Multiplicities of hadrons in deep-inelastic scattering of muons on nucleons at COMPASS

Nicolas Pierre

NNT : 2019SACLS183

Multiplicities of Hadrons in Deep-Inelastic Scattering of Muons on Nucleons at COMPASS

Thèse de doctorat de Johannes Gutenberg Universität Mainz et de l'Université Paris-Saclay, préparée à l'Université Paris-Sud

École doctorale n°576 : Particules, Hadrons, Énergie, Noyau, Instrumentation, Imagerie, Cosmos et Simulation (PHENIICS)
Spécialité de doctorat: Physique Hadronique

Thèse présentée et soutenue à Saclay, le 3 juillet 2019, par

M. Nicolas Pierre

Composition du Jury :

Dr. Franck Sabatié Directeur de recherche, Université Paris-Saclay (– CEA/IRFU/DPhN)	Président
Prof. Dr. Friedrich Klein Professeur, Univesität Bonn (– PI)	Rapporteur
Prof. Andrea Bressan Professeur, Università degli studi di Trieste	Rapporteur
Prof. Dr. Hubert Spiesberger Professeur, Johannes Gutenberg Universität Mainz (– THEP)	Examinateur
Dr. Fabienne Kunne Directeur de recherche, Université Paris-Saclay (– CEA/IRFU/DPhN)	Directrice de thèse
Prof. Dr. Eva-Maria Kabuß Professeur, Johannes Gutenberg Universität Mainz (– KPH)	Co-Directrice de thèse

Thèse de doctorat

MULTIPLICITIES OF HADRONS IN DEEP-INELASTIC SCATTERING OF MUONS ON NUCLEONS AT COMPASS

Dissertation
zur Erlangung des Grades
"Doktor der Naturwissenschaften"

am Fachbereich 08 - Physik, Mathematik, Informatik
DER JOHANNES GUTENBERG-UNIVERSITÄT
in Mainz

und

Dissertation
zur Erlangung des Grades
"Docteur en Physique"

am UNIVERSITÉ PARIS-SACLAY
in Gif-Sur-Yvette

NICOLAS PIERRE
geboren in Strasbourg

Paris, 10. Juni 2019

Nicolas Pierre:
Multiplicities of hadrons in Deep-Inelastic Scattering of muons on nucleons at COMPASS, Dissertation zur Erlangung des Grades "Doktor der Naturwissenschaften" am Fachbereich 08 - Physik, Mathematik und Informatik der Johannes-Gutenberg-Universität
Mainz **D77**
und
Dissertation zur Erlangung des Grades "Docteur en Physique" am Université Paris-Saclay
Gif-Sur-Yvette
July 2019

BERICHTERSTATTER:
Prof. Dr. Eva-Maria Kabuß
Dr. Fabienne Kunne

TAG DER MÜNDLICHEN PRÜFUNG: 3. Juli 2019

ABSTRACT

The spin structure of the nucleon is studied at the COMPASS experiment. This subject is of special interest since the surprising finding of the European muon collaboration (EMC) that the contribution from the quark spins to the nucleon spin is rather small. This finding started a series of experiments looking to measure various contribution of to the nucleon spin. The question of the polarization of the sea quark is an important topic in the hadronic physics, especially for the strange quark polarization.

In order to better describe the quark polarization, a precise knowledge of the quarks Fragmentation Functions (FFs) into hadrons, which describes the final state hadronisation of quark q into hadron h, is mandatory. The FFs can be extracted from hadron multiplicities produced in Deep Inelastic Scattering (DIS). At the COMPASS experiment, a 160 GeV/c muon beam is scattered off a fixed pure proton target (lH_2).

This thesis presents the measurement of charged hadron, identified pion, kaon and proton multiplicities from DIS data collected in 2016. It also details the improvements of the analysis with the DJANGOH event generator, which is used to better describe the inclusive and semi-inclusive radiative corrections in DIS, which are then used as correction factors to the multiplicities, and the adaptation of DJANGOH to the COMPASS Monte-Carlo chain. The radiative correction factors to the multiplicities are computed in a three dimensional binning in x, y and z. For the first time a solid determination of radiative effects dependent on the z variable has been achieved.

The data cover a large kinematic range : $Q^2 > 1$ (GeV/c)^2, $y \in$ [0.1,0.7], $x \in$ [0.004,0.4], $W \in$ [5,17] GeV/c^2 and $z \in$ [0.2,0.85]. The charged hadron multiplicities are obtained in a 3-dimensional (x, y, z) binning yielding a total of 540 bins. Several corrections factors that are applied to the multiplicities are discussed : acceptance correction, vector meson corrections, RICH efficiency unfolding, electron contamination correction and radiative corrections. The multiplicities, which represent about 1800 data points in total, provide an important input for global QCD fit of world data at NLO, aiming at FFs determination and complete the previous results for the charged hadron multiplicities using the data taken in 2006 of muon scattering off an isoscalar target (^{6}LiD).

The quark FFs into kaons are particularly important as they are necessary to better constrain the strange quark polarization. The K^+ and K^- multiplicities were used to extract the favoured D^K_{fav}, unfavoured D^K_{unf} and strange D^K_s quark FFs with a fit at LO. The result of the fit points out that there is only a weak sensitivity to the strange quark of these measurements. The fit gives too much contribution to the favoured and unfavoured fragmentation functions at the expense of the strange fragmentation function. This needs a more detailed investigation and also a fit in NLO perturbative QCD.

RÉSUMÉ

Un des buts de la collaboration COMPASS est l'étude de la structure de spin du nucléon. Ce sujet présente un intérêt particulier depuis les découvertes surprenantes de la European muon collaboration (EMC) selon lesquelles la contribution des spins de quarks au spin de nucléons est plutôt faible. Cette découverte a lancé la recherche des diverses contributions au spin du nucléon. La question de la polarisation des quarks de la mer est un sujet capital en physique hadronique, en particulier pour la polarisation du quark étrange. Afin de mieux contraindre la polarisation des quarks, une connaissance précise des fonctions de fragmentation (FFs), qui expriment l'hadronisation d'un quark q en un hadron h dans l'état final, est nécessaire. Les FFs peuvent être extraites depuis les multiplicités de hadrons produites en Diffusion Inélastique Profonde (DIS). Les données ont été prises à COMPASS avec un faisceau de muons de 160 GeV/c diffusant sur une cible de protons pure (lH^2).

La présente thèse présente les mesures des multiplicités de hadron chargé, pion, kaon et proton identifiés, faites à partir des données SIDIS collectées en 2016. Elle détaille aussi les améliorations apportées au générateur d'événement DJANGOH dans le but d'améliorer la description des corrections radiative inclusive et semi-inclusive dans DIS qui sont ensuite utilisées comme facteurs de corrections aux multiplicités et l'adaptation de DJANGOH à la chaîne Monte-Carlo de COMPASS. Les facteurs de correction radiative aux multiplicités sont calculés selon un binning en trois dimensions en x, y et z. Pour la première fois, une détermination claire des effets radiatifs en fonction de la variable z a été effectuée.

Les données couvrent une large plage cinématique: $Q^2 > 1$ $(GeV/c)^2$, $y \in [0.1,0.7]$, $x \in [0.004,0.4]$, $W \in [5,17]$ GeV/c^2 et $z \in [0.2,0.85]$. Les multiplicités de hadrons chargés sont obtenues dans un binning en 3 dimensions (x,y,z), ce qui donne un total de 540 bins. De multiples facteurs de correction des multiplicités sont abordés : la correction d'acceptance, la correction des mésons vecteurs, l'unfolding du RICH, la correction due à la contamination par les électrons et les corrections radiatives. Les multiplicités, qui représentent environ 1800 points de données au total, fournissent une contribution importante au fit global QCD des données mondiales à NLO visant à déterminer les FFs et complètent les résultats précédents sur les multiplicités de hadrons chargés en utilisant les données prises en 2006 de diffusion de muons sur une cible isoscalaire (^{6}LiD).

Les FF de quark en kaons sont particulièrement recherchées car elles sont nécessaires pour mieux contraindre la polarisation du quark étrange. Les multiplicités K^+ et K^- ont été utilisées pour extraire la fragmentation de quark favorisée D^K_{fav}, non favorisée D^K_{unf} et étrange D^K_s avec un fit à LO. Le résultat du fit montre qu'il y a une mauvaise sensibilité au quark étrange de ces mesures. Ce fit donne trop de poids aux fonctions de fragmentation favoured et unfavoured aux dépens de la fonction de fragmentation strange. Cela nécessite une enquête plus détaillée et un fit de QCD perturbative à NLO.

ZUSAMMENFASSUNG

Am COMPASS-Experiment wird die Spinstruktur des Nukleons untersucht. Diese ist von besonderen Interesse seit der überraschenden Entdeckung durch die European Muon Collaboration (EMC), dass nur ein kleiner Teil des Nukleonspins von den Spins der Quarks stammt. Mit dieser Entdeckung begann die Suche nach den verschiedenen Bestandteilen. Diese sind durch die Spins der Quarks und Gluonen sowie deren Bahndrehimpulsen gegeben, wobei der aktuelle wird von den Beitrag der Quarkspins etwa 30% beträgt. Die Frage nach der Polarisation der Seequarks ist ein wichtiges Thema in der Hadronphysik, insbesondere für die Strangequarkpolarisation.

Um die Quarkpolarisation genauer bestimmt zu können, ist eine genaue Kenntnis der Quarkfragmentationsfunktionen (FFs) in Hadronen erforderlich, die die Hadronisierung von einen Quark q in ein Hadron h beschreiben. Die FFs können aus Hadronenmultiplizitäten extrahiert werden, die durch inklusive tiefinelastische Streuung (DIS) erzeugt wurden. Beim COMPASS-Experiment wird ein 160 GeV/c Myonstrahl an einem ruhenden reinen Protontarget (lH_2) gestreut.

In dieser Arbeit wird die Messung von Multiplizitäten geladener Hadronen, identifizierter Pionen, Kaonen und Protonen anhand der 2016 gesammelten DIS-Daten vorgestellt. Außerdem werden die Verbesserungen der Analyse durch die Verwendung des DJANGOH-Ereignisgenerators zur Bestimmung der inklusiven und semi-inklusiven Strahlungskorrekturen in DIS erläutert. Diese Korrekturfaktoren werden an die Multiplizitäten eingebracht. Die Anpassung von DJANGOH an die COMPASS Monte-Carlo Kette wird beschrieben. Die Strahlungskorrekturfaktoren für die Multiplizitäten werden in einer dreidimensionalen Einteilung von x,y und z-Intervallen. Hiermit wurde zum ersten Mal eine Bestimmung der Strahlungseffekte in Abhängigkeit von der z-Variablen durchgeführt.

Die Daten decken einen großen kinematischen Bereich ab: $Q^2 > 1$ (GeV/c)^2, $y \in$ [0.1,0.7], $x \in$ [0.004,0.4], $W \in$ [5,17] GeV/c^2 und $z \in$ [0.2,0.85]. Die geladenen Hadronmultiplizitäten werden in eine dreidimensionalen (x,y,z)-Einteilung bestimmt, was insgesamt 540 Intervalle ergibt. Verschiedene Korrekturfaktoren, die auf die Multiplizitäten angewendet werden, werden diskutiert: Akzeptanzkorrektur, Vektormesonkorrektur, RICH-Akzeptanzentfaltung, Elektronkontaminationskorrektur und Strahlungskorrektur. Die Multiplizitäten, die insgesamt etwa 1800 Datenpunkte repräsentieren, liefern einen wichtigen Input für die globale NLO pQCD-Anpassung der Weltdaten, mit denen FFs bestimmt werden. Außerdem werden die vorherigen Ergebnisse der Myonstreuung an einem isoskalaren Target (^{6}LiD) aus dem Jahr 2006 vervollständigt.

Besonders gefragt sind die Quark-FFs in Kaonen, um die Strangequarkpolarisation besser einschränken zu können. Die K^+-und-K^- Multiplizitäten werden verwendet, um die favourisierte D^K_{fav}, die nicht favourisierte D^K_{unf} und die Strangequark D^K_s Quark-FFs mit einer Anpassung in LO pQCD zu extrahieren. Das Ergebnis der Anpassung weist darauf hin, dass die Daten nicht sehr empfindlich auf den Strangequarkbeitrag sind. Die Anpassung führt zu einem überraschend grossen Beitrag von favourisierter und nicht favourisierter fragmentation auf Kosten der Strangefragmentationsfunktion. Daher sind in Zukunft weitere Untersuchungen und eine Anpassung in NLO perturbative QCD nötig.

À ma famille et mes amis.

ACKNOWLEDGEMENTS

First and foremost, I would like to thank CEA/IRFU and particularly the DPhN, the University of Mainz and particularly the KPH, and the COMPASS collaboration. That is with them that this whole story started.

I would like to then thank my two thesis directors, Fabienne Kunne and Eva-Maria Kabuß, for giving me the opportunity to become a PhD student, for always being present to guide me and always conforting me in my choices.

I would like to express my gratitude to the members of my jury. I would like to thank Franck Sabatié for agreeing to head the jury and also for the great support he has been during my year-ish in Saclay. The discussion with Andrea Bressan were always interesting and elevating. The moments with Fritz Klein at the Bonn meeting every Christmas were always a pleasure. I would eventually express my deepest appreciation to Hubert Spiesberger that helped me so much with the theory behind all the radiative corrections and was so nice and pedagogical in any circumstance.

Ich danke meinen COMPASS-Mainz Kollegen Johannes, Moritz, Malte und Nick. Es war wirklich eine Freude, mit Ihnen in Mainz und anderswo zusammen zu sein. Ich war der 'Petit Français' des Teams mit allen Klischees, aber Sie haben mich trotzdem als Teil des Teams begrüßt.

Je voudrais aussi remercier mes collègue de COMPASS à Saclay, que ce soit Damien, Stéphane ou Yann. Les discussions que nous avons eu sur la physique ou même des thèmes plus variés ont toujours été un plaisir. Merci à toi Nicole d'avoir été comme ma troisième directrice de thèse dans les derniers mois, j'aurais été dans la panade sans ton aide !

Special thanks for all the COMPASS collaboration members. Thanks to Riccardo, Jan, Misha, Andrea, Bakur, it was always great to discuss with you in and outside work. Thanks to Marcin for the always high-spirited and brilliant discussions we had, live of by email. You are never short of ideas and that is pretty impressive.

Ich möchte mich bei Manfred und Margrit bedanken, die mich für 2 Jahre in ihrem Haus in Mainz begrüßen durften. Sie haben meinen Aufenthalt in Deutschland deutlich verbessert!

J'aimerais remercier l'ensemble des précaires du DPhN. On était tous dans la même galère, on en est tous sorti en se serrant les coudes dans la bonne humeur. Toutes ces discussions entre midi et deux, ces sorties etc. Si je devais faire un petit résumé non exhaustif : les sorties de la fête de la musique avec Zoé, Jordan, Mehdi et Julien, la finale de la Coupe du Monde avec les deux zigotos précités, les passages au bars en ajoutant Nabil, Robin, Vladimir et Saba, les éclats de rire avec Nancy et Pierre, les discussions dans le bus avec Adrien, les moments pause dans notre 'salon de thé' avec Benjamin et Ève, les sorties 'footing' avec Aurélie (un jour je ferai un meilleur chrono !).. J'en ai sûrement oublié, mais vous étiez vraiment les meilleurs !

Un grand merci à Loïc, mon compagnon de sport mais aussi bien plus que ça. Avec toi on ne s'ennuie jamais et je suis bien content que tu sois revenu des États-Unis, même si c'était pour piquer mon bureau !

Merci à Hervé, Maxime et Franscesco pour les partage de café et les rires engendrés par la comparaison de nos situations dans nos collaborations.

Un grand merci à Danielle et Isabelle qui ont toujours été là pour m'aider dans les démarches administratives. Merci à Nicolas pour avoir eu la patience de répondre à toutes mes demandes en tant que Mac user.

J'aimerais adresser un merci tout particulier à la fameuse Saclay-Team, renommée dans la collaboration toute entière et bientôt regrettée. Thanks Po-Ju for your good mood and

high quality discussion. Merci Brian pour des discussions techniques jamais inintéressantes et moins techniques qui ne le sont pas moins.

Le dernier membre de la Saclay-Team, c'est Antoine et je voulais te remercier tout particulièrement. Tu es littéralement la personne avec qui j'ai passé le plus de temps cette dernière année, que ce soit dans notre bureau ou en dehors. Des discussions jamais lassantes, des barres de rire à foison, je n'aurais pu avoir meilleur co-bureau !

Merci aussi à Emiko que j'ai d'abord connu à travers Antoine puis à travers Marco. Merci aussi à Marco, même si tu t'es exilé loin de nous, on a passé quand même un paquet d'années et de temps à faire de la physique ensemble.

Je voulais remercier toute la petite troupe de l'OJS, vous avez toujours été un vecteur de motivation et une deuxième petite famille, l'impact que vous avez eu sur cette thèse est d'évidence non négligeable.

Merci encore à Laurie et Camille, toujours prêtes à passer de bons moments dans la capitale et à Laura, pour tout le support que tu m'as apporté. Merci aussi à toi Romane, tu as toujours été derrière moi soeurette !

Merci à Claire qui m'a toujours poussée à donner le maximum de moi-même et a toujours été là dans les hauts comme dans les bas.

Merci à Clémence pour ce soutien dans la dernière ligne droite, tu as amené de la stabilité à un moment où les événements professionnels s'enchaînaient à vitesse grand V.

Merci à vous Papa et Maman, c'est vous qui m'avez donné le goût des sciences et de la physique et qui avez veillé à ce que je ne me perde pas en chemin. Si j'en suis là c'est grâce à vous.

Un véritable merci à mes frères, Alexandre et Maxime, car sans la 'Boule Magique' et nos retrouvailles à Paris, ma thèse aurait sûrement été moins joyeuse et intéressante. Vous êtes des chefs.

CONTENTS

LIST OF FIGURES

ACRONYMS

AKK Albino-Kniehl-Kramer

DSEHS de Florian-Sassot-Epele-Hernandez Pinto-Stratmann

DIS Deep Inelastic Scattering

FF Fragmentation Function

HKNS Hirai-Kumano-Nagai-Sudoh

JAM Jefferson Lab Angular Momentum (Sato et al.)

KKP Kniehl-Kramer-Poter

KRE Kretzer

LSS Leader-Sidorov-Stamenov

MC Monte Carlo

PDF Parton Distribution Function

QCD Quantum ChromoDynamics

QED Quantum ElectroDynamics

QPM Quark Parton Model

SIDIS Semi-Inclusive measurement of Deep Inelastic Scattering

Part I

INTRODUCTION AND THEORETICAL PART

INTRODUCTION

The internal structure of the nucleon was first observed at the Standford Linear Accelerator Center (SLAC) in 1969 [1] by scattering high energy electron on nucleons. In this process, electrons interact with the internal components of the nucleon with an exchange of a virtual photon γ^*. Some years before, theories about substructure of hadrons were developed by Gell-Mann (*quarks*) [2] and Zweig (*aces*) [3] describing the nucleon as a system composed by three point-like objects. In response to the aforementionned SLAC results, Feynman developed the parton model [4], a model analoguous to the Gell-Mann's quark model. Fusing both models together, the *Quark Parton Model* (QPM) was born, describing the nucleon as being composed by three *valence quarks* and by quark-antiquark pairs (*sea quarks*). Following the SLAC discovery, other experiments performed lepton-nucleon scattering but at higher energies and with different types of leptons (e.g. muons or neutrinos). They found that in addition to quarks there should be other components of the nucleon. At the same time, the *Quantum ChromoDynamics* (QCD) was developped and introduced gluons as the vector bosons of the strong interaction but also the concept of confinement, which explains why one cannot observe isolated quarks.

This lepton-nucleon scattering, named *Deep Inelastic Scattering* (DIS), has a final state composed by the scattered lepton and a hadronic system. If both the scattered lepton and at least one hadron are detected in the final state, corresponding to a *Semi-Inclusive* measurement (SIDIS), the corresponding cross-section is the one of a hard scattering process (lepton-quark scattering in *perturbative* QCD (pQCD)) convoluted with two probability densities : the *Parton Distribution Functions* (PDFs) and the *Fragmentation Functions* (FFs). The PDFs are parametrizing the partonic structure of, in our case, the nucleons, while the FFs are parametrizing the hadronisation process, which is the formation of hadrons out of quarks that were scattered in the process. These quantities are expected to be universal in the sense that they are process independent. The PDFs of u, $\bar{u}$, d and $\bar{d}$ quarks are currently well determined with high precision measurements from experiment, while the ones of s and $\bar{s}$ are still not well constrainted. They are affected by large incertainties coming notably from the poor knowledge of strange quark into hadron FFs. They are measured from three different processes : electron-positron annihilation, SIDIS and hadron-hadron collisions. The FFs are useful in the determination of the spin structure functions through SIDIS. While they are well known for the first generation of quarks, they are still not well determined for higher mass quark with a discrepancy up to a factor 3 between parametrizations for the strange quark. They thus constitute the largest uncertainty for the determination of the strange quark polarization in SIDIS[5].

One of the goals of the COMPASS experiment at CERN is to study the nucleon spin structure. To reach it SIDIS data were taken using a 160 GeV μ^+/μ^- beam and a pure proton (lH_2) target, complementing the data already taken using a polarized 160 GeV muon beam and a polarized target (^6LiD or NH_3). COMPASS has already been analyzing measurements with an isoscalar ^6LiD target that allow to better constrain FFs through world data QCD fit. In order to continue to contribute with new measurements, new extraction of charged hadrons multiplicities (averaged number of hadron produced per DIS events) from COMPASS data with a pure proton target was decided.

The work described in this thesis includes the entire analysis chain of extraction of charged unidentified and identified hadrons (pion, kaon, proton) multiplicities.

The first part will go through the theoretical framework used for the analysis. The kinematic variables used to describe the DIS processes and the corresponding cross-sections are

presented. The QPM and its QCD improved version is discussed. The PDFs and FFs are introduced and a state of the art picture of the current knowledge of the FFs is drawn.

The second part describes the COMPASS experiment. The main components of the COMPASS apparatus are shortly described while the Ring Imaging CHerenkov (RICH) detector is more thourougly adressed as one of the main detector used in the analysis. The determination of RICH performance for particle identification and misidentification is also discussed.

The third part is about the DJANGOH Monte-Carlo generator, which allows generation of radiative events within our Monte-Carlo simulation and more broadly radiative corrections estimation.

The fourth and final part is focused on the extraction of charged unidentified (h^+/h^-) and identified ($\pi^+/\pi^-, K^+/K^-, p/\bar{p}$) multiplicities from COMPASS data as well as leading order extraction of FFs from the aforementionned kaon multiplicities.

1 THEORETICAL FRAMEWORK

DIS is one physics process to study FFs. A DIS process involves a high-energy lepton l interacting with a nucleon N, producing new particles in the final state X ($l + N \rightarrow l' + X$). In an *inclusive* measurement, only the scattered lepton l' is measured. In a *semi-inclusive* measurement, in addition to the scattered lepton, at least one hadron of the final state is detected ($l + N \rightarrow l' + h + X$). In an *exclusive* measurement, all particles from the final state are detected.

The theoretical framework of DIS and SIDIS, which allows us to extract the FFs from hadron multiplicities, are introduced. The DIS and SIDIS cross-sections and kinematic variables are discussed. The *Quark Parton Model* (QPM) model used to interpret the DIS and SIDIS results and to describe the nucleon structure is described, as well as its extended version with *Quantum ChromoDynamics* (QCD). The PDFs and FFs are defined, as well as how they are extracted. A closer look on FFs is then taken: how they are extracted from SIDIS data and what the existing parametrizations are.

1.1 DEEP INELASTIC SCATTERING

The deep inelastic scattering process in first order QED is depicted in Fig. 1. The incoming lepton l exchanges a virtual photon γ^* with the nucleon N. The nucleon absorbs the energy of the virtual photon and fragments into a final state X. The scattered lepton is represented by l'. This process description is also known as the one photon exchange approximation.

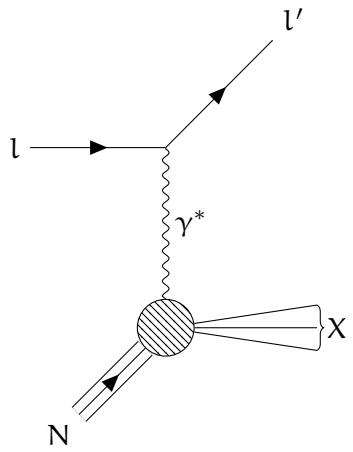


Figure 1: Deep inelastic scattering diagram.

The kinematics of a DIS event are fixed by the 4-momentum vector of l ($\mathbf{l}$ = (E,$\vec{l}$)), l' ($\mathbf{l'}$ = (E$'$,$\vec{l}'$)) and N ($\mathbf{P}$ = (M,$\vec{0}$)). The 4-momentum vector for the virtual photon is calculated as $\mathbf{q}$ = $\mathbf{l}$ - $\mathbf{l'}$ = (ν = E - E$'$, $\vec{q} = \vec{l} - \vec{l}'$). One needs only two Lorentz invariant variables to describe inclusive DIS [6]. One is the invariant mass of the virtual photon Q^2:

$$Q^2 = -\mathbf{q}^2 \overset{lab}{\approx} 4EE'\sin^2\left(\frac{\theta}{2}\right). \tag{1}$$

Q^2 gives a measure of the scale at which the nucleon structure is probed: the larger Q^2 is, the deeper the probing of the nucleon is performed. θ is the angle between the incoming and outgoing leptons.

The other variable x measures the elasticity of the interaction:

$$x = \frac{Q^2}{2\mathbf{P}\cdot\mathbf{q}} \overset{lab}{=} \frac{Q^2}{2M\nu} = \frac{Q^2}{Q^2 + (W^2 - M^2)}, \tag{2}$$

where $W^2 = (\mathbf{P}+\mathbf{q})^2 \overset{lab}{=} M^2 + 2M\nu - Q^2$ is the invariant mass of the hadronic final state. x is comprised between 0 and 1. If $x = 1$ ($W^2 = M^2$) the interaction is elastic, if $x < 1$ ($W^2 > M^2$) then it is inelastic.

Other Lorentz invariants are given in Table 7.

Table 1: DIS kinematic variables. The lepton mass is neglected. For a fixed target experiment, these quantities can be expressed and used in the laboratory frame.

