). The results are integrated over all the last kinematic variable (z or P_T).

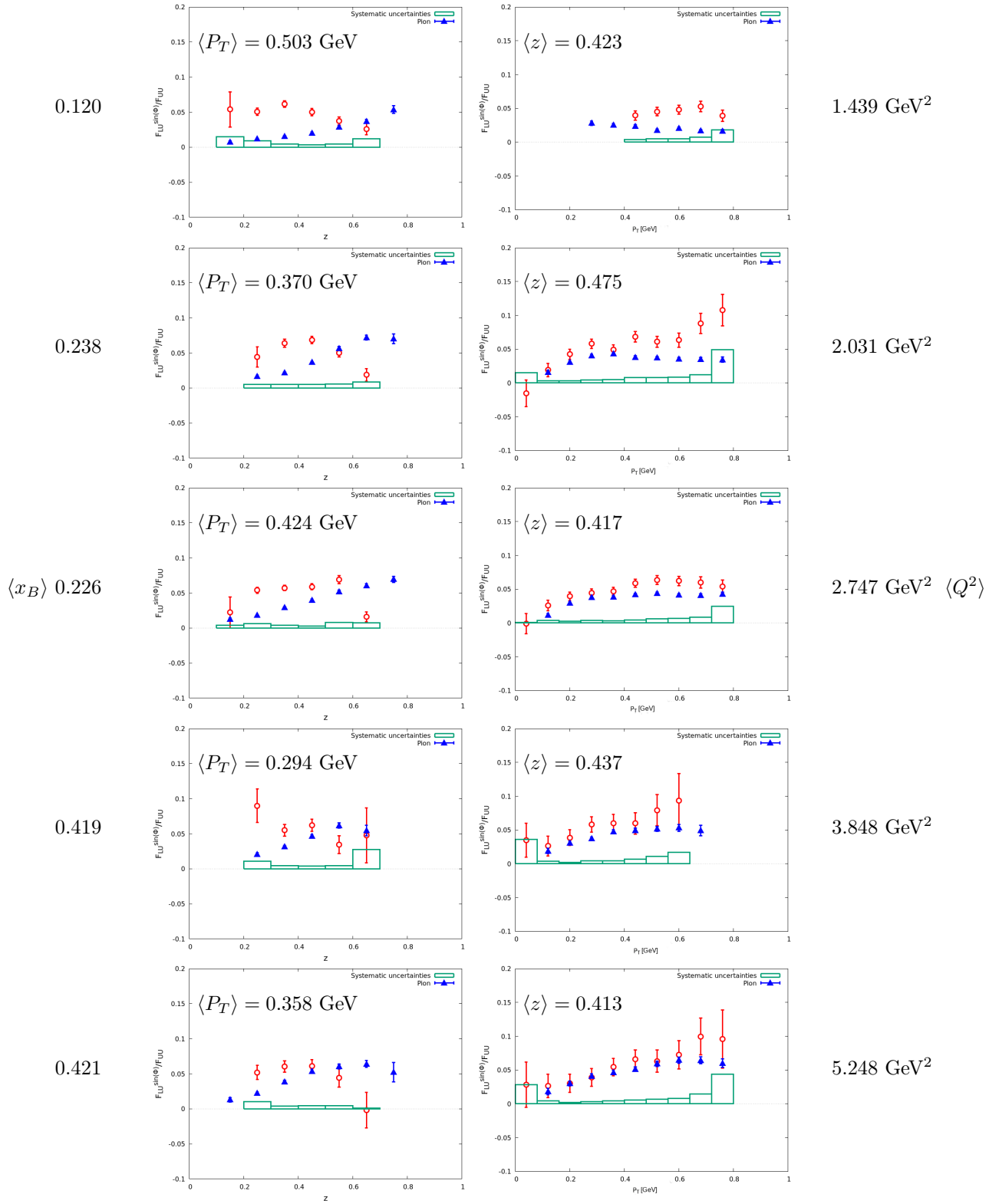


Figure 3.139: Final results after subtracting the pion asymmetries in bins of z (left) and P_T (right) for low x_B and low Q^2 (top), medium x_B and low Q^2 (upper middle), medium x_B and high Q^2 (middle), high x_B and low Q^2 (lower middle) and high x_B and low Q^2 (bottom) for K^+ (x_B-Q^2-binning: Figure 3.137). The results are integrated over all the last kinematic variable (z or P_T).

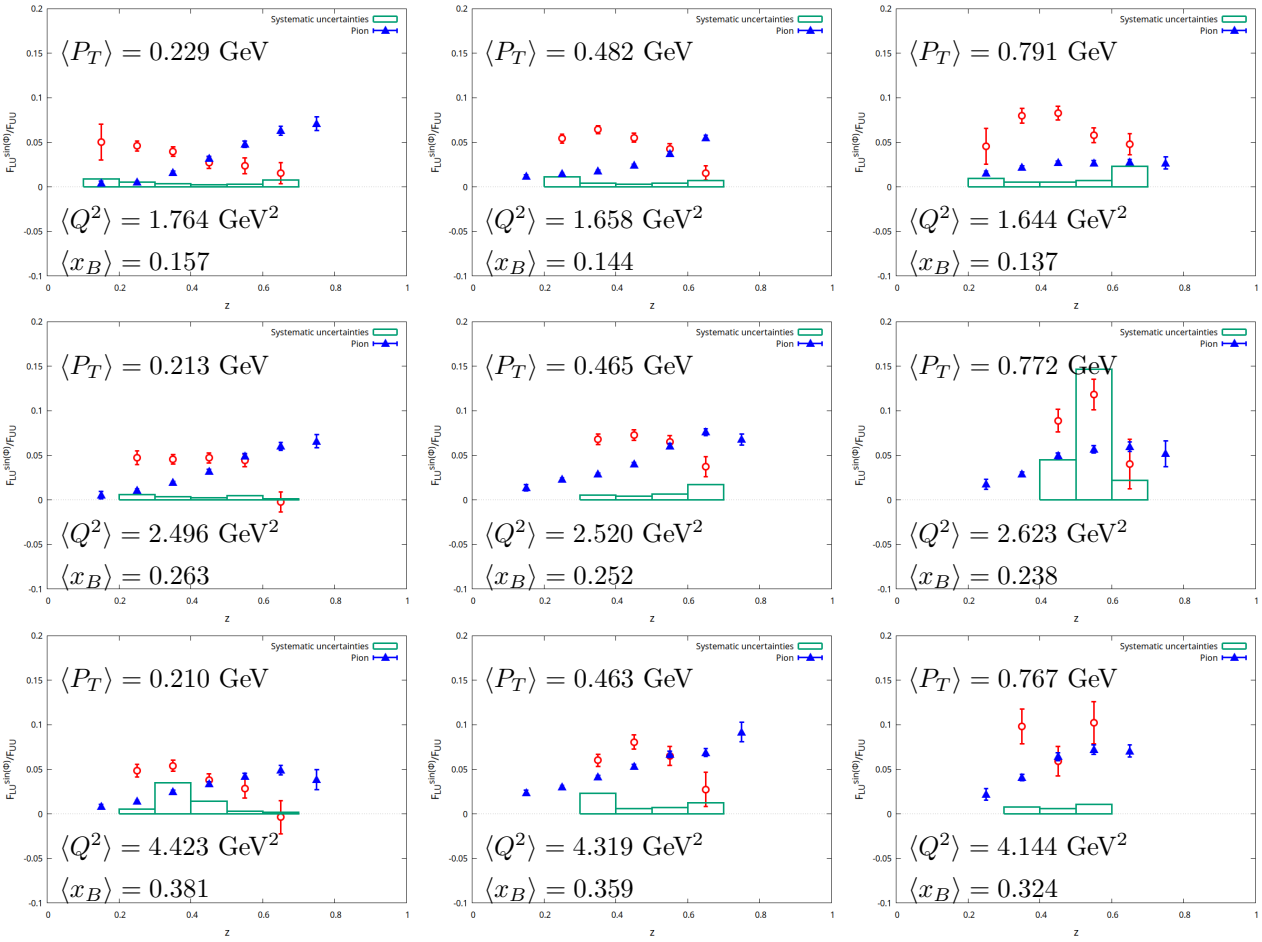


Figure 3.140: Final results after subtracting the pion asymmetries in bins of z for low x_B and low Q^2 (top), high x_B and low Q^2 (middle) and high x_B and high Q^2 (bottom) for low (left), medium (middle) and high (right) P_T for K^+ (x_B-Q^2-binning: Figure 3.137)

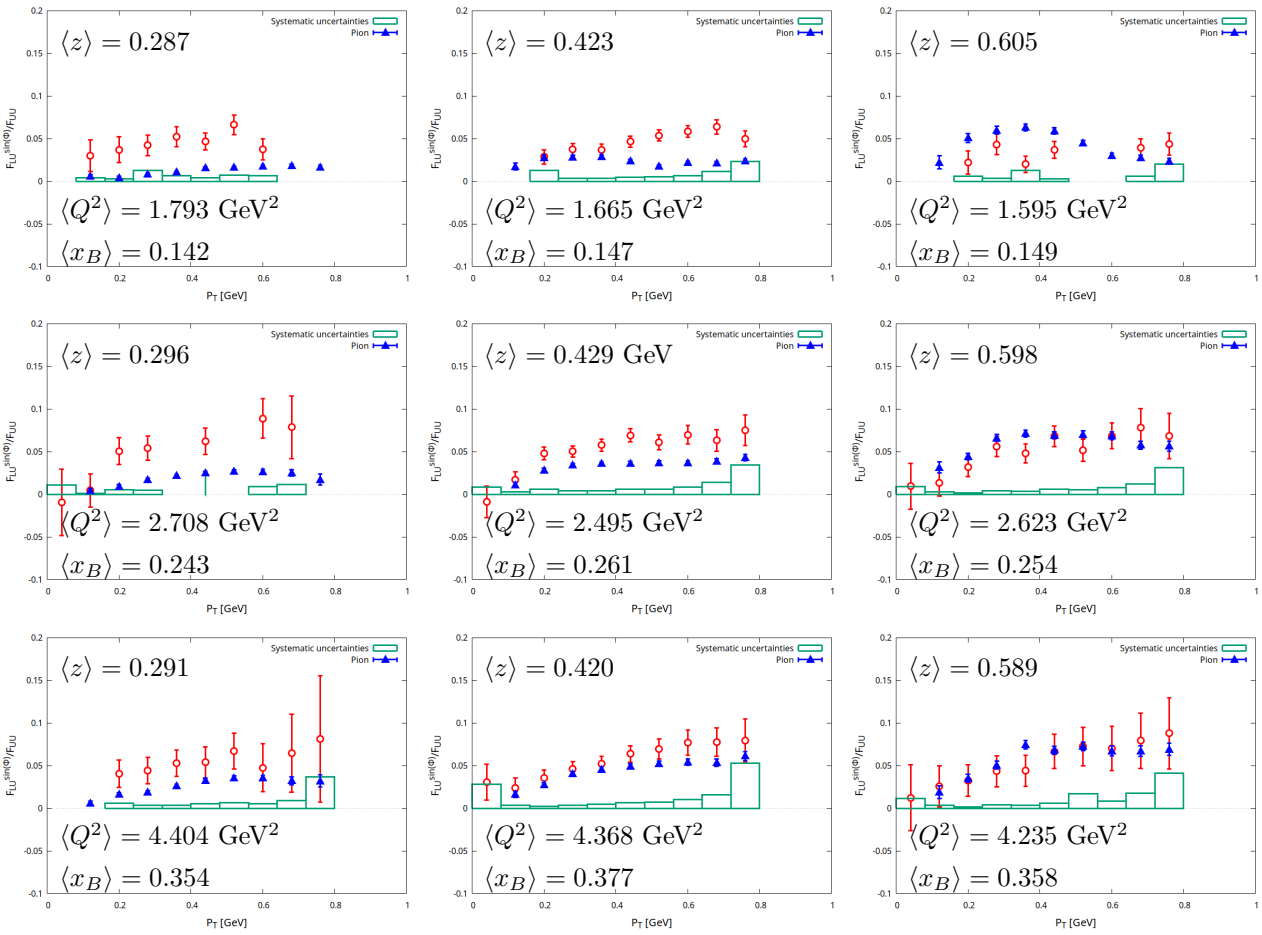


Figure 3.141: Final results after subtracting the pion asymmetries in bins of P_T for low x_B and low Q^2 (top), high x_B and low Q^2 (middle) and high x_B and high Q^2 (bottom) for low (left), medium (middle) and high (right) z for K^+ (x_B-Q^2-binning: Figure 3.137)

3.8.2 Investigation of the falling behavior at high z

The sudden drop at high z, present in all kinematic bins for K^+ could be the result of exclusive reactions in the SIDIS sample, although we expect the exact opposite, a sharp rise, as high z should correspond to low M_X. To demonstrate this, the missing mass is plotted against z, alongside the missing mass spectrum on Figure 3.142 for the K^+ SIDIS sample using the loosest PID ($p < 5$ GeV) without a cut on M_X or z. We can observe 6 peaks, which are marked on the figure. Four of them can be identified as different Λ resonances and with the exception of the one with the highest mass, they are all removed from the final sample. A small Σ peak is also present, besides a neutron peak. The latter one is coming from the pion contamination, which is extremely high at high z. The missing mass spectrum is also shown using the tight ML and the RICH PID. In these cases we can observe that they span the whole missing mass range and only small amount of them contains the $\Lambda(1890)$, thus a valid comparison can be made and the effects of the pion contamination can be estimated.

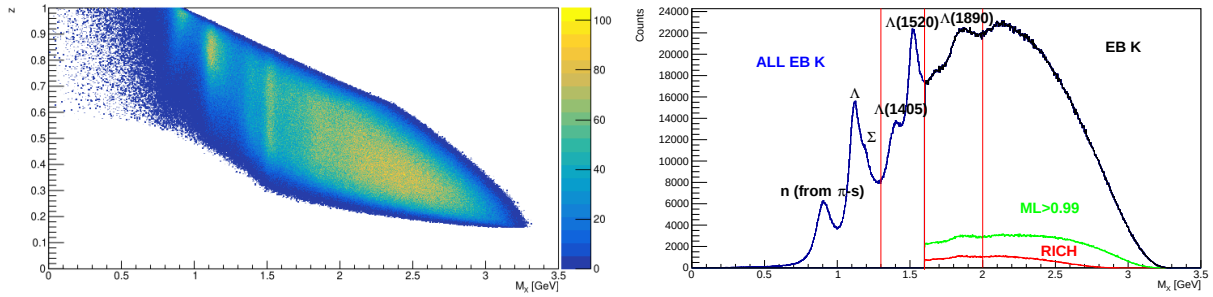
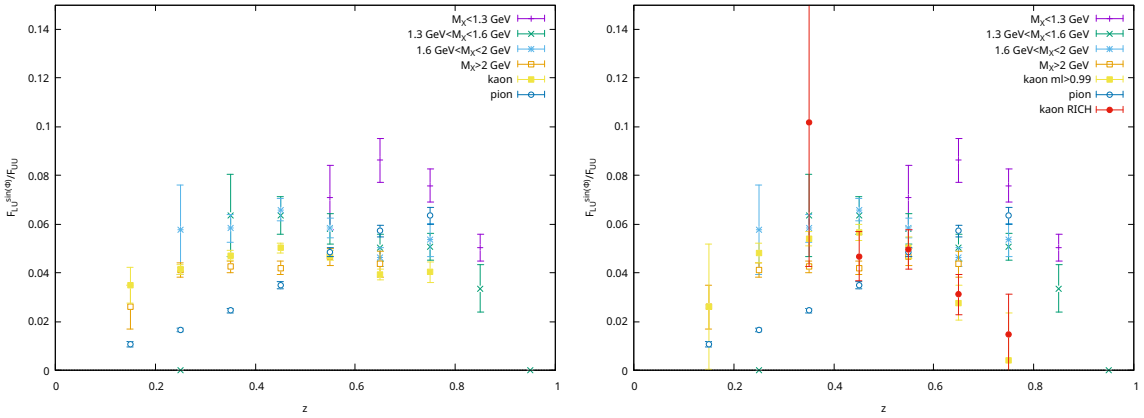


Figure 3.142: z as a function of missing mass for the K^+ SIDIS sample using the loosest PID ($p < 5$ GeV) without a cut on M_X or z on the left. The missing mass spectrum is shown for the K^+ SIDIS sample using the loosest PID ($p < 5$ GeV) without a cut on M_X in blue, with the cut on M_X using the loosest PID in black, using the strict ML PID in green and using the RICH PID in red in the right. The peaks are marked with the possible resonances they are corresponding to, and the bin borders in M_X used for the study described in this section are marked with vertical red lines.

The z dependence of the asymmetries was extracted in bins of missing mass. The M_X bin borders are shown as red vertical lines in Figure 3.142. As the averages of the other kinematic variables change significantly using different M_X cuts in a given z bins, the extracted values are not directly comparable, instead one should concentrate on the tendencies. Figure 3.143a compares the asymmetries extracted in M_X bins with the pion and with the kaon asymmetries. As expected, the asymmetry is higher when exclusive channels are included. But in all bins we can observe a falling trend at high z. By comparing the yellow and the orange points, one can conclude that the effect of the $\Lambda(1890)$ on the asymmetries is negligible. The falling trend becomes much more prominent after subtracting the pion asymmetries or by using a very strict PID, like the strict ML or the RICH, as both the pion asymmetries and the contamination increase with z. This comparison is shown in Figure 3.143b. By using these strict PIDs we can also exclude the possibility, that this behavior is an artifact of the pion subtraction.



(a) EB PID (yellow)

(b) ML>0.99 (yellow) and RICH (red) PID

Figure 3.143: "Raw" $F_{LU}^{\sin\phi}/F_{UU}$ values as a function of z in bins of M_X for K^+ using p <5 GeV compared with pions and kaons. The kaon asymmetries used for comparison are extracted with the p <5 GeV condition using the M_X >1.6 GeV cut using different PIDs.

This study shows, that the falling trend in z is not coming from exclusive reactions and it is a new behavior not predicted by theoretical calculations so far. As Figure 3.134 shows, this trend was also present in the results of the previous measurements, but it was not unequivocal due to the high statistical uncertainties.

3.8.3 Negative Kaons

The K^- shows exactly the same behavior, that was described in [46] for π^-: It is close to zero for the electronic variables, but it shows sign changes in z and P_T. Since they suspected that the sign change comes from π^--s produced in meson or baryon decays, so they investigated if the behavior is changed as a result of increasing the missing mass cut. Interestingly they found that while the z and P_T behavior does not change, the x_B and Q^2 values do change sign. This study was repeated for negative kaons and pions and shown in Figure 3.145. The asymmetries were extracted in 3 different missing mass ranges: 1.6 GeV$< M_X <$ 2.3 GeV, $M_X >$ 1.6 GeV and $M_X >$ 2.3 GeV. For π^- the above described behavior was reproduced, while for kaons we don't see any M_X dependence, except at small Q^2 and x_B. But if we also consider the systematic uncertainty of the pion contamination, which is not shown on the figures and higher than the statistical one in those bins, then there is no difference in those datapoints. With the higher missing mass cut, the π^- values have the same behavior as K^- in all kinematic variables, except z. If we compare the kaons to the pions with high missing mass cut, then we can observe the same things as in case of the K^+ and draw the same conclusion about the sea s-quarks, although the tendency here is not conspicuous due to the higher uncertainties. The only difference from pions seems to be the z dependence. As we can observe a sign change and an overall rising trend for the pions, the kaon values remain negative and they even fall with increasing z, just like for the positive kaons. This is much more prominent at high P_T and seems to be independent of x_B and Q^2. To understand the underlying physics, further studies are needed.

In the negative case, the comparison of the pion and kaon asymmetries are more complicated, since the d quark can originate from a sea or a valence quark of the proton, but the s quark of the kaon always originates from a sea quark. One of the possible explanations of the missing mass dependence of the π^- asymmetry is that different quarks are involved in the process. As the kaons don't have this behavior and they behave throughout the whole missing mass range as the pions at high missing mass, could support this idea, but much more detailed studies are needed to have a conclusion. But nonetheless this is a good example that the high precision multidimensional binning can shed light on new previously unseen phenomena.

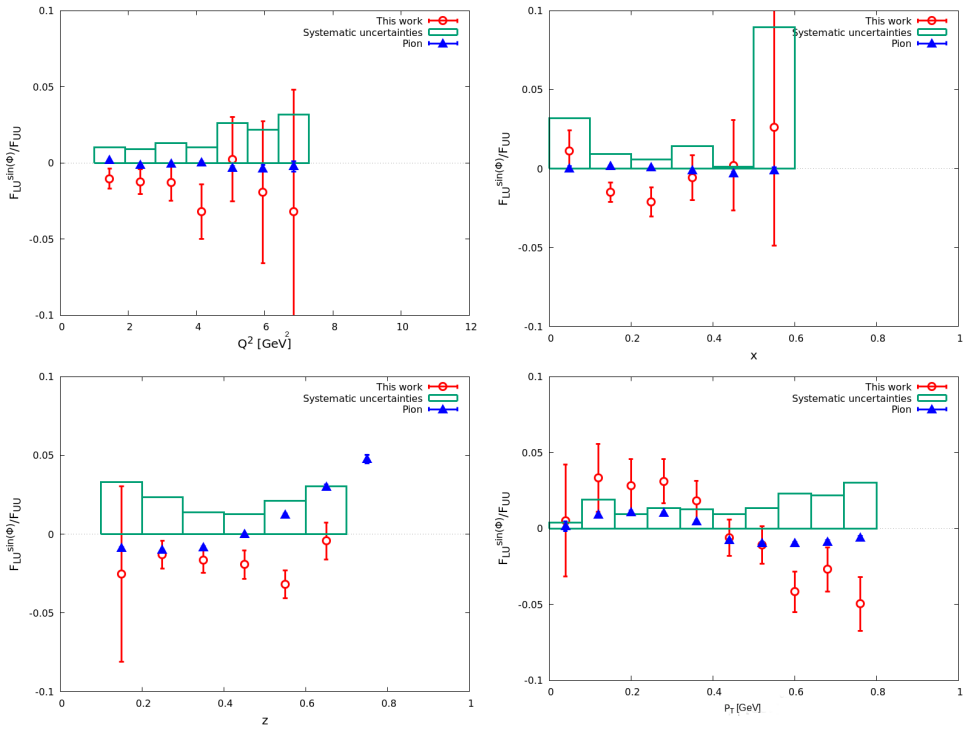


Figure 3.144: Final results after subtracting the pion asymmetries compared with pions in bins of Q^2 (top left), x_B (top right), z (bottom left) and P_T (bottom right) for K^-. The results are integrated over all the other kinematic variables. ($\langle x_B \rangle = 0.201$, $\langle Q^2 \rangle = 2.349$ GeV^2, $\langle z \rangle = 0.466$, $\langle P_T \rangle = 0.446$ GeV)

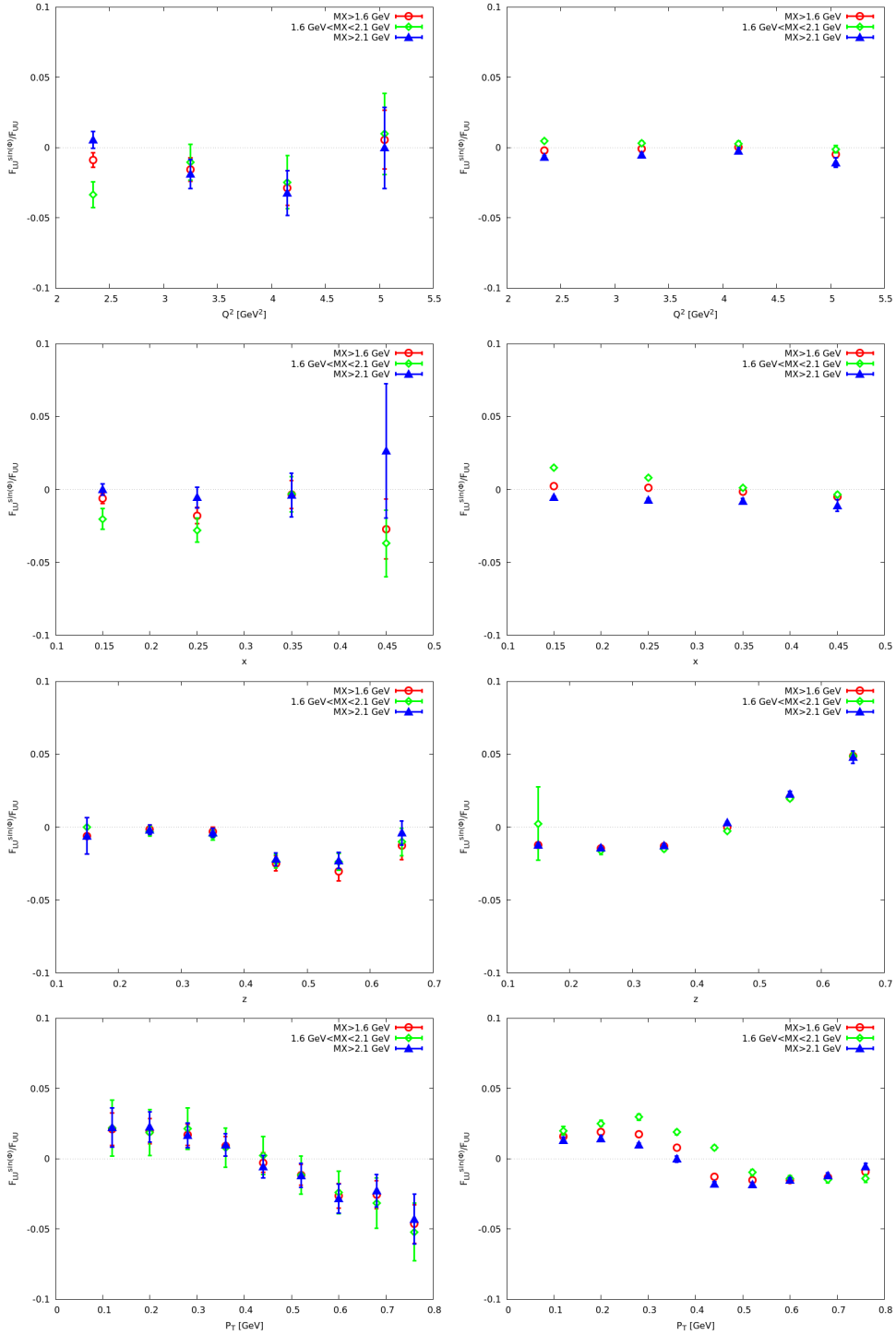


Figure 3.145: Final results after subtracting the pion asymmetries for K^- (left) compared with pions (right) in bins of Q^2 (top), x_B (upper middle), z (lower middle) and P_T (bottom) using different missing mass ranges. The results are integrated over all the other kinematic variables.