Variable	Description
$Q^2 = -\mathbf{q}^2 \overset{lab}{\approx} 4EE'\sin^2(\frac{\theta}{2})$	Interaction scale
$\nu = \frac{\mathbf{P}\cdot\mathbf{q}}{M} \overset{lab}{=} E - E'$	Energy transfer from the lepton l to γ^*
$x = \frac{Q^2}{2\mathbf{P}\cdot\mathbf{q}} \overset{lab}{=} \frac{Q^2}{2M\nu}$	Fraction of the nucleon momentum $\mathbf{P}$ carried by the parton struck by γ^*
$\nu = \frac{\mathbf{P}\cdot\mathbf{q}}{\mathbf{P}\cdot\mathbf{l}} \overset{lab}{=} \frac{\nu}{E}$	Fraction of the incoming lepton energy transferred to γ^*
$s = (\mathbf{P}+\mathbf{l})^2 \overset{lab}{\approx} M^2 + 2ME$	Center-of-mass energy squared
$W^2 = (\mathbf{P}+\mathbf{q})^2 \overset{lab}{=} M^2 + 2M\nu - Q^2$	Invariant mass of the hadronic final state

1.1.1 *Cross section calculation for the inclusive DIS process*

The deep inelastic cross section, in the one photon exchange approximation, can be written in terms of the lepton-photon coupling tensor $L_{\mu\nu}$ and the hadronic coupling tensor $W^{\mu\nu}$ and the proton propagator $\sim \frac{1}{q^4}$ [7]:

$$\frac{d\sigma}{dE'd\Omega} = \frac{\alpha^2}{2Mq^4}\frac{E'}{E}L_{\mu\nu}W^{\mu\nu}, \tag{3}$$

α is the fine structure constant. The leptonic and hadronic tensors can be split in a symmetric and antisymmetric parts [8]:

$$L_{\mu\nu}(l,s;l') = 2L^{(S)}_{\mu\nu}(l;l') + iL^{(A)}_{\mu\nu}(l,s;l') \tag{4}$$

where $L_{\mu\nu}$ is given for point-like fermions by QED:

$$\begin{aligned} L^{(S)}_{\mu\nu} &= l'_\mu l_\nu + l'_\nu l_\mu - g_{\mu\nu}(\vec{l}'\vec{l} - m^2), \\ L^{(A)}_{\mu\nu} &= -m\epsilon_{\mu\nu\sigma\rho}s^\sigma q^\rho, \end{aligned} \tag{5}$$

and

$$W^{\mu\nu}(q;P,s) = W^{\mu\nu\ (S)}(q;P) + iW^{\mu\nu\ (A)}(q;P,S), \tag{6}$$

where, assuming the parity and time reversal invariances, the hadron tensor can be expressed as:

$$\begin{aligned} &\frac{1}{2M}W^{\mu\nu\ (S)}(q;P) = \\ &W_1(P\cdot q, q^2)\left(-g^{\mu\nu} - \frac{q^\mu q^\nu}{q^2}\right) + \frac{W_2(P\cdot q, q^2)}{M^2}\left(P^\mu - \frac{P\cdot q}{q^2}q^\mu\right)\left(P^\nu + \frac{P\cdot q}{q^2}q^\nu\right), \end{aligned} \tag{7}$$

$$\frac{1}{2M}W^{\mu\nu\ (A)}(q;P,S) = \epsilon_{\mu\nu\sigma\rho}q^{\sigma}G_1(P\cdot q,q^2)MS^{\rho} + \frac{G_2(P\cdot q,q^2)}{M}(P\cdot q)S^{\rho} - (S\cdot q)P^{\rho}. \tag{8}$$

The lepton and nucleon polarizations[1] are given by s and S, respectively. The Minkowski metric is $g_{\mu\nu}$[2] and m is the lepton mass. The functions $W_1(P\cdot q,q^2)$, $W_2(P\cdot q,q^2)$, $G_1(P\cdot\ q,q^2)$ and $G_2(P\cdot q,q^2)$ are the spin averaged and spin dependent structure functions parametrizing the internal structure of the nucleon. They can be expressed as dimensionless functions :

$$\begin{aligned} MW_1(P\cdot q,Q^2) &= F_1(x,Q^2),\\ \nu W_2(P\cdot q,Q^2) &= F_2(x,Q^2),\\ \frac{(P\cdot q)^2}{\nu}G_1(P\cdot q,Q^2) &= g_1(x,Q^2),\\ \nu(P\cdot q)G_2(P\cdot q,Q^2) &= g_2(x,Q^2). \end{aligned} \tag{9}$$

Going back to Eq. 3 and using the symmetric and antisymmetric parts of the tensors:

$$\frac{d\sigma}{dE'd\Omega} = \frac{\alpha^2}{2Mq^4}\frac{E'}{E}\left[L_{\mu\nu\ (S)}W^{\mu\nu\ (S)} - L_{\mu\nu\ (A)}W^{\mu\nu\ (A)}\right]. \tag{10}$$

After averaging over all possible spin configurations in the initial state and summing in the final state lepton, one obtains the unpolarized DIS cross-section in terms of the structure functions F_1 and F_2, neglecting the leptonic mass :

$$\frac{d\sigma^{unpolarized}}{dxdQ^2} = \frac{4\pi\alpha}{Q^4}\left[y^2F_1(x,Q^2) + \left(\frac{1-y}{x} - \frac{My}{2E}\right)F_2(x,Q^2)\right]. \tag{11}$$

1.2 QUARK PARTON MODEL

The *Quark Parton Model* [9, 10] is developed in the infinite momentum frame where the nucleon has a very large momentum along a certain direction and is composed by point-like spin-1/2 particles called partons. In this case, the transverse momentum of these partons can be neglected. In DIS, the virtual photon interacts with the parton, which carries a fraction ξ of the 4-momentum $\mathbf{P}$ of the nucleon and the invariant mass of the initial and final states are respectively $(\xi\mathbf{P}+\mathbf{q})^2$ and 0. This yelds:

$$(\xi\mathbf{P}+\mathbf{q})^2 = 0 \Rightarrow 2\xi\mathbf{P}\cdot\mathbf{q} + \mathbf{q}^2 = 0 \Rightarrow \xi = \frac{Q^2}{2\mathbf{P}\cdot\mathbf{q}}, \tag{12}$$

which is equal to Bjorken x, thus Bjorken x is interpreted as the momentum fraction carried by the struck quark.

Within this model, since gluons do not carry any electric charge, the DIS interaction can only involve quarks and it has to be noted that the spectator quarks are not affected by the interaction. The hadronic tensor is given by [7]:

$$W^{\mu\nu} = \sum_{q,s} e_q^2 n_q(x,s;S)\frac{1}{\mathbf{P}\cdot\mathbf{q}}\left[2x\mathbf{P}^{\mu}\mathbf{P}^{\nu} + \mathbf{P}^{\nu}\mathbf{q}^{\mu} + \mathbf{P}^{\mu}\mathbf{q}^{\nu} - g^{\mu\nu}\mathbf{P}\cdot\mathbf{q}\right], \tag{13}$$

where $n_q(x,s;S)$ is the density of quarks q with charge e_q and spin s, the nucleon spin being given by S. In this model, the structure functions for spin-1/2 partons are given by [7]:

$$\begin{aligned} F_1(x) &= \frac{1}{2}\sum_q e_q^2 q(x)\\ F_2(x) &= x\sum_q e_q^2 q(x), \end{aligned} \tag{14}$$

1 Properties of the covariant spin 4-vector: $s\cdot k = 0$ and $s\cdot s = -1$. Similar for S

2 Signature of the metric is $(-+++)$

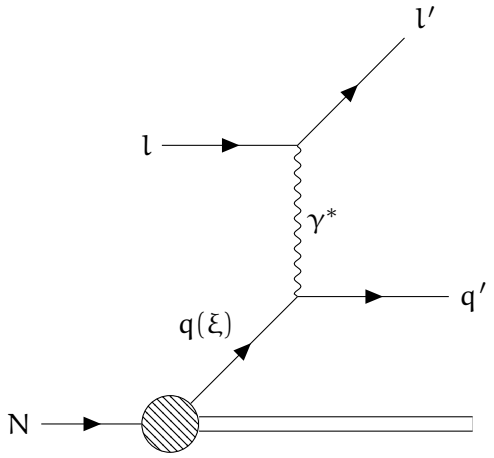


Figure 2: DIS in the QPM. The lepton scatters off a single quark, the remaining quarks are only spectating the process.

where $q(x)$ are the spin averaged Parton Distribution Functions (PDFs). The sums run over all quark and antiquark flavors. Eq. 14 can be reformulated as the Callan-Gross relation [11]:

$$F_1(x) = \frac{1}{2x} F_2(x).$$

If partons are point like and spin-1/2, the Q^2 dependence is lost in the QPM infinite momentum frame.

Thus the result for the spin averaged DIS cross-section in the QPM is [12]:

$$\frac{d^2\sigma}{dxdy} \overset{QPM}{=} \frac{8\pi\alpha^2 ME}{Q^2} \left[\frac{1}{2}y^2 + \left(1 - y - \frac{y^2\gamma^2}{4}\right)\right] x \sum_q e_q^2 q(x) \tag{15}$$

1.2.1 *Scaling violation*

The structure function F_2 has been measured by several collaborations covering a wide x - Q^2 kinematic range [13]. It is constructed from PDFs using additional coefficient functions. The measured values are depicted as a function of Q^2 and in bins of x in Fig. 3. Scaling is only visible in a small x region between 0.1 and 0.4. Outside this region the structure function F_2 has mostly a logarithmic dependence on Q^2. At small x, F_2 increases with Q^2, while at large x, F_2 decreases. From the momentum sum rule a conclusion was made that there should be a missing contribution from the force carriers: the gluon contribution. In order to take into account this contribution, the Quantum ChromoDynamics frame (QCD) was developed as the theory describing the interaction of the quarks and gluons and embedded in the QPM.

1.2.2 *QCD-improved QPM*

The Q^2 dependence mentionned in previous subsection can be estimated by introducing quark interactions in the framework of QCD [6, 14]. Quantum ChromoDynamics is a non-abelian gauge theory based on a symmetry group SU(3), which describes the interaction of quarks and gluons. The charge of this theory is called colour and the force carriers are the gluons, which are also coloured particles. The internal nucleon dynamic is due to the gluon emission and absorption and quark-antiquark pair creation from gluons. This creates a cloud of gluons and virtual $q\bar{q}$ pairs known as sea quarks.

The QCD coupling constant α_s depends on the scale of the interaction. At low energies quarks or gluons are always forming colorless particles, which are named hadrons: this is

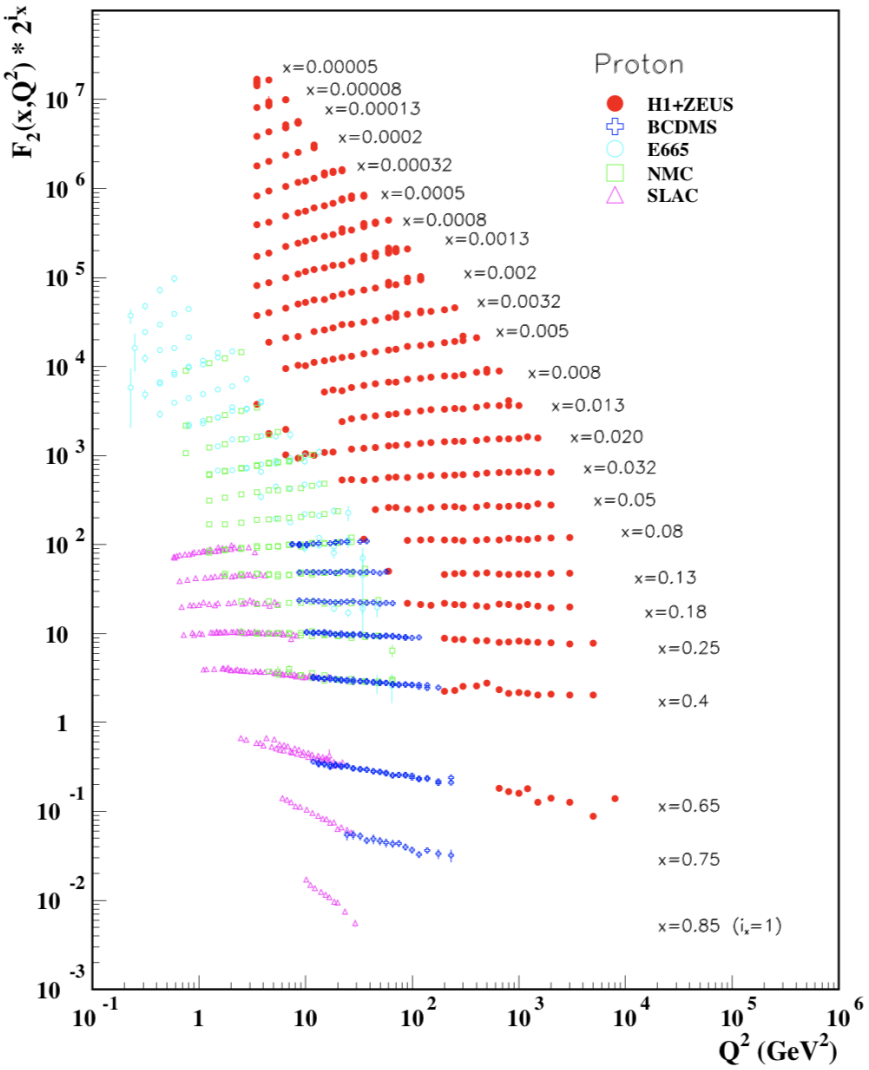


Figure 3: The proton structure function F_p^2 measured in electromagnetic scattering of electrons and positrons on protons in the kinematic domain of the HERA data (collider experiments H1 and ZEUS for $Q^2 \geqslant 2$ GeV^2), and for electrons (SLAC) and muons (BCDMS, E665, NMC) on a fixed target. Figure taken from [13].

called confinement. At high energies quarks or gluons are free particles: this is asymptotic freedom.

Depending on the energy regime, a process can be labeled as a hard ($\alpha_s \sim 0$) or soft process ($\alpha_s large$). Hard processes can be described within the perturbative QCD (pQCD) framework, while soft processes can only be parametrized from experimental data. As in DIS the scale variable is often chosen as Q^2, the DIS cross-section is factorized [15] in terms of soft and hard processes for $Q^2 > 1$ GeV^2, where α_s is small enough: the hard process is described by the lepton-quark cross-section σ_q convoluted with the soft process parametrized by the PDFs. These two regimes differ by the factorisation scale Λ that is mostly chosen as Q^2.

The resolution of the virtual photon probe is proportional to $1/Q^2$ (see Fig. 4 at fixed x). At $Q^2 \sim 0$, the virtual photon sees the nucleon as a point-like particle. As Q^2 increases, the virtual photon starts to resolve the nucleons constituents. At large Q^2 the virtual photon is able to resolve point-like quarks. The first QCD correction to the QPM concerns the gluon emission by the initial and the final quark.

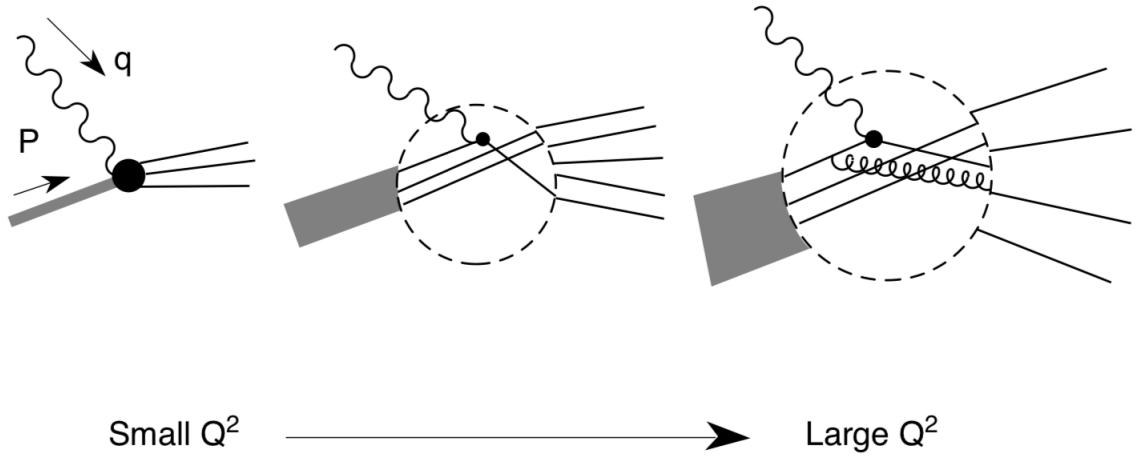


Figure 4: Resolution of the photon probe versus Q^2. Figure taken from [14].

The Q^2 dependence can be calculated using the Dokshiter-Gribov-Lipatov-Altarelli-Parisi (DGLAP) equations [16–19]:

$$\frac{dq_i(x,Q^2)}{d\ln Q^2} = \frac{\alpha_s(Q^2)}{2\pi}\sum_j \int_x^1 P_{ij}(x/\xi, \alpha_s(Q^2))q_i(\xi,Q^2). \tag{16}$$

Here, the splitting functions $P_{ij}(x/\xi)$ [20] are the probability that a quark or gluon of type j and momentum fraction ξ is the parent of i with momentum fraction x. A similar equation holds for the gluon distribution. If the PDFs are known at a given scale Q_0^2, they can be evolved to any given Q^2 using these equations.

1.3 DETERMINATION OF PARTON DISTRIBUTION FUNCTIONS

The PDFs are non-perturbative quantities and thus cannot be calculated from a theoretical framework. A global fit to world data is the only way to quantify them. It is possible to fit measurement coming from different processes because PDFs are universal quantities, i.e. they are process independent. The world data consists mostly of lepton-nucleon DIS but collider experiments (pp or $p\bar{p}$) or neutrino scattering can be used for special contributions, e.g. for gluons or strange quarks. As experiments cover different kinematic ranges, this allows one to determine the PDFs in a large (x,Q^2) space.

For the fit to be performed, a functional form has to be provided at an initial scale Q_0^2. Often the form $xq_i(x,Q_0^2) = x^\alpha(1-x)^\beta$, where q_i are partons, is used with additional terms refining the fit, reaching a number of free parameters from 10 to 25. The DGLAP equations are then used to evolve the PDFs to a given Q^2. An example of a fit done by the MMHT group [21] at Next to Leading Order (NLO) for different Q^2 values is shown in Fig. 5.

1.4 SEMI-INCLUSIVE DEEP INELASTIC SCATTERING

SIDIS is the semi-inclusive measurement of DIS. In the final state, at least one hadron and the scattered lepton are detected ($l + N \to l' + h + X$) and a new invariant variable z is introduced, which corresponds to the energy fraction of the virtual photon held by the hadron h:

$$z = \frac{\mathbf{P}\cdot\mathbf{p}_h}{\mathbf{P}\cdot\mathbf{q}} \overset{lab}{=} \frac{E_h}{\nu}. \tag{17}$$

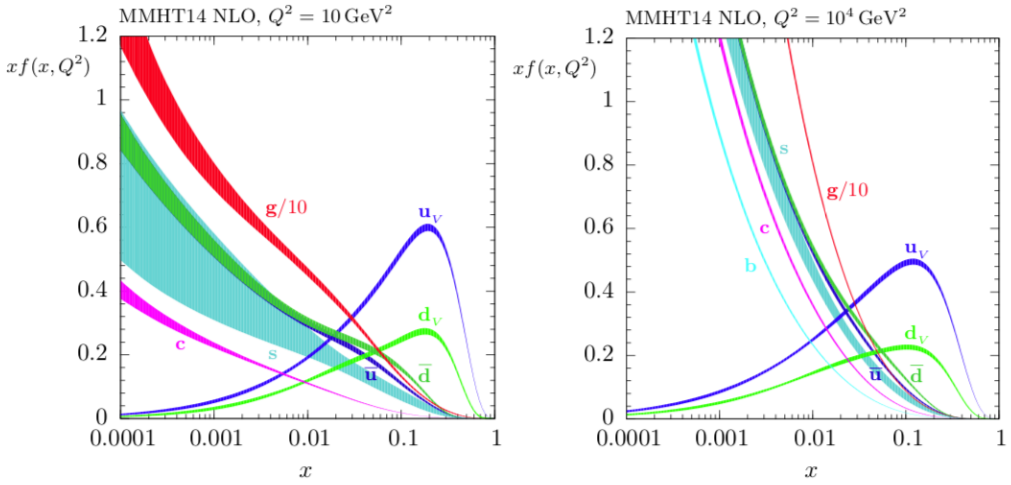


Figure 5: Unpolarized PDFs at Next to Leading Order (NLO) from MMHT group at Q^2 = 10 GeV^2 (left) and Q^2 = 10^4 GeV^2 (right) with associated 68% confidence-level uncertainty bands. Figure taken from [21].

The semi-inclusive cross section reads [22]:

$$\frac{d\sigma}{dxdydz} = \frac{8\pi\alpha^2 ME}{Q^4}\left[xy^2 H_1(x,Q^2,z) + (1-y)H_2(x,Q^2,z)\right], \tag{18}$$

where H_1 and H_2 are structure functions related to F_1 and F_2 [12, 22]:

$$\sum_h \int_0^1 H_i(x,Q^2,z)dz = F_i(x,Q^2) \quad , \quad i \in [\![1,2]\!]. \tag{19}$$

Additional variables used to describe the hadron kinematics are given in Table. 2.

Table 2: SIDIS kinematic variables.

Variable	Description
$\mathbf{p} = (E_h, \vec{p}_h)$	Hadron 4-momentum vector
$p_{h\parallel}$	Component of $\vec{p}_h$ along $\vec{q}$
$p_{h\perp}$	Transverse component of $\vec{p}_h$ with respect to $\vec{q}$
θ_h	Angle between $\vec{q}$ and $\vec{p}_h$
Φ_h	Angle between the scattering plane and the hadron production plane
$z = \frac{E_h}{\nu}$	Energy fraction of the virtual photon transferred to the hadron h
$\eta = \frac{1}{2}\ln\left(\frac{\lvert\mathbf{p}\rvert+p_L}{\lvert\mathbf{p}\rvert-p_L}\right)$	Pseudorapidity

1.4.1 *SIDIS in QPM*

The factorization Ansatz is also valid for SIDIS measurement thus the hadron production can be described as a convolution of three independent processes: the soft part $q(x)$ that are the PDFs, the hard process σ_q describing the absorption of the virtual photon γ^* by the quark q and the soft part $D_q^h(z)$ characterize the fragmentation of the quark q into a hadron h.

The structure functions $H_i(x,Q^2,z)$ contain the information on what happens to the struck quark after the interaction with the virtual photon. The fragmentation function (FF) $D_q^h(z,Q^2)$

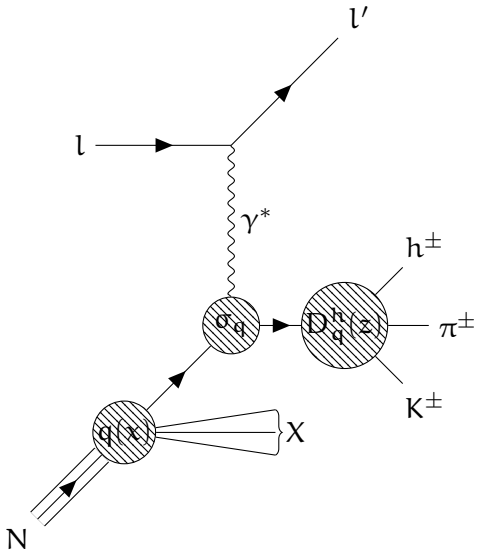


Figure 6: Factorization in SIDIS.

is defined as the probability for a quark of flavour q to fragment into a hadron h with a fraction of energy z. The expression of the spin averaged SIDIS cross section can be expressed within QPM at LO in terms of PDFs and FFs [12, 23]:

$$\frac{d^3\sigma}{dxdydz} \overset{LO}{=} \frac{8\pi\alpha^2 ME}{Q^2}\left[\frac{1}{2}y^2 + \left(1 - y - \frac{y^2\gamma^2}{4}\right)\right] x \sum_q e_q^2 q(x) D_q^h(z). \tag{20}$$

1.5 FRAGMENTATION FUNCTIONS

When computing the cross-section of a given process $A + B \to h + X$, this cross-section is found to be a convolution of three different terms (Eq. 21): one non-perturbative term involving the PDFs (probability to obtain parton a from nucleus A $f_{a/A}(x_a, Q^2)$), one hard cross-section term for perturbative calculation ($d\sigma_{a,b\to c}(x_a, x_b, Q^2)$) and a last non-perturbative term involving the FFs (probability to obtain hadron h from parton c $D_c^h(x_c, Q^2)$).

$$d\sigma_{A+B\to h+X} = \sum_{a,b,c} \left[f_{a/A}(x_a, Q^2) f_{b/B}(x_b, Q^2)\right] \otimes \left[d\sigma_{a,b\to c}(x_a, x_b, Q^2)\right] \otimes \left[D_c^h(x_c, Q^2)\right] \tag{21}$$

Fig. 6 illustrates this factorization in SIDIS. The extraction of FFs can also be done from electron-positron annihilation and hadron-hadron collisions measurements. The universality of FFs has been experimentally tested by Kniehl, Kramer and Pötter [24]. Different ideas have been developed to model how quarks confine together to make a hadron.

1.5.1 *Lund String Fragmentation Model*

In the Lund String Model [25], the hadron production is explained by the creation of quark-antiquark pairs $q\bar{q}$. The strong interaction between partons is represented by a string. The energy inside the string is linear function of the distance between two stringed partons. At some point the energy is large enough to create a new $q\bar{q}$ pair and the string breaks. All unpaired remnants have new strings and the process repeats until there are only hadrons. The hadronization scheme in the Lund model in the center of mass frame is illustrated in Fig. 7 (a). The virtual photon is absorbed e.g. by a u quark and in consequence the u quark is ejected from the nucleon. A new $q\bar{q}$ pair is created by the string breaking e.g. $d\bar{d}$. The

remaining u quark binds with the $\bar{d}$ quark to form a π^+ with a given z as shown in Fig. 7 (b). The remaining system repeats fragmentation process until the energy is smaller than the available energy ν. In addition baryon creation by di-$q\bar{q}$ pair formation is introduced.

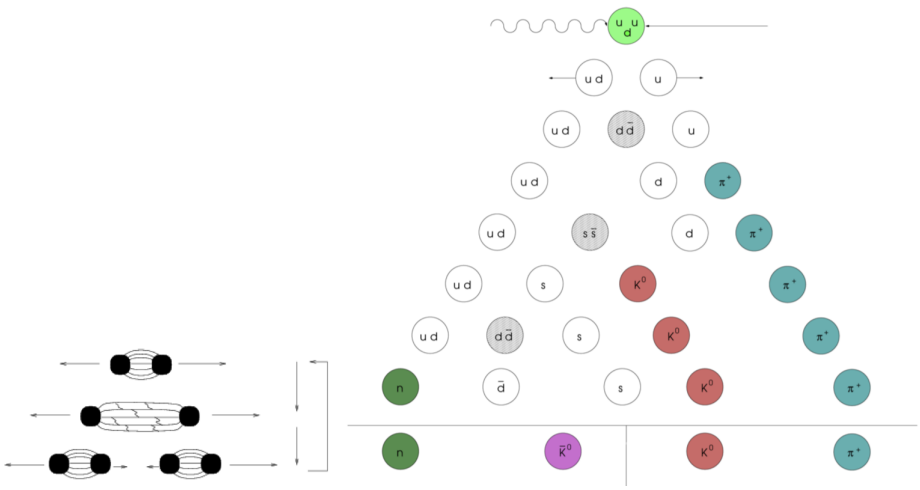


Figure 7: The fragmentation process in the Lund model. In (b), the produced K^0, $\overline{K^0}$, π^+ or n could also be an excited state. Figure taken from [23].

1.5.2 *Quark Fragmentation Regions*

Up to this point only the fragmentation of the struck quark was considered. The spectator quarks, which are not involved in the scattering process, have also to hadronize. This phenomenon is happening in two distinct p_h regions: the target fragmentation region, where the final hadron h has a small momentum in the rest frame of the target, and the current fragmentation region, where the product $\mathbf{P} \cdot \mathbf{p}_h$ grows with Q^2. At low energies there is a large overlap of these regions, while at high energies they start to separate. This hadron production can contaminate the SIDIS measurement of current fragmentation. To deal with this issue, Berger [12] came with a criterion based on the pseudorapidity of the final state η, which is the measurement of the longitudinal momentum. The sign of η is linked to the different regions: if $\eta > 0$ the hadron moves towards the direction of the virtual photon and is a current hadron, else is $\eta < 0$ the hadron is a target remnant (Fig. 8). Defining $p(k)$ to be the probability that k hadrons of some specific type decay from one cluster, one finds that the *fully inclusive* correlation function has the form:

$$C(y_1, y_2) = \frac{\langle k(k-1)\rangle}{\langle k\rangle} \left(\frac{1}{\sigma}\frac{d\sigma}{dy}\right)_{y\sim 0} G(y_1 - y_2), \tag{22}$$

when y_1 and y_2 are in the central region and the averages are:

$$\begin{aligned} \langle k\rangle &= \sum kp(k) \\ \langle k-1\rangle &= \sum k(k-1)p(k). \end{aligned} \tag{23}$$

The Gaussian function:

$$G(y_1 - y_2) = \frac{1}{2\delta\sqrt{\pi}} \exp\left[-\frac{(y_1 - y_2)}{4\delta^2}\right], \tag{24}$$

has an effective *correlation length* of 2δ. As the typical hadronic correlation length in pseudorapidity is $\delta \sim 2$, a separation criterion, the Berger criterion, is that $\Delta\eta = \eta_{max} - \eta_{min} \geqslant 2\delta$

or in terms of DIS kinematics variables $W \gtrsim 7.4$ GeV. It is important to select the kinematic region so that selected hadrons are dominated by the struck quark fragmentation.

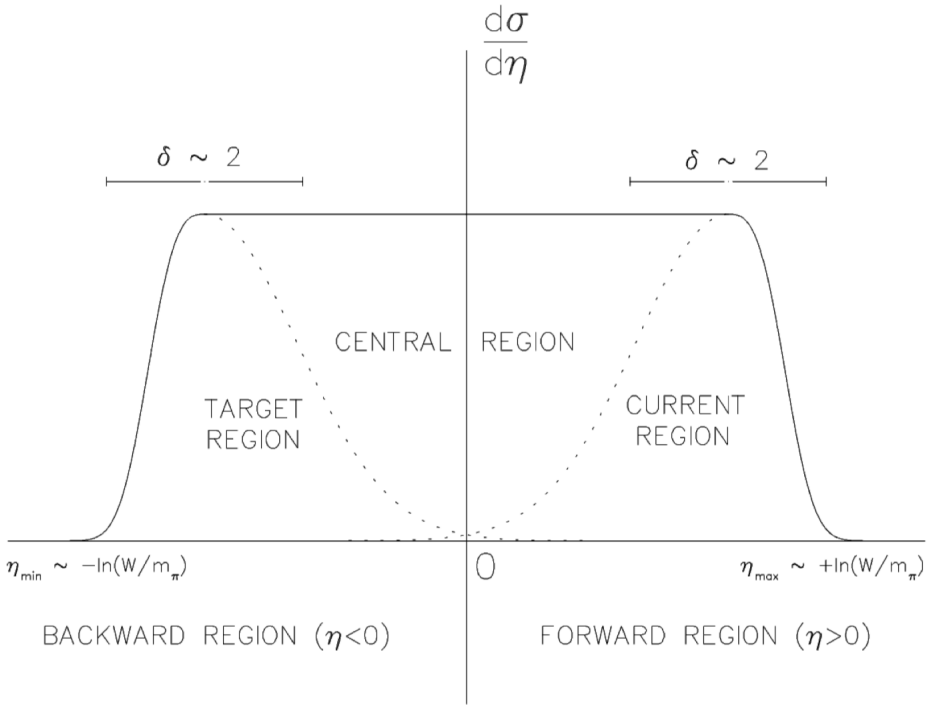


Figure 8: Hadronic pseudorapidity (η) distribution at very high energies. Figure taken from [26].