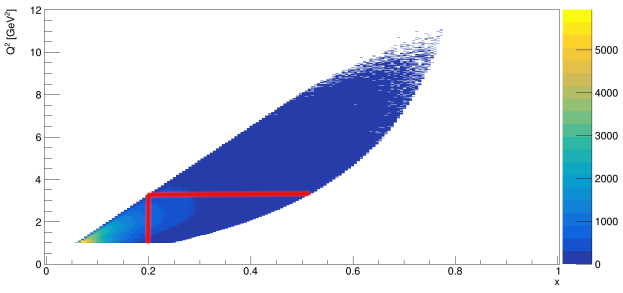


Figure 3.146: Bin borders on the Q^2-x_B-plane for the e^-K^-X dataset

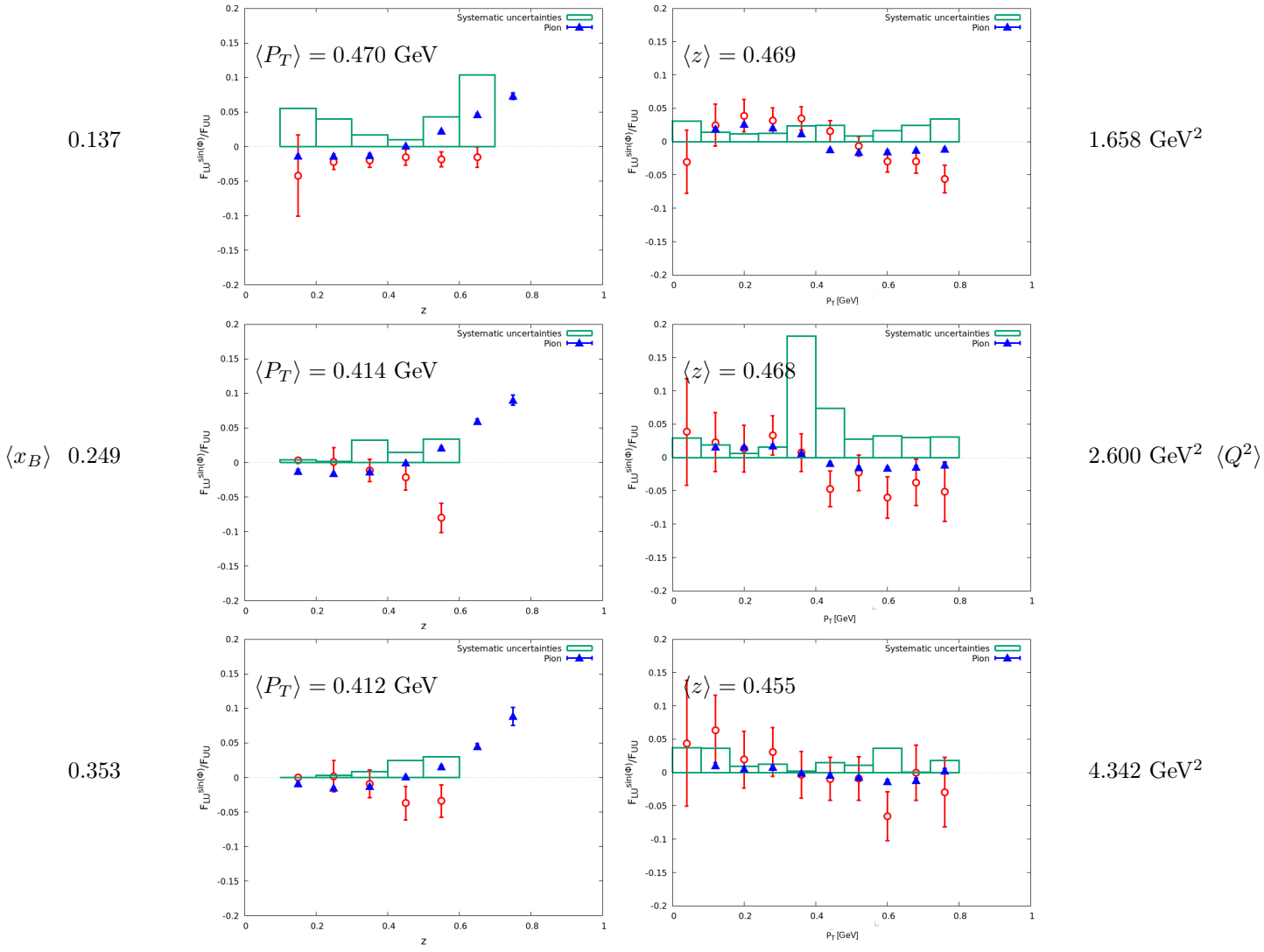


Figure 3.147: Final results after subtracting the pion asymmetries in bins of z (left) and P_T (right) for low x_B and low Q^2 (top), high x_B and low Q^2 (middle) and high x_B and high Q^2 (bottom) for K^- (x_B-Q^2-binning: Figure 3.146). The results are integrated over all the last kinematic variable (z or P_T).

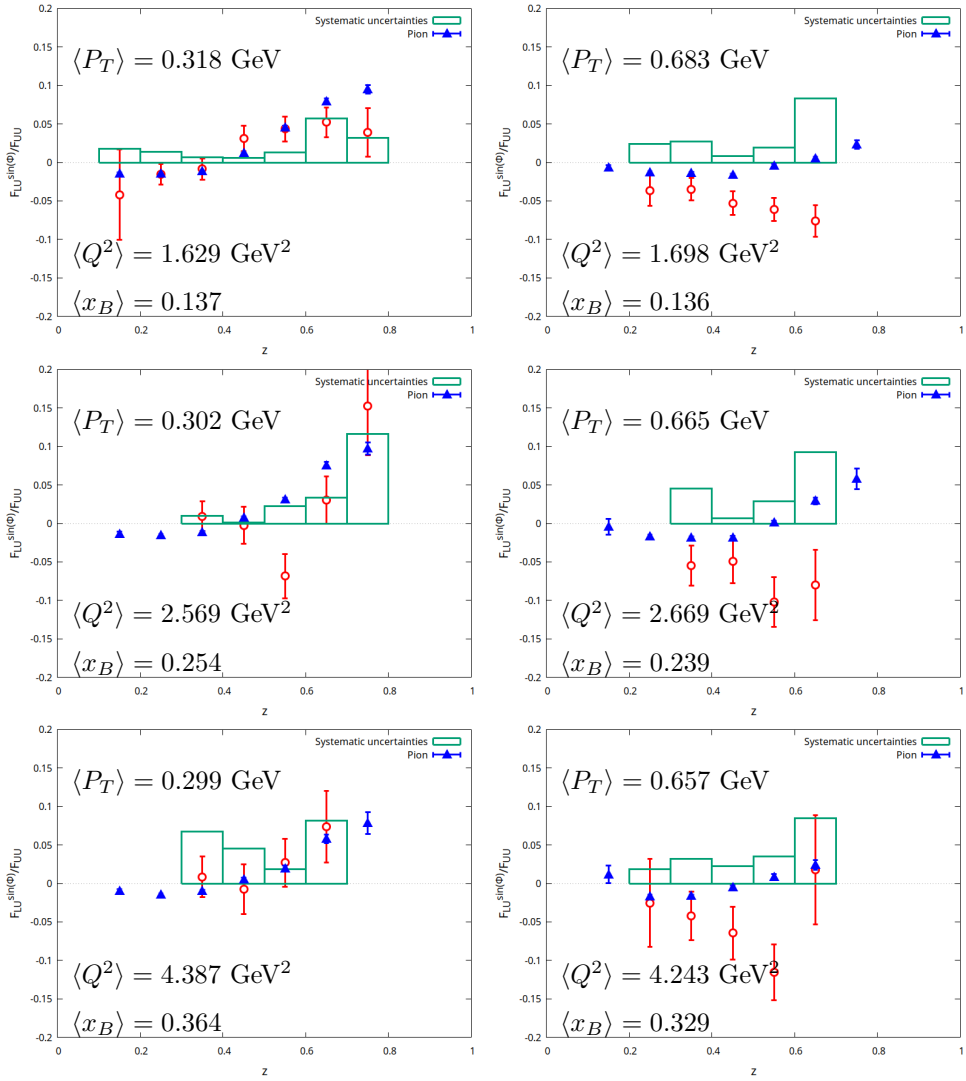


Figure 3.148: Final results after subtracting the pion asymmetries in bins of z for low x_B and low Q^2 (top), high x_B and low Q^2 (middle) and high x_B and high Q^2 (bottom) for low (left) and high (right) P_T for K^- (x_B-Q^2-binning: Figure 3.146)

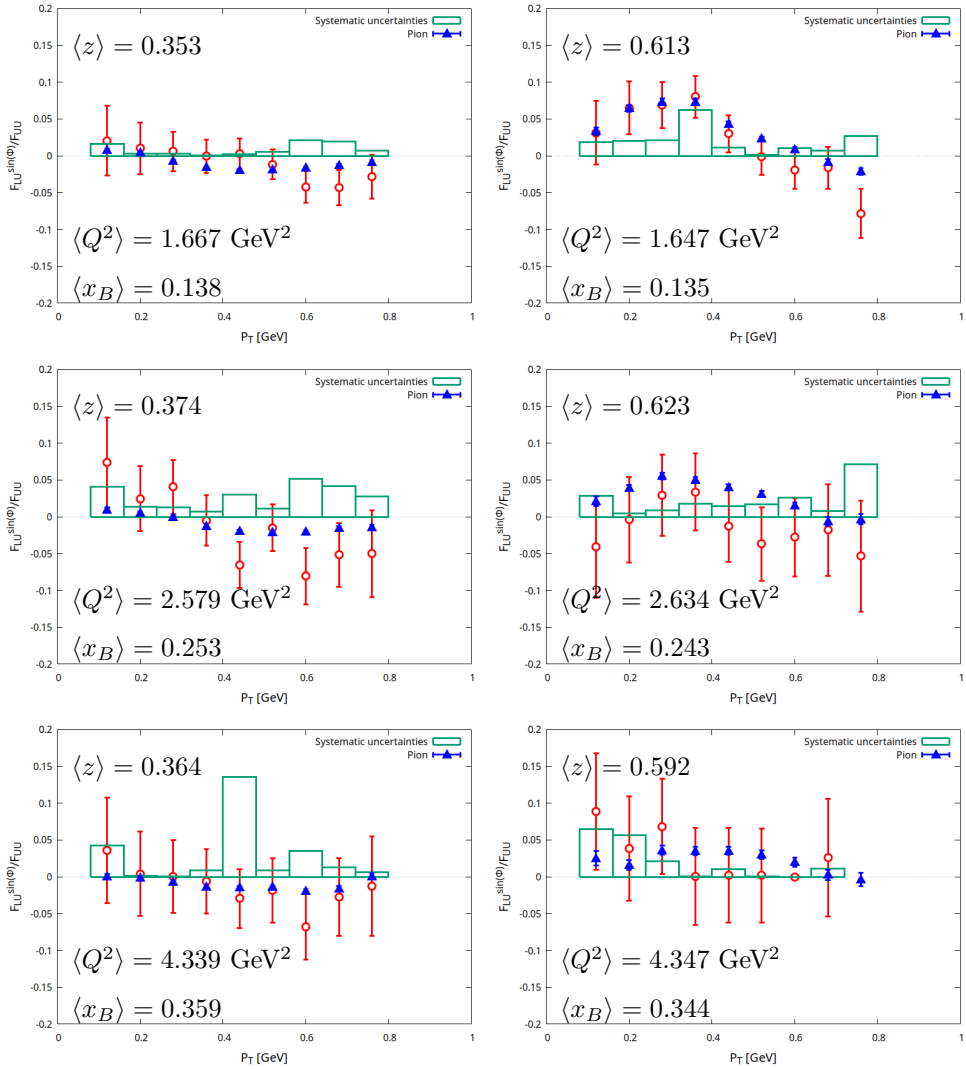


Figure 3.149: Final results after subtracting the pion asymmetries in bins of P_T for low x_B and low Q^2 (top), high x_B and low Q^2 (middle) and high x_B and high Q^2 (bottom) for low (left) and high (right) z for K^- (x_B-Q^2-binning: Figure 3.146)

The data presented in this work will help to further constrain the TMDs and FFs in global fits. They will serve as a valuable input, as they have much finer binning and smaller uncertainties compared to previous measurements and also this is the first fully multidimensional analysis for kaons.

Chapter 4

Conclusion and summary

The cluster finding algorithm for the $\overline{\text{P}}$ANDAs EMC was reworked and the Monte Carlo matching for resonances and neutral final states in PandaRoot was optimized. The effect of these improvements was demonstrated by performing a feasibility study for a predicted hybrid charmonium state. The possibility to utilize natural algorithms for optimizing the cuts during the analysis to increase the statistical significance was also explored. The conclusion of these studies was that with all of these improvements it will be possible to discover this exotic state once the $\overline{\text{P}}$ANDA experiment will be fully operational. In the best case scenario, the discovery will be possible within one year of data taking and if that happens it will be a forerunner of exploring a new exotic type of hadronic states.
A multidimensional study of charged kaon SIDIS beam spin asymmetry over a wide range of kinematics with CLAS12 was also presented. The main challenge of this analysis was to achieve a clean kaon sample, as the pion contamination was high, while still keeping the statistics as high as possible. This goal was reached by improving the PID with machine learning techniques and by performing a bin by bin defolding of the pion asymmetries. With the presented methods, the analysis has a much finer binning and smaller uncertainties compared to previous measurements and also it is the first fully multidimensional analysis for kaons. It can be used to further constrain the TMDs and FFs in global fits and thus provide a better understanding of the 3D structure of the proton.
As demonstrated by this work, the use of machine learning can be a very useful tool when doing data analysis in the field of hadron physics. By utilizing traditional approaches in parallel one can test the efficiency of the ML method from the one side. From the other side it also gives a better understanding of the data and of the way of thinking required to solve the given problem. This way, it makes the selection and construction of the input variables more efficient and also helps to find the ideal ML method or the ideal DNN layout. ML techniques were applied for three different problems in this work. The first one was the optimization of the cluster finding algorithm of the $\overline{\text{P}}$ANDAs EMC by developing a DNN to pair digis to be put into the same cluster. In this case traditional algorithms proved to be more effective. The second case was the optimization of the set of cuts used to increase the statistical significance of the signal. In some cases the improvement was more than a factor of two. The last case was a use of another DNN to differentiate charged kaons and pions by using all available detector information and thus achieving a better separation than achievable with just a single subsystem, according to the design parameters. Although the improvement in this case was significant (the average contamination was reduced from $\sim$50% to $\sim$15%, by keeping 58% of the statistics), it was not enough per se. But it provided an excellent tool to estimate the uncertainty of the pion subtraction based on MC data and to help verify this procedure, thus making the analysis possible.

Appendix

I Results of the 1 dimensional binning

i K^+

$<x_B>$	$<Q^2>$ [GeV^2]	$<z>$	$<P_T>$ [GeV]	$<W>$ [GeV]	$<y>$	$<M_X>$ [GeV]	$<\varepsilon>$	$F_{LU}^{\sin\phi}/F_{UU}$	Stat. unc.	Sys. unc.
z binning:										
0.170	2.352	0.150	0.220	3.512	0.705	2.953	0.527	0.0379	0.0167	0.0071
0.206	2.578	0.250	0.350	3.301	0.644	2.629	0.607	0.0511	0.0033	0.0065
0.229	2.590	0.350	0.414	3.111	0.585	2.332	0.677	0.0575	0.0030	0.0040
0.236	2.481	0.450	0.438	2.990	0.542	2.100	0.724	0.0585	0.0033	0.0031
0.227	2.348	0.550	0.442	2.974	0.529	1.945	0.739	0.0508	0.0037	0.0053
0.209	2.256	0.650	0.437	3.059	0.549	1.833	0.721	0.0182	0.0047	0.0083
0.183	2.035	0.750	0.386	3.142	0.561	1.715	0.717	0.0126	0.0120	1.2612
Q^2 binning:										
0.147	1.450	0.461	0.476	3.128	0.532	2.168	0.740	0.0514	0.0029	0.0037
0.222	2.350	0.452	0.419	3.043	0.549	2.137	0.718	0.0559	0.0032	0.0046
0.293	3.250	0.445	0.398	2.965	0.570	2.091	0.691	0.0583	0.0040	0.0047
0.356	4.150	0.439	0.383	2.904	0.597	2.053	0.657	0.0553	0.0052	0.0033
0.414	5.050	0.433	0.370	2.840	0.623	2.011	0.620	0.0483	0.0072	0.0066
0.469	5.950	0.428	0.356	2.762	0.645	1.959	0.587	0.0524	0.0110	0.0110
0.522	6.850	0.421	0.340	2.672	0.665	1.899	0.556	0.0700	0.0166	0.0194
0.572	7.750	0.413	0.319	2.577	0.683	1.838	0.525	0.0052	0.0299	0.0075
0.620	8.650	0.401	0.292	2.474	0.701	1.776	0.494	0.0034	0.0597	0.0534
0.664	9.550	0.384	0.263	2.376	0.718	1.722	0.464	-0.1929	0.1471	0.5152
P_t binning:										
0.284	2.927	0.461	0.040	2.869	0.530	2.109	0.734	0.0050	0.0091	0.0049
0.276	2.839	0.453	0.120	2.883	0.529	2.124	0.735	0.0224	0.0048	0.0033
0.266	2.740	0.450	0.200	2.894	0.527	2.122	0.738	0.0365	0.0042	0.0021
0.253	2.626	0.449	0.280	2.929	0.531	2.127	0.735	0.0446	0.0039	0.0036
0.238	2.515	0.447	0.360	2.986	0.542	2.146	0.724	0.0447	0.0039	0.0036
0.225	2.426	0.446	0.440	3.039	0.553	2.152	0.713	0.0552	0.0043	0.0049
0.214	2.352	0.448	0.520	3.087	0.563	2.144	0.703	0.0565	0.0045	0.0055
0.203	2.288	0.452	0.600	3.138	0.575	2.128	0.691	0.0582	0.0050	0.0060
0.193	2.226	0.458	0.680	3.189	0.588	2.105	0.677	0.0625	0.0059	0.0087
0.183	2.168	0.466	0.760	3.243	0.602	2.076	0.662	0.0526	0.0068	0.0240
x_B binning:										
0.050	1.150	0.438	0.606	3.562	0.654	2.475	0.606	0.0490	0.0063	0.0066
0.150	1.708	0.454	0.476	3.214	0.570	2.243	0.698	0.0524	0.0029	0.0024
0.250	2.565	0.457	0.404	2.954	0.533	2.068	0.734	0.0579	0.0033	0.0047
0.350	3.575	0.448	0.362	2.771	0.531	1.951	0.730	0.0568	0.0041	0.0056
0.450	4.784	0.434	0.328	2.630	0.553	1.865	0.701	0.0532	0.0061	0.0071

0.550	6.212	0.415	0.294	2.493	0.589	1.785	0.654	0.0398	0.0108	0.0162
0.650	7.836	0.390	0.247	2.349	0.635	1.705	0.588	-0.0076	0.0346	0.0701
0.750	9.483	0.358	0.188	2.206	0.696	1.637	0.494	-0.1621	0.4177	0.4385

Table 1: Final result of the one-dimensional study with z, Q^2, P_T or x_B as the main bin variable for K^+. The columns $<x_B>$, $<Q^2>$, $<z>$, $<P_T>$, $<W>$, $<y>$, $<M_X>$ and $<\varepsilon>$ provide the mean value of the stated variable or the bin middle if it is the main bin variable. The columns Stat. and Sys. provide the statistical and systematic uncertainty of $F_{LU}^{\sin\phi}/F_{UU}$

ii K^-

$<x_B>$	$<Q^2>$ [GeV^2]	$<z>$	$<P_T>$ [GeV]	$<W>$ [GeV]	$<y>$	$<M_X>$ [GeV]	$<\varepsilon>$	$F_{LU}^{\sin\phi}/F_{UU}$	Stat. unc.	Sys. unc.
z binning:										
0.136	1.888	0.150	0.243	3.587	0.701	3.016	0.535	-0.0253	0.0558	0.0330
0.185	2.334	0.250	0.379	3.352	0.648	2.662	0.603	-0.0129	0.0088	0.0234
0.216	2.400	0.350	0.434	3.110	0.575	2.323	0.690	-0.0164	0.0080	0.0140
0.215	2.405	0.450	0.474	3.124	0.582	2.180	0.675	-0.0193	0.0091	0.0126
0.198	2.411	0.550	0.489	3.272	0.627	2.117	0.625	-0.0319	0.0088	0.0213
0.190	2.292	0.650	0.456	3.267	0.616	1.941	0.646	-0.0044	0.0116	0.0302
0.170	1.910	0.750	0.344	3.190	0.568	1.740	0.712	0.0470	0.0258	0.5937
0.196	1.853	0.850	0.741	2.972	0.502	1.624	0.778	0.4813	0.3511	2.6385
Q^2 binning:										
0.129	1.450	0.490	0.475	3.308	0.587	2.238	0.678	-0.0103	0.0065	0.0104
0.205	2.350	0.482	0.472	3.180	0.594	2.154	0.667	-0.0123	0.0083	0.0089
0.276	3.250	0.481	0.449	3.076	0.605	2.084	0.650	-0.0127	0.0120	0.0132
0.340	4.150	0.483	0.432	2.998	0.626	2.028	0.620	-0.0320	0.0180	0.0103
0.399	5.050	0.484	0.413	2.916	0.646	1.971	0.590	0.0024	0.0277	0.0261
0.455	5.950	0.484	0.398	2.828	0.664	1.910	0.561	-0.0192	0.0467	0.0218
0.509	6.850	0.479	0.371	2.727	0.680	1.850	0.534	-0.0320	0.0801	0.0319
0.563	7.750	0.469	0.346	2.619	0.695	1.791	0.509	0.1487	0.1580	0.7558
0.611	8.650	0.459	0.322	2.513	0.711	1.737	0.479	0.0142	0.0293	0.0630
P_T binning:										
0.236	2.504	0.501	0.040	3.002	0.545	2.130	0.721	0.0052	0.0369	0.0039
0.237	2.524	0.498	0.120	3.011	0.550	2.132	0.715	0.0333	0.0222	0.0190
0.233	2.526	0.489	0.200	3.043	0.560	2.155	0.703	0.0281	0.0174	0.0094
0.223	2.464	0.480	0.280	3.092	0.573	2.187	0.689	0.0311	0.0144	0.0133
0.214	2.407	0.475	0.360	3.138	0.584	2.202	0.677	0.0184	0.0128	0.0125
0.208	2.385	0.474	0.440	3.172	0.594	2.196	0.666	-0.0061	0.0120	0.0096
0.201	2.364	0.476	0.520	3.212	0.605	2.185	0.653	-0.0109	0.0123	0.0135
0.194	2.335	0.482	0.600	3.256	0.618	2.165	0.640	-0.0417	0.0134	0.0229
0.187	2.306	0.489	0.680	3.300	0.630	2.139	0.625	-0.0270	0.0146	0.0219
0.180	2.269	0.500	0.760	3.349	0.644	2.105	0.609	-0.0496	0.0178	0.0300
x_B binning:										
0.050	1.158	0.513	0.509	3.609	0.670	2.408	0.583	0.0111	0.0132	0.0318
0.150	1.763	0.489	0.493	3.316	0.606	2.238	0.656	-0.0149	0.0061	0.0092
0.250	2.744	0.480	0.446	3.053	0.574	2.075	0.686	-0.0212	0.0093	0.0058
0.350	3.812	0.473	0.396	2.854	0.569	1.952	0.686	-0.0058	0.0143	0.0143
0.450	5.017	0.463	0.353	2.687	0.582	1.852	0.664	0.0021	0.0286	0.0010
0.550	6.375	0.443	0.304	2.524	0.607	1.765	0.629	0.0261	0.0747	0.0896
0.650	7.891	0.414	0.252	2.366	0.644	1.692	0.575	0.3609	0.2991	0.3579

Table 2: Final result of the one-dimensional study with z, Q^2, P_T or x_B as the main bin variable for K^-. The columns $<x_B>$, $<Q^2>$, $<z>$, $<P_T>$, $<W>$, $<y>$, $<M_X>$ and $<\varepsilon>$ provide the mean value of the stated variable or the bin middle if it is the main bin variable. The columns Stat. and Sys. provide the statistical and systematic uncertainty of $F_{LU}^{\sin\phi}/F_{UU}$

II Results of the 3 dimensional binning

i K^+ with 3 bins in $x_B - Q^2$

$<x_B>$	$<Q^2>$ [GeV^2]	$<z>$	$<P_T>$ [GeV]	$<W>$ [GeV]	$<y>$	$<M_X>$ [GeV]	$<\varepsilon>$	$F_{LU}^{\sin\phi}/F_{UU}$	Stat. unc.	Sys. unc.
$x - Q^2$ bin 1:										
z binning:										
0.141	1.945	0.150	0.237	3.576	0.706	3.009	0.527	0.0422	0.0191	0.0085
0.142	1.818	0.250	0.391	3.455	0.656	2.756	0.597	0.0497	0.0040	0.0067
0.145	1.726	0.350	0.480	3.332	0.609	2.501	0.654	0.0594	0.0037	0.0044
0.148	1.644	0.450	0.530	3.221	0.569	2.255	0.700	0.0556	0.0040	0.0039
0.150	1.594	0.550	0.551	3.161	0.547	2.048	0.724	0.0430	0.0045	0.0045
0.148	1.603	0.650	0.551	3.181	0.553	1.885	0.720	0.0223	0.0059	0.0102
P_T binning:										
0.167	1.878	0.460	0.040	3.207	0.579	2.370	0.687	-0.0129	0.0217	0.0150
0.163	1.808	0.451	0.120	3.185	0.569	2.363	0.697	0.0238	0.0094	0.0062
0.161	1.734	0.452	0.200	3.152	0.554	2.322	0.714	0.0285	0.0066	0.0028
0.156	1.681	0.450	0.280	3.159	0.552	2.308	0.718	0.0392	0.0057	0.0038
0.150	1.654	0.446	0.360	3.199	0.562	2.320	0.707	0.0342	0.0051	0.0028
0.147	1.638	0.445	0.440	3.231	0.571	2.310	0.697	0.0440	0.0052	0.0040
0.145	1.631	0.447	0.520	3.251	0.577	2.281	0.691	0.0531	0.0054	0.0057
0.143	1.630	0.451	0.600	3.274	0.584	2.244	0.684	0.0490	0.0055	0.0053
0.141	1.634	0.456	0.680	3.299	0.592	2.202	0.675	0.0572	0.0065	0.0080
0.139	1.642	0.464	0.760	3.333	0.604	2.156	0.663	0.0443	0.0071	0.0202
$x - Q^2$ bin 2:										
z binning:										
0.220	2.974	0.150	0.190	3.383	0.691	2.840	0.542	0.0303	0.0481	0.0048
0.237	2.770	0.250	0.319	3.141	0.602	2.498	0.660	0.0530	0.0062	0.0060
0.255	2.593	0.350	0.387	2.924	0.527	2.188	0.743	0.0562	0.0042	0.0039
0.263	2.468	0.450	0.419	2.805	0.485	1.976	0.785	0.0627	0.0044	0.0032
0.258	2.419	0.550	0.429	2.801	0.481	1.850	0.791	0.0604	0.0054	0.0065
0.247	2.473	0.650	0.429	2.902	0.512	1.766	0.762	0.0194	0.0078	0.0089
P_T binning:										
0.271	2.485	0.476	0.040	2.767	0.476	2.015	0.793	-0.0032	0.0140	0.0032
0.269	2.473	0.464	0.120	2.770	0.476	2.026	0.793	0.0151	0.0075	0.0022
0.267	2.459	0.459	0.200	2.774	0.476	2.021	0.793	0.0441	0.0061	0.0029
0.265	2.461	0.458	0.280	2.789	0.480	2.010	0.790	0.0513	0.0056	0.0039
0.261	2.472	0.456	0.360	2.816	0.488	2.002	0.782	0.0539	0.0058	0.0038
0.256	2.485	0.453	0.440	2.852	0.499	1.993	0.773	0.0677	0.0066	0.0056
0.251	2.508	0.454	0.520	2.897	0.513	1.981	0.759	0.0556	0.0069	0.0053
0.246	2.541	0.459	0.600	2.948	0.529	1.963	0.743	0.0696	0.0086	0.0072
0.241	2.584	0.465	0.680	3.005	0.548	1.942	0.724	0.0676	0.0113	0.0094
0.237	2.633	0.474	0.760	3.063	0.568	1.912	0.702	0.0690	0.0151	0.0315
$x - Q^2$ bin 3:										
z binning:										
0.289	4.012	0.150	0.166	3.282	0.708	2.749	0.511	0.0090	0.0415	0.0017
0.343	4.386	0.250	0.300	3.051	0.656	2.422	0.581	0.0572	0.0061	0.0067
0.372	4.415	0.350	0.374	2.899	0.612	2.170	0.635	0.0588	0.0050	0.0040
0.378	4.347	0.450	0.412	2.842	0.592	2.002	0.660	0.0584	0.0055	0.0030
0.368	4.277	0.550	0.434	2.876	0.597	1.888	0.655	0.0525	0.0076	0.0058
0.346	4.208	0.650	0.443	2.976	0.622	1.794	0.626	0.0055	0.0128	0.0025
P_T binning:										

0.395	4.433	0.439	0.040	2.778	0.579	2.062	0.674	0.0315	0.0170	0.0300
0.393	4.429	0.435	0.120	2.789	0.582	2.068	0.671	0.0317	0.0091	0.0047
0.390	4.423	0.432	0.200	2.800	0.584	2.064	0.669	0.0357	0.0075	0.0020
0.385	4.404	0.432	0.280	2.822	0.589	2.057	0.663	0.0451	0.0072	0.0034
0.378	4.383	0.433	0.360	2.854	0.597	2.051	0.654	0.0496	0.0074	0.0034
0.368	4.349	0.433	0.440	2.894	0.606	2.042	0.643	0.0615	0.0082	0.0049
0.358	4.308	0.437	0.520	2.937	0.616	2.025	0.632	0.0695	0.0099	0.0066
0.347	4.264	0.443	0.600	2.987	0.629	2.006	0.617	0.0713	0.0121	0.0073
0.336	4.212	0.452	0.680	3.038	0.641	1.978	0.602	0.0764	0.0151	0.0108
0.325	4.159	0.462	0.760	3.088	0.653	1.940	0.587	0.0775	0.0207	0.0354