1.5.3 *Scaling and Q^2 evolution*

For the FFs extracted from e^+e^- annihilation displayed in Fig. 9, the scaling is present for a wide $x = 2p_h/\sqrt{s}$ range (Fig. 9 (a)). At low x ($x < 0.1$) the FFs increases with the total center-of-mass energy $\sqrt{s}$ (Fig. 9 (b)), while at large x, the FFs are shifted towards lower values for large Q^2 (similar behaviour as PDFs). Here $\sqrt{s}$ has the same role has Q^2. Scaling violation is observed [13].

The evolution of the fragmentation functions is also described by DGLAP equations [16–19]:

$$\frac{dD_q^h(z,Q^2)}{d\ln Q^2} = \frac{\alpha_s(Q^2)}{2\pi} \sum_j \int_x^1 P_{qj}\left(z/\xi, \alpha_s(Q^2)\right) D_q^h(\xi, Q^2) \frac{d\xi}{\xi}. \tag{25}$$

In Fig. 10 the process contributing to the Q^2-evolution is illustrated : the fragmentation of a quark q_i through its own hadronization after emmiting a gluon G ($P_{qq}D_{q_i}^h$), through the hadronization of a gluon G ($P_{Gq}D_G^h$), the fragmentation of a gluon splitting into a quark-antiquark pair and following hadronization of the quark in hadron ($P_{qG}D_{q_i}^h$) and eventually the gluon fragmentation via the three-gluon self-interaction ($P_{GG}D_G^h$).

1.5.4 *Fragmentation Function Symmetries*

One FF $D_q^h(z,Q^2)$ is introduced for each flavour q and each hadron species h. Considering only the light quarks (u,ū,d,d̄,s and s̄), in case the mass threshold for heavy quarks is higher than the covered kinematic domain, implies that for charged hadrons one has to measure twelve different fragmentation functions for positive and negative hadrons. Nevertheless,

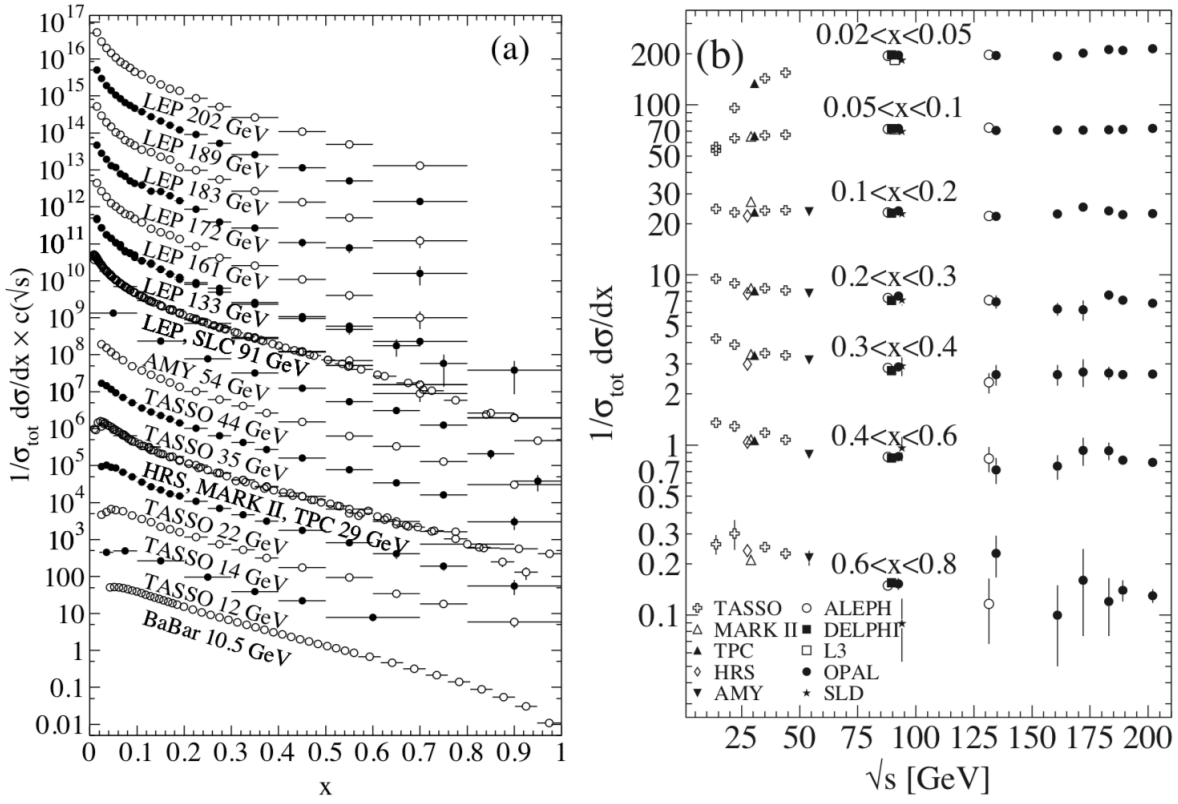


Figure 9: The e^+e^- fragmentation function for all charged particles for different center of mass energy $\sqrt{s}$ versus x (a) and for various range of x versus $\sqrt{s}$ (b). Figures taken from [13].

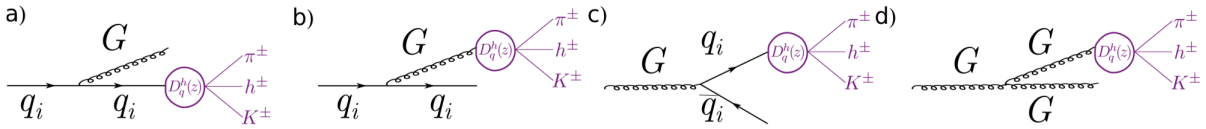


Figure 10: The fragmentation of the quark q_i decaying into a hadron h while emitting a gluon G ($P_{qq}D^h_{q_i}$) (a), the fragmentation of the quark q_i through a gluon G ($P_{Gq}D^h_G$) (b), the fragmentation of the gluon G via the creation of a $q_i\bar{q}_i$ pair and the decay of q_i ($P_{qG}D^h_{q_i}$) (c) and the fragmentation of the gluon G via a three gluon vertex ($P_{GG}D^h_G$) (d). Figure taken from [27].

within the QCD-improved QPM, symmetries as isospin or charge-conjugation can be used to reduce the number of independent fragmentation functions.

The FFs can be split into two main categories. If a quark fragments into a hadron h and the quark is a valence quark of h, the FF is said to be favoured (D^h_{fav}). If it is a sea quark of h, the FF is unfavoured (D^h_{unf}).

For pions charge-conjugation symmetry reduces the number of independent fragmentation functions to six. The application of the isospin (viz. $D^{\pi^+}_u = D^{\pi^+}_{\bar{d}}$ for π^+) and SU(3) symmetries lowers this number further to two:

$$\begin{aligned} D^h_{fav} : D^{\pi^+}_u &= D^{\pi^-}_{\bar{u}} = D^{\pi^-}_d = D^{\pi^+}_{\bar{d}}, \\ D^h_{unf} : D^{\pi^-}_u = D^{\pi^+}_{\bar{u}} = D^{\pi^+}_d = D^{\pi^-}_{\bar{d}} &\overset{SU(3)\,sym.}{=} D^{\pi^\pm}_s = D^{\pi^\mp}_{\bar{s}}. \end{aligned} \tag{26}$$

For kaons charge-conjugation symmetry reduces the number of independent fragmentation functions to six. The application of the isospin (viz. $D^{\pi^+}_u = D^{\pi^+}_{\bar{d}}$ for π^+) and SU(3) symmetries lower the number to three independent FFs: the favoured, grouping the kaons

valence quarks u and $\bar{u}$ FFs, the strange, grouping the kaons valence quarks s and $\bar{s}$ FFs and the unfavoured, grouping the kaons sea quark FFs:

$$\begin{aligned} D^h_{fav} : D^{K^+}_u &= D^{K^-}_{\bar{u}} \\ D^h_{str} : D^{K^+}_{\bar{s}} &= D^{K^-}_s \\ D^h_{unf} : D^{K^+}_{\bar{u}} = D^{K^-}_u = D^{K^+}_s = D^{K^-}_{\bar{s}} &= D^{K^\pm}_{\bar{d}} = D^{K^\mp}_{\bar{d}} \end{aligned} \tag{27}$$

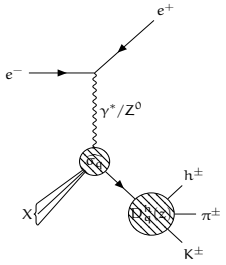
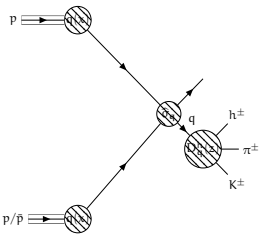
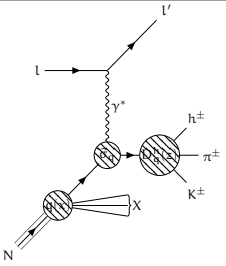
1.6 STATE OF THE ART OF THE FRAGMENTATION FUNCTIONS

1.6.1 *Measurements*

The production of hadrons from the the struck quark cannot be computed as final state hadron masses are of order or smaller than Λ_{QCD}. In order to extract FFs reliably from the Q^2 dependence of the measured hadron production, a large kinematic range is needed. Thus data taken at different energies are used. Three different processes are so far used to extract quark fragmentation functions: electron-positron annihilation (SIA), lepton-nucleon (SIDIS) and hadron-hadron collisions (pp or $p\bar{p}$). A summary of the aspects of the different processes can be found in Table. 3. The SIA data (LEP [28–30], SLAC [31], BaBar [32] and BELLE [33]) provide the cleanest access to the FFs, since the cross-section of the process does not involve PDFs and are well calculated up to NNLO. But due to the dependence of the cross-section on e_q^2, D^h_q and $D^h_{\bar{q}}$ cannot be separated and there is a limited access to gluon FF D^h_g. The data from hadron-hadron collisions (UA5 [34], UA1 [35], ALICE [36], CMS [37, 38], ATLAS [39], RHIC [40–42]) give access to D^h_q, $D^h_{\bar{q}}$ and D^h_g, but allow no direct access to z.

Data from SIDIS can be compared to data from previously presented processes for the current fragmentation region (see Section 1.5). SIDIS data have the advantage that factorization has been proven to all orders of α_s. In addition they cover a wide range in Q^2 in a single measurement compared to SIA. The experiments providing inputs for the SIDIS process are EMC [43] and COMPASS [44, 45] using muon beam and E00-108 [46] and HERMES [47] using electron beam. All experiments measured with proton and deuteron targets.

Table 3: Fragmentation functions access for different processes.

	e^+e^- annihilation	pp/$p\bar{p}$ collision	DIS
			
Dependence	$\hat{\sigma} \otimes FF$	$\hat{\sigma} \otimes PDF \otimes PDF \otimes FF$	$\hat{\sigma} \otimes PDF \otimes FF$
Separate D^h_q / $D^h_{\bar{q}}$	✗	✓	✓
Access parton kinematics	✓	✗	✓
Theoretical calculation	LO, NLO, NNLO	LO, NLO	LO, NLO

1.6.2 *Accessing the fragmentation functions in SIDIS*

Hadron multiplicities are defined as the cross-section ratio:

$$M^h(x, Q^2, z) = \frac{d\sigma^{lN\to l'hX}}{d\sigma^{lN\to l'X}dz} = \frac{d\sigma^h(x, Q^2, z)/dxdQ^2dz}{d\sigma^{DIS}(x, Q^2)/dxdQ^2}, \tag{28}$$

equivalent to the number of hadrons produced per DIS events allowing to access FFs by measuring hadron multiplicities.

Using the expressions of the DIS and SIDIS cross-sections (Eqs. 11 and 20) one obtains in the QPM:

$$M^h(x, Q^2, z) = \frac{\sum_q e_q^2 q(x, Q^2) \otimes D_q^h(z, Q^2)}{\sum_q e_q^2 q(x, Q^2)} \overset{LO}{=} \frac{\sum_q e_q^2 q(x, Q^2) D_q^h(z, Q^2)}{\sum_q e_q^2 q(x, Q^2)}. \tag{29}$$

As PDFs and FFs depend on different variables x and z one can write the convolution as a single product. By measuring $M^h(x, Q^2, z)$ for positive and negative hadrons, one can distinguish D_q^h and $D_{\bar{q}}^h$. The procedure of FFs extraction from COMPASS multiplicity measurement is described in Chapter 13.

1.6.3 *Global fits of multiplicity data and parametrizations of FFs*

As the FFs are universal quantities, a global QCD fit of available data on multiplicities from SIA, SIDIS and $pp/p\bar{p}$ collisions can be performed to give a general parametrization of the FFs. There are different parametrization available in the literature. Some parametrizations are only based on SIA data: KKP [24], KRE [48] and HKNS [49], when the AKK [50] parametrization uses in addition some hadron-hadron scattering data. One only uses SIDIS data: LSS [51]. The newest parametrizations from DSEHS [52, 53] include all three types of data and JAM [54] parametrizations include SIA+SIDIS. Each parametrization has its own set of assumptions based on symmetries and its different parametrization of D_q^h. A summary of the assumptions of the different groups can be found in Table 4. Only DSEHS and JAM will be described in more details in the following as they are the latest ones (or have been updated recently).

Table 4: Parametrization of FFs for pions and kaons.

Parametrization	Year	Data			# FFs fitted	
		SIDIS	$pp/p\bar{p}$	SIA	π	K
KKP [24]	2000	✗	✗	✓	5	5
KRE [48]	2001	✓	✗	✓	2	3
HKNS [49]	2007	✗	✗	✓	2	2
AKK [50]	2008	✗	✓	✓	3	5
LSS [51]	2014	✓	✗	✗	3	3
DSEHS [52, 53]	2017	✓	✓	✓	4	4
JAM [54]	2018	✓	✗	✓	3	3

DSEHS parametrization

DSEHS (previously DSS) was the first group, who determined individual FFs for quarks and antiquarks and the first to try to fit data coming from three different processes alltogether. The functional form they use for D_q^h is the following:

$$D_i^h(z,Q_0) = \frac{N_i^h z^{\alpha_i^h}(1-z)^{\beta_i^h}\left[1+\gamma_i^h(1-z)^{\delta_i^h}\right]}{B\left[2+\alpha_i^h, 1+\beta_i^h\right]+\gamma_i^h B\left[2+\alpha_i^h, 1+\beta_i^h+\delta_i^h\right]}, \tag{30}$$

where N_i^h, α_i^h, β_i^h, γ_i^h and δ_i^h are the fit parameters and B is the Euler beta function. For pions the two independent favoured FFs are related by a proportionality factor k. Moreover, isospin symmetry is considered only for the unfavoured FF ($D_{\bar{u}}^{\pi^+} = D_{\bar{d}}^{\pi^+}$) and the fragmentation of a strange quark into pion is related to the unfavoured FFs with z-dependent factor ($D_{\bar{s}}^{\pi^+} = D_s^{\pi^+} = N_s z^{\alpha_s} D_{\bar{u}}^{\pi^+}$). Thus four FFs are fitted for pions: $u+\bar{u}$, $d=\bar{u}$, $s+\bar{s}$ and g. For kaons, $D_{u+\bar{u}}^{K^+}$ and $D_{s+\bar{s}}^{K^+}$ are fitted independently to account for the fact that phenomenologically it is expected that the formation of secondary $s\bar{s}$, required to form K^+ from a u, should be suppressed. Previous fits from DSS showed that $D_{s+\bar{s}}^{K^+} > D_{u+\bar{u}}^{K^+}$, highlighting this fact. For the unfavoured FFs all distributions have the same functional form $D_{\bar{u}}^{K^+} = D_s^{K^+} = D_{\bar{d}}^{K^+} = D_{\bar{d}}^{K^+}$ as the data are unable to discrimate between flavours. Four FFs are fitted for kaons, as for the pions: $u+\bar{u}$, $s+\bar{s}$, $\bar{u} = d = \bar{d} = s$ and g. The FFs for heavy quarks (c,b) are also considered above their $\overline{MS}$ mass thresholds. The parameters are determined from a standard χ^2 minimization:

$$\chi^2 = \sum_{i=1}^{m}\left[\left(\frac{1-\mathcal{N}_i}{\delta\mathcal{N}}\right)+\sum_{j=1}^{m_i}\frac{(\mathcal{N}_i T_j - E_j)^2}{\delta E_j^2}\right], \tag{31}$$

where m is the number of datasets with m_i points each, E_j are the data points and δE_j their error and T_j the theoretical estimate for a given set. The normalisation factor $\mathcal{N}_i$ is defined as $\delta\chi^2/\delta\mathcal{N}_i = 0$.

Table 5: DSEHS FFs hypotheses for pions and kaons.

	Pions
Favoured	$D_u^{\pi^+} = N_{\pi^+} D_{\bar{d}}^{\pi^+} = D_{\bar{d}}^{\pi^-} = k_{\pi^-} D_{\bar{u}}^{\pi^-}$
Unfavoured	$D_{\bar{u}}^{\pi^+} = D_{\bar{d}}^{\pi^+}$
Unfavoured strange	$D_{\bar{s}}^{\pi^+} = D_s^{\pi^+}$
	$D_{\bar{d}}^{\pi^-} = D_u^{\pi^-} = D_{\bar{s}}^{\pi^-} = D_s^{\pi^-} = k_{\pi^-} D_u^{\pi^-}$
Gluons	$D_g^{\pi^+} = D_g^{\pi^-}$
	Kaons
Favoured	$D_u^{K^+}, D_{\bar{u}}^{K^-}$
Unfavoured	$D_{\bar{u}}^{K^+} = D_{\bar{d}}^{K^+} = D_{\bar{d}}^{K^+} = D_s^{K^+}$
	$D_u^{K^-} = D_{\bar{d}}^{K^-} = D_{\bar{d}}^{K^-} = D_{\bar{s}}^{K^-}$
Strange	$D_{\bar{s}}^{K^+}, D_s^{K^-}$
Gluons	$D_g^{K^+}, D_g^{K^-}$

The favoured, unfavoured and gluon FFs from DSS at LO for $Q^2 = 10\ \mathrm{GeV}^2$ are shown as function of z for π^+ in Fig. 11 and K^+ in Fig. 12.

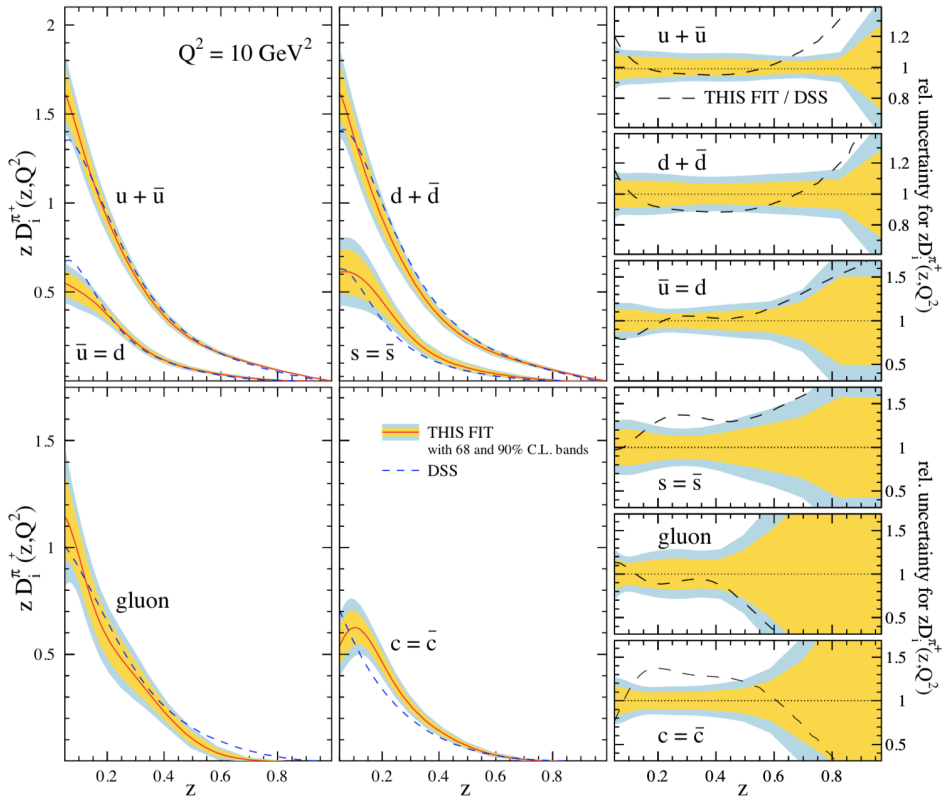


Figure 11: Individual FFs for positively charged pions $zD^{\pi^+}(z, Q^2)$ at $Q^2 = 10\,\mathrm{GeV}^2$ (solid lines) along with uncertainty estimates at 68% and 90% C.L. indicated by the inner and outer shaded bands, respectively. The panels on the right-hand-side show the corresponding relative uncertainties. Also shown is a comparison to previous DSS07 global analysis [55] (dashed lines). Figure taken from [52].

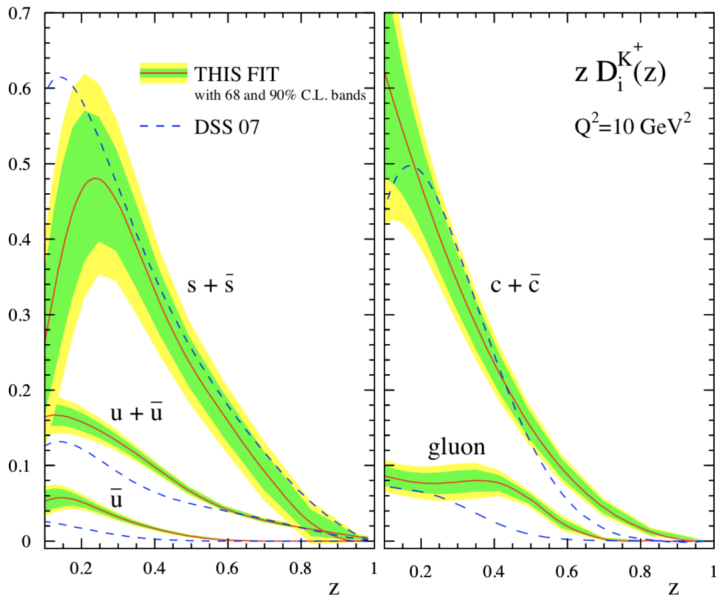


Figure 12: Individual FFs for positively charged kaons $zD^{K^+}(z, Q^2)$ at $Q^2 = 10\,\mathrm{GeV}^2$ (solid lines) along with uncertainty estimates at 68% and 90% C.L. indicated by the inner and outer shaded bands, respectively. Also shown is a comparison to previous DSS07 global analysis [55] (dashed lines). Figure taken from [53].

JAM parametrization

JAM is a Monte-Carlo based combined fit of PDFs and FFs using Bayesian statistics. In order to address some of the questions raised by the recent ambiguities in the strange quark FFs and their impact on the Δs determination, they go beyond the standard fitting paradigm by performing the first Monte Carlo (MC) analysis of PDFs and FFs, extending the methodology of the iterative Monte-Carlo (IMC) approach already used for the analysis of spin-dependent PDFs [56] to the case of FFs (Fig. 13). The IMC approach allows for a full exploration of the parameter space using Monte-Carlo sampling together with data resampling techniques and cross validation of the fit. In consequence it reduces considerably any bias introduced by fine-tuning or fixing specific parameters that are not well constrained by the data.

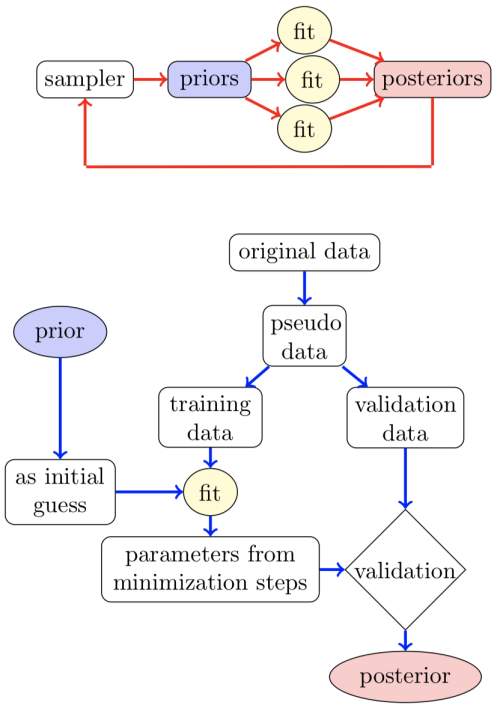


Figure 13: Workflow of the iterative Monte Carlo fitting strategy. In the upper diagram (red lines) an iteration begins at the prior sampler and a given number of fits are performed generating an ensemble of posteriors. After the initial iteration, with a flat sampler, the generated posteriors are used to construct a multivariate Gaussian sampler for the next iteration. The lower diagram (with blue lines) summarizes the workflow that transforms a given prior into a final posterior. Figure taken from [56].

The functional they use for D_q^h is the following:

$$D_i^h(z, Q_0; \mathbf{a}) = M \frac{z^\alpha (1-z)^\beta}{B(2+\alpha, 1+\beta)} \tag{32}$$

where $\mathbf{a} = M, \alpha, \beta, \gamma$ is the vector of shape parameters to be fitted and B is the Euler beta function. The denominator is chosen so that the coefficient M corresponds to the average momentum fraction z. Isospin symmetry is considered for all partons. Using $D_{q^\pm}^{h^+}(z, Q^2) = D_q^{h^+}(z, Q^2) \pm D_{\bar{q}}^{h^+}(z,$ allows JAM to consider two templates functions for the FFs $D_{u^+}^{\pi^+} = D_{d^+}^{\pi^+}$, $D_{u^+}^{K^+}$ and $D_{s^+}^{K^+}$, which contain both favoured and unfavoured distributions and only one template function for the rest of unfavoured distributions viz. $D_{\bar{u}}^{\pi^+} = D_{\bar{d}}^{\pi^+}$, $D_{\bar{s}}^{\pi^+} = (1/2) D_{s^+}^{\pi^+}$, $D_{\bar{u}}^{K^+} = (1/2) D_{d^+}^{K^+}$ and $D_{\bar{s}}^{K^+}$ along with the heavy quarks and gluons[3].

The favoured and unfavoured FFs from JAM at LO for $Q^2 = 5$ GeV^2 are plotted as a function of z for π^+ and K^+ (Fig. 14). The strange quark fragmentation into K^+ from JAM17

3 The choice of the factor 1/2 is motivated by data.

is compared with DSS07 and HKNS results. While part of the disagreement between fits can be explained by the different parametrizations used, the large uncertainties on the data and the PDFs also play a role.

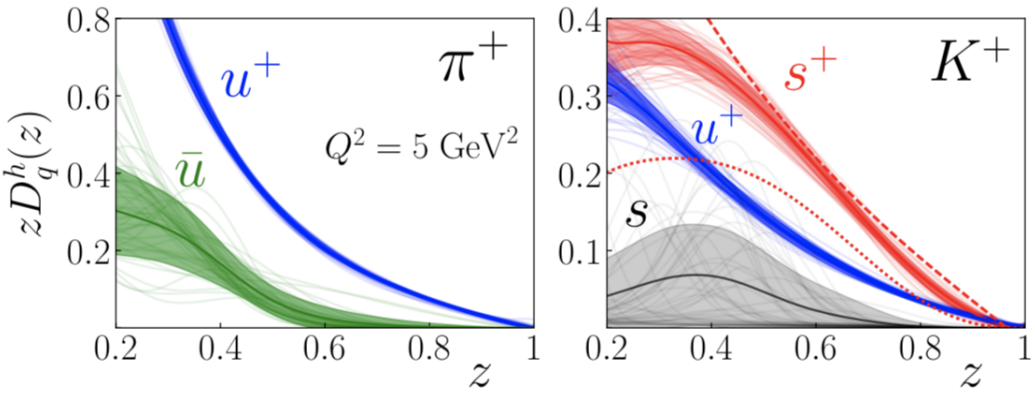


Figure 14: Fragmentation functions zD_q^h to π^+ (left panel) and K^+ (right panel) for u^+ (blue), $\bar{u}$ (green), s^+ (red) and s (grey) at $Q^2 = 5$ GeV^2 for the JAM17 analysis, compared to $s^+ \to K^+$ from DSS07 (dashed line) and HKNS (point line). Figure taken from [54].

1.7 SUMMARY

The DIS process is a very interesting channel for the study of the nucleon structure. The spin averaged PDFs, mostly determined from inclusive DIS, are well constrained in a wide kinematic domain for the first generation of quarks. Going to higher masses, large uncertainties still subsist e.g. for s and $\bar{s}$. In SIDIS one gets additional access to FFs, which are universal quantities and parametrize quark hadronization.

Several groups have already issued parametrization of quark FFs based on LO and NLO analyses of various data sets. They differ significantly in the strange quark sector. This is why COMPASS did a new measurement and why the analysis presented in this was done.