Table 3: Final result of the 3-dimensional study with z or P_T as the main bin variable with 3 $x_B - Q^2$ bins for K^+. The columns $<x_B>$, $<Q^2>$, $<z>$, $<P_T>$, $<W>$, $<y>$, $<M_X>$ and $<\varepsilon>$ provide the mean value of the stated variable or the bin middle if it is the main bin variable. The columns Stat. and Sys. provide the statistical and systematic uncertainty of $F_{LU}^{\sin\phi}/F_{UU}$

ii K^+ with 5 bins in $x_B - Q^2$

$<x_B>$	$<Q^2>$ [GeV^2]	$<z>$	$<P_T>$ [GeV]	$<W>$ [GeV]	$<y>$	$<M_X>$ [GeV]	$<\varepsilon>$	$F_{LU}^{\sin\phi}/F_{UU}$	Stat. unc.	Sys. unc.
$x - Q^2$ bin 1:										
z binning:										
0.121	1.679	0.150	0.249	3.622	0.709	3.049	0.524	0.0538	0.0249	0.0151
0.118	1.541	0.250	0.416	3.526	0.665	2.810	0.586	0.0505	0.0050	0.0091
0.119	1.464	0.350	0.512	3.421	0.625	2.563	0.637	0.0614	0.0047	0.0045
0.121	1.407	0.450	0.568	3.335	0.593	2.326	0.676	0.0501	0.0052	0.0031
0.122	1.374	0.550	0.595	3.286	0.575	2.116	0.696	0.0372	0.0058	0.0044
0.122	1.367	0.650	0.596	3.278	0.571	1.925	0.702	0.0257	0.0082	0.0118
P_T binning:										
0.134	1.657	0.442	0.040	3.402	0.632	2.542	0.630	0.0409	0.0433	0.0741
0.132	1.595	0.436	0.120	3.376	0.621	2.532	0.638	0.0163	0.0187	0.0089
0.121	1.407	0.441	0.440	3.338	0.594	2.402	0.674	0.0394	0.0070	0.0036
0.120	1.400	0.443	0.520	3.350	0.597	2.367	0.670	0.0453	0.0064	0.0048
0.119	1.396	0.447	0.600	3.361	0.600	2.322	0.667	0.0481	0.0068	0.0051
0.118	1.397	0.453	0.680	3.376	0.605	2.271	0.662	0.0529	0.0078	0.0074
0.117	1.404	0.460	0.760	3.399	0.613	2.218	0.653	0.0392	0.0083	0.0179
$x - Q^2$ bin 2:										
z binning:										
0.236	2.203	0.250	0.285	2.826	0.476	2.217	0.807	0.0443	0.0144	0.0052
0.243	2.123	0.350	0.372	2.740	0.446	2.034	0.831	0.0638	0.0059	0.0049
0.242	2.030	0.450	0.413	2.687	0.426	1.887	0.847	0.0686	0.0052	0.0051
0.235	1.981	0.550	0.418	2.696	0.426	1.786	0.848	0.0499	0.0055	0.0054
0.223	1.960	0.650	0.396	2.762	0.442	1.706	0.836	0.0188	0.0089	0.0087
P_T binning:										
0.257	2.152	0.501	0.040	2.659	0.426	1.903	0.846	-0.0151	0.0197	0.0151
0.252	2.105	0.484	0.120	2.663	0.424	1.920	0.847	0.0191	0.0097	0.0031
0.247	2.059	0.476	0.200	2.670	0.424	1.921	0.848	0.0425	0.0074	0.0031
0.243	2.033	0.475	0.280	2.678	0.424	1.905	0.848	0.0580	0.0068	0.0047
0.240	2.021	0.472	0.360	2.692	0.427	1.887	0.846	0.0494	0.0068	0.0049
0.236	2.017	0.468	0.440	2.714	0.432	1.867	0.843	0.0683	0.0078	0.0080
0.231	2.010	0.469	0.520	2.744	0.440	1.842	0.837	0.0609	0.0082	0.0082
0.225	2.000	0.474	0.600	2.778	0.449	1.809	0.831	0.0633	0.0104	0.0086
0.218	1.980	0.479	0.680	2.817	0.459	1.774	0.823	0.0881	0.0148	0.0124
0.209	1.939	0.483	0.760	2.860	0.469	1.736	0.815	0.1078	0.0235	0.0492
$x - Q^2$ bin 3:										
z binning:										
0.208	2.853	0.150	0.200	3.429	0.701	2.880	0.530	0.0221	0.0221	0.0037
0.222	2.822	0.250	0.340	3.278	0.648	2.618	0.602	0.0539	0.0042	0.0065
0.226	2.766	0.350	0.430	3.214	0.624	2.423	0.635	0.0572	0.0038	0.0039
0.228	2.729	0.450	0.487	3.181	0.611	2.240	0.652	0.0588	0.0044	0.0030
0.228	2.710	0.550	0.519	3.169	0.605	2.061	0.659	0.0690	0.0058	0.0077
0.226	2.693	0.650	0.520	3.173	0.606	1.883	0.659	0.0160	0.0070	0.0073
P_T binning:										
0.236	2.825	0.441	0.040	3.161	0.611	2.358	0.651	-0.0008	0.0149	0.0008
0.236	2.813	0.435	0.120	3.159	0.609	2.359	0.653	0.0258	0.0079	0.0039
0.235	2.795	0.434	0.200	3.159	0.608	2.347	0.655	0.0394	0.0063	0.0024
0.232	2.772	0.435	0.280	3.171	0.610	2.334	0.653	0.0447	0.0055	0.0036
0.228	2.748	0.435	0.360	3.185	0.613	2.318	0.649	0.0471	0.0055	0.0033
0.227	2.733	0.436	0.440	3.191	0.614	2.286	0.648	0.0587	0.0059	0.0046

0.225	2.718	0.440	0.520	3.196	0.615	2.241	0.647	0.0634	0.0065	0.0061
0.224	2.707	0.446	0.600	3.203	0.616	2.187	0.645	0.0621	0.0068	0.0066
0.221	2.693	0.455	0.680	3.212	0.618	2.125	0.642	0.0599	0.0082	0.0084
0.219	2.677	0.467	0.760	3.226	0.622	2.058	0.638	0.0538	0.0099	0.0246
$x - Q^2$ bin 4:										
z binning:										
0.436	4.295	0.250	0.211	2.523	0.500	1.960	0.770	0.0900	0.0239	0.0110
0.430	3.977	0.350	0.299	2.465	0.468	1.822	0.800	0.0550	0.0085	0.0047
0.418	3.780	0.450	0.320	2.464	0.457	1.747	0.810	0.0623	0.0086	0.0037
0.405	3.734	0.550	0.316	2.512	0.467	1.692	0.803	0.0345	0.0127	0.0044
0.389	3.747	0.650	0.285	2.595	0.488	1.645	0.785	0.0479	0.0392	0.0278
P_T binning:										
0.429	3.859	0.447	0.040	2.435	0.454	1.790	0.812	0.0350	0.0252	0.0361
0.427	3.842	0.444	0.120	2.438	0.454	1.787	0.812	0.0264	0.0146	0.0041
0.425	3.840	0.439	0.200	2.447	0.457	1.781	0.810	0.0386	0.0116	0.0022
0.421	3.833	0.435	0.280	2.461	0.460	1.768	0.808	0.0582	0.0116	0.0044
0.417	3.832	0.433	0.360	2.484	0.465	1.752	0.804	0.0602	0.0132	0.0043
0.410	3.845	0.433	0.440	2.519	0.474	1.735	0.796	0.0598	0.0159	0.0068
0.403	3.852	0.435	0.520	2.557	0.484	1.712	0.788	0.0793	0.0228	0.0108
0.394	3.842	0.439	0.600	2.597	0.494	1.682	0.779	0.0935	0.0398	0.0168
0.384	3.805	0.440	0.680	2.640	0.504	1.651	0.771	0.0718	0.0794	0.0266
0.371	3.722	0.433	0.760	2.681	0.511	1.625	0.765	0.0894	0.3322	0.0420
$x - Q^2$ bin 5:										
z binning:										
0.382	5.313	0.150	0.124	3.077	0.708	2.561	0.503	-0.0123	0.1249	0.0031
0.419	5.373	0.250	0.274	2.879	0.653	2.272	0.581	0.0518	0.0101	0.0100
0.426	5.321	0.350	0.366	2.827	0.634	2.110	0.606	0.0602	0.0080	0.0042
0.424	5.217	0.450	0.414	2.812	0.624	1.974	0.620	0.0609	0.0093	0.0044
0.416	5.114	0.550	0.428	2.826	0.623	1.852	0.622	0.0442	0.0134	0.0047
0.404	5.013	0.650	0.402	2.865	0.629	1.738	0.615	-0.0019	0.0254	0.0009
P_T binning:										
0.431	5.302	0.438	0.040	2.794	0.625	2.072	0.619	0.0283	0.0336	0.0282
0.430	5.299	0.433	0.120	2.797	0.625	2.074	0.618	0.0266	0.0173	0.0042
0.429	5.288	0.431	0.200	2.801	0.626	2.063	0.618	0.0304	0.0136	0.0018
0.428	5.278	0.432	0.280	2.805	0.626	2.041	0.617	0.0389	0.0132	0.0031
0.426	5.262	0.433	0.360	2.814	0.627	2.014	0.616	0.0544	0.0128	0.0044
0.422	5.221	0.434	0.440	2.824	0.628	1.981	0.615	0.0662	0.0137	0.0055
0.417	5.167	0.436	0.520	2.838	0.629	1.941	0.614	0.0632	0.0162	0.0067
0.410	5.113	0.443	0.600	2.861	0.633	1.896	0.609	0.0726	0.0210	0.0080
0.403	5.068	0.453	0.680	2.885	0.637	1.843	0.603	0.0996	0.0272	0.0144
0.396	5.044	0.460	0.760	2.920	0.646	1.795	0.592	0.0959	0.0429	0.0439

Table 4: Final result of the 3-dimensional study with z or P_T as the main bin variable with 5 $x_B - Q^2$ bins for K^+. The columns $<x_B>$, $<Q^2>$, $<z>$, $<P_T>$, $<W>$, $<y>$, $<M_X>$ and $<\varepsilon>$ provide the mean value of the stated variable or the bin middle if it is the main bin variable. The columns Stat. and Sys. provide the statistical and systematic uncertainty of $F_{LU}^{\sin\phi}/F_{UU}$

iii K^-

$<x_B>$	$<Q^2>$ [GeV^2]	$<z>$	$<P_T>$ [GeV]	$<W>$ [GeV]	$<y>$	$<M_X>$ [GeV]	$<\varepsilon>$	$F_{LU}^{\sin\phi}/F_{UU}$	Stat. unc.	Sys. unc.
$x-Q^2$ bin 1										
z binning:										
0.118	1.637	0.150	0.252	3.625	0.701	3.050	0.536	-0.0419	0.0588	0.0554
0.133	1.714	0.250	0.401	3.475	0.656	2.770	0.597	-0.0219	0.0109	0.0397
0.141	1.637	0.350	0.476	3.298	0.594	2.475	0.674	-0.0202	0.0101	0.0172
0.140	1.658	0.450	0.539	3.340	0.610	2.331	0.647	-0.0151	0.0115	0.0099
0.134	1.684	0.550	0.558	3.440	0.643	2.222	0.610	-0.0182	0.0108	0.0433
0.135	1.644	0.650	0.540	3.391	0.622	2.007	0.643	-0.0153	0.0143	0.1038
0.137	1.537	0.750	0.432	3.247	0.567	1.759	0.716	0.0325	0.0282	0.4107
0.129	1.347	0.850	0.503	3.178	0.534	1.624	0.753	0.0885	0.3263	0.4851
P_T binning:										
0.141	1.571	0.504	0.040	3.232	0.564	2.291	0.708	-0.0302	0.0477	0.0312
0.139	1.554	0.509	0.120	3.255	0.572	2.292	0.697	0.0248	0.0314	0.0141
0.139	1.580	0.500	0.200	3.280	0.583	2.313	0.684	0.0391	0.0243	0.0121
0.139	1.614	0.490	0.280	3.314	0.597	2.338	0.668	0.0320	0.0191	0.0126
0.138	1.633	0.486	0.360	3.341	0.607	2.340	0.655	0.0346	0.0175	0.0233
0.138	1.645	0.484	0.440	3.352	0.612	2.321	0.650	0.0154	0.0159	0.0244
0.138	1.668	0.483	0.520	3.370	0.620	2.299	0.641	-0.0070	0.0151	0.0086
0.138	1.687	0.486	0.600	3.390	0.627	2.266	0.632	-0.0301	0.0158	0.0167
0.137	1.701	0.490	0.680	3.413	0.636	2.228	0.622	-0.0299	0.0176	0.0242
0.135	1.715	0.501	0.760	3.445	0.647	2.182	0.608	-0.0559	0.0210	0.0339
$x-Q^2$ bin 2										
z binning:										
0.219	2.969	0.150	0.208	3.384	0.686	2.833	0.550	0.0030	0.0019	0.0039
0.235	2.779	0.250	0.366	3.159	0.609	2.492	0.651	0.0008	0.0205	0.0016
0.254	2.576	0.350	0.419	2.916	0.525	2.169	0.746	-0.0115	0.0158	0.0326
0.259	2.492	0.450	0.451	2.847	0.502	1.993	0.764	-0.0217	0.0186	0.0146
0.245	2.624	0.550	0.488	3.007	0.555	1.959	0.707	-0.0801	0.0213	0.0337
0.239	2.700	0.650	0.483	3.082	0.579	1.842	0.689	-0.0120	0.0254	0.0818
0.236	2.621	0.750	0.401	3.057	0.564	1.696	0.712	0.1263	0.0611	1.5947
P_T binning:										
0.268	2.484	0.498	0.040	2.783	0.480	1.983	0.789	0.0384	0.0799	0.0292
0.266	2.498	0.488	0.120	2.799	0.486	2.001	0.784	0.0232	0.0440	0.0191
0.263	2.514	0.476	0.200	2.831	0.497	2.024	0.772	0.0133	0.0349	0.0059
0.259	2.528	0.465	0.280	2.864	0.508	2.043	0.761	0.0335	0.0295	0.0157
0.255	2.542	0.457	0.360	2.893	0.517	2.044	0.752	0.0075	0.0282	0.1824
0.250	2.565	0.456	0.440	2.935	0.530	2.042	0.738	-0.0469	0.0265	0.0740
0.246	2.603	0.460	0.520	2.987	0.547	2.031	0.721	-0.0227	0.0266	0.0280
0.241	2.647	0.469	0.600	3.044	0.567	2.014	0.700	-0.0600	0.0309	0.0326
0.235	2.690	0.481	0.680	3.104	0.587	1.990	0.677	-0.0374	0.0344	0.0302
0.232	2.742	0.496	0.760	3.160	0.607	1.956	0.654	-0.0514	0.0448	0.0310
$x-Q^2$ bin 3										
z binning:										
0.281	3.944	0.150	0.180	3.309	0.707	2.770	0.513	-0.0002	0.0020	0.0003
0.337	4.312	0.250	0.343	3.068	0.658	2.417	0.578	0.0018	0.0228	0.0034
0.371	4.314	0.350	0.413	2.877	0.603	2.134	0.647	-0.0094	0.0199	0.0082
0.361	4.389	0.450	0.479	2.954	0.630	2.052	0.611	-0.0371	0.0243	0.0247
0.347	4.436	0.550	0.500	3.041	0.655	1.965	0.581	-0.0341	0.0237	0.0301
0.339	4.262	0.650	0.466	3.031	0.641	1.819	0.603	0.0575	0.0396	0.3918

0.327	3.880	0.750	0.387	2.980	0.605	1.677	0.656	-0.0624	0.1403	1.0079
P_T binning:										
0.391	4.486	0.498	0.040	2.816	0.591	1.992	0.660	0.0439	0.0943	0.0371
0.387	4.461	0.493	0.120	2.830	0.595	2.001	0.656	0.0637	0.0521	0.0365
0.379	4.448	0.486	0.200	2.867	0.605	2.024	0.643	0.0195	0.0425	0.0092
0.372	4.416	0.478	0.280	2.897	0.613	2.038	0.634	0.0309	0.0366	0.0124
0.365	4.391	0.471	0.360	2.929	0.621	2.044	0.625	-0.0033	0.0350	0.0023
0.356	4.371	0.471	0.440	2.967	0.631	2.039	0.612	-0.0094	0.0323	0.0149
0.347	4.325	0.472	0.520	3.011	0.641	2.029	0.600	-0.0090	0.0330	0.0113
0.338	4.307	0.482	0.600	3.051	0.652	2.000	0.587	-0.0654	0.0364	0.0361
0.328	4.259	0.493	0.680	3.098	0.664	1.969	0.572	-0.0004	0.0414	0.0003
0.318	4.203	0.503	0.760	3.145	0.675	1.935	0.557	-0.0294	0.0519	0.0178

Table 5: Final result of the 3-dimensional study with z or P_T as the main bin variable with 3 $x_B - Q^2$ bins for K^-. The columns $<x_B>$, $<Q^2>$, $<z>$, $<P_T>$, $<W>$, $<y>$, $<M_X>$ and $<\varepsilon>$ provide the mean value of the stated variable or the bin middle if it is the main bin variable. The columns Stat. and Sys. provide the statistical and systematic uncertainty of $F_{LU}^{\sin\phi}/F_{UU}$

III Results of the 4 dimensional binning

i K^+

$<x_B>$	$<Q^2>$ [GeV^2]	$<z>$	$<P_T>$ [GeV]	$<W>$ [GeV]	$<y>$	$<M_X>$ [GeV]	$<\varepsilon>$	$F_{LU}^{\sin\phi}/F_{UU}$	Stat. unc.	Sys. unc.
$x-Q^2$ bin 1										
P_T bin 1										
0.144	1.983	0.150	0.215	3.561	0.703	3.006	0.531	0.0504	0.0203	0.0087
0.152	1.920	0.250	0.238	3.413	0.649	2.805	0.606	0.0460	0.0057	0.0052
0.156	1.824	0.350	0.254	3.277	0.598	2.573	0.667	0.0397	0.0052	0.0039
0.160	1.710	0.450	0.273	3.143	0.549	2.320	0.720	0.0272	0.0064	0.0024
0.161	1.613	0.550	0.290	3.044	0.512	2.083	0.758	0.0238	0.0087	0.0028
0.160	1.599	0.650	0.303	3.046	0.511	1.901	0.763	0.0154	0.0117	0.0079
0.155	1.626	0.750	0.310	3.119	0.532	1.739	0.750	0.0056	0.0242	0.5570
P_T bin 2										
0.121	1.725	0.150	0.364	3.661	0.721	3.025	0.506	-0.0242	0.0515	0.0091
0.137	1.767	0.250	0.464	3.471	0.658	2.734	0.595	0.0542	0.0051	0.0113
0.143	1.697	0.350	0.499	3.328	0.606	2.503	0.659	0.0644	0.0045	0.0045
0.147	1.621	0.450	0.519	3.213	0.565	2.275	0.705	0.0552	0.0049	0.0031
0.149	1.573	0.550	0.536	3.150	0.542	2.067	0.730	0.0427	0.0056	0.0045
0.147	1.585	0.650	0.549	3.173	0.549	1.898	0.725	0.0157	0.0078	0.0074
P_T bin 3										
0.120	1.652	0.250	0.707	3.606	0.696	2.647	0.543	0.0453	0.0201	0.0094
0.134	1.679	0.350	0.768	3.435	0.640	2.376	0.619	0.0798	0.0083	0.0055
0.139	1.635	0.450	0.814	3.320	0.599	2.138	0.668	0.0828	0.0077	0.0052
0.140	1.623	0.550	0.833	3.297	0.591	1.973	0.678	0.0580	0.0083	0.0071
0.138	1.650	0.650	0.840	3.352	0.610	1.837	0.657	0.0481	0.0119	0.0232
$x-Q^2$ bin 2										
P_T bin 1										
0.220	2.972	0.150	0.186	3.380	0.691	2.840	0.544	0.0283	0.0489	0.0049
0.241	2.755	0.250	0.222	3.103	0.590	2.518	0.674	0.0472	0.0077	0.0057
0.262	2.581	0.350	0.234	2.872	0.511	2.226	0.759	0.0455	0.0053	0.0034
0.273	2.442	0.450	0.246	2.731	0.464	2.002	0.805	0.0471	0.0057	0.0025
0.268	2.371	0.550	0.258	2.718	0.455	1.864	0.815	0.0444	0.0073	0.0049
0.254	2.407	0.650	0.269	2.819	0.484	1.773	0.790	-0.0023	0.0111	0.0011
P_T bin 2										
0.212	3.053	0.150	0.345	3.495	0.729	2.871	0.486	-0.0038	0.0012	0.0025
0.232	2.785	0.250	0.437	3.185	0.616	2.474	0.643	0.0633	0.0092	0.7893
0.251	2.587	0.350	0.476	2.941	0.531	2.164	0.739	0.0680	0.0058	0.0055
0.259	2.464	0.450	0.495	2.825	0.490	1.967	0.781	0.0730	0.0059	0.0043
0.255	2.428	0.550	0.509	2.828	0.488	1.846	0.785	0.0651	0.0071	0.0068
0.243	2.497	0.650	0.522	2.939	0.524	1.764	0.751	0.0373	0.0110	0.0170
0.232	2.632	0.750	0.521	3.092	0.576	1.682	0.698	0.0182	0.0350	1.8164
P_T bin 3										
0.216	2.982	0.250	0.689	3.421	0.701	2.476	0.529	0.0039	0.0876	0.0014
0.240	2.604	0.450	0.779	3.025	0.555	1.904	0.716	0.0889	0.0126	0.0447
0.236	2.597	0.550	0.798	3.044	0.560	1.807	0.711	0.1182	0.0170	0.1468
0.230	2.713	0.650	0.812	3.160	0.602	1.739	0.662	0.0402	0.0277	0.0222
$x-Q^2$ bin 3										
P_T bin 1										
0.289	4.016	0.150	0.165	3.280	0.708	2.748	0.511	0.0128	0.0418	0.0024
0.357	4.478	0.250	0.214	2.998	0.645	2.421	0.594	0.0485	0.0071	0.0056

0.388	4.485	0.350	0.227	2.831	0.597	2.189	0.653	0.0541	0.0064	0.0350
0.396	4.404	0.450	0.237	2.764	0.573	2.023	0.681	0.0380	0.0069	0.0143
0.386	4.322	0.550	0.249	2.790	0.575	1.903	0.681	0.0283	0.0105	0.0031
0.362	4.262	0.650	0.262	2.898	0.602	1.808	0.652	-0.0037	0.0184	0.0017
0.329	3.919	0.750	0.267	2.977	0.606	1.687	0.654	0.0030	0.0638	0.3020
P_T bin 2										
0.245	3.596	0.150	0.343	3.454	0.736	2.832	0.471	-0.0023	0.0009	0.0013
0.363	4.379	0.350	0.474	2.936	0.621	2.158	0.626	0.0601	0.0068	0.0229
0.372	4.331	0.450	0.493	2.870	0.598	1.999	0.653	0.0807	0.0079	0.0061
0.362	4.269	0.550	0.509	2.902	0.604	1.889	0.648	0.0651	0.0106	0.0074
0.340	4.202	0.650	0.524	3.004	0.630	1.792	0.618	0.0275	0.0193	0.0126
0.312	3.833	0.750	0.508	3.054	0.625	1.671	0.630	0.0135	0.0772	1.3421
P_T bin 3										
0.264	3.769	0.250	0.687	3.372	0.721	2.427	0.493	0.0836	0.1075	0.0104
0.317	4.153	0.350	0.740	3.136	0.668	2.105	0.567	0.0981	0.0196	0.0077
0.332	4.177	0.450	0.784	3.050	0.643	1.919	0.601	0.0591	0.0167	0.0059
0.323	4.139	0.550	0.808	3.092	0.653	1.829	0.587	0.1023	0.0235	0.0108
0.305	4.023	0.650	0.817	3.169	0.671	1.741	0.564	-0.0507	0.0415	0.6202
$x - Q^2$ bin 1										
z bin 1										
0.159	1.937	0.320	0.120	3.341	0.627	2.695	0.634	0.0301	0.0185	0.0042
0.158	1.892	0.320	0.200	3.320	0.617	2.660	0.645	0.0371	0.0154	0.0026
0.152	1.804	0.320	0.280	3.304	0.605	2.622	0.660	0.0421	0.0119	0.0128
0.146	1.756	0.320	0.360	3.336	0.612	2.618	0.652	0.0522	0.0115	0.0065
0.143	1.723	0.320	0.440	3.355	0.616	2.593	0.646	0.0467	0.0101	0.0041
0.140	1.706	0.320	0.520	3.372	0.621	2.558	0.642	0.0662	0.0115	0.0070
0.138	1.702	0.321	0.600	3.402	0.630	2.527	0.631	0.0375	0.0122	0.0067
0.134	1.700	0.321	0.680	3.444	0.644	2.499	0.615	0.0733	0.0157	0.5309
0.130	1.707	0.322	0.760	3.508	0.665	2.484	0.586	0.0240	0.0213	0.0181
z bin 2										
0.166	1.926	0.422	0.040	3.255	0.597	2.473	0.670	0.0029	0.0265	0.0030
0.163	1.843	0.419	0.120	3.217	0.581	2.442	0.684	0.0220	0.0124	0.0486
0.161	1.755	0.421	0.200	3.168	0.560	2.388	0.707	0.0286	0.0082	0.0127
0.156	1.693	0.422	0.280	3.167	0.555	2.365	0.715	0.0372	0.0071	0.0035
0.151	1.661	0.421	0.360	3.204	0.564	2.368	0.705	0.0371	0.0066	0.0037
0.147	1.643	0.421	0.440	3.234	0.573	2.358	0.695	0.0464	0.0064	0.0045
0.145	1.635	0.421	0.520	3.252	0.578	2.330	0.690	0.0536	0.0066	0.0056
0.143	1.632	0.423	0.600	3.272	0.584	2.295	0.684	0.0583	0.0068	0.0065
0.141	1.634	0.425	0.680	3.294	0.591	2.253	0.677	0.0638	0.0082	0.0112
0.139	1.640	0.427	0.760	3.325	0.601	2.212	0.666	0.0499	0.0092	0.0231
z bin 3										
0.170	1.733	0.600	0.040	3.057	0.524	2.017	0.745	-0.0198	0.0506	0.0185
0.166	1.655	0.598	0.120	3.023	0.509	1.993	0.762	0.0203	0.0208	0.0633
0.162	1.598	0.595	0.200	3.021	0.504	1.985	0.768	0.0221	0.0137	0.0061
0.157	1.579	0.593	0.280	3.054	0.512	1.992	0.760	0.0431	0.0120	0.0033
0.153	1.569	0.592	0.360	3.096	0.525	1.999	0.748	0.0200	0.0095	0.0129
0.149	1.569	0.591	0.440	3.138	0.538	1.999	0.735	0.0367	0.0100	0.0029
0.145	1.587	0.590	0.600	3.206	0.560	1.970	0.713	0.0328	0.0099	0.3517
0.143	1.604	0.590	0.680	3.245	0.573	1.950	0.699	0.0390	0.0107	0.0057
0.141	1.625	0.590	0.760	3.289	0.588	1.928	0.682	0.0436	0.0129	0.0201
$x - Q^2$ bin 2										
z bin 1										