2

RENORMALIZATION AND QED RADIATIVE CORRECTIONS

In Chapter 1 only 1-photon exchange in QED was discussed. However experimental measurements contain all possible processes e.g. multiple real photon emission. They can be described by the following framework.

2.1 DIVERGENCES, REGULARIZATION AND RENORMALIZATION IN QED

When calculating Feynman loop diagrams in QED at some point the integrals will diverge. To cure these divergences we have to go through the process of renormalization. Renormalization is a formal manipulation, embedded inside the quantum field theory formalism, which allows us to calculate finite testable expectation values and scattering amplitudes. This implies that there may be high energy virtual particles contributing to the loop diagrams and they would impact the divergence of the integral. Thus we could say that renormalization is a procedure allowing to calculate reasonably the effects of the low-energy physics independently to what happens at high energies.

To describe the renormalization process, consider as an example the electron self-energy. Using the full electron propagator $G_F(p)$ including order-by-order in the perturbation theory the corrections to the Feynman-Green function [57] yields :

$$\begin{aligned} G_F(p) &= \quad + \quad + \mathcal{O}(e_0^4) \\ &= \frac{i}{\not{p} - m_0 + i\epsilon} + G_F^{(1)}(p) + \mathcal{O}(e_0^4) \end{aligned} \tag{33}$$

where e_0 and m_0 are the *bare* charge and mass parameters and $G_F^{(1)}(p)$ is a divergent term. The renormalization procedure then follows three steps:

1. **Regularization**: set a new *finite* integral $G_F^{(1)}(p, \Lambda)$ with a dependence with the *cut-off scale*[1] Λ yielding:

$$G_F^{(1)}(p, \Lambda) \xrightarrow{\Lambda \to \infty} G_F^{(1)}(p) \tag{34}$$

   This integral has a divergent and finite part:

$$G_F^{(1)}(p, \Lambda) = I_{div}(p, \Lambda) + I_{fin}(p, \Lambda) \tag{35}$$

   The finite part $I_{fin}(p, \Lambda)$ leads to physically measurable effects and is the *radiative correction*.

2. **Renormalization**: if the theory is *renormalizable* the divergent part can be combined with the tree-level propagator:

$$\begin{aligned} G_F^{(1)}(p, \Lambda) &= \frac{i}{\not{p} - m_0 + i\epsilon} + I_{div}(p, \Lambda) + I_{fin}(p, \Lambda) + \mathcal{O}(e_0^4) \\ &= \frac{iZ_2(\Lambda)}{\not{p} - m(\Lambda) + i\epsilon} + I_{fin}(p, \Lambda) + \mathcal{O}(e_0^4) \end{aligned} \tag{36}$$

   Thanks to *wavefunction renormalization* $Z_2(\Lambda)$ and a renormalized mass parameter $m(\Lambda)$, the divergent terms can be integrated into the tree-level propagator.

1 A cut-off is often introduced in text books to explain the procedure. However, all modern practical calculations are performed with *dimensional regularization*.

3. **Removing Λ-dependence**: By taking $\Lambda \to \infty$, the bare parameters e_0 and m_0 become singular allowing the physical e and m to be finite. The perturbation expansion is a series in e and not in e_0.

This procedure is well defined and gives consistent finite physics results, whatever regularization is chosen. By definition, a theory is said to be *renormalizable* if all divergences can be removed by renormalization of a finite number of parameters in the Lagrangian.

2.2 QED RADIATIVE CORRECTIONS

Radiative corrections had a key role in the development of QED: they enable one to calculate cross-sections with extremely high precision that has until now not been contradicted by any experiment. When a process like DIS involves charged particles, a more general hadron current has to be used than in Chapter 1 leading to the emission of more than one photon. Among these photons there are also real photons emitted at the incoming and outgoing particles.

If one were to compute the cross section of the process $1+2 \to 3+4$ with no photon emission (Born level process), one would find a different result $\sigma_{1\gamma}$ than from the measurement σ_{meas}. To obtain a more accurate result, one has to consider the processes $1+2 \to 3+4+\gamma_1+\gamma_2+..+\gamma_n$. Taking these corrections to Born level process into account, one obtains:

$$\sigma_{meas} = (1+\delta_{RC})\sigma_{1\gamma}, \tag{37}$$

where δ_{RC} are the radiative corrections.

In this thesis, I will not discuss the corrections beyond first order. These first order corrections are also known as order α ($\mathcal{O}(\alpha)$) corrections. They comprise:

- Lepton radiation
- Hadron radiation
- Interference of lepton/hadron radiation (two-photon exchange)
- Vacuum polarization
- Weak corrections

The goal is to quantify the effect of radiative corrections. The measured cross-section can be expressed as the convolution of the 1γ cross-section times a function called the radiator function which takes into account the radiative effects [58]:

$$d\sigma^{meas}(p,q) = \int \frac{d^3k}{2k^0} R(l,l',k)\, d\sigma^{1\gamma}(p,-q,k). \tag{38}$$

A similar relation holds also for the structure functions:

$$F_n^{meas}(x,Q^2) = \int d\tilde{x}\, d\tilde{Q}^2 R_n(x,Q^2,\tilde{x},\tilde{Q}^2) F_n^{1\gamma}(\tilde{x},\tilde{Q}^2). \tag{39}$$

The previous formulas are valid for emission of one additional photon but can be extended to include higher-order multi-photon emissions. As one has access to both observed quantities and radiator function, the determination of the 1γ cross-sections or structure functions from measured ones can be done by unfolding using an iterative procedure. The principal drawbacks of such a method is that the solution is ill-defined: there is no unique solution, there

are large uncertainties and the process is numerically unstable. To stabilize the calculation of the convolution table, i.e. the *folding*, partial fractioning is used on the radiator function:

$$R(l, l', k) = \frac{I}{k.l} + \frac{F}{k.l'} + \frac{C}{\tilde{Q}^2}. \tag{40}$$

The partial fractioning is splitting the radiator function in three:

- Initial state radiation (I fraction)
- Final state radiation (F fraction)
- Compton peak (C fraction)

For each of them, an observation can be made. For initial state radiation (ISR), $k.l$ is small for $\angle(l_{in}, \gamma) \to 0$, for final state radiation (FSR), $k.l'$ is small for $\angle(l_{out}, \gamma) \to 0$ and eventually for Compton peak Q^2 is small for $p_T(l_{out}) \simeq p_T(\gamma)$.

For ISR and FSR, the photon is emitted within narrow cones with width of the order of $\sqrt{\frac{m_t}{E_t}}$. The radiated photon can be collinear, but it is not always collinear. The photon can be collinear independent of whether the radiating particle has a finite mass or is massless. Collinear radiation leads to a divergence if the radiating particle is massless.

Two additional notes have to be made:

- As $E^2_{\gamma,max} \propto Q^2 \frac{1-x}{x}$, the largest radiation in energy are at large Q^2 and small x. Radiation is suppressed at small Q^2 and large x. There are also large negative corrections from uncancelled virtual contributions.
- As $\tilde{Q}^2_{min} = \frac{x^2}{1-x} M^2_N$, the case where $\tilde{Q}^2_{min} \ll Q^2$ is possible.

All the preceding explanations did concern lepton radiation. One should also address the question of the hadron corrections (quark line radiation). These corrections are infrared divergent (radiation of soft photons and gluons) but they cancel with loops. The emission of the photon/gluon can be collinear and gives rise to correction of type $\frac{\alpha}{2\pi} \log(m^2_q)$. For quarks, the approximation $m_q \approx 0$ is giving rise to divergent corrections. One way to solve this issue is to factorize and absorb the divergences into the PDFs and the FFs:

$$d\sigma = \sum_f d\hat{\sigma}_f (1 + \delta_f(Q^2, m^2_q)) q_f(x) = \sum_f d\hat{\sigma}_f \hat{q}_f(x, Q^2). \tag{41}$$

2.2.1 *Characterization and impact of radiative corrections in analysis*

The description of a radiative event is given by the following: an event is called radiative as soon as it contains one real radiated photon which is emitted at the lepton line (Fig. 15).

In the following, we will only consider these corrections (Fig. 16):

- Internal Bremsstrahlung (from both incoming and outgoing leptons) (b,c)
- Vertex correction (d)
- Vacuum polarization (e)

Correction to the quark line are not included in calculations, as explained in Section 8.1. If we call σ_{Born} the cross-section of the tree-level diagram and $\sigma_{Born+\mathcal{O}(\alpha)}$ the cross-section of tree-level plus the first order correction enumerated above, the definition of the radiative corrections factor η is:

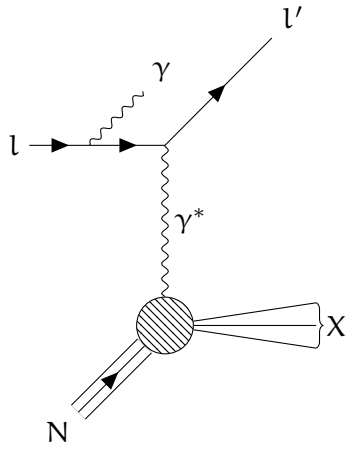


Figure 15: Typical diagram of a radiative event. One can note that the pair (Q^2, ν) at the vertex (called hadronic) is not the same as the one calculated using the incoming and outgoing lepton (called leptonic). The relation between the two pairs is drawn by: $\nu_{had} = \nu_{lep} - E_\gamma, Q^2_{had} = Q^2_{lep} + 2E_\gamma(\nu_{lep}\sqrt{\nu^2_{lep} + Q^2_{lep}}cos\theta_\gamma)$

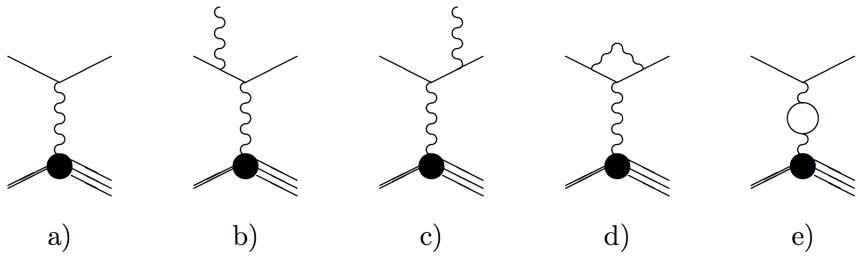


Figure 16: List of the diagrams used for the calculation of the radiative corrections. From left to right, tree level, internal bremsstrahlung (incoming and outgoing leptons), vertex correction and vacuum polarization.

$$\eta(x, y) = \frac{\sigma_{Born}(x, y)}{\sigma_{Born+\mathcal{O}(\alpha)}(x, y)} \tag{42}$$

Obviously, the emission of a real photon is modifying the kinematic variables of the event. Let us take the case of one DIS event and a second one, which is exactly like the first one except there is an ISR. The kinematics at the DIS vertex are not the same as measured from the lepton variables. In the case of multiplicities, this discrepancy in the kinematic variables induces bin migration of some hadrons. Applying the correction factor η to the multiplicities is redirecting the hadrons to the right (x,y) bins.

2.2.2 *About emission of radiative photons*

There are two privileged angles (Fig. 17) for emission of a real photon:

- One in the direction of the incident lepton (s-peak)
- One in the direction of the outgoing lepton (p-peak)

In the case of muons, note that the s and p peaks are much less pronounced than for electrons. This knowledge will later be useful to verify the consistency of the new results.

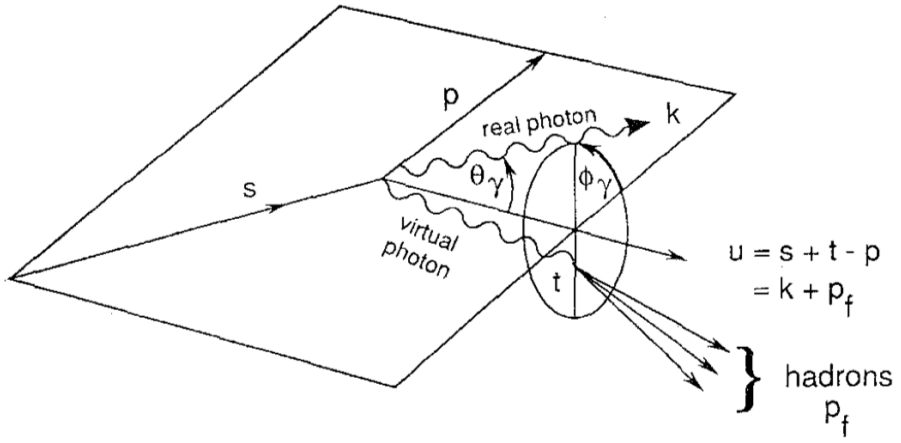


Figure 17: Angles characterizing the emission of a radiative photon. The plane is defined by the incoming lepton (s) and the outgoing lepton (p). θ_γ is the polar angle and Φ_γ the azimuthal angle. The different 4-momenta are: s for incident lepton, p for scattered lepton, t for target proton, k for real photon, p_f for hadronic final state. Figure taken from [59].

2.2.3 *About the radiative tail*

For elastic scattering, radiation of a real photon still happen, modifying the lepton variables of the events: this is the so-called *radiative tail*. The contribution σ_{tails} of processes where real photon of energy larger than the cutoff parameter Δ are emitted yields [59] :

$$\frac{d^2\sigma_{tails}}{d\nu d\Omega} = \frac{d^2\sigma(\omega > \Delta)}{d\nu d\Omega} = \int_{M^2}^{M^2_{j,max}} \frac{d^2\sigma_{j,r} M_j^2}{d\nu d\Omega} dM_j^2, \tag{43}$$

where M denotes a target mass, $M_{j,max} = \sqrt{M^2 - Q^2 + 2M(\nu - \Delta)}$ and $\sigma_{j,r}$ denotes the radiative tail from the j^{th} mass level. M_j is an effective mass of the hadronic state. The integration over M_j means in Eq. 43 that all the final hadronic states contribute to the cross section measured in a kinematic point (Q^2, ν), as seen in Fig. 18, e.g. elastic ($M_j = M$), resonance production ($M_j = M_{res}$) or deep inelastic tails i.e. tails from the continuum. This kinematic point is called the radiative tail.

2.3 SUMMARY

The renormalization is a necessary procedure when it comes to compute cross-sections as it allows to cure the divergences from the calculation and gives consistant results when compared to measurements. From this renormalization procedure arise radiative corrections which imply the emission of a real photon in the final state. These corrections can be splitted in three groups: from the Initial State Radiation (ISR), from the Final State Radiation (FSR) and from the Compton Peak. The emission of a real photon implies that there is a difference between the hadronic and leptonic variables that must be taken into account. Radiative correction factor are computed to measure the effect of the kinematic bin migration and correct the data accordingly.

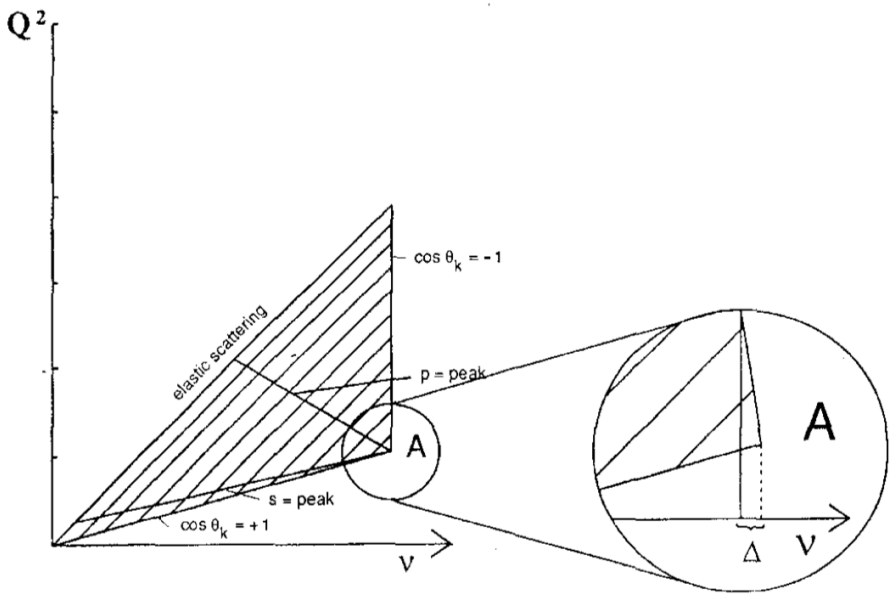


Figure 18: Range of kinematical variables from which the radiative tails contribute to the cross section measured at the point $A(Q^2, \nu)$. Figure taken from [59].

Part II

EXPERIMENTAL PART

3

THE COMPASS EXPERIMENT AT CERN

In this chapter a description of the COMPASS experiment is provided. The general features of the spectrometer are given in Section 3.1. The beam and target are presented in Section 3.2. The descriptions of the detectors used for the tracking and the ones used for the particle identification are done in Section 3.4, respectively. The trigger system is discussed in Section 3.6. The last sections describe data acquisition and reconstruction.

3.1 GENERAL OVERVIEW

COMPASS is a high energy, high rate fixed-target experiment at the Super Proton Synchrotron (SPS) at CERN. It is dedicated to the study of hadron structure and hadron spectroscopy with high intensity muon and hadron beams.

In order to cover the necessary large range in Q^2 and x for the available beam energy the COMPASS spectrometer as shown in Fig. 19 covers a large momentum and angular range. This is obtained by usinf a two-stage spectrometer for detecting outgoing particles.

The apparatus is divided in three parts: the first part is dedicated to the detection of the incoming beam and is located upstream the target location. The second and third part are located downstream of the target and represent a length of 50 meters. The second part called the *Large Angle Spectrometer* (LAS) is built around the magnet SM1. The LAS has been designed to provide a 180 mrad acceptance. The *Small Angle Spectrometer* (SAS), built around the magnet SM2, measures the particles emitted at small angles ($\pm$ 30 mrad).

In 2016, the data taking was performed with a 160 Gev/c muon beam scattering off a liquid H_2 target.

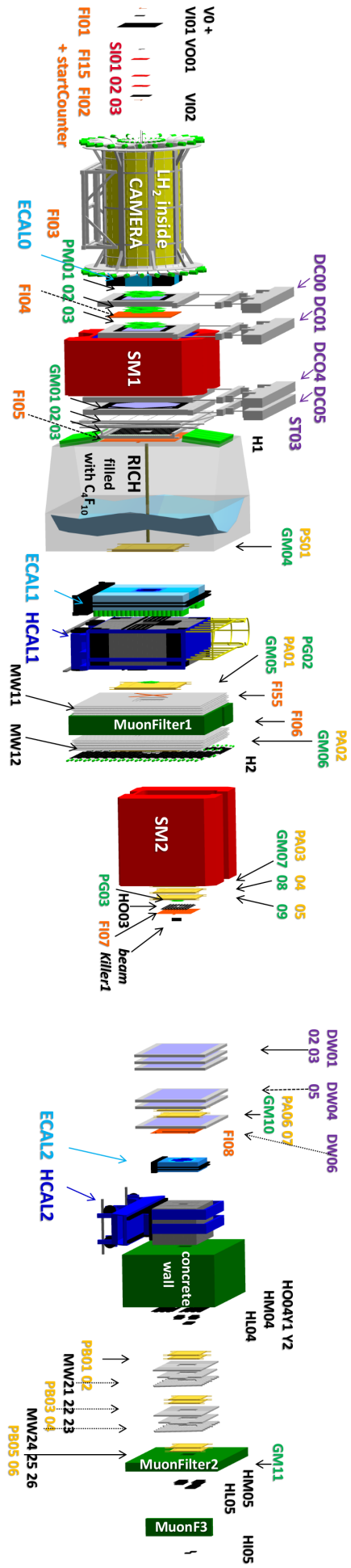
3.2 BEAM

The muon beam used by COMPASS is obtained from a primary proton beam accelerated in the SPS to 400 GeV/c. The proton beam interacts with T6 target, a 50 cm thick beryllium target, producing mainly pions and kaons. The spill time, which is the time window, within which the proton beam is delivered to the T6 target, was of 4.8 s. In each cycle of 36 s there were two spills.

3.2.1 *The M2 Beam Line*

The hadrons produced at T6 are transported in the 600 m long decay channel of the M2 beamline [61, 62]. During this time, 5% of the pions and kaons are decaying into muons and neutrinos. At the end of this 600 m decay section, the remaining hadrons are stopped by a hadron absorber and the muons are focused. A system of magnets is then used to select and focus the muons of 160 GeV/c.

The beam has transverse dimensions of $\sigma_x \times \sigma_y \sim 8 \times 8$ mm^2 and an angular divergence of $\sigma_{\theta_x} \times \sigma_{\theta_y} \sim 0.5 \times 1$ mrad^2 in the experimental area. At each spill, $2 \cdot 10^8$ muons enter the experimental area. The beam is accompanied by a muon halo that extends transversely up to several meters of distance with respect to the beam line. The intensity of this halo decreases with the distance. The halo near the beam line as measured by a 30×30 cm^2 dedicated veto counter with a 4 cm diameter central hole represents about 16% of the muon beam. The far



(a) LAS (b) SAS

Figure 19: COMPASS 2016/2017 muon setup side view. The two stages of the spectrometer at different scales in this drawing. Taken from [60].

halo or low intensity halo is measured by a large veto counter with a central hole of 30×30 cm^2. It represents about 7% of the muon beam.

3.2.2 *The Beam Momentum Station*

The Beam Momentum Station (BMS) illustrated in Fig. 20 is used for the determination of the incident muon momentum. It consists of six scintillators hodoscopes (BM01-BM06) located asymmetrically upstream and downstream a bending magnet (B6: three consecutive dipole magnets) surrounded by four quadrupoles (Q29-Q32).

The BMS system was designed to measure the momentum of more than 10^8 individual particles per spill with a relative precision of 0.5%. To eliminate the ambiguities in the reconstruction of particle trajectories, their time of transit is measured with a resolution of 50 ps.

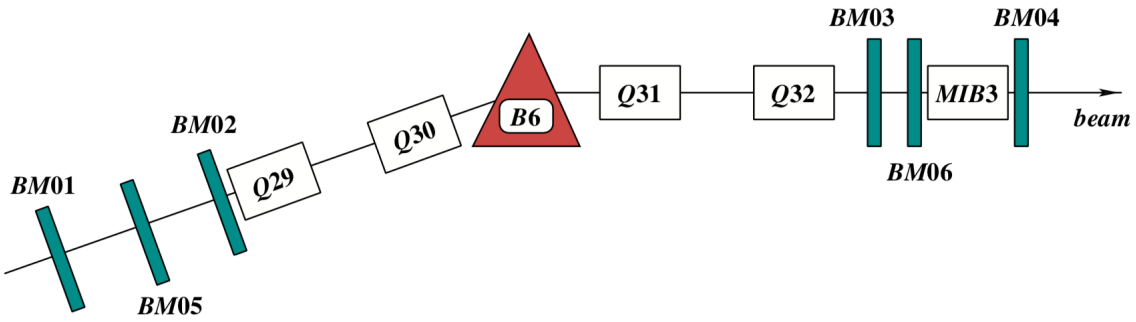


Figure 20: Layout of the Beam Momentum Station for the COMPASS muon beam. Taken from [63].

3.3 TARGET

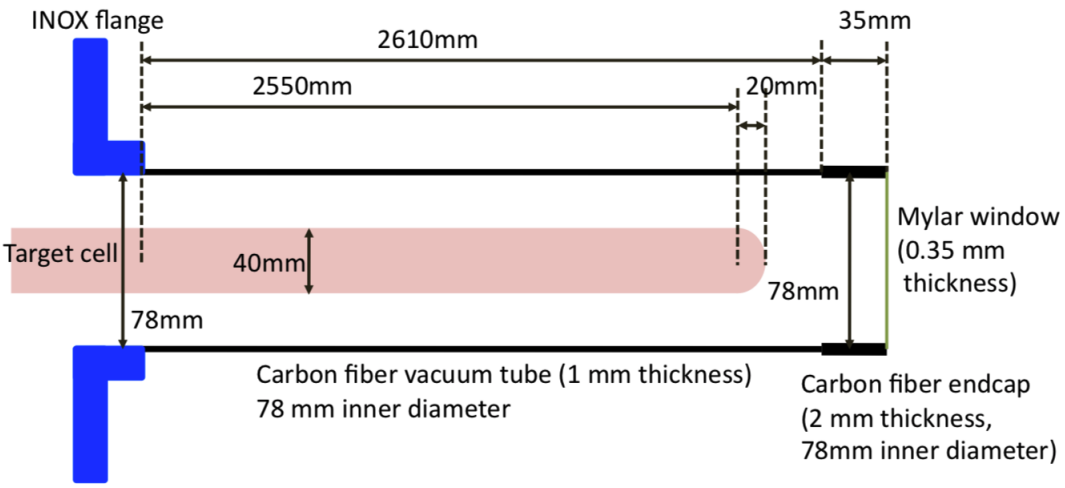


Figure 21: Target geometry for the 2016/2017 setup.

The target is liquid H_2 contained in a 2.5 meter long cylinder (Fig. 21). The target material is contained in a mylar tube with a diameter of 4 cm. The total volume of the target cell and the liquid hydrogen system is located in a cryostat made of carbon fiber. The operation temperature of hydrogen is 18 K with a pressure of 1020 mbar. The material for the target cell was chosen in order to minimise proton absorption.

3.4 TRACKING DETECTORS

The *Large Angle Spectrometer* (LAS) and *Small Angle Spectrometer* (SAS) are each equipped with different type of tracking detectors. Depending on the direction coverage area of each detectors they are classified as: *very small area trackers*, *small area trackers* and *large area trakers*

3.4.1 *Very small area trackers*

The very small area trackers cover the transverse beam size up to $\sim$ 3 cm. In this region the particle rate is very high (10^5/s/mm^2 in the center of the muon beam), hence the tracking detectors must have an excellent time and position resolutions.

The scintillating fibre detectors are used at several locations of the experiment and cover areas between $\sim$ 16 cm^2 and 144 cm^2. They are fabricated from 0.5-1 mm diameter fibres and reach a time resolution better than 500 ps. All along the apparatus there are 9 stations composed by two or three scintillating fiber detectors. The third detector is always pivoted by 45° with respect to the others.

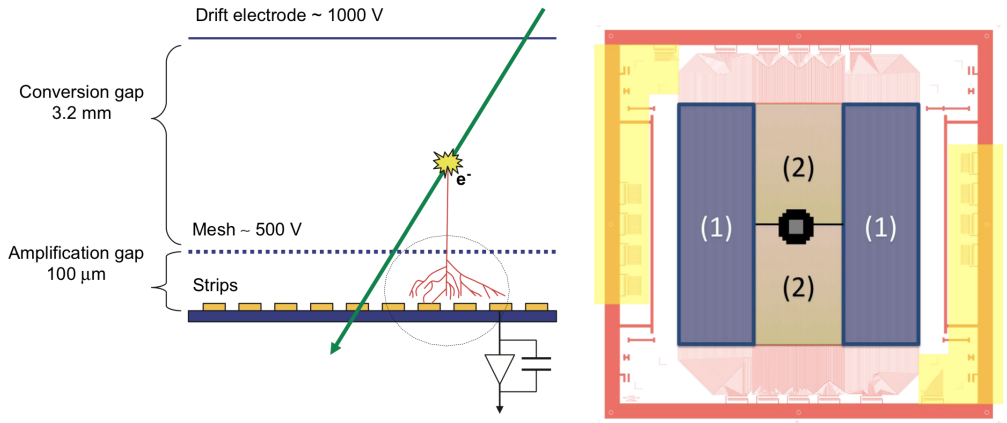
The silicon detector size is 5x7 cm^2 with a space and time resolution of $\sim$ 10 μm and $<$ 2.5 ns. The 3 silicon detectors stations are located upstream the target. The stations are each composed by two silicon detectors, the second one being rotated by 5° with respect to the other, each detector measuring perpendicular views.

3.4.2 *Small area trackers*

The radial region between 2.5 cm and 20 cm is covered by two types of gaseous detectors : PixelMicromegas (MICRO MEsh GAseous Structure) and (Pixel)GEM (Gas Electron Multiplier) detectors. These detector have a high rate capability ($\sim 10^4$/s/mm^2) and good spatial resolution ($<$ 100 μm). They also present a minimal material budget.

The principle of the PixelMicromegas is explained in Fig. 22 (a). The particle ionizes the gas in the conversion gap, the produced electrons drift in a moderate field of 1.5 kV/cm to prevent secondary ionization, towards the amplification gap. The field in the amplification area is large enough to accelerate the electrons to produce an avalanche. The conversion and amplification gaps are separated by a *micromesh*, which collects the positive ions produced during the avalanche in a short period of time ($<$ 100 ns). This feature is possible because of the small width of the gap ($\sim$ 100 μm). The PixelMicromegas have an active area of 40 $\times$ 40 cm^2 with a central area made of 1280 pixels of 400 μ m $\times$ 2.5 mm and 400 μ m $\times$ 6.5 [64] as shown in Fig. 22 (b). All PixelMicromegas detectors operate with a detection efficiency of 98% and with a spatial resolution of better than 100 μm. There are three PixelMicromegas stations located at LAS. They are composed by four detectors each with different directions: horizontal (X), vertical (Y), and two (U,V) rotated by $\pm$ 45° with respect to the vertical. Each plane has an active area of 40 $\times$ 40 cm^2.

A GEM is a 50 μm thin polyimide foil with Cu cladding on both sides, into which numerous microholes ($\sim 10^4$/cm^2) with a diameter of 70 μm have been chemically etched using lithographic techniques. A high voltage (several 100 V) is applied between the surfaces of the foil to generate the avalanche multiplication of electrons through the holes. The fast signal is induced by the electron cloud emerging from the last GEM foil on an anode segmented into two sets of 768 orthogonal strips (pitch of 400 μm). The COMPASS GEM detection principle is shown in Fig. 23: it consists of three GEM amplification stages separated by thin grids of 2 mm height (transfer gap). Using several GEMs in a stack allows to split the gas gain over several GEMs and reduce the voltage across the two sides of the used GEM foils thus achieving the same gain while reducing the probability for discharges. The COMPASS PixelGEM



(a) (b)

Figure 22: (a) COMPASS MicroMegas detection principle. Taken from [63]. (b) PixelMicromegas detector geometry. (1) Tracks of 40 cm × 480 µm; (2) Tracks of 20 cm × 400 µm; In the center is the pixellized area. Taken from [64].

are GEMs with a pixellized central area. A GEM station is composed by 2 detectors oriented by 45° relatively to each other. All in all there are 11 GEM stations and 2 PixelGEM stations located after SM1 and at SAS. The active area for the GEM is 31×31 cm^2 and the central area of 5 cm diameter can be activated to align the detector with low intensity beams. The PixelGEM have an active area of 10×10 cm^2 with a central area of 3.2×3.2 cm^2 filled with 1 mm^2 squared pixels. The detectors efficiency is ~ 97% with a spatial and time resolution of about ~ 70 µm and 12 ns, respectively.

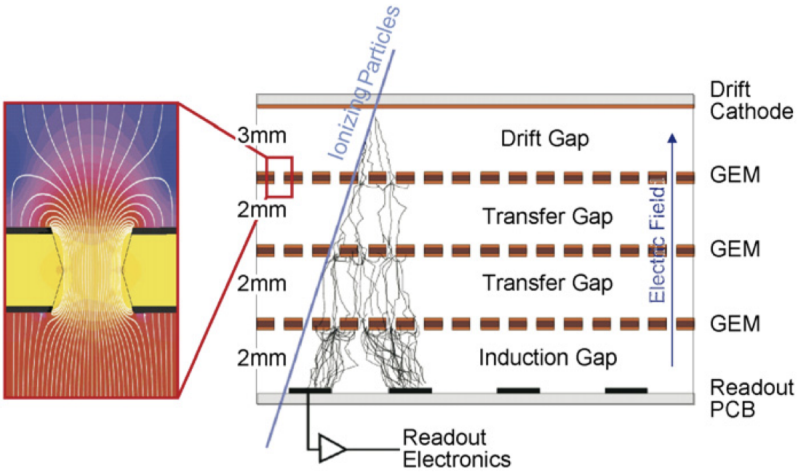


Figure 23: COMPASS GEM detection principle. Taken from [63].

3.4.3 *Large area trackers*

The large area trackers cover all the remaining spectrometer acceptance with a good spatial resolution. As the particle rate in the region covered by the large area trackers is small in comparison to the central region (10^2/s/mm^2), the use of detectors such as drift chambers (DCs and W4/5) [65], straw drift tubes [66] and multiwire proportional chambers (MWPCs) is possible. These detectors have large active size area (~ m^2) with a central dead area of few cm^2.

Each DC consists of eight layers of wires with four different inclinations: horizontal, vertical and rotated $\pm$ 20° with respect to the vertical direction (X, Y, U and V). Two consecutive planes with the same inclination are staggered by 3.5 mm to disimbiguate left-right and the ordering of the planes with different orientations is such to minimize the fake track combinations. The detectors are filled with a gas mixture of Ar, C_2H_6 and CF_4 at a volume ratio 9 : 9 : 2. Two of the DC are located before SM1 and have an active area of 180×127 cm^2 . the last two are located downstream SM1 and have a larger active area of 204×204 cm ^{2}. All these DCs have a central dead area of 30 cm. The central dead area can be activated for alignment needs with a low intensity beam. The average resolution of a DC is 270 µm and the efficiency above 95%.

The W4/5 detectors have an active area of 5 x 2.5 m^2, and consist of 4 anode wire layers with a wire pitch of 4 cm. The anode wires are separated by layers of cathode wires with a pitch of 2 mm. The diameter of the anode wire is 20 µm and of the potential wires, 200 µm. A CF_4-based gas mixture, $Ar/CF_4/CO_2$ (85/10/5), is used.

A straw detector station consists in 3 straw detectors with different orientations: horizontal, vertical and rotation by 10° with respect to the vertical. The only station used is located between SM1 and SM2. Each detector is composed by two layers of straw tubes with the same orientation. The straw tubes consist in two layers of thin plastic film, one coated with carbon loaded Kapton, the other one with aluminised Kapton foil. The active area for the straw detector is 320×280 cm^2 and have a central dead area of 20×20 cm. The average resolution is of 190 µm.

There are three types of MWPCs in COMPASS, which differ by the number of layers, the size of the dead area for the beam and the combination of the measured projections (X, Y, U and V). The active area is of $178 \times (90 - 180)$ cm^2. All layers have a wire length of about 1 m, a wire diameter of 20 µm and a pitch of 2 mm and are enclosed on both sides by graphite-coated Mylar foils. The central deadarea of each detector increases with respect to the detector position, from 16 to 22 cm. The average spatial resolution of the MWPC is of 1.6 mm.

Table 6: Table with the characteristics of a selection of tracking detectors.

Detector type	Active area	Spacial resolution	Time resolution
Scintillating Fibre	$(3.9)^2$ - $(12.3)^2$ cm^2	130 - 210 µm	400 ps
Silicon Micro-strip	5 x 7 cm^2	8 - 11 µm	2.5 ns
(Pixel)GEM	31 x 31 cm^2	70 µm	12 ns
PixelMicromegas	40 x 40 cm^2	90 µm	9 ns
MWPC	178 - $(90 - 120)$ cm^2	1.6 µm	N/A
DC	180 - 127 cm^2	190 - 500 µm	N/A
Straws	280 - 323 cm^2	190 µm	N/A

3.5 PARTICLE IDENTIFICATION

Following the nature of the particles, several techniques are used to identify them. Two types of calorimeters are used to measure the energy of the hadrons, photons and electrons: hadron calorimeters (HCAL1 and HCAL2) separate hadrons and muons and electromagnetic calorimeters (ECAL0, ECAL1 and ECAL2) detect and identify photons. Two muon wall detectors (MW1 and MW2) are used together with a hadron absorber for muon identification (muon filter). A RICH detector allows to separate between pions, kaons and protons in the

momentum range from 3 to 50 GeV. While the RICH will be further described in a dedicated chapter (Chapter 5), the other identification detectors will be briefly described in the following subsections.

3.5.1 *Hadron Calorimeters*

A hadron calorimeter allows to separate hadron and muon tracks using the energy deposit. Contrary to a hadron, which deposits almost all its energy via a hadron shower, the muon suffers energy loss only depositing a small energy fraction. HCAL1 and HCAL2 are sampling calorimeters with a modular structure with iron and scintillator plates and are located before the muon filters. Their threshold depends on the energies: for HCAL1 for hadrons with momenta above 5 GeV/c it is almost constant and close to 100% when for HCAL2 the same efficiency is reached for hadrons with momenta above 10 GeV/c.

3.5.2 *Electromagnetic Calorimeters*

The electromagnetic calorimeters are used to measure the energy of electrons and photons. ECAL0 is made of radiation-hard Shashlyk-type lead/scintillator modules. ECAL1 and ECAL2 are formed by blocks of lead glass connected to photomultipliers with light guides. An electromagnetic shower is initiated when the incoming electrons or photon reach the calorimeter. This electromagnetic shower produces Cherenkov radiation inside the lead glass and this light intensity is proportional to the energy deposited. The inner-most part of ECAL2 has Shashlyk modules.

3.5.3 *Muon Identification with Muon Walls and Muon Filters*

An efficient way to identify muons is to use an absorber surrounded by two tracking detectors. With a radiation length large enough to absorb all hadrons, particles detected behind the absorber are considered muons. At COMPASS, this is done in the LAS with the Muon Wall 1 (MW1) and the Muon Filter 1 (MF1). In the SAS, the Muon Wall 2 (MW2) in combination with the Muon Filter 2 (MF2) identify the muons. At the very end of the spectrometer, the Muon Filter 3 (MF3) is the last muon filter detector. The three muon filters are made of iron or concrete. The MW1 system consists of Mini Drift Tubes. The tubes are made of 0.6 mm thick aluminum tubes surrounding a 50 μm thick tungsten wire. The muon filter surrounded by the MW1 system is made of 60 cm of iron. The active areas are 4845 x 4050 mm^2 (hole: 1445 x 880 mm^2) and 4730 x 4165 mm^2 (hole: 1475 x 765 mm^2) for the X and Y planes. The gas mixture of MW1 is Ar/CO_2 (70/30). The MW2 system in the SAS has two identical stations of layers of drift tubes. Each of the two stations consists of 6 layers with an active area of 4470 x 2020 mm^2. A gas mixture of Ar/CH_4 (75/25) is used. The stainless steel drift tubes have an inner diameter of 29 mm and a wall thickness of 0.5 mm and the wires are 50 μm thick. In the central region MWPCs complements the coverage of the muon acceptance.

3.6 THE TRIGGER SYSTEM

The trigger system [67] has the task to select physic event candidates in a high rate environment. It is composed by scintillator hodoscope, complemented by scintillator veto detectors to suppress halo muons and by calorimeters to select events with hadron production.

Table 7: COMPASS triggers with the muon beam in 2016.

Trigger name	Components
Middle Trigger (MT)	HM04, HM05
Ladder Trigger (LT)	HL04, HL05
Outer Trigger (OT)	HO03, HO04
LAS Trigger (LAST)	H1, H2

Depending on the event kinematics two different algorithms are used to select the scattered muons kinematics. For events with $Q^2 > 0.5$ (GeV/c^2) the vertical scattering is measured using two hodoscopes stations (vertical target pointing) as illustrated in Fig. 24.

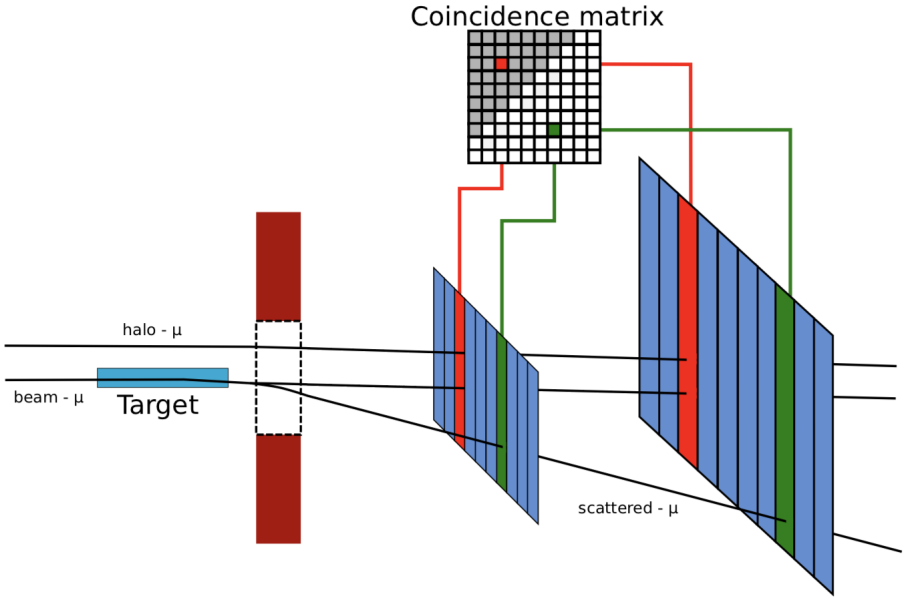


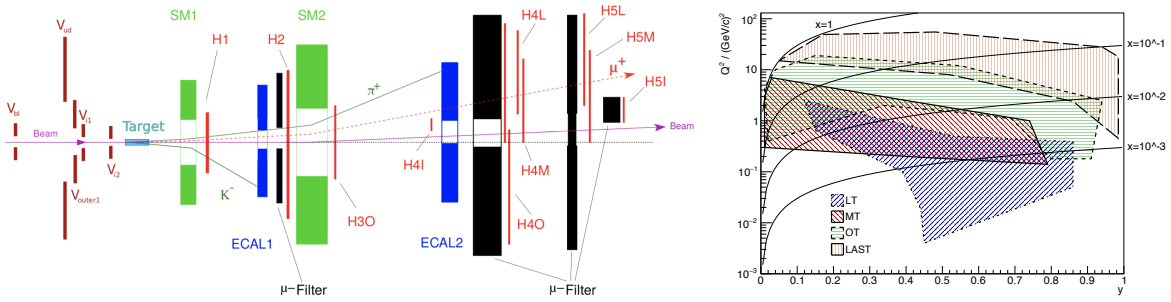
Figure 24: Concept of the trigger. The scattered muon leads to a coincidence in the activated area of the coincidence matrix while the halo muon fails to do so.

The vertical component θ_y of the scattered muon is determined by using two hodoscopes with horizontal strips located at different positions along the beam direction. If θ_y is compatible with a vertex in the target position the trigger system validates the event. The y-z plane is selected since the particle track is not deflected by the dipole magnet in y-direction. In cases where the scattering angle of the muon is too low to be measured ($Q^2 < 0.5$ (GeV/c)^2), the bending angle of the magnet is used to determine the momentum of the muon and thus trigger measures the energy loss relative to the beam energy.

The kinematic range covered by the trigger system is shown in Fig. 25. The trigger system is optimized to select DIS events. The lowest Q^2 events are covered by the *ladder* trigger (LT) followed by *middle* trigger (MT), *outer* trigger (OT) and *LAS* trigger (LAST) with increasing Q^2.

3.7 DATA ACQUISITION

The data acquisition system (DAQ) [63] is in charge of managing the information coming from more than 250000 spectrometer electronic channels and building events. At COMPASS the typical event size is 45 kB at a trigger rate of about 10 kHz. The pipeline used in the DAQ



(a) (b)

Figure 25: (a) Main elements of the trigger system. (b) Trigger system kinematic coverage.

is illustred in Fig. 26. First the analog signals are coming from the detectors are preamplified if necessary, then they are digitized mostly directly at the front-end by Analog to Digital Converters (ADCs) or Time to Digital Converters (TDCs) according to the type of detectors the front-ends are coupled to. The data are then transferred to the readout driver modules CATCH (COMPASS Accumulate, Transfer and Control Hardware) of GeSiCA (GEM and Silicon Control and Acquisition) upon the arrival of a trigger signal provided by the Trigger Control System (TCS). CATCH and GeSiCA combine the data from up to 16 cards (ADC or TDC) and transmit them via an optical S-Link to the computers named *Readout Buffer* (ROBs, maximum through output 160 MB/s), where they are stored in 512 MB spill buffer cards. During the 4.8 s of a spill the data are written to memory, during the rest of the full SPS cycle (36 s) they are read through a PCI interface. In this way the required bandwidth is reduce by a factor of three. The events are built by 12 event builders and are then written to multiple 1 GB large files (chunks) labeled by the run number and their consecutive chunk number. Finally the data are transferred to the CERN central data recording facility (CASTOR).

3.8 EVENT RECONSTRUCTION

The offline reconstruction of the events stored in CASTOR is performed by the COMPASS software CORAL[1] [63]. CORAL is also used for the reconstruction of events generated by the Monte-Carlo simulation tool TGEANT (see Chapter 9). CORAL is written in C++ and has a modular structure. The scheme of the steps followed by the reconstruction program is shown in Fig. 27. First the information on the fired detectors channels is extracted. This is known as decoding and in the MC case digitization. In general there are more than one detector channels fired by the same particle. In that case a clustering algorithm is applied : the neighbouring detector channels that were fired are grouped together and the coordinate of the cluster in the apparatus reference system is computed. At this stage the detector calibration and position are used to extract the information. The CORAL output is stored in a ROOT Tree called mDST (mini Data Summary Tape).

The physics information is extracted from the mDST using the software package PHAST[2]. PHAST gives access to the reconstructed event information and it provides a set of algorithm to compute the relevant physics variables of each event. The PHAST outputs are stored again in a ROOT Tree. These files are significantly smaller than the mDSTs and are used for the final physics analysis.

In COMPASS, the experimental data are organized into several levels. The basic level are the events collected in one *spill* provided by the SPS. A *run* is the equivalent of 200 spills. As

1 COMPASS Reconstruction Algorithm Library
2 PHysics Analysis Software Tools

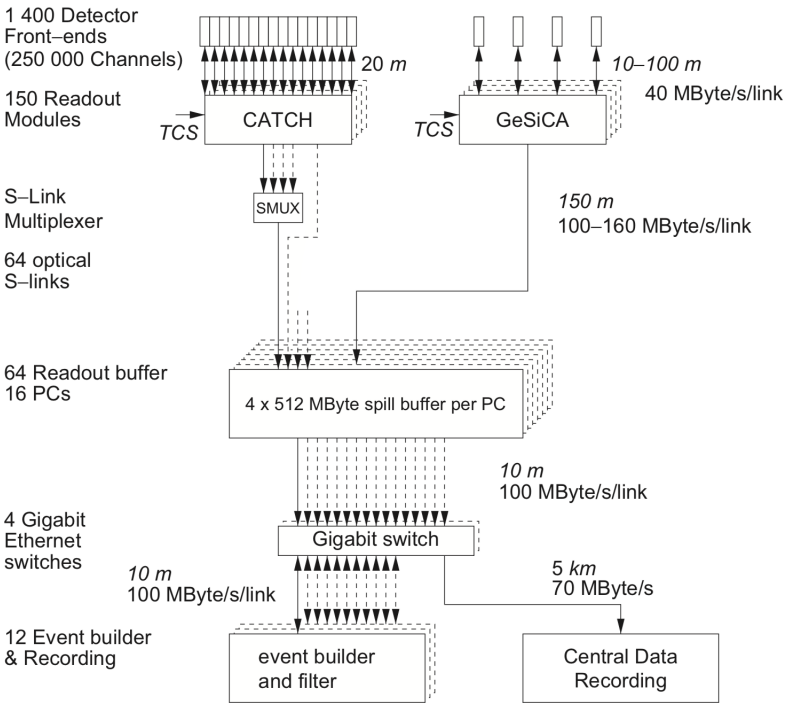


Figure 26: General architecture of the DAQ system. Digitized data from the detector front-end are combined on the CATCH and GeSiCA modules. The storage of the data during the spill and the event building is performed locally. The data are recorded at the CERN computing center. Taken from [63].

there are machine development and/or realignment each week, the data are then structured in *weeks* (also called *period*) containing multiple runs.

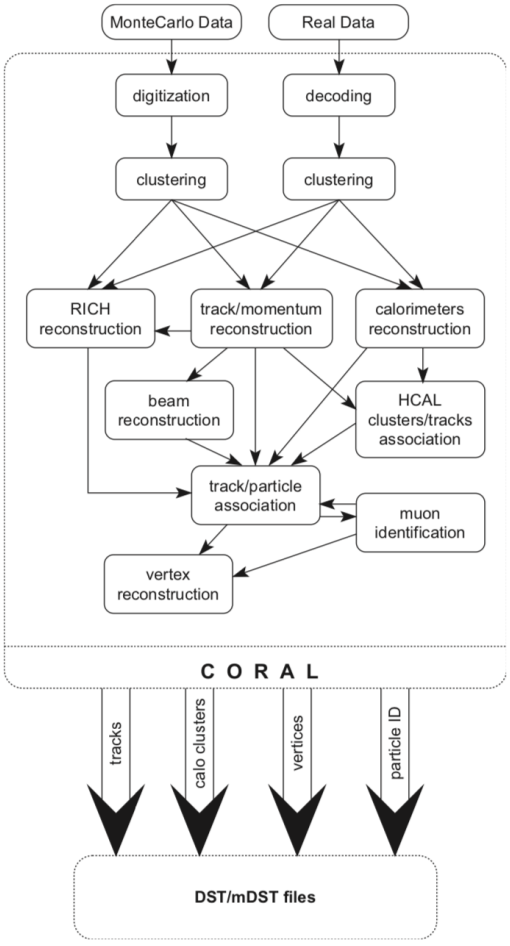


Figure 27: Schematic representation of the COMPASS reconstruction software. Taken from [63].

4

RICH DETECTOR

The particle identification is an important step in the hadron multiplicity extraction. In the COMPASS spectrometer it is performed by a large Ring Imaging Cherenkov detector (RICH) capable of separating pions, kaons and protons in a wide momentum range (~2 GeV/c to ~55 GeV/c) and an angular aperture of 0.01-0.4 radians.

In this chapter the RICH detection principle is presented as well as the description of its main components: the gas and mirror system, the photon detectors, the readout electronics and the data reconstruction.

4.1 CHERENKOV EFFECT

When a charged particle is moving through a transparent medium with a speed v greater than the speed of light ($v_{light} = c/n$, n being the medium refractive index), a radiation known as *Cherenkov radiation* is produced by the medium.

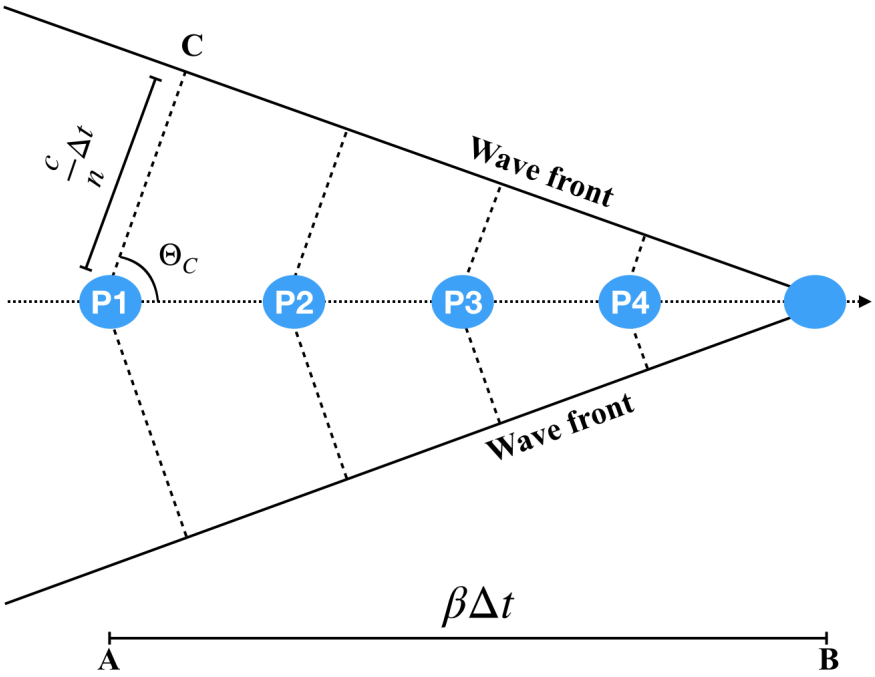


Figure 28: Cherenkov radiation geometry.

The Cherenkov radiation produced by a particle with a mass M_h and momentum p_h is emitted in a narrow cone around a particular angle Θ_C with respect to the particle track (Fig. 28). The wavelength of these radiations goes from visible to UV.

The coherence between waves (emitted between A and B) is achieved, when the particle traverses $\overline{AB}$ at the same time as the radiation travels from A to C. The opening angle Θ_C is defined geometrically in Eq. 44 with β being the particle velocity over the speed of light.

$$\cos\Theta_C = \frac{c/n\Delta t}{\beta c\Delta t} = \frac{1}{n\beta} \tag{44}$$

Some limit cases can be devised:

1. Threshold limit: if $\beta \leqslant 1/n$ no Cherenkov radiation will be emitted.
2. Maximum emission angle: $\cos\Theta_C = \frac{1}{n}$ is reached for ultra-relativistic particles ($\beta = 1$).

In order to perform particle identification with a RICH detector, two variables have to be measured: Θ_C and p_h. The angle can be measured detecting the emitted photons. Different techniques can be used to collect and transport the produced photons to the location of the light detectors. The resulting image in the detector plane is a ring, only for specific techniques, if one does a proper optical image. In such a case the ring has a radius proportional to Θ_C. p_h is measured independently by the spectrometer. The particle mass can thus be calculated by:

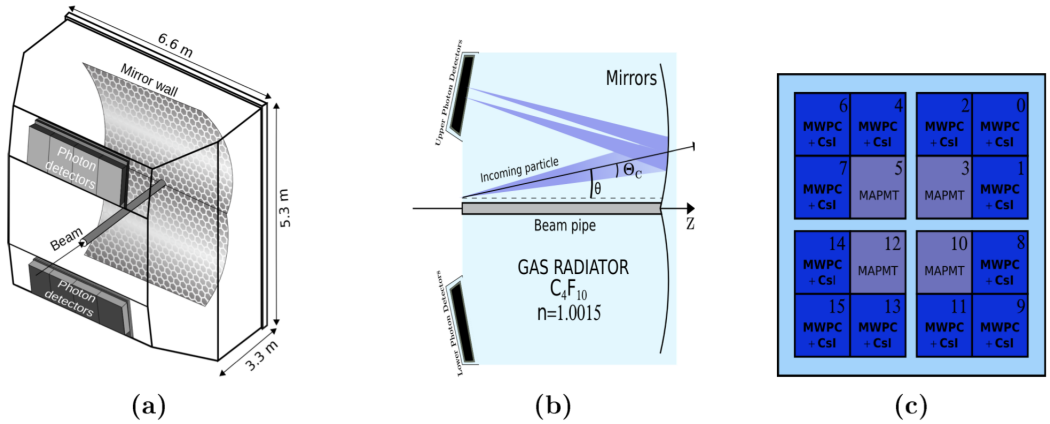
$$M_h = p_h\sqrt{n^2\cos^2\Theta_C - 1} \tag{45}$$

4.2 THE COMPASS RICH DETECTOR

The COMPASS RICH detector is designed to distinguish between pions, kaons and protons at high intensities. The momentum range covers the pion Cherenkov threshold (~ 2.67 GeV/c) to ~ 55 GeV/c.

The RICH is a large size detector (~ 3 x 5 x 6 m^3, see Fig. 29) filled with a gaseous radiator. Two spherical mirror systems reflect the photons into an array of photon detectors sensitive to a large wavelength range, from visible to far UV, placed outside the spectrometer acceptance, one above and one below the beam line. The goal is to count as many photons as possible with a good spacial resolution. The whole structure of the detector vessel is built mainly in thin aluminium in order to minimize the material budget.

Until 2004, MultiWire Proportional Chambers (MWPC) equipped with solid-state CsI photo-cathodes were used to detect Cherenkov photons. The gains of the MWPC operation was limited. The first stage of the electronic readout was characterized by a long integration time, which was a limiting factor in the COMPASS environment as there is a high-rate uncorrelated background due to the large muon halo beam. Moreover, the long base-line restoration time generated a non-negligible dead time. To overcome these limitations, the central region that covers 25% of the photo-detection surface was replaced with MultiAnode PhotoMultiplier Tubes (MAPMT). They are intrisically fast and have better time resolutions.



(a) (b) (c)

Figure 29: (a) Artistic view of the COMPASS RICH detector. (b) Basic functioning of the RICH detector. (c) Photon detector disposition (not to scale).

4.2.1 *Gas System*

One of the principal elements of a RICH detector is the radiator. At COMPASS it is perfluorobutane (C_4F_{10}). It has a refractive index of $n \approx 1.0015$ and a low chromaticity[1] (dn/dE_γ of about $5 \cdot 10^{-5}$ eV^{-1}). These characteristics allow the particle identification (PID) to be performed in the aforementioned wide momentum range.

The propagation of the Cherenkov photons in the vessel can be affected by the presence of water vapor and oxygen (high UV light absorption cross section). In order to remove these impurities, the gas is constantly circulating and filtered at a constant pressure (1 mbar higher than the atmospheric pressure) in a dedicated gas system [69]. The overpressure of the vessel is needed to prevent air contamination and to avoid mechanical stress to the detector, given its large size. Other circulation system (known as *fast circulation* system) allows a reshuffling of the gas inside the vessel: as perfluorobutane has a density of 11.21 kg/m^3 it avoids stratification that may cause a gradient in the value of the refractive index from top to bottom.

In order to absorb the photon emitted by the muon beam, a 10 cm diameter pipe filled with helium is positioned inside the vessel along the beam.

4.2.2 *Mirror System*

The RICH optical system covers an area of ~ 21 m^2 and consists of two spherical surfaces, each one containing 58 spherical mirrors of different shapes (34 hexagons and 24 pentagons). The mirror pattern is shown in Fig. 30. All the mirrors have a reflectance above 80% in the UV region.

The mirror system has a radius of curvature of 6.6 m. The photon image is focused outside the spectrometer acceptance where the photon detectors are located. As the radii of the curvature have a scatter of 1% ($R = 6600 \pm 66$ mm), the reflected image is slightly blurred. This effect is more pronounced for particles at large angles [70]: this aberration contributes to the dispersion of the photon angle with respect to the angle of emission, which affects the detection resolution [71].