0.256	2.689	0.321	0.040	2.966	0.546	2.378	0.724	-0.0094	0.0391	0.0108
0.256	2.640	0.321	0.120	2.940	0.535	2.347	0.735	0.0044	0.0197	0.0010
0.256	2.627	0.321	0.200	2.933	0.532	2.321	0.739	0.0507	0.0155	0.0056
0.254	2.636	0.321	0.280	2.950	0.536	2.308	0.734	0.0542	0.0143	0.0049
0.247	2.622	0.321	0.440	2.988	0.547	2.255	0.724	0.0622	0.0153	0.0213
0.234	2.724	0.322	0.600	3.135	0.596	2.271	0.670	0.0888	0.0231	0.0091
0.227	2.808	0.323	0.680	3.233	0.630	2.292	0.627	0.0787	0.0367	0.0113
0.221	2.897	0.325	0.760	3.334	0.668	2.311	0.576	-0.0417	0.0606	0.0191
z bin 2										
0.276	2.525	0.434	0.040	2.760	0.477	2.074	0.791	-0.0089	0.0187	0.0082
0.273	2.489	0.430	0.120	2.756	0.474	2.067	0.795	0.0169	0.0092	0.0027
0.271	2.463	0.429	0.200	2.756	0.472	2.052	0.797	0.0479	0.0076	0.0061
0.268	2.463	0.429	0.280	2.768	0.475	2.037	0.794	0.0502	0.0065	0.0039
0.264	2.470	0.428	0.360	2.795	0.483	2.027	0.787	0.0577	0.0071	0.0039
0.259	2.479	0.427	0.440	2.830	0.493	2.016	0.778	0.0692	0.0078	0.0058
0.254	2.498	0.427	0.520	2.872	0.505	2.001	0.767	0.0609	0.0084	0.0062
0.249	2.530	0.429	0.600	2.924	0.522	1.986	0.751	0.0698	0.0106	0.0087
0.243	2.572	0.432	0.680	2.982	0.541	1.969	0.732	0.0631	0.0125	0.0138
0.239	2.621	0.436	0.760	3.042	0.561	1.946	0.710	0.0749	0.0179	0.0345
z bin 3										
0.266	2.364	0.589	0.040	2.724	0.456	1.821	0.815	0.0095	0.0268	0.0090
0.265	2.380	0.588	0.120	2.738	0.460	1.825	0.811	0.0131	0.0152	0.0029
0.262	2.382	0.585	0.200	2.754	0.465	1.826	0.806	0.0321	0.0116	0.0019
0.259	2.385	0.584	0.280	2.774	0.471	1.822	0.802	0.0560	0.0116	0.0042
0.256	2.407	0.583	0.360	2.806	0.481	1.818	0.792	0.0480	0.0109	0.0033
0.252	2.435	0.583	0.440	2.847	0.494	1.814	0.780	0.0682	0.0120	0.0058
0.247	2.469	0.582	0.520	2.896	0.510	1.809	0.765	0.0519	0.0130	0.0051
0.243	2.508	0.582	0.600	2.949	0.527	1.800	0.748	0.0685	0.0153	0.0075
0.238	2.563	0.582	0.680	3.010	0.548	1.792	0.726	0.0779	0.0225	0.0124
0.234	2.627	0.581	0.760	3.075	0.571	1.782	0.700	0.0683	0.0266	0.0314
$x - Q^2$ bin 3										
z bin 1										
0.385	4.514	0.320	0.040	2.855	0.606	2.277	0.641	0.0561	0.0396	0.0517
0.385	4.527	0.320	0.120	2.856	0.607	2.268	0.641	0.0678	0.0206	0.0536
0.384	4.508	0.320	0.200	2.855	0.605	2.248	0.643	0.0405	0.0163	0.0058
0.380	4.493	0.321	0.280	2.876	0.610	2.237	0.638	0.0442	0.0153	0.0033
0.370	4.452	0.321	0.360	2.916	0.619	2.234	0.627	0.0530	0.0155	0.0035
0.359	4.403	0.321	0.440	2.964	0.630	2.230	0.615	0.0539	0.0184	0.0051
0.344	4.333	0.321	0.520	3.025	0.644	2.230	0.598	0.0669	0.0209	0.0066
0.326	4.251	0.322	0.600	3.107	0.664	2.242	0.573	0.0476	0.0281	0.0050
0.306	4.113	0.323	0.680	3.196	0.684	2.255	0.545	0.0644	0.0458	0.0092
0.285	3.961	0.325	0.760	3.290	0.706	2.267	0.516	0.0811	0.0740	0.0371
z bin 2										
0.402	4.456	0.421	0.040	2.752	0.573	2.073	0.681	0.0306	0.0211	0.0280
0.398	4.445	0.420	0.120	2.764	0.576	2.076	0.678	0.0241	0.0114	0.0034
0.395	4.435	0.419	0.200	2.777	0.579	2.071	0.675	0.0356	0.0092	0.0022
0.390	4.411	0.419	0.280	2.799	0.583	2.065	0.670	0.0462	0.0085	0.0037
0.383	4.395	0.419	0.360	2.830	0.591	2.057	0.661	0.0521	0.0090	0.0049
0.374	4.361	0.419	0.440	2.869	0.600	2.048	0.651	0.0638	0.0095	0.0063
0.364	4.327	0.420	0.520	2.909	0.609	2.033	0.640	0.0694	0.0117	0.0070
0.354	4.287	0.421	0.600	2.957	0.621	2.015	0.627	0.0770	0.0146	0.0104
0.342	4.241	0.424	0.680	3.010	0.634	1.995	0.611	0.0776	0.0169	0.0156

0.330	4.190	0.428	0.760	3.063	0.647	1.966	0.595	0.0797	0.0248	0.0529
z bin 3										
0.383	4.297	0.585	0.040	2.799	0.576	1.862	0.681	0.0123	0.0388	0.0113
0.379	4.290	0.585	0.120	2.814	0.580	1.866	0.676	0.0257	0.0238	0.0035
0.377	4.300	0.584	0.200	2.827	0.584	1.863	0.671	0.0323	0.0185	0.0018
0.372	4.294	0.584	0.280	2.852	0.591	1.861	0.664	0.0433	0.0185	0.0038
0.365	4.277	0.584	0.360	2.885	0.599	1.859	0.654	0.0441	0.0182	0.0032
0.357	4.256	0.583	0.440	2.923	0.609	1.854	0.642	0.0666	0.0201	0.0059
0.349	4.226	0.584	0.520	2.964	0.620	1.844	0.630	0.0725	0.0226	0.0172
0.339	4.195	0.583	0.600	3.014	0.632	1.835	0.614	0.0702	0.0261	0.0084
0.329	4.165	0.584	0.680	3.060	0.645	1.817	0.599	0.0793	0.0324	0.0176
0.318	4.117	0.582	0.760	3.112	0.658	1.802	0.581	0.0879	0.0417	0.0410

Table 6: Final result of the 4-dimensional study with z or P_T as the main bin variable with 3 $x_B - Q^2$ bins and 3 P_T or z bins for K^+. The columns $<x_B>$, $<Q^2>$, $<z>$, $<P_T>$, $<W>$, $<y>$, $<M_X>$ and $<\varepsilon>$ provide the mean value of the stated variable or the bin middle if it is the main bin variable. The columns Stat. and Sys. provide the statistical and systematic uncertainty of $F_{LU}^{\sin\phi}/F_{UU}$

ii K^-

$<x_B>$	$<Q^2>$ [GeV^2]	$<z>$	$<P_T>$ [GeV]	$<W>$ [GeV]	$<y>$	$<M_X>$ [GeV]	$<\varepsilon>$	$F_{LU}^{\sin\phi}/F_{UU}$	Stat. unc.	Sys. unc.
$x-Q^2$ bin 1										
P_T bin 1										
0.118	1.637	0.150	0.252	3.625	0.701	3.050	0.536	-0.0418	0.0588	0.0181
0.134	1.704	0.250	0.341	3.460	0.650	2.799	0.605	-0.0152	0.0133	0.0139
0.142	1.599	0.350	0.345	3.255	0.577	2.522	0.695	-0.0083	0.0138	0.0065
0.142	1.603	0.450	0.359	3.270	0.584	2.378	0.677	0.0312	0.0167	0.0062
0.135	1.647	0.550	0.370	3.399	0.627	2.286	0.629	0.0436	0.0162	0.0127
0.135	1.622	0.650	0.376	3.362	0.611	2.059	0.657	0.0522	0.0195	0.0569
0.139	1.535	0.750	0.354	3.228	0.560	1.770	0.723	0.0391	0.0316	0.0323
0.118	1.310	0.850	0.261	3.261	0.557	1.624	0.732	0.1336	0.2852	0.7271
P_T bin 2										
0.132	1.743	0.250	0.584	3.518	0.675	2.682	0.573	-0.0367	0.0195	0.0242
0.140	1.689	0.350	0.654	3.357	0.617	2.410	0.645	-0.0348	0.0148	0.0270
0.138	1.712	0.450	0.716	3.408	0.636	2.283	0.618	-0.0529	0.0153	0.0080
0.133	1.721	0.550	0.744	3.481	0.659	2.159	0.592	-0.0612	0.0148	0.0192
0.134	1.670	0.650	0.741	3.426	0.636	1.943	0.626	-0.0760	0.0206	0.0828
0.133	1.546	0.750	0.695	3.308	0.587	1.720	0.692	0.0305	0.0573	0.5358
$x-Q^2$ bin 2										
P_T bin 1										
0.219	2.969	0.150	0.208	3.384	0.686	2.833	0.550	0.0033	0.0019	0.0004
0.237	2.765	0.250	0.332	3.138	0.601	2.497	0.661	0.0116	0.0228	0.0943
0.260	2.546	0.350	0.338	2.863	0.508	2.178	0.765	0.0094	0.0193	0.0100
0.267	2.433	0.450	0.337	2.762	0.474	1.994	0.793	-0.0024	0.0242	0.0016
0.251	2.552	0.550	0.350	2.924	0.527	1.971	0.738	-0.0684	0.0289	0.0221
0.243	2.666	0.650	0.362	3.035	0.562	1.863	0.708	0.0306	0.0306	0.0339
0.237	2.613	0.750	0.343	3.049	0.561	1.700	0.716	0.1525	0.0635	0.1160
P_T bin 2										
0.225	2.861	0.250	0.563	3.282	0.653	2.465	0.596	-0.0619	0.0529	0.1632
0.241	2.651	0.350	0.617	3.044	0.568	2.145	0.701	-0.0549	0.0259	0.0455
0.242	2.611	0.450	0.676	3.016	0.558	1.989	0.707	-0.0489	0.0284	0.0069
0.234	2.743	0.550	0.718	3.146	0.603	1.937	0.657	-0.1019	0.0321	0.0291
0.231	2.760	0.650	0.718	3.172	0.610	1.801	0.653	-0.0800	0.0455	0.0930
0.234	2.669	0.750	0.756	3.106	0.584	1.667	0.686	-0.0907	0.1737	0.0914
$x-Q^2$ bin 3										
P_T bin 1										
0.342	4.351	0.250	0.314	3.047	0.653	2.418	0.584	0.0086	0.0251	0.6247
0.384	4.358	0.350	0.329	2.815	0.587	2.137	0.666	0.0085	0.0265	0.0678
0.377	4.435	0.450	0.344	2.879	0.610	2.072	0.635	-0.0072	0.0323	0.0452
0.360	4.518	0.550	0.353	2.988	0.643	2.000	0.596	0.0270	0.0313	0.0185
0.349	4.323	0.650	0.352	2.989	0.631	1.838	0.616	0.0738	0.0463	0.0817
0.325	3.869	0.750	0.325	2.981	0.604	1.680	0.658	0.0313	0.1492	0.0245
P_T bin 2										
0.296	4.016	0.250	0.560	3.229	0.693	2.414	0.533	-0.0253	0.0569	0.0189
0.337	4.203	0.350	0.621	3.031	0.643	2.126	0.599	-0.0421	0.0317	0.0318
0.336	4.318	0.450	0.690	3.073	0.660	2.020	0.574	-0.0645	0.0342	0.0225
0.328	4.314	0.550	0.717	3.121	0.673	1.912	0.559	-0.1152	0.0362	0.0354
0.318	4.129	0.650	0.714	3.122	0.662	1.778	0.576	0.0180	0.0709	0.0845
0.336	3.957	0.750	0.837	2.968	0.609	1.647	0.647	-0.1914	0.0741	0.1462
$x-Q^2$ bin 1										

z bin 1										
0.142	1.507	0.396	0.120	3.177	0.545	2.446	0.727	0.0207	0.0475	0.0163
0.142	1.539	0.390	0.200	3.208	0.558	2.465	0.713	0.0102	0.0352	0.0032
0.142	1.587	0.390	0.280	3.252	0.576	2.477	0.692	0.0059	0.0266	0.0027
0.142	1.620	0.390	0.360	3.283	0.588	2.472	0.677	-0.0005	0.0225	0.0003
0.142	1.643	0.390	0.440	3.297	0.595	2.448	0.669	0.0030	0.0204	0.0021
0.142	1.665	0.392	0.520	3.323	0.605	2.422	0.657	-0.0116	0.0201	0.0053
0.141	1.686	0.395	0.600	3.350	0.615	2.389	0.646	-0.0423	0.0213	0.0215
0.140	1.701	0.397	0.680	3.378	0.625	2.352	0.634	-0.0435	0.0239	0.0196
0.138	1.731	0.404	0.760	3.420	0.641	2.317	0.615	-0.0285	0.0297	0.0072
z bin 2										
0.140	1.612	0.601	0.040	3.289	0.585	2.153	0.685	-0.0432	0.0660	0.0208
0.137	1.592	0.601	0.120	3.318	0.594	2.166	0.673	0.0313	0.0431	0.0185
0.137	1.616	0.599	0.200	3.345	0.606	2.177	0.659	0.0651	0.0359	0.0202
0.136	1.643	0.596	0.280	3.380	0.619	2.191	0.642	0.0689	0.0310	0.0210
0.134	1.647	0.592	0.360	3.405	0.628	2.195	0.631	0.0801	0.0285	0.0624
0.133	1.647	0.593	0.440	3.416	0.632	2.176	0.627	0.0298	0.0253	0.0112
0.134	1.670	0.591	0.520	3.426	0.637	2.154	0.621	-0.0015	0.0240	0.0013
0.134	1.687	0.590	0.600	3.434	0.641	2.126	0.616	-0.0190	0.0255	0.0100
0.134	1.701	0.589	0.680	3.449	0.647	2.096	0.609	-0.0161	0.0284	0.0073
0.133	1.701	0.588	0.760	3.467	0.653	2.061	0.602	-0.0782	0.0332	0.0265
$x - Q^2$ bin 2										
z bin 1										
0.278	2.450	0.410	0.040	2.704	0.455	2.059	0.813	0.0270	0.1294	0.0127
0.275	2.463	0.404	0.120	2.727	0.463	2.076	0.807	0.0735	0.0615	0.0412
0.271	2.476	0.398	0.200	2.762	0.475	2.093	0.795	0.0248	0.0438	0.0140
0.265	2.487	0.393	0.280	2.801	0.487	2.104	0.782	0.0405	0.0367	0.0132
0.260	2.501	0.391	0.360	2.840	0.500	2.103	0.770	-0.0049	0.0343	0.0069
0.254	2.519	0.391	0.440	2.883	0.513	2.096	0.757	-0.0653	0.0315	0.0300
0.250	2.560	0.392	0.520	2.936	0.530	2.088	0.740	-0.0149	0.0318	0.0111
0.244	2.604	0.395	0.600	2.997	0.551	2.079	0.718	-0.0805	0.0380	0.0519
0.237	2.643	0.401	0.680	3.065	0.574	2.069	0.693	-0.0516	0.0431	0.0419
0.233	2.707	0.412	0.760	3.131	0.597	2.046	0.665	-0.0499	0.0588	0.0273
z bin 2										
0.256	2.522	0.595	0.040	2.871	0.507	1.898	0.762	0.0408	0.1054	0.0249
0.255	2.543	0.595	0.120	2.891	0.515	1.905	0.754	-0.0407	0.0691	0.0283
0.251	2.571	0.595	0.200	2.936	0.530	1.920	0.738	-0.0040	0.0583	0.0047
0.247	2.602	0.593	0.280	2.977	0.544	1.935	0.723	0.0295	0.0554	0.0087
0.246	2.624	0.592	0.360	2.999	0.552	1.926	0.715	0.0337	0.0525	0.0181
0.242	2.659	0.589	0.440	3.044	0.567	1.932	0.699	-0.0128	0.0488	0.0142
0.238	2.685	0.589	0.520	3.083	0.580	1.923	0.685	-0.0367	0.0499	0.0168
0.235	2.717	0.589	0.600	3.121	0.593	1.907	0.671	-0.0278	0.0533	0.0257
0.233	2.752	0.588	0.680	3.157	0.606	1.885	0.656	-0.0178	0.0623	0.0078
0.230	2.781	0.587	0.760	3.192	0.618	1.857	0.641	-0.0534	0.0757	0.0713
$x - Q^2$ bin 3										
z bin 1										
0.412	4.481	0.399	0.040	2.711	0.563	2.062	0.692	0.0772	0.1339	0.0493
0.407	4.459	0.394	0.120	2.726	0.566	2.072	0.688	0.0359	0.0717	0.0425
0.396	4.427	0.392	0.200	2.776	0.580	2.098	0.673	0.0042	0.0572	0.0014
0.384	4.403	0.389	0.280	2.827	0.594	2.117	0.656	0.0006	0.0492	0.0004
0.374	4.373	0.389	0.360	2.873	0.605	2.121	0.643	-0.0060	0.0434	0.0087
0.363	4.341	0.391	0.440	2.925	0.618	2.121	0.627	-0.0293	0.0398	0.1355