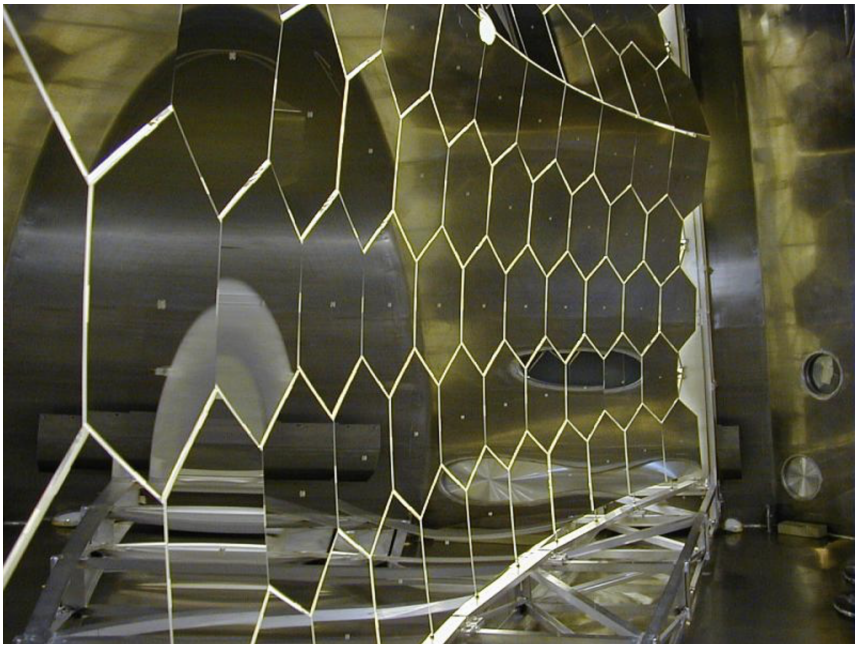


Figure 30: COMPASS RICH detector optical system.

1 Dependence of the refractive index of a dielectric medium on the photon wavelength [68]

4.2.3 *Photon Detectors*

The photon detector array consists of two symmetric parts with respect to the beam line, each one is composed of 8 modules located at the mirror focal plane. The modules in the external regions are MultiWire Proportional Chambers (MWPC) equipped with solid state CsI photocathodes [72]. The central area is composed by MultiAnode Photomultiplier Tubes (MAPMT) [73] coupled to individual telescopes of fused silica lenses. The use of two different detector types employing different different photon converters results in the detection of photons in two wavelength regions: < 200 nm for MWPCs and ~ 200 - 650 nm for MAPMT. The low momentum particles are mainly detected by the outer part (MWPC), while the high momentum ones are detected by the central part (MAPMT). Only the central part is used in the following analysis.

The spherical mirrors will focus all the photons emitted parallel in the same point. Thus the Cherenkov light cone of our particle will result in a ring at the detector plane. The distribution of photons in the detectors for a physics event is shown in Fig. 31.

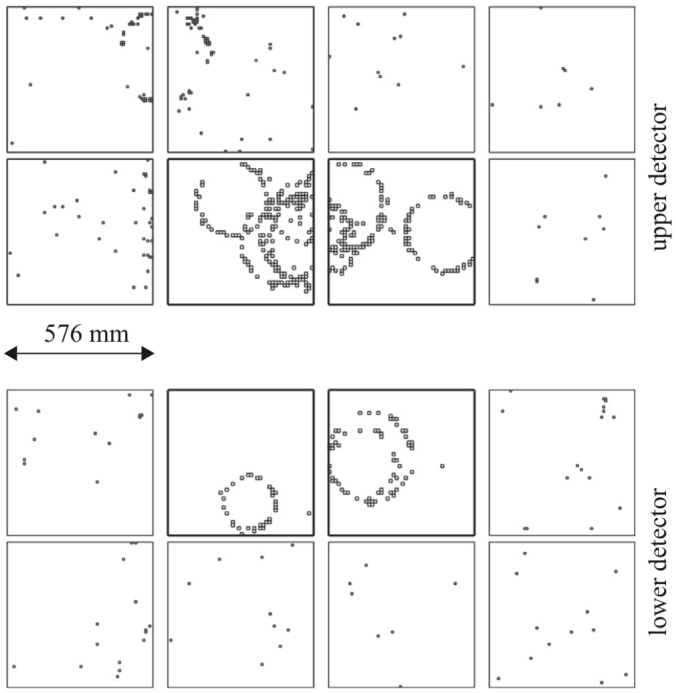


Figure 31: An event from the online event display of COMPASS RICHONE. The 16 squares represent the detectors areas ; the four central ones are equipped with MAPMTs. the small squares represent the hits with signal amplitudes larger than a threshold, individually set for each channel. Figure taken from [61]

4.2.4 *RICH Infos Reconstruction*

RICHONE is a package contained in the CORAL software, which is in charge of RICH information reconstruction. The reconstruction is divided in several parts, the first being decoding the data and clustering. Then the reconstruction of the Cherenkov angle for each individual photon is done. It is possible to perform a ring reconstruction which is used for studies on the apparatus. The particle identification (PID) is based on a maximum likelihood calculation. The PID will be explained more thouroughly afterwards.

Decoding and clustering

There are two different types of photon detectors and they have different decoding systems and clustering algorithms. For the MWPCs, if more than one channel fires, a clustering is done. When the pad with the highest pulse height is found, all the adjacent pads with a smaller signal are included in the cluster [71]. The mean position of each active pad is evaluated in the cluster, weighting the signal with their maximum pulse height, to determine the center of gravity of the cluster. For the MAPMT, decoding the signal is enough to read the time information coming from the PMT that was hit. As the probability of having correlated hits in adjacent area is negligible, the MAPMT data does not need clustering [74].

The cluster or hit position is used to determine the trajectory of the photon. In addition, the time information coming from the MAPMT is used to reject out-of-time photons while the amplitude information from the MWPC serves to reduce the background both from out-of-time photons and from electronic noise [71].

Cherenkov angle and ring reconstruction

The ring reconstruction begins with the selection of a particle tracks. Then one looks for the photons around this track. The trajectory of each Cherenkov photon is calculated with respect to the plane containing the particle track and its virtual reflection in the mirror in order to reconstruct Θ_C [75]. All the photons emitted by one particle are expected to have the same angle Θ_C and to be uniformly distributed in ϕ. The photons emitted by other particles or from background have on the contrary a flat Θ_C distribution. The emitted photon with the same (Θ_C,ϕ) pair are reflected on the same location at the focal surface (neglecting any spherical aberration), resulting in a ring image of the photon detector. Since the emission point of the photon along the particle trajectory is not known, the middle point between the detector and the mirror is taken. A good determination of the track trajectory parameters and the momentum of the particle are mandatory in order to extract Θ_C with good precision.

To characterize the RICH, determining its angular resolution for instance, the ring reconstruction of the emitted photons is needed. The ring reconstruction is based on the search of a peak in the Θ_C distribution. Small intervals of $\pm 3\sigma$ (σ being the single photon resolution, σ_{MAPMT} = 2.0 mrad and σ_{MWPC} = 2.5 mrad) on an overall range of 0 to 70 mrad are considered. The interval with the maximum number of entries is used to define the ring. This procedure associates a ring to each track and in order to reject tracks with only background photons a minimal amount of photons per ring is required (four photons for the MAPMT part) [71]. The resolution of the Cherenkov angle measurement provided by each single photon as a function of the particle momentum is illustrated in Fig. 32.

The measured values of θ_C as a function of p_h for the RICH detector are shown in Fig. 33. In the low momentum region, the RICH detector is only sensitive to electrons, muons and pions. The bands corresponding to kaons and protons start to be visible respectively at $p_h \approx$ 9.45 GeV/c and $p_h \approx$ 17.95 GeV/c. For high momentum values above 40 GeV/c, saturation of the Cherenkov angle is observed or pions and kaons. The final particle identification is performed using likelihood methods and is described in the following chapter.

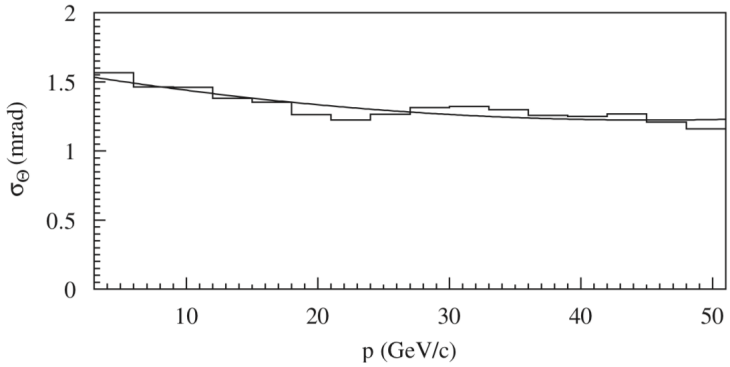


Figure 32: Resolution of the Cherenkov angle for the reconstructed ring images, provided by each single photon, versus the particle momentum for a sample of identified pions. Figure taken from [63].

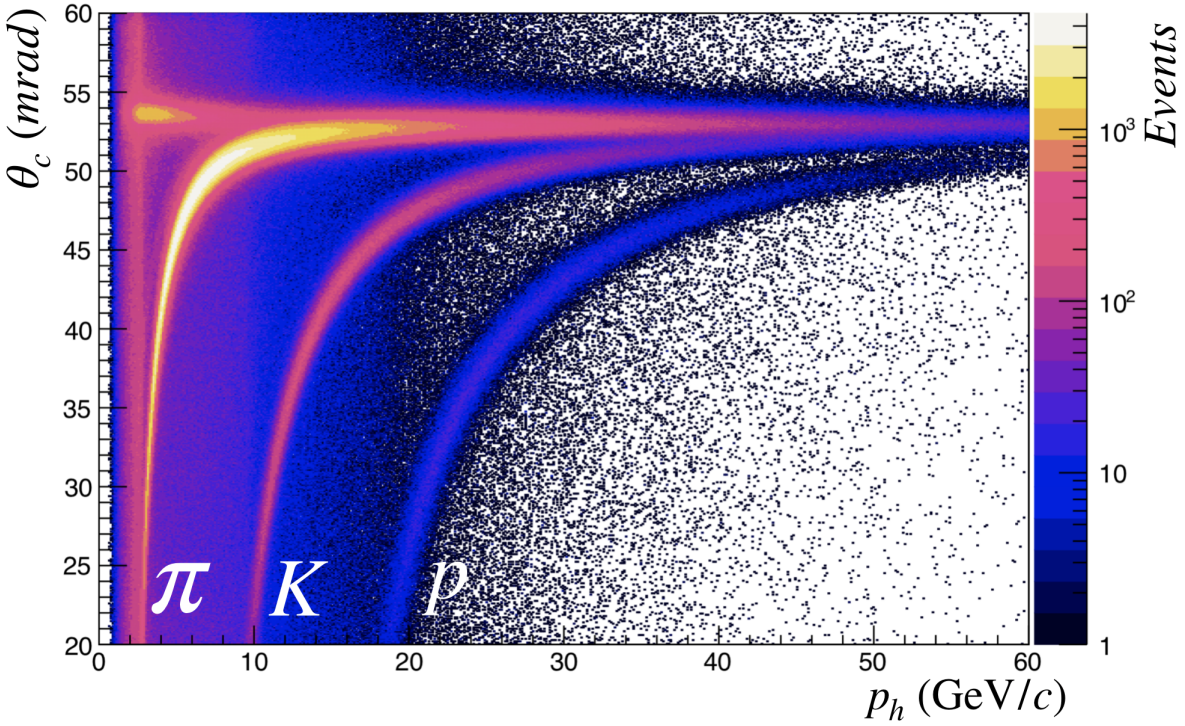


Figure 33: Measured Cherenkov angle Θ_C as a function of p_h. π threshold is about 2.67 GeV/c, K threshold about 9.45 GeV/c and p threshold about 17.95 GeV/c, respectively.

5

PERFORMANCE STUDY AND PARTICLE IDENTIFICATION

For particle tracks with a measured momentum using the RICH information enables us to determine the particle mass and thus identify the particle. As the detector efficiency is not perfect, the misidentification probability is not zero ; one thus has to determine the identification performance of the RICH detector. The RICH performance study relies on extracting the identification and misidentification probabilities for pions, kaons and protons. From samples of pure pions, kaons and protons the RICH detector response is measured: the hadrons are identified using RICH information and the identification and misidentification probabilities are calculated in the hadron (p_h,θ_h) phase space for pions, kaons and protons. The method used for identification is a likelihood estimation for different hypotheses. At first order, the mass assignment corresponds to the highest likelihood, but further requirements can be added to improve the result.

5.1 DETERMINATION OF RICH DETECTOR PERFORMANCE

The identification and misidentification efficiency is given by the ratio of the number of particles correctly and wrongly identified, respectively, out of a pure sample of specific hadron type species over the total number of hadrons composing the pure sample:

$$\epsilon(t \to i) = \frac{N(t \to i)}{N(t)} \tag{46}$$

$\epsilon(t \to i)$ is the probability that a particle t is identified as a particle i, $N(t \to i)$ is the number of particles t identified as i and $N(t)$ is the total number of hadron t of the pure sample. The identification ($\epsilon(t \to t)$) and misidentification ($\epsilon(t \to i)$) efficiencies are properties of the RICH and can be displayed in an efficiency matrix with the identification efficiencies on the diagonal and the misidentification ones off-diagonal:

$$M_R = \begin{bmatrix} \epsilon(\pi \to \pi) & \epsilon(K \to \pi) & \epsilon(p \to \pi) \\ \epsilon(\pi \to K) & \epsilon(K \to K) & \epsilon(p \to K) \\ \epsilon(\pi \to p) & \epsilon(K \to p) & \epsilon(p \to p) \end{bmatrix}. \tag{47}$$

5.1.1 *Selection of* Φ, K^0 *and* Λ

To obtain the efficiency matrix, three pure hadron samples are needed. The RICH performance analysis is based on the study of the pion, kaon and proton samples originating from Φ, K^0 and Λ decays.

For the determination of the RICH efficiency, it is necessary to have a source of events where the true kind of the particle passing the RICH is known. That kind of events is obtained using two body particle decays, namely the decay of a K^0 into two pions ($K^0 \to \pi^+\pi^-$), the Φ decay into two kaons ($\Phi \to K^+K^-$), the Λ decay into a pion and a proton ($\Lambda \to p\pi^-$). In order to select events with such decays, scattering events with a scattered muon are selected. The following cuts are thus applied to the data:

- Exclude bad spills
- Select best primary vertex with incoming and scattered muon

- Check if primary vertex is inside one of the target cells
- Extrapolated track of the incoming muon should cross all target cells
- $0.1 \leqslant y \leqslant 0.9$

Different selection criteria have to be used for K^0, Λ and Φ decays. In the case of K^0 mesons and Λ baryons, the particles decay by the weak force. Therefore, the decay length is long enough to produce a secondary vertex, which can be separated from the primary one. The Φ mesons decays by the strong force. This results in a very short decay length and it is not possible to separate the secondary vertex from the primary one.

5.1.2 K^0 *and* Λ *selection*

For K^0 mesons the decay into π^+ and π^- with a branching ration of $(69.20 \pm 0.05)\%$ [13] and in the case of Λ and $\bar{\Lambda}$ baryons the decay into a proton and a pion with a branching ratio of $(63.9 \pm 0.5)\%$ [13] is used. In both decays the reconstruction of the secondary vertex is possible. The following cuts are applied to select these decays:

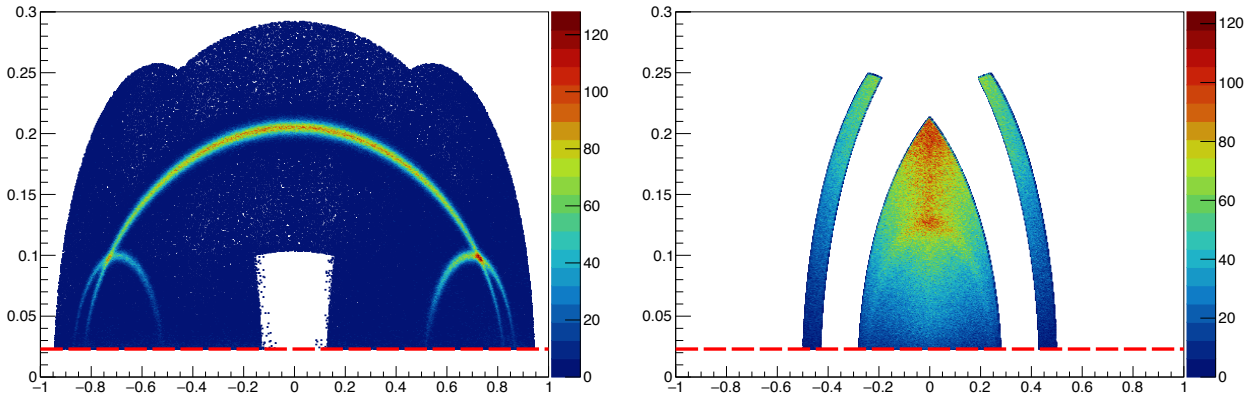
1. Selection of good secondary vertices
   - Loop over all vertices
   - Vertex is not primary one
   - Exactly two opposite charged outgoing particles
   - The tracks should not be connected to any other primary vertex to ensure that they belong to a secondary vertex
   - Primary and secondary vertex separated by more than two times the reconstruction accuracy
2. Select good hadron tracks
   - Both particles should not have crossed more than 10 radiation length in order to supress the muons from the sample.
   - Last measured position (Z_{last}) behind SM1 to ensure a measured momentum
   - Transverse momentum with respect to the mother particle larger than 23 MeV to suppress electrons from photon conversion
   - Check that the decaying particle is connected to the primary vertex ($\theta_p \leqslant 0.01$)
3. Additional cuts
   - $p_h \geqslant 1$ GeV/c
   - Mass difference smaller than 150 MeV/c^2 between the K^0/Λ mass and the invariant mass of the two decay hadrons assuming the correct masses

The same cuts except for the mass cuts are used for K^0 and Λ candidates. The transverse momentum of K, Λ or dacay product is shown as a function of the ratio of the longitudinal momentum ratio of two particles:

$$\alpha = \frac{p_{L,1} - p_{L,2}}{p_{L,1} + p_{L,2}}, \tag{48}$$

in the Armenteros plots in Fig. 34 (a) for K^0 and Λ and (b) for Φ.

The three visible arcs are produced by the decay of the K^0 mesons and the Λ baryons. The decay of K^0 mesons in two particles with the same mass results in the symmetric arc, whereas



(a) K^0 and Λ (b) Φ

Figure 34: Armenteros plots. The cut on the transverse momemtum is illustrated by the red line.

the decay of Λ baryons into two particles with different masses result in the two smaller arcs on the left and right side. All particles below the red dashed line, which corresponds to the 23 MeV limit, are rejected in order to supress tracks from electron coming from photon conversion.

5.1.3 Φ *selection*

The branching ratio of the Φ decay into two kaons is $(48.9 \pm 0.5)\%$ [13]. The Φ meson decay length is too short to separate the primary and decay vertex. Therefore, all outgoing particles from a primary vertex are taken into account for the search of possible Φ mesons. The selection steps are similar the selection of the K^0/Λ candidates:

1. Selection of possible good events with Φ mesons
   - At least three outgoing particles including scattered muon
   - Loop over all outgoing particles
   - Oppositely charged pairs of hadrons (none is a muon)
2. Select good hadron tracks
   - Last measured position (Z_{last}) behind SM1 to ensure a measured momentum
   - Transverse momentum with respect to the mother particle larger than 23 MeV to suppress electrons from photon conversion
3. Additional cuts
   - $9\ \mathrm{GeV/c} \leqslant p_h \leqslant 55\ \mathrm{GeV/c}$
   - Mass difference smaller than 120 $\mathrm{MeV/c^2}$ between the Φ mass and the invariant mass of the two decay hadrons assuming the kaon mass.

The selection of Φ meson candidates results in a large combinatorial background. The three regions in the Armenteros plots in Fig. 34 (b) are produced by the decay of the Φ mesons. The decay of Φ mesons in two particles with the same mass results in symmetric regions. All particles below the red dashed line are rejected as in the K^0/Λ case.

5.2 LIKELIHOOD METHOD

The likelihood (LH) method is a statistical method that can be used to find a suitable model for describing a data set or to estimate the values of the parameters used in this model [76, 77].

The first step is to define the LH function. Let $\mathbf{s}$ be a sample of elements s_i, $i \in [\![1, l]\!]$:

$$\mathbf{s} = (s_1, s_2, .., s_l)\,, \tag{49}$$

which follows a distribution with the probability density $f(s|\theta)$. This probabilty density is determined by a set of parameters $\theta = \theta_1, .., \theta_m$. In order to fit the sample with the model some requirements must be met.

1. The sample can be considered as a l-dimensional random variable and can be assigned a probability density $g(\mathbf{s})$:
$$g(\mathbf{s}) = g\,(s_1, s_2, .., s_l)\,. \tag{50}$$
2. The sample is random.
   (a) The s_i are independent:
   $$g(\mathbf{s}) = g_1\,(s_1) \cdot g_2\,(s_2) \cdot .. \cdot g_l\,(s_l)\,. \tag{51}$$
   (b) Each element s_i follows the probability density of the distribution:
   $$g(s_i) = f(s|\theta). \tag{52}$$

If these conditions are fulfilled, the LH function $L\,(s_1, .., s_l|\theta)$ is defined as:

$$L\,(s_1, .., s_l|\theta) = \prod_{i=1}^{l} f(s_i|\theta), \tag{53}$$

and states that the probability of occurence of the sample is equal to the product of the occurence of each element of the sample. The probability density of the sample is normalized to its domain of definition Ω:

$$\int_\Omega L\,(s_1, .., s_l|\theta)\,ds_1..ds_l = 1. \tag{54}$$

In order to select from all possible sets of parameters in the parameter space Θ the estimator $\hat{\theta}$, which gives the best description of the true description, one applies the Maximum Likelihood Estimation (MLE) method. The method can be formulated as follows:

$$\hat{\theta} \in \left\{ \underset{\theta \in \Theta}{\arg\max}\, L\,(s_1, .., s_l|\theta) \right\}, \tag{55}$$

if a maximum exists. This means that the maximum of the LH function has to be found in relation to the parameters. Having found the maximum, one has also found the best estimate of the parameters. Since the LH function can yield very small values as a probability density, it is common to define the logarithmic LH function instead:

$$\mathscr{L}\,(s_1, .., s_l|\theta) = \ln L\,(s_1, .., s_l|\theta) = \sum_{i=1}^{l} \ln f(s_i|\theta). \tag{56}$$

The maximization condition for several parameters $\theta = \theta_1, .., \theta_m$ is thus:

$$\frac{\delta \mathscr{L}(s_1, .., s_l|\delta\theta)}{\delta\theta_j} = \frac{\delta}{\delta\theta_j} \sum_{i=1}^{l} \ln f(s_i|\theta) = 0 \quad \text{for} \quad \theta = \hat{\theta}. \tag{57}$$

In many physics systems one can find geometrical or kinematic constraints such as the sum of the impulses equal to zero and the sum of the energies equal to twice the photon energy in the electron-positron annihilation in the center-of-mass system. These conditions can be used to eliminate one of the parameters of the system. However, this is often not desirable since this elimination is often only possible through complicated algorithm or the equivalent treatment of the parameters after the adaptation is no longer guaranteed. In order to consider the constraints further, they can be included in the LH functions as functions $c_k(\theta)(k = 1, .., z_c)$, analogous to the method of Lagrange multipliers:

$$\mathscr{L}(s_1, .., s_l|\delta\theta) = \ln L(s_1, .., s_l|\theta) = \sum_{i=1}^{l} \ln f(s_i|\theta) - \sum_{k=1}^{z_c} \lambda_k c_k(\theta) \tag{58}$$

The overall z_c Lagrange multipliers λ_k are treated as additional parameters. Thus the z_c conditions of the constraints still come to the maximization conditions in Eq. 57:

$$\frac{\delta \mathscr{L}}{\delta \lambda_k} = c_k(\theta) = 0. \tag{59}$$

Since the LH method is used not only for particle identification by the RICH detector but also for the fits to determine the RICH efficiencies, the extended Maximum Likelihood Estimation (eMLE) method is discussed further below. This method is used primarily for problems, in which the fit also supplies the number of expected events and these are to be adapted to the observed events. For example this is the case if there are n events that results from a sum of several sources n_i. The constraint that follows is that $n = \sum_{i=1}^{j} n_i$. This condition can now be used as an additional factor in the LH function. The factor then corresponds to a Poisson distribution, which describes the probability that n events are also observed at an expected value λ:

$$L(s_1, .., s_l|\theta) = \frac{\lambda^n e^{-\lambda}}{n!} \prod_{i=1}^{n} f(s_i|\theta). \tag{60}$$

For the logarithmic LH function follows:

$$\mathscr{L}(s_1, .., s_l|\theta) = n \ln \lambda - \lambda + \sum_{i=1}^{n} \ln f(s_i|\theta) \tag{61}$$

The term $-\ln(n!)$ is irrelevant for the following maximization and has been omitted. With the help of the following simplification:

$$n \ln \lambda + \sum_{i=1}^{n} \ln f(s_i|\theta) = \sum_{i=1}^{n} (\ln f(s_i|\theta) + \ln \lambda) = \sum_{i=1}^{n} \ln(\lambda f(s_i|\theta)), \tag{62}$$

is it possible to define a function $g(s_i|\theta) = \lambda f(s_i|\theta)$, which is normalized by λ:

$$\int_\Omega g(s_i|\theta) ds_1..ds_l = \lambda \int_\Omega f(s_i|\theta) ds_1..ds_l = \lambda. \tag{63}$$

Thus Eq. 61 can be rewritten into the common form of an extended LH function:

$$\mathscr{L}(s_1, .., s_l|\theta) = \sum_{i=1}^{n} \ln g(s_i|\theta) - \int_\Omega g(s_i|\theta) ds_1..ds_l. \tag{64}$$

It follows that $\mathscr{L}(s_1, .., s_l|\theta)$ becomes maximal, when the additional term equals the number of actual events n.

5.3 RICH PARTICLE IDENTIFICATION

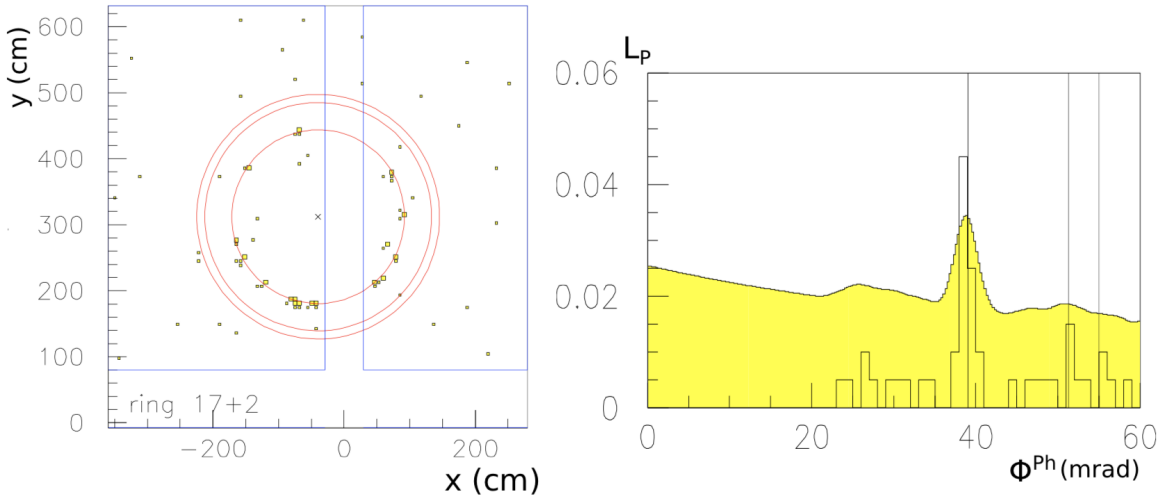
Particle identification using the LH method is accomplished by finding the maximum LH function value. The LH function describes the radial photon distribution. An essential step is to determine the radial photon distribution (see Fig. 35 (a)) as precisely as possible. This depends on the detector geometry, the accuracy in determining the trajectory of the particle to be identified, as well as on the accuracy with which the parameters describing the photon position in the ring plane can be determined. The two parameters used are the reconstructed angles at which the photon (Ph) was emitted relative to the particle path Θ^{Ph} and the reconstructed azimuthal angle around the particle path Φ^{Ph} [78–80]. The LH function for describing the photon distribution is given as:

$$L_{N_{Ph}} = \prod_{k=1}^{N_{Ph}} \left[(1-\epsilon) G\left(\Theta_k^{Ph}, \Phi_k^{Ph}\right) + \epsilon B\left(\Theta_k^{Ph}\right) \right], \tag{65}$$

where

$$G\left(\Theta_k^{Ph}, \Phi_k^{Ph}\right) = \frac{1}{\sqrt{2\pi}\cdot\sigma_{\theta,k}^{Ph}} e^{-\frac{1}{2}\frac{\left(\Theta_k^{Ph}-\Theta_k^{Ph}\right)^2}{\left(\sigma_{\theta,k}^{Ph}\right)^2}} \cdot \frac{\Theta_k^{Ph}}{\Theta^M} \tag{66}$$

is a Gaussian distribution with which the signal (see Fig. 35 (b)) can be described. The standard deviation $\sigma_\theta^{Ph}(\Phi_P^{Ph}, \beta)$ originates from the accuracy with which the photon distribution can be determined. This in turn depends on the azimuthal photon angle (Φ_P^{Ph}) in the plane of the photodetectors, which is characterized by the index P, and on the particle velocity relative to the speed of light β. Θ^M is the mass hypothesis for the considered particle. It is determined by the expected Cherenkov angle for a particular particle type or particle mass with β. Possible particles are electrons, pions, kaons and protons.



(a) Photon distribution in the detector plane

(b) LH-fit of the radial photon distribution

Figure 35: Figure (a) shows the photon distribution in the plane of the photodetectors. The x marks the projection of the particle path. The red rings correspond, from the inside to the outside, to the distribution that would be generated by a proton, a kaon or a pion. The yellow marks are the detected photons. In (b) the result of the LH-fit of the photon distribution with the proton hypothesis is shown.

The RICH particle identification efficiency is studied in the momentum range of 3 GeV/c $\leqslant p \leqslant$ 50 GeV/c. In this range, pions and kaons are emitting Cherenkov light, while up to ~ 17 GeV protons are still below the threshold of:

$$p_{thr,i} = \frac{m_i}{\sqrt{n^2 - 1}} \tag{67}$$

where n is the refractive index. This is shown in Fig. 36 where the reconstructed Cherenkov angle is shown as a function of the hadron momentum. The identification of pions, kaons and protons above the momentum threshold is done by comparing the likelihood values with one another. The identification of these particles is done using likelihood cuts. Using the likelihood values, the particle identification is done by comparing these values with one another. In the simplest case, the highest one determines the particle type. This method is used in the case of pions. In the case of kaons, stricter likelihood cuts are applied to suppress misidentified pions as the goal of the selection is to obtain a clean pion and kaon sample. The likelihood cuts for protons require its likelihood to be the largest one. Below the momentum threshold, protons do not emit Cherenkov light. Therefore, the likelihood values are used to test, whether the detected light is consistent with random noise in the detector (background). In order to avoid possible problems due to the uncertainty on the reconstructed momentum or the uncertainty of the refractive index of the RICH gas, a region of $\pm$ 5 GeV/c around the proton threshold is used, where both hypothesis are applied for proton identification. The likelihood cuts are listed in Table 8. As the electrons cannot be distinguished from pions for momentum above 8 GeV, they are not separated at this stage of the analysis. This contamination is dealt with the Monte-Carlo later in the analysis (Chapter 11).

Table 8: Likelihood cuts for pion, kaon and protons

	PION	KAON	PROTON		
MOMENTUM	$p > p_{\pi,thr}$	$p > p_{K,thr}$	$p \leqslant p_{p,thr}$		$p > p_{p,thr}$
	π	K	p	$\bar{p}$	$p/\bar{p}$
LH(π)/LH(2^{nd})	> 1.02	—	—	—	—
LH(π)/LH(bg)	> 2.02	—	< 2.2	< 2.1	$< 1.$
LH(K)/LH(2^{nd})	—	> 1.08	—	—	—
LH(K)/LH(bg)	—	> 2.08	< 2.9	< 2.8	$< 1.$

5.4 METHOD OF UNFOLDING

The particle identification efficiency of the RICH is studied as a function of the hadron phase space. Here we use the hadron momentum and the polar angle at the entrance of the RICH, as already studied before, for example in Reference [81]. A fine binning is used for the momentum dependence, since the Cherenkov effect depends strongly on this variable. For the dependence on the polar angle, a coarse binning can be used, since only a weak dependence is observed. The binning used is:

- Momentum p_h (GeV/c): $\{3, 5, 7, 10, 12, 13, 15, 17, 19, 22, 25, 27, 30, 35, 40, 50\}$
- Angle θ_h (rad): $\{0, 0.01, 0.04, 0.12, 0.3\}$

For each bin, the elements of the efficiency matrix M_R are determined separately for positive and negative particles. The elements of this matrix contain the probability for a particle t to be identified as a particle of type i, for example a pion that is correctly identified as

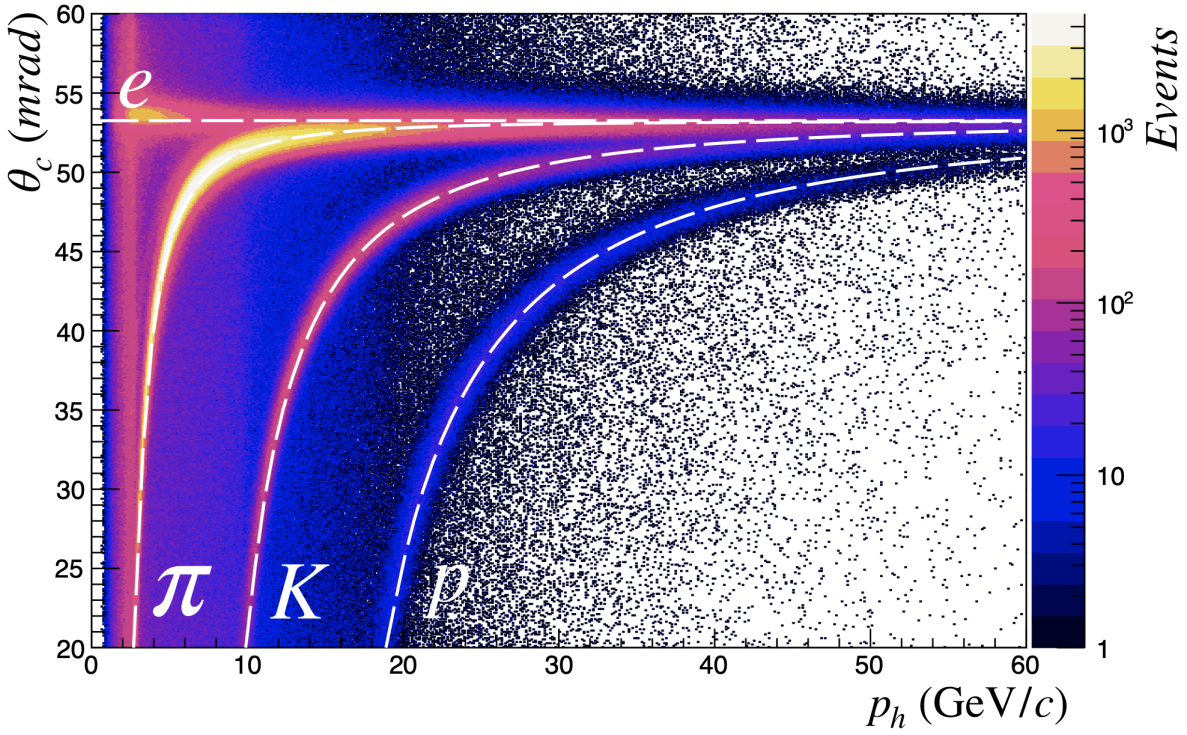


Figure 36: Comparison of reconstructed Cherenkov angle as a function of the momentum with the calculated Cherenkov angle for each particle type using the refractive index of the RICH gas. This plot is not used for the PID.

pion or wrongly as a kaon. In the case of positive pions, the events from the K^0 sample are used, where the negative hadron is identified as a pion using the likelihood cuts shown in Table 8. Therefore, the second particle has to be a pion too, if the decaying particle was a K^0. Using the RICH, the particle type is determined for the second particle, which results in the number $N(\pi^+ \to i)$. An equivalent procedure is used for positive kaons and protons using the Φ and Λ samples. In order to obtain these numbers for the negative particles, the same samples are used but this time performing the identification of the positive particle in the first place. The numbers $N(t \to i)$ are extracted using a fit, which is described here for the K^0 sample, where the negative pion is already identified. The events are put into five different groups, depending on the particle type determined by the RICH:

- All events (RICH not used for second particle)
- Events where π^+ is identified as π^+
- Events where π^+ is identified as K^+
- Events where π^+ is identified as p
- Events where π^+ is not identified

For each of these groups, the invariant K^0 mass spectra are shown in Fig. 37, for a selected momentum bin. The number of events in the peak and the background are determined by a simultaneous fit of all five spectra. These spectra are described using two Gaussian distributions with the same mean for the signal, f_{sig}, and a polynomial to describe the background, f_{bg}. Their expressions are given in Table 9. The two Gaussian distributions account for the different resolutions of the two spectrometer stages. The fitted function for each of the groups is given by

$$f(x) = N_{sig} \cdot f_{sig} + N_{bg} \cdot f_{bg}, \tag{68}$$

where N_{sig} is the amount of K^0 and N_{bgd} the amount of background events. Here, the same widths, σ_1 and σ_2, of the two Gaussian distributions was used for all five spectra.

Table 9: Functional form for the descriptin of the mass spectra for K^0, Φ and Λ candidates from the clean samples. The symbol G represents a Gaussian distribution and the symbol *BW* a relative Breit-Wigner distribution.

SAMPLE	SIGNAL	BACKGROUND
K^0	$\delta G(\mu,\sigma_1)+(1-\delta)G(\mu,\sigma_2)$	$1+ax+b(2x^2-1)+c(4x^3-3x)$
Φ	$BW(\mu,\sigma_1)\otimes G(\mu,\sigma_2)$	$(x-t)^n\cdot exp(-a(x-t))$ with $t=2\cdot m_K$
Λ	$\delta G(\mu,\sigma_1)+(1-\delta)G(\mu,\sigma_2)$	$(x-t)^n\cdot exp(-a(x-t))$ with $t=m_p+m_\pi$

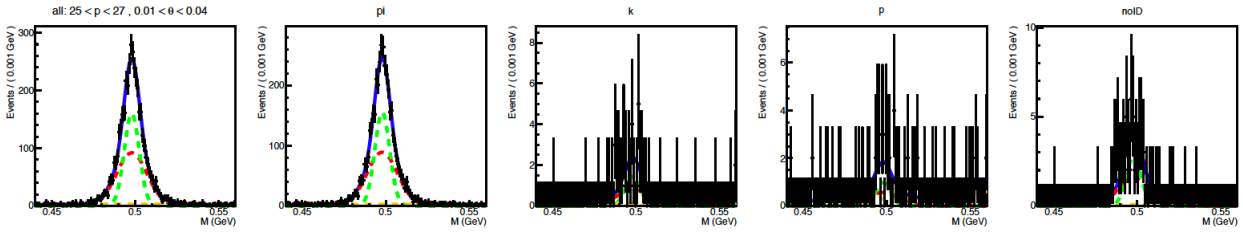


Figure 37: Mass spectra for K^0 candidates with an identified π^- for various hypotheses for the second hadron (from left to right: all, π, K, p, no ID). The momentum of the positive hadron is in the range of [25,27] GeV/c and in the angle in the range [0.01,0.04] rad.

Also the ratio δ of the amount of events in both Gaussian distributions is assumed to be the same. The shape of the background is the same for all spectra, except for the one where the pion is identified as a proton. In this case, a possible background contribution due to decays from Λ baryons into a pion and an proton can be found. This results in a slightly different background shape. The integral of the background remains as an independent parameter in all five cases. In order to ensure that the sum of all efficiencies ($\varepsilon(\pi^+\to\pi^+)+\varepsilon(\pi^+\to K^+)+\varepsilon(\pi^+\to p)+\varepsilon(\pi^+\to noID)$) is 100%, an additional constraint is introduced to the fit.

$$N^{all}(K^0)=N^{\pi}(K^0)+N^{K}(K^0)+N^{p}(K^0)+N^{noID}(K^0), \tag{69}$$

where $N_i(K^0)$ (i = π, K, p, noID) is the number of K^0 obtained from the histogram where the pion is identified as i. This results in 16 free parameters of the fit.

The main difference between the fits of K^0, Φ and Λ samples is the description of the signal and the background, while the same method is used. The functions describing both are also given in Table 9. Again the parameters describing the shape are the same in all five spectra and the fit parameters describing the integrals of the functions are used as free parameters, except for the parameter of the mass spectrum including all events. This results in 15 free parameters for the fit of the Φ sample and in 15 free parameters for the fit of the Λ sample. Examples of the fits performed for the Φ and Λ samples are shown in Figs. 38 and 39. The figures show the results for the same momentum bin (25 GeV/c $< p_h <$ 27 GeV/c) and angular bin (0.01 rad $< \theta <$ 0.04 rad), which was also shown for the K^0 sample.

5.5 CALCULATION OF THE EFFICIENCIES AND UNCERTAINTIES

The elements of the efficiency matrix M_R are determined from fitted numbers of signal events,

$$\epsilon(t\to i)=\frac{N(t\to i)}{N(t)} \tag{70}$$

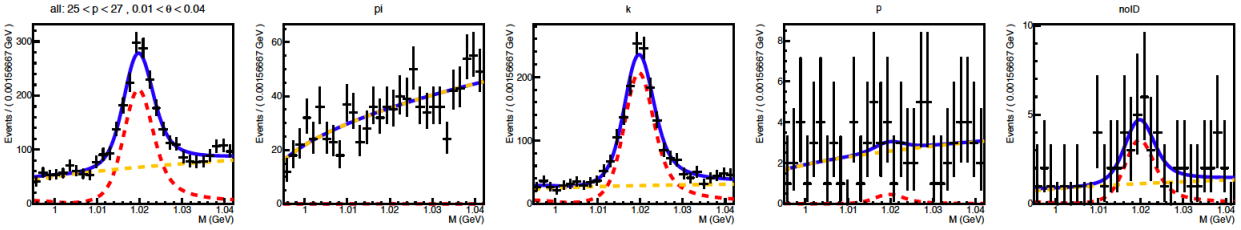


Figure 38: Mass spectra for Φ candidates with an identified K^- for various hypotheses for the second hadron (from left to right: all, π, K, p, no ID). The momentum of the positive hadron is in the range of [25,27] GeV/c^2 and in the angle in the range [0.01,0.04] rad.

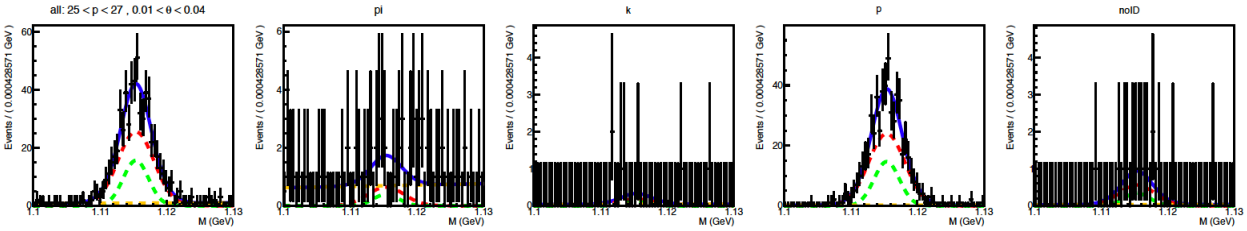


Figure 39: Mass spectra for Λ candidates with an identified π^- for various hypotheses for the second hadron (from left to right: all, π, K, p, no ID). The momentum of the positive hadron is in the range of [25,27] GeV/c^2 and in the angle in the range [0.01,0.04] rad.

Here, $N(t)$ is given by the sum of all $N(t \to i)$. As the nominator and denominator are correlated, the uncertainty has to be determined via error propagation taking into account the covariance matrix of the fit,

$$\Delta\epsilon = \sqrt{\sum_{j=1}^{m} \left(\frac{\delta\epsilon}{\delta N(i \to j)} \right)^2 \cdot u_j + 2 \sum_{j=1}^{m-1} \sum_{k=j+1}^{m} \left(\frac{\delta\epsilon}{\delta N(i \to j)} \frac{\delta\epsilon}{\delta N(i \to k)} \cdot u(j,k) \right)} \tag{71}$$

where u_j are the diagonal elements of the covariance matrix, $u(i,j)$ are the off diagonal elements and ϵ is one of the elements of the efficiency matrix. The summations are done over all possible particle types i.e. pion, kaon, proton and non identified.

5.6 RESULTS

The results for the RICH particle identification efficiency are shown in Figs. 40 to 41 for π^+ and K^+ for the various particle types and charges. The plots for the other species can be found in Appendix A. In each figure, the momentum dependence for the different angular bins is shown. The efficiencies are weakly dependent on the angle, while it is more strongly correlated with the momentum, especially in the region near the threshold.

The RICH performs a correct identification of pions in more than 95% of the cases for momenta below 30 GeV/c^2 and the probability for a misidentification of a pion as a kaon is below $\sim$ 1%. For kaons, near the threshold, a strong momentum dependence of the efficiencies is observed. Therefore a cut of 12 GeV is chosen in the following analysis. At higher momenta the correct identification is given in $\sim$ 95% of the cases and the probability for a misidentification of a kaon as a pion is below $\sim$ 2%. For protons, the momentum dependence around threshold level is even stronger. Below the threshold, protons are identified correctly in 50% of the cases. Above the threshold numbers rise to $\sim$ 95%.

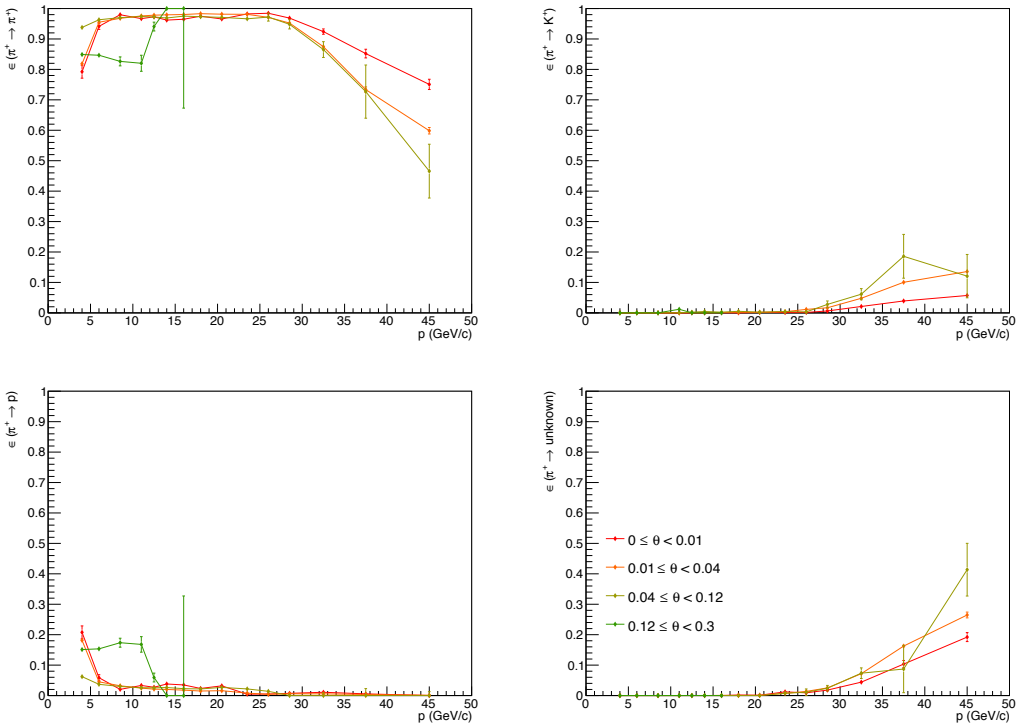


Figure 40: Identification probabilities $\epsilon(p \to j)$ for π^+.

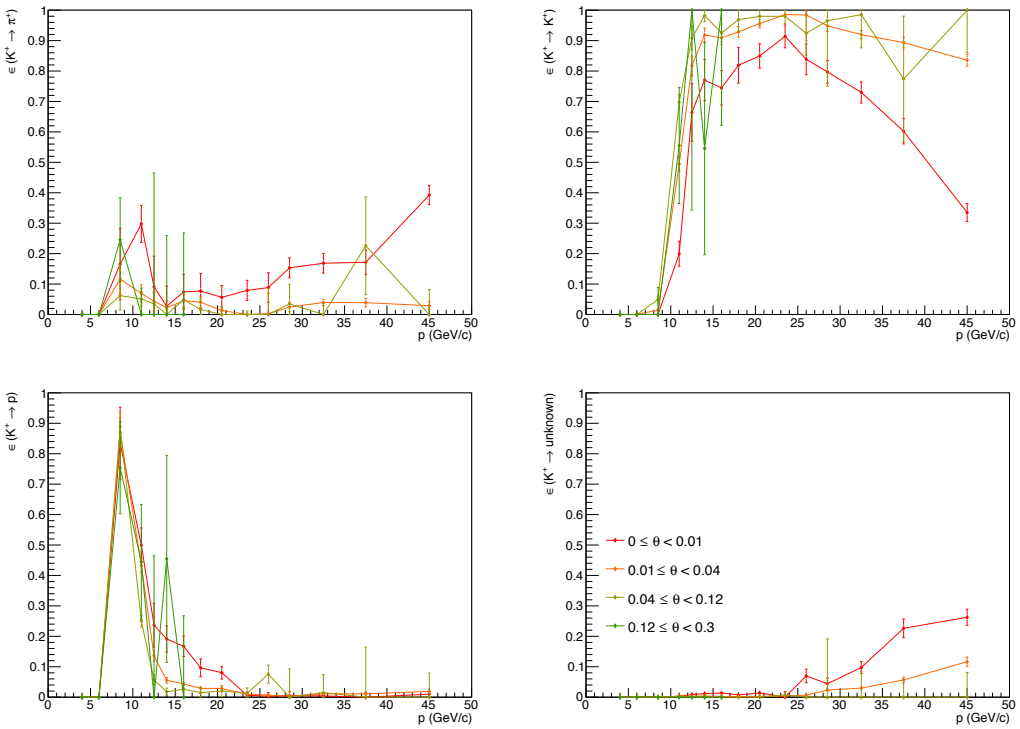


Figure 41: Identification probabilities $\epsilon(p \to j)$ for K^+.

5.7 PROBLEM AT HIGH z

At some point it was discovered that at high momenta and high z ($35\,\mathrm{GeV/c} < p_h < 40\,\mathrm{GeV/c}$, $z > 0.7$) a contamination of the kaon sample by misidentified pions was observed, which was not accounted for in the efficiency matrix (Fig. 42). Instead of the expected separation between pions and kaons at $LH_\pi = LH_K$, pions are found at $LH_\pi < LH_K$. We are now not looking at the clean samples but to the normal data.

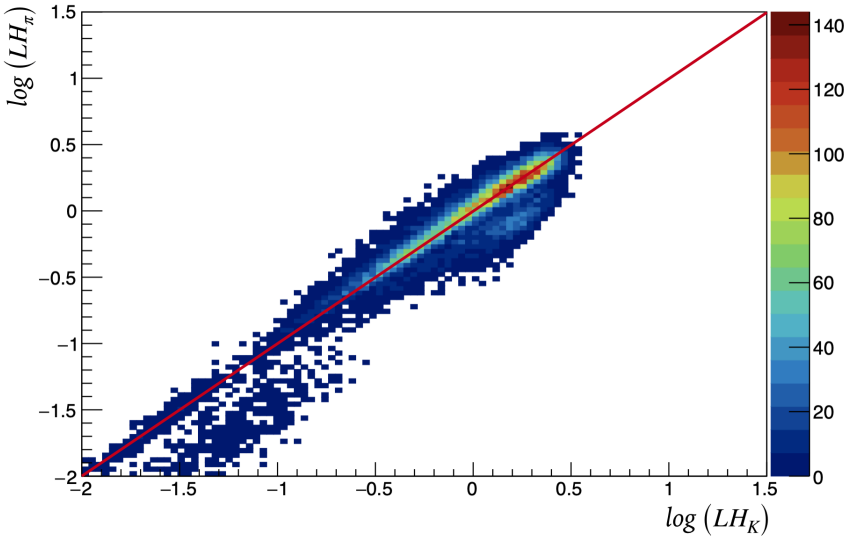


Figure 42: Likelihood for pions as a function of the likelihood for kaons using 2016 data.

The larger the likelihood values are, the larger the effect is. This behaviour is also present in 2006, 2007 and 2011 data (Fig. 43). When investigating, it was shown that the probability for the misidentification of pions as kaons differs from the value given in the RICH tables in this kinematic region using the p_T spectra [82].

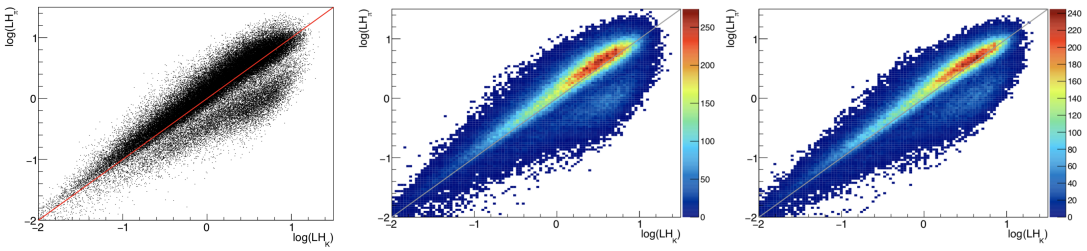


Figure 43: Likelihood for pions as a function of the likelihood for kaons using (from left to right) 2006, 2007 and 2011 data.

In order to check if the non-linearities are taken correctly into account in the RICH tables, the likelihood values for pions and kaons are compared using the K^0 sample for high momenta and high z. This comparison is shown in Fig. 44, highlighting the fact that the problematic region is not covered by the K^0 sample and therefore the RICH table are not valid in this kinematic region.

In order to obtain correct values for the RICH tables, a different sample for pions is needed. This new sample is obtained using the ρ^0 decay into two pions. The sample contains, in contrast to the K^0 sample, events at high momenta and high z, which cover high likelihood values. Though this method has been working for the 2006 data it was not conclusive with 2016 data probably due to the lower statistics available for the current analysis compared to what was used in 2006.

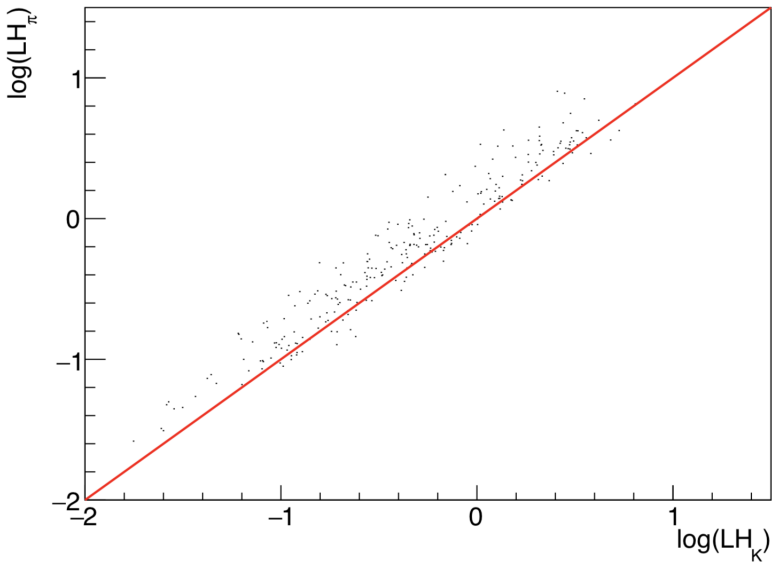
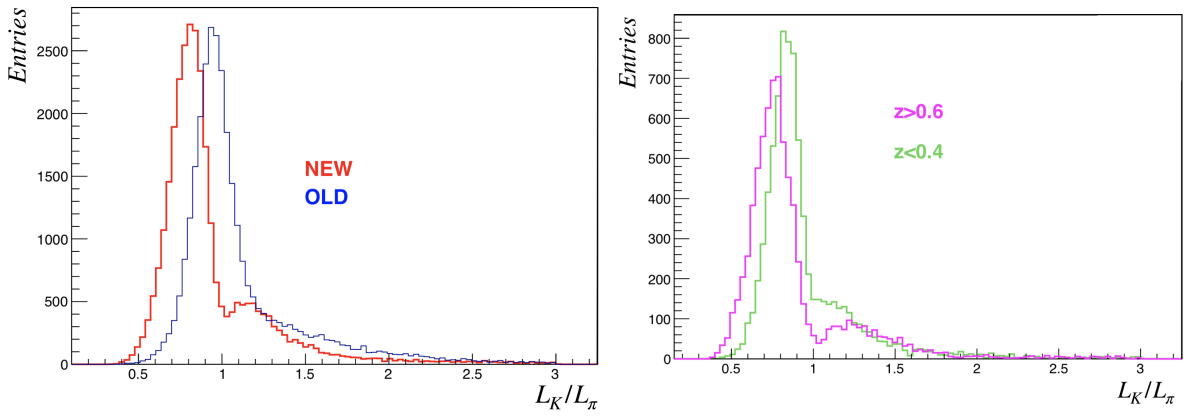


Figure 44: Likelihood for pions as a function of the likelihood for kaons using the 2011 K^0 sample. Figure taken from [83].



(a) (b)

Figure 45: Figure (a) displays the likelihood ratio L_K/L_π before (blue curve) and after (red curve) the new alignement of the RICH. The separation minimum is visible at $L_K/L_\pi = 1$ with the new alignement. Figure (b) compares the likelihood ratio L_K/L_π for $z > 0.6$ (magenta curve) and $z < 0.4$ (green curve). While previously likelihoods were perfoming worse with z, the trend is as expected with the new alignement. Figures taken from [84].