0.352	4.307	0.392	0.520	2.973	0.631	2.111	0.613	-0.0184	0.0438	0.0091
0.343	4.297	0.401	0.600	3.024	0.645	2.090	0.596	-0.0678	0.0448	0.0349
0.331	4.247	0.408	0.680	3.079	0.659	2.070	0.577	-0.0277	0.0528	0.0129
0.322	4.224	0.421	0.760	3.130	0.673	2.033	0.559	-0.0127	0.0678	0.0062
z bin 2										
0.370	4.492	0.594	0.040	2.918	0.619	1.924	0.628	0.0231	0.1526	0.0139
0.366	4.463	0.595	0.120	2.936	0.624	1.927	0.623	0.0886	0.0792	0.0649
0.361	4.472	0.592	0.200	2.968	0.634	1.942	0.610	0.0382	0.0707	0.0567
0.356	4.432	0.590	0.280	2.984	0.637	1.938	0.607	0.0683	0.0645	0.0213
0.351	4.416	0.587	0.360	3.007	0.643	1.935	0.598	0.0005	0.0662	0.0007
0.347	4.414	0.585	0.440	3.029	0.649	1.922	0.591	0.0021	0.0640	0.0100
0.339	4.350	0.581	0.520	3.063	0.656	1.916	0.581	0.0019	0.0636	0.0009
0.333	4.319	0.580	0.600	3.085	0.661	1.890	0.576	0.0000	0.0010	0.0000
0.326	4.272	0.579	0.680	3.117	0.669	1.866	0.566	0.0259	0.0798	0.0116

Table 7: Final result of the 4-dimensional study with z or P_T as the main bin variable with 3 $x_B - Q^2$ bins and 2 P_T or z bins for K^-. The columns $<x_B>$, $<Q^2>$, $<z>$, $<P_T>$, $<W>$, $<y>$, $<M_X>$ and $<\varepsilon>$ provide the mean value of the stated variable or the bin middle if it is the main bin variable. The columns Stat. and Sys. provide the statistical and systematic uncertainty of $F_{LU}^{\sin\phi}/F_{UU}$

Bibliography

[1] Cliff Burgess and Guy Moore, The Standard Model (2006)

[2] Jenny Regina, Time for Hyperons Development of Software Tools for Reconstructing Hyperons at PANDA and HADES (2022)

[3] N. Brambilla et al., hep-ph/0412158 (2004).

[4] Stefan Diehl, Optimization of the Influence of Longitudinal and Lateral Non-Uniformity on the Performance of an Electromagnetic Calorimeter (2015)

[5] EMC, Phys. Lett. B, 206, 364-370 (1988)

[6] A. W. Thomas, Phys.Rev.Lett., 101:102003 (2008)

[7] A. Bacchetta, M. Diehl, K. Goeke et al., JHEP 0702, 093 (2007)

[8] Barbara Pasquini, Simone Rodini, Phys. Lett. B, 788, 414-424 (2019)

[9] PANDA collaboration, Physics Performance Report for PANDA: Strong Interaction Studies with Antiprotons (2009)

[10] M. Kotulla et al., Progress Report, PANDA, Strong Interaction Studies with Antiprotons (2005)

[11] A.Lundborg, T.Barnes, U.Wiedner, Phys.Rev. D73, 096003 (2006)

[12] Nora Brambilla, Wai Kin Lai, Jorge Segovia, Jaume Tarrús Castellà, Antonio Vairo, Phys. Rev. D 99, 014017 (2019)

[13] Richard F. Lebed, Ryan E. Mitchell, Eric S. Swanson, Progress in Particle and Nuclear Physics 93, 143 (2017)

[14] R.A. Briceno, T.D. Cohen, S. Coito, J.J. Dudek, E. Eichten, C.S. Fischer et al., Chinese Phys. C 40 042001 (2015)

[15] Agnes Lundborg, AIP Conference Proceedings 717, 431 (2004)

[16] Philip R.Page, Physics Letters B 402 (1997)

[17] F.E. Close, I. Dunietz, P.R. Page, S. Veseli, H. Yamamoto, arXiv:hep-ph/9708265 (1997)

[18] The PANDA Collaboration, arXiv:0907.0169 (2009)

[19] Martin Johannes Galuska, Master Thesis:Simulation of X(3872) Decays Using the PandaRoot Framework (2011)

[20] PANDA wiki page: https://panda-wiki.gsi.de/Main/WebHome [1]

[21] Stefano Spataro, J. Phys.: Conf. Ser. 331 032031 (2011)

[22] Stefano Spataro, J. Phys.: Conf. Ser. 396 022048 (2012)

[23] https://hpc.gsi.de/virgo/platform/software.html

[24] Marcel Tiemens, PhD Thesis: Online cluster-finding algorithms for the PANDA electromagnetic calorimeter

[25] Virgo Cluster webpage: https://hpc.gsi.de/virgo/preface.html [1]

[26] David J. Lange, NIM A, 462, 152-155 (2001)

[27] A. Capella, U. Sukhatme, C.-I. Tan and J. Tran Thanh Van, Physics Reports 236, 225-329 (1994)

[28] Binsong Ma and the PANDA Collaboration J. Phys.: Conf. Ser. 503 012008 (2014)

[29] Private communication with Tobias Stockmanns

[30] Private communication with Ben Salisbury

[31] T. A. Armstrong et al., Phys. Rev. Lett. 69, 2337 (1992)

[32] M. Ablikim, et al., arXiv:hep-ex/0503030 (2005)

[33] R.A Briere, et al, arXiv:hep-ex/0505101 (2005)

[34] P.A. Zyla et al. (Particle Data Group), Prog. Theor. Exp. Phys. 2020, 083C01 (2020)

[35] S. Diehl et al., Phys. Rev. Lett. 128, 062005 (2022)

[36] Robert Hofstadter and Robert W. McAllister, Phys. Rev. 98, 217 (1955)

[37] M. Gell-Mann, Phys. Lett. Vol. 8 Iss. 3 Pages 214-215 (1964)

[38] G. Zweig, CERN-TH 412, 80 p. (1964)

[39] J. D. Bjorken and E. A. Paschos, Phys. Rev. 185, 1975 (1969).

[40] R. P. Feynman, Phys. Rev. Lett. 23, 1415 (1969)

[41] E. P. Wigner, Phys. Rev. 40, 749 (1932).

[42] Moyal, J. E., CUP 45 (1): 99–124. (1949)

[43] W. Gohn, Probing the Proton's Quark Dynamics in Semi-inclusive Pion Electroproduction (2012)

[44] W. Gohn, H. Avakian, K. Joo, M. Ungaro et al. (CLAS collab.), Phys. Rev. D 89, 072011 (2014).

[45] A. Bacchetta, U. D'Alesio, M. Diehl, and C. A. Miller, Phys. Rev. D 70, 117504 (2004).

[46] S. Diehl, SciPost Phys. Proc. 8, 149 (2022)

[47] J. Levelt and P. J. Mulders, Phys. Lett. B 338, 357 (1994).

[48] Andreas Metz and Anselm Vossen, arXiv:1607.02521 (2016)

[49] A. V. Efremov, K. Goeke, and P. Schweitzer, Phys. Rev. D 67, 114014 (2003).

[50] Y. Ohnishi and M. Wakamatsu, Phys. Rev. D 69, 114002 (2004).

[51] A. Afanasev and C. E. Carlson, arXiv:hep-ph/0308163 (2003).

[52] F. Yuan, Phys. Lett. B 589, 28 (2004).

[53] A. Bacchetta, P. J. Mulders, and F. Pijlman, Phys. Lett. B 595, 309 (2004).

[54] A. Metz and M. Schlegel, Eur. Phys. J. A 22, 489 (2004).

[55] F. Pijlman, arXiv:hep-ph/0604226 (2006).

[56] Z. Lu and I. Schmidt, Phys. Lett. B 712, 451 (2012).

[57] W. Mao and Z. Lu, Phys. Rev. D 87, 014012 (2013).

[58] D. Boer and P. J. Mulders, Phys. Rev. D 57, 5780 (1998).

[59] D. W. Sivers, Phys. Rev. D 43, 261 (1991).

[60] R. Jaffe and X.-D. Ji, Nucl. Phys. B375, 527 (1992).

[61] M. Burkardt, AIP Conf. Proc. 1155, 26 (2009).

[62] J. C. Collins, Nucl. Phys. B396, 161 (1993).

[63] D. Riser, https://doi.org/10.2172/1575058 (2019)

[64] A. Airapetian et al. (HERMES collab.), arXiv:1903.08544 (2019)

[65] Christoph W. Leemann, David R. Douglas, Geoffrey A. Krafft, Annu. Rev. Nucl. Part., Vol. 51:413-450 (2001)

[66] J. Arrington, M. Battaglieri, A. Boehnlein et al., arXiv:2112.00060 (2021)

[67] V.D. Burkert et al. (CLAS collab.), NIM A, 959, 163419 (2020).

[68] N. Baltzell et al. (CLAS collab.), NIM A, 959, 163421 (2020).

[69] Hashir Rashad, Mohamed Nuhman. Deeply Virtual Compton Scattering at Hall A, Jefferson Lab. United States. https://doi.org/10.25777/z7th-my96

[70] Sangbaek Lee, Measurement of the Deeply Virtual Compton Scattering Cross Section from the Proton at 10.6 GeV using the CLAS12 Detector (2022)

[71] CLAS12 webpage: https://www.jlab.org/physics/hall-b/clas12 [1]

[72] Florian Hauenstein, Update on DC Calibrations, CLAS Collaboration meeting (2022 Nov.): https://indico.jlab.org/event/597/contributions/10939/attachments/8601/12365/DCupdate_clasmeeting_Nov_03_2022.pdf [1]

[73] M. Contalbrigo, The CLAS12 RICH Reconstruction, Workshop on Kaons with CLAS12 (2022 Dec.): https://agenda.infn.it/event/33338/contributions/185664/attachments/100361/139631/Contalbrigo_RICH_reco.pdf [1]

[74] Rene Brun, Fons Rademakers NIM A, 389, Pages 81-86 (1997)

[75] S. Boyarinov et al. (CLAS collab.), NIM A, 966, 163698 (2020).

[76] CLAS12 RG-A wiki: https://clasweb.jlab.org/wiki/index.php/Run_Group_A#tab=Fiducial_Studies [1]

[77] V. Ziegler, N.A. Baltzell, F. Bossù, NIM A, 959, 163472 (2020)

[78] V Gyurjyan et al., J. Phys.: Conf. Ser. 331 032013 (2011)

[79] S. Diehl et al., Phys. Rev. Lett. 125, 182001 (2020)

[80] A. Hoecker et al., arXiv:physics/0703039v4 (2007)

[81] CLAS12 RICH Technical Design Report (2013): http://www.lnf.infn.it/~mirazita/RICH/RICH_TDR.pdf [1]

[82] CLASDIS github: https://github.com/JeffersonLab/clasdis [1]

[83] L. Mankiewicz, A. Schäfer and M. Veltri, Comput. Phys. Commun. Vol. 71, Iss. 3, P. 305-318 (1992)

[84] Torbjörn Sjöstrand, arXiv:1907.09874 (2019)

[85] CLAS SIDIS group wiki: https://clasweb.jlab.org/wiki/index.php/SIDIS_Analysis_Group#tab=Simulations [1]

[86] M. Ungaro et al., Nucl. Instrum. Meth. A 959, 163422444 (2020).

[87] S. Agostinelli et al. (GEANT4), Nucl. Instrum. Meth. A 442 506, 250 (2003).

[88] Private communication with F. X. Girod

[89] Private communication with Marco Contalbrigo

[90] Gabriel Pecar, Pion/Kaon and Kaon/Pion mis-identifications studies using RICH, Workshop on Kaons with CLAS12 (2022 Dec.): https://agenda.infn.it/event/33338/contributions/186292/attachments/100420/139715/Workshop-pi-ki-Gabe.pdf [1]

[91] Screen-shots taken of JLab electronic logbook entries made by Nathan Baltzell

[92] W. Mao and Z. Lu, Eur. Phys. J. C 73, 2557 (2013).

[93] W. Mao and Z. Lu, Eur. Phys. J. C 74, 2910 (2014).

[94] M. Anselmino et al., Nucl. Phys. B, Proc. Suppl. 191, 98 (2009).

[95] P. Schweitzer, Phys. Rev. D 67, 114010 (2003).

[1] Accessed successfully on April 18. 2023.

Erklärung der Urheberschaft

Ich erkläre: Ich habe die vorgelegte Dissertation selbstständig und ohne unerlaubte fremde Hilfe und nur mit den Hilfen angefrtigt, die ich in der Dissertation angegeben habe.
Alle Textstellen, die wörtlich oder sinngemäß aus veröffentlichen Schriften entnommen sind, und alle Angaben, die auf mündlichen Auskünften beruhen, sind als solche kenntlich gemacht.
Bei den von mir durchgeführten und in der Dissertation erwähnten Untersuchungen habe ich die Grundsätze guter wissenschaftlicher Praxis, wie sie in der "Satzung der Justus-LiebigUniversität Gießen zur Sicherung guter wissenschaftlicher Praxis" niedergelgt sind, eingehalten.

Áron Kripkó

Gießen, Mai 2023