These non-linearities have been adressed recently by the COMPASS RICH group. They have been reporting that the mirrors had to be realigned. This leads to a change of the refractive index and subsequently improving the pion versus kaon likelihood picture, apparently curing the non-linearities. Further tests must still be done in order to quantify the improvement but preliminary studies [84] displayed in Fig. 45 show that the misidentification problem adressed in this section is much reduced. The π peak is now below $L_K/L_\pi = 1$ and the separation minimum close to $L_K/L_\pi = 1$ is visible. In addition, while previously likelihoods were performing worse at high z, now the trend is as expected. This new alignement of the RICH could not be taken into account in the following analysis.

5.8 COMPARISON OF THE EFFICIENCIES OF 2006 TO 2016

A comparison of the RICH efficiencies for 2006, 2011 and 2016 is shown in Figs. 46 to 51. The first data taking with the refurbished RICH was done in 2006. All these effiencies are extracted using the same likelihood cuts. The results point to a good stability of the RICH through time, even improving for some efficiencies, like pion identification for example. All data support the good identification of all species in the selected momentum range highlighting the excellent separation performed by the RICH.

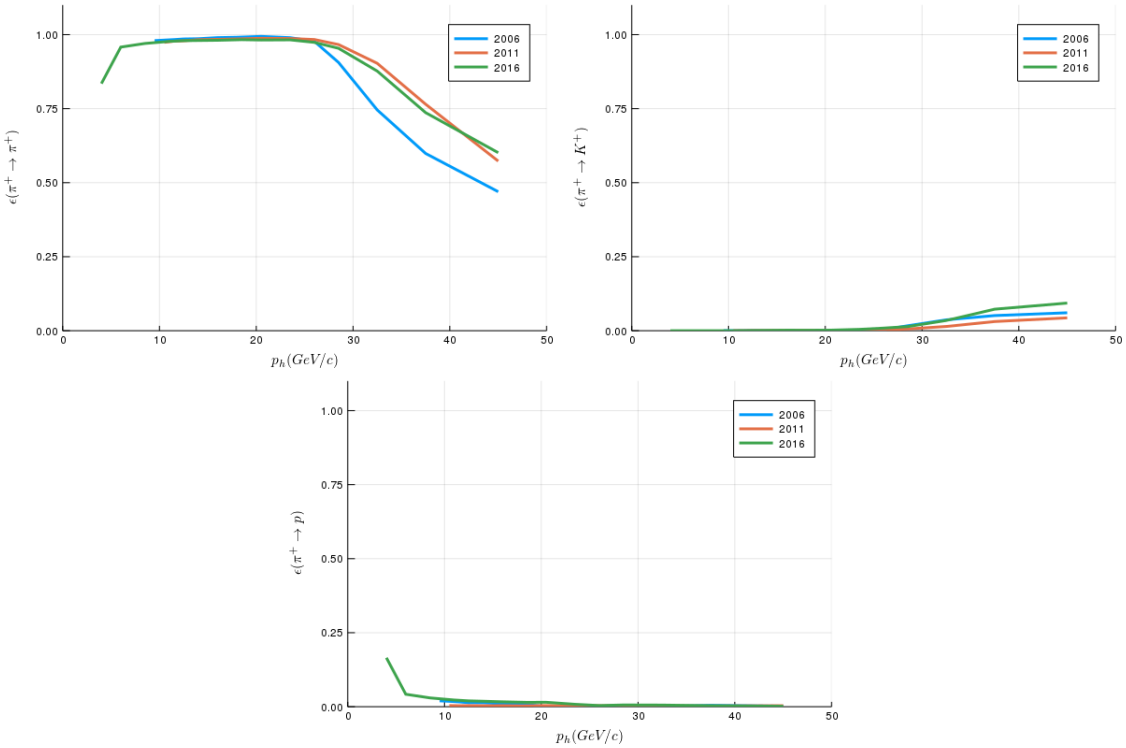


Figure 46: Identification probabilities $\epsilon(p \to j)$ for π^+ in the bin 0.01 rad $< \theta_h <$ 0.04 rad for 3 different years: 2006 (blue), 2011 (orange) and 2016 (green).

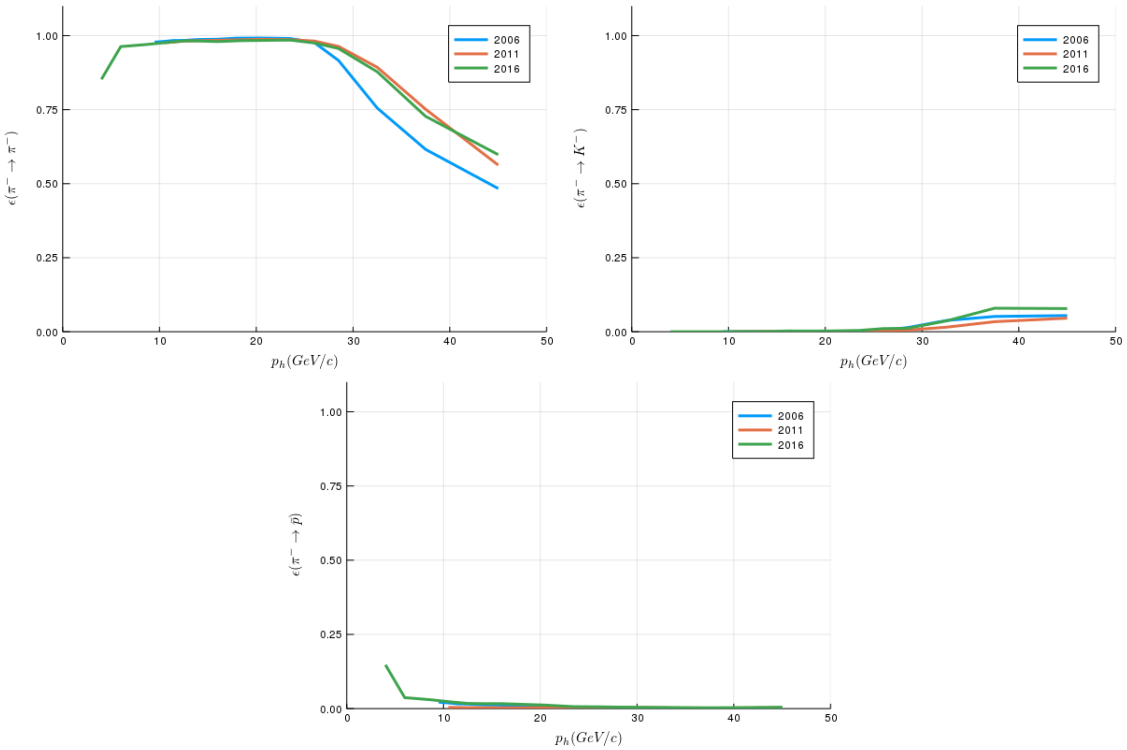


Figure 47: Same as Fig. 46 for $\epsilon(p \to j)$.

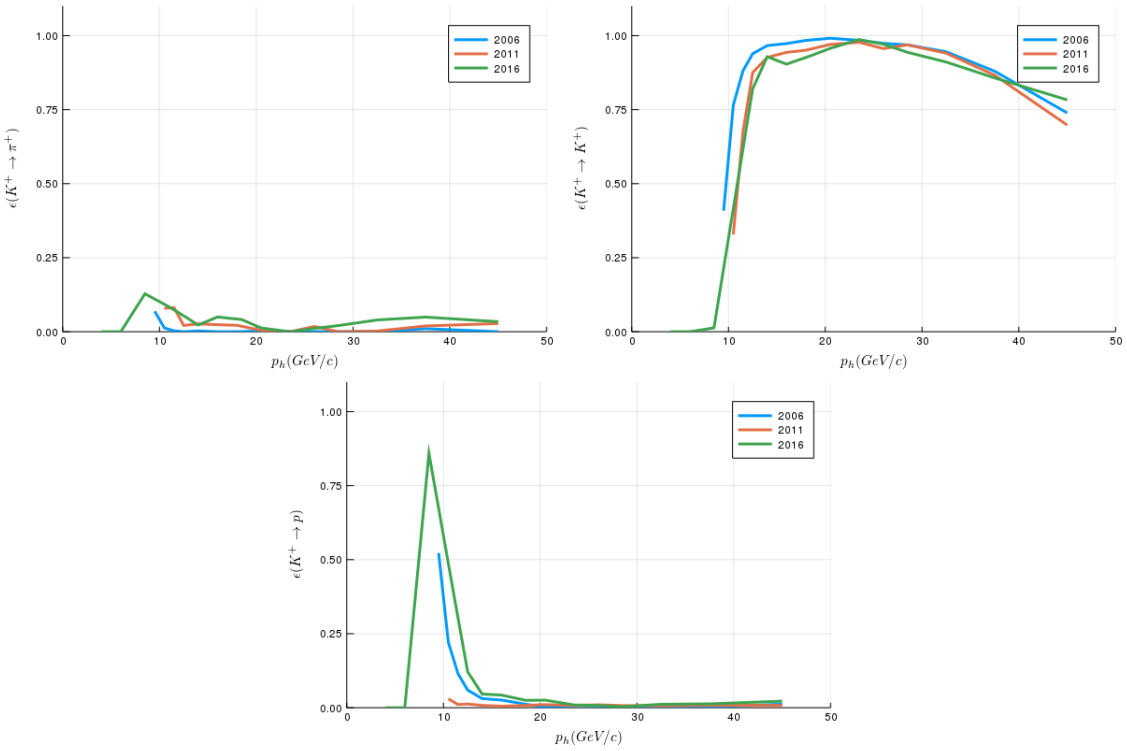


Figure 48: Same as Fig. 46 for $\epsilon(p \to j)$.

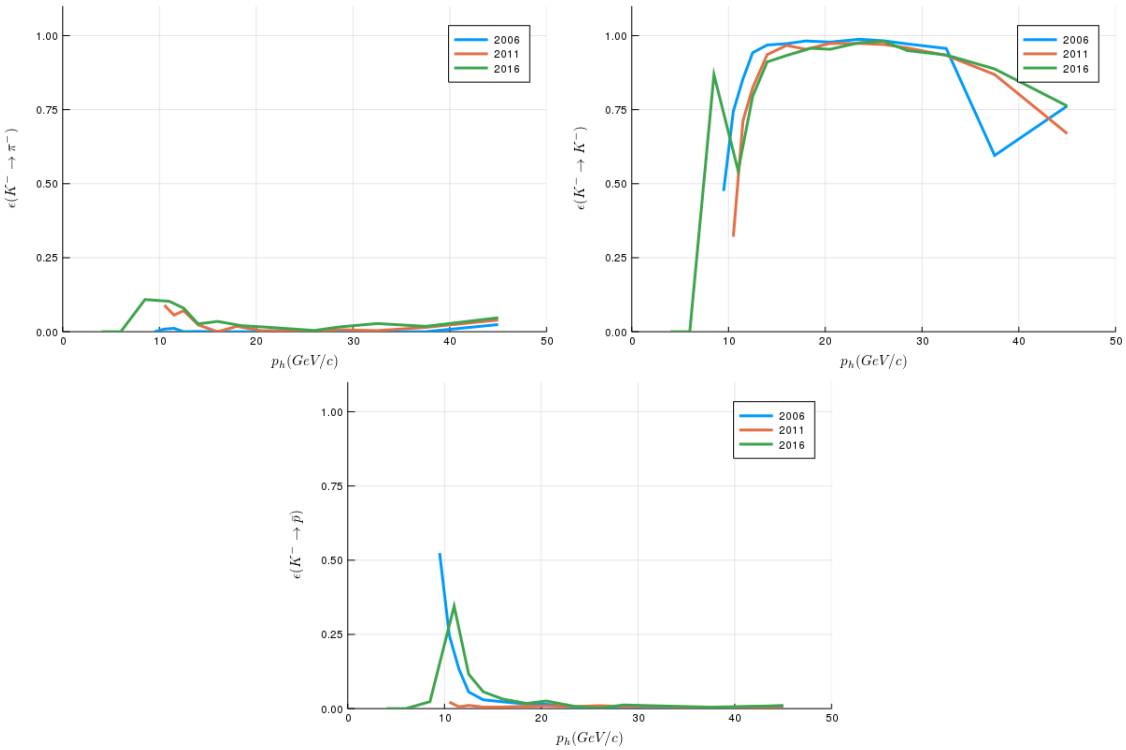


Figure 49: Same as Fig. 46 for $\epsilon(\mathrm{p} \to \mathrm{j})$.

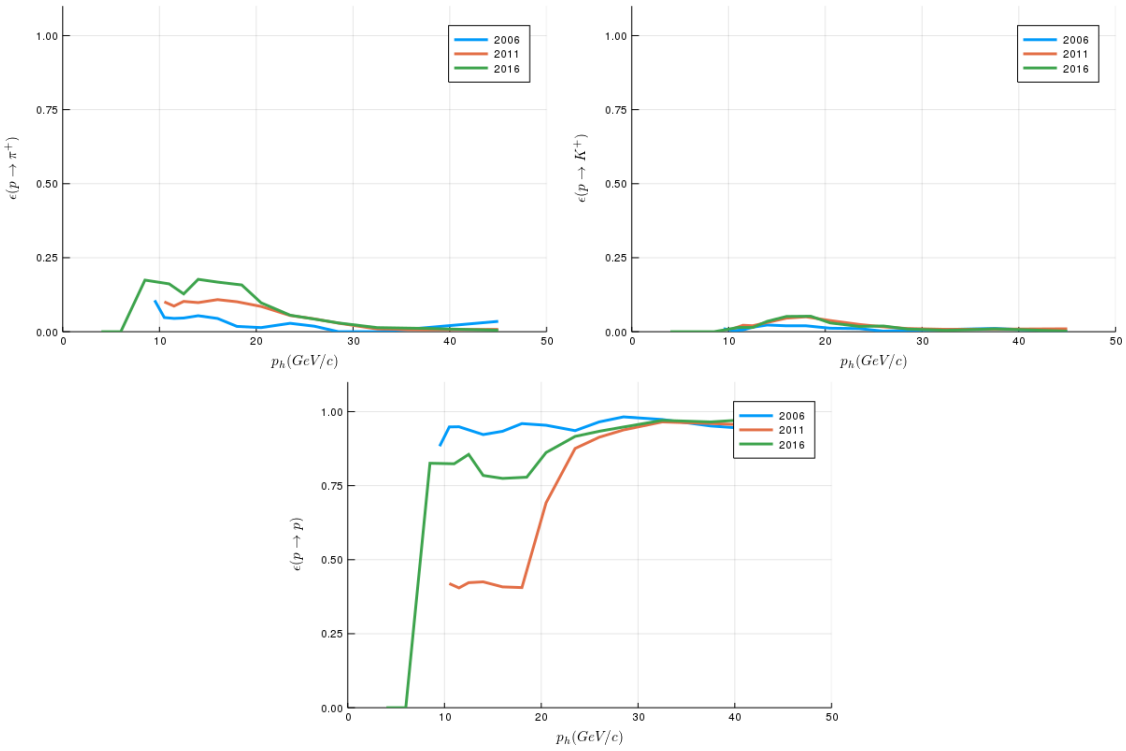


Figure 50: Same as Fig. 46 for $\epsilon(\mathrm{p} \to \mathrm{j})$.

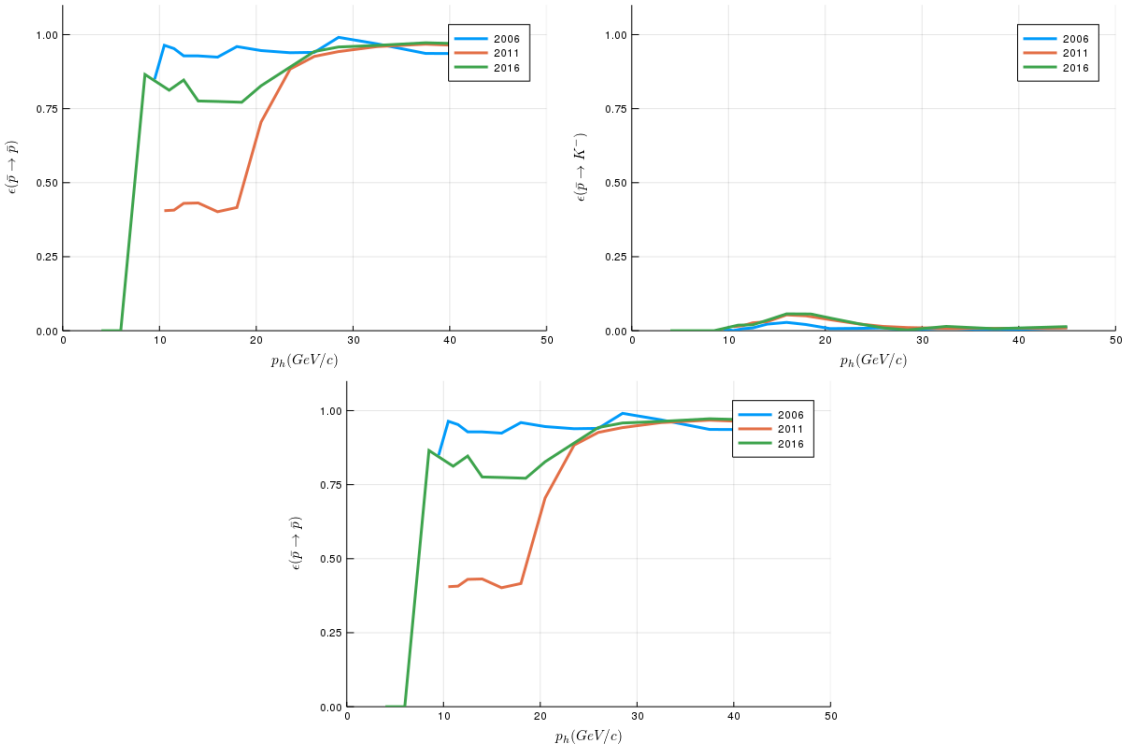


Figure 51: Same as Fig. 46 for $\epsilon(p \to j)$ for $\bar{p}$.

5.9 SUMMARY

The RICH detector performances are determined from real data, using samples of π, K and p coming from the decay into two charged particles of K^0, Φ and Λ. In order to take into account the dependence of the RICH performance with the hadron kinematics, the performances are extracted in bins of the particle momentum p_h and the track polar angle θ_h at the RICH entrance.

High identification probabilities are reached for p_h below 30 GeV/c. For pions, it reaches values larger than 97% and for kaons and protons, it reaches values larger than 90% except for the regions around the kaon and proton thresholds of about 9.45 and about 17.95 GeV/c, respectively. The identification probability values drop for larger p_h values in the pion and kaon case due to the Cherenkov angle saturation ($\beta \to 1$). As a consequence, the misidentification probabilities are larger in this region.

Part III

DJANGOH: A MONTE-CARLO GENERATOR WITH RADIATIVE CORRECTIONS

6 CALCULATION OF RADIATIVE CORRECTIONS

In the past, COMPASS has been using two programs for radiative corrections estimation : one is TERAD, a program that does analytic calculations of the (x, y)-dependent radiative correction factors, the other is the RADGEN event generator, which in addition to the (x, y)-dependent radiative correction factors allows one to take into account a kinematic smearing caused by the radiated photon. A third program is discussed within this thesis: DJANGOH. DJANGOH can compute radiative correction factors in bins of (x, y, z) and generate events with real photon emission.

6.1 TERAD

The TERAD program is based on the calculations described in Refs [59, 85, 86]. These calculations are also referred to as the Dubna radiative correction scheme. A model-independent approach and the QPM are used for computing the deep inelastic processes. An example of the former are radiative corrections to the leptonic current. The QPM is used for all other corrections, i.e. for computing the double photon exchange, real photon emission from quark lines, quark self-energy and weak loop corrections. Also implemented were the $\mathcal{O}(\alpha^2)$ corrections, corresponding to α^4 contributions to the cross-section. The diagrams corresponding to the $\mathcal{O}(\alpha^2)$ corrections are shown in Fig. 52.

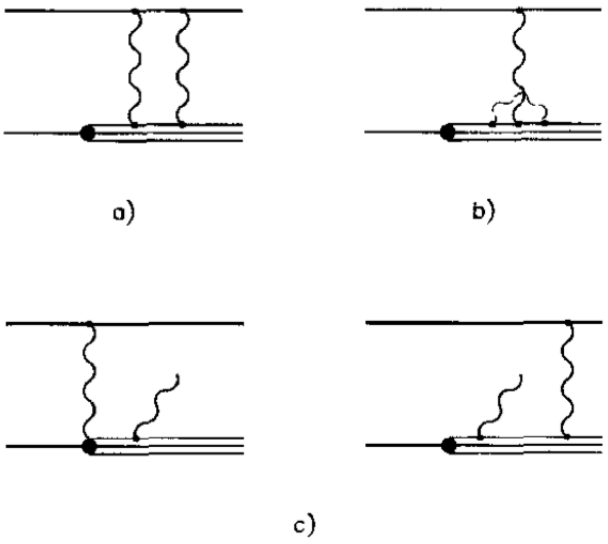


Figure 52: $\mathcal{O}(\alpha^2)$ corrections implemented inside TERAD. Double photon exchange (a) and hadron current corrections (b,c) in the Dubna scheme.

6.2 RADGEN

The RADGEN generator [87] contains two patches, which can be considered as independent generators. The first one is based on POLRAD2.0 [88] and is used for polarized lepton-nucleon scattering. The second one is based on FERRAD3.5 [89] and deals with the unpolarized case. POLRAD2.0 is based on the method of covariant cancellation of infrared di-

vergences developed by Bardin and Shumeiko [90]. FERRAD3.5 is calculating the radiative corrections to DIS of unpolarized particles in accordance with the analytical fomulae given by Mo and Tsai [91]. RADGEN is calculating the total radiative corrections at the lowest order. The Monte-Carlo generator is only considering single photon exchange and pure QED corrections.

Event generation with possible photon radiation is performed as follows. The event is generated with the kinematics of the scattered lepton and a event weight is calculated from these kinematics. According to their weight in the total cross-section, an appropriate scattering channel (non-radiative, (quasi)elastic or inelastic radiative tail) is chosen. If the selected channel is radiative, a photon is emitted and the kinematic variables and the event weight are recalculated.

6.3 DJANGOH

6.3.1 *Presentation of DJANGOH*

First a quick summary of the DJANGOH generator is given :

- DJANGOH [92] is at first a Monte-Carlo event simulation tool for neutral and charged current ep interactions at HERA with the event generators HERACLES and DJANGO6.
- DJANGOH was then modified to also simulate μp interactions at the COMPASS experiment.
- The emphasis is put on the inclusion of QED radiative corrections (single photon emission from the lepton or the quark line, self energy correction, complete set of one-loop weak corrections). The background from radiative elastic scattering $\mu p \to \mu p \gamma$ is also included.
- HERACLES is treating the lp scattering by means of structure function parametrizations or parton distribution functions in the quark-parton model framework.
- DJANGO6 is simulating deep inelastic scattering including both QED and QCD radiative effects.
- DJANGOH is an interface to LEPTO [93], ARIADNE [94] (for parton cascades), PYTHIA [95] (LUND string fragmentation in JETSET [96] for hadronic final state) and SOPHIA [97] (for low-mass hadronic final states).

DJANGOH is able to perform:

- Generation of lp scattering with and without fragmentation for the final state with radiative events.
- Calculation of cross-sections (radiative, born)
- Calculation of radiative correction factors (inclusive, semi-inclusive)
- Generation of events as an event generator in a Monte-Carlo chain

6.3.2 *Radiative scattering*

The treatment of the non-radiative part of the cross-section is straightforward: the Monte-Carlo program HERACLES is used to generate events at the parton level, which in turn are

fragmented and hadronized by DJANGO6 using routines from LEPTO and JETSET. For that reason, in the following we only consider the radiative process:

$$l(p_l) + p(p_p) \to l'(p_{l'}) + \gamma(k) + X(p_X). \tag{72}$$

The particle momenta are given in parentheses. The standard DIS variables (the lepton variables) are defined as the ones given by the inclusive lepton measurement:

$$\begin{aligned} q_{lep} &= p_l - p_{l'}, \quad Q^2_{lep} = -q^2_{lep} \\ x_{lep} = \frac{Q^2_{lep}}{2p_p \cdot q_{lep}}, \quad y_{lep} &= \frac{p_p \cdot q_{lep}}{p_p \cdot p_l}, \quad W^2 = (p_p + q_{lep})^2 = (p_x + k)^2. \end{aligned} \tag{73}$$

The invariant mass W includes the contribution of the radiative photon. At a given center-of-mass energy $\sqrt{s} = p_p + p_l$, only two variables in Eq. 73 are independent of the relations:

$$Q^2_{lep} = x_{lep} y_{lep} s, \quad W^2 = (1 - x_{lep}) y_{lep} s + m_p^2. \tag{74}$$

The kinematics can also be defined through the hadronic state. If the real photon can be identified, one can measure the hadron variables:

$$\begin{aligned} q_{had} &= p_l - p_{l'}, \quad Q^2_{had} = -q^2_{had} \\ x_{had} = \frac{Q^2_{had}}{2p_p \cdot q_{had}}, \quad y_{had} &= \frac{p_p \cdot q_{had}}{p_p \cdot p_l}, \quad W^2_{had} = (p_p + q_{had})^2 = (p_x + k)^2. \end{aligned} \tag{75}$$

Note that the variable y_{had} is defined in terms of $p_p \cdot p_l$ rather than in terms of $p_p \cdot (p_l - k)$. The hadron variables obey relations in analogy to Eq. 76:

$$Q^2_{had} = x_{had} y_{had} s, \quad W^2 = (1 - x_{had}) y_{had} s + m_p^2. \tag{76}$$

In HERACLES, cuts can be imposed on the lepton variables x_{lep}, y_{lep} and Q^2_{lep} and, in addition, on the mass of hadronic final state W_h.

6.3.3 *Factorization*

HERACLES is used to generate variables for the semi-inclusive reaction. The momenta p_l and p_p are input quantities, thus k and $p_{l'}$ output quantities of HERACLES. A given event is then classified according to the channels of HERACLES as being either non-radiative, with leptonic initial-state radiation (ISR), with leptonic final-state radiation (FSR) or a Compton event. For radiative events a rescaled Born-term-like is defined so that in any case we are left with:

$$L(p_L) + p(p_p) \to L'(p_{L'}) + X(p_X). \tag{77}$$

In general both L and L' will be virtual particles. Since our considerations are restricted to $\mathcal{O}(\alpha^3)$, only one of them will be virtual.

Initial state radiation

In case the event is classified as resulting from leptonic ISR, we imagine the two-step process:

$$l(p_l) \to \widetilde{l}(p_{\widetilde{l}}) + \gamma(k), \quad \widetilde{l}(p_{\widetilde{l}}) + p(p_p) \to l'(p_{l'}) + \gamma(k) + X(p_X), \tag{78}$$

thus $L(p_L) = \widetilde{l}(p_{\widetilde{l}})$ and $L'(p_{L'}) = l'(p_{l'})$. Here $p_{\widetilde{l}} = p_l - k$ and DIS variables for the $\widetilde{l}p$ subprocess yield:

$$\begin{aligned} \widetilde{s} &= (p_{\widetilde{l}} + p_p)^2 = s - 2k \cdot (p_l + p_p), \quad \widetilde{Q}^2 = -\widetilde{q}^2 = -(p_{\widetilde{l}} - l')^2 \\ \widetilde{x} = \frac{\widetilde{Q}^2}{2p_p \cdot \widetilde{q}} = \frac{\widetilde{Q}^2}{ys - 2p_p \cdot k}, \quad \widetilde{y} &= \frac{\widetilde{Q}^2}{\widetilde{x}s}, \quad \widetilde{W}^2 = (p_p + \widetilde{q})^2 = (1 - \widetilde{x})\widetilde{y}\widetilde{s} + m_p^2. \end{aligned} \tag{79}$$

The tilde variables coincide with the hadron variables except for s and y.

Final state radiation

A similar rescaling of variables is done for events with leptonic FSR, imagining the process as a hard scattering followed by a decay:

$$l(p_l) + p(p_p) \to \widetilde{l'}(p_{\widetilde{l'}}) + \gamma(k) + X(p_X), \quad \widetilde{l'}(p_{\widetilde{l'}}) \to l'(p_{l'}) + \gamma(k), \tag{80}$$

where $p_{\widetilde{l'}} = p'_l - k$ and:

$$\begin{aligned} \widetilde{s} = s, \quad \widetilde{Q}^2 = -\widetilde{q}^2 = -(p_l - p_{\widetilde{l'}})^2 \\ \widetilde{x} = \frac{\widetilde{Q}^2}{2p_p \cdot \widetilde{q}} = \frac{\widetilde{Q}^2}{ys - 2p_p \cdot k'}, \quad \widetilde{y} = \frac{\widetilde{Q}^2}{\widetilde{x}s}. \end{aligned} \tag{81}$$

Compton events

Compton events are characterized by typically small Q^2 and small W_{had}. For this reason they are not fragmented and no hadronic final state is generated for them. The number of Compton events is small, typically a fraction of a percent if a cut on the mass of the hadronic final state of $W_h \geqslant W_{had}^{min} \sim 2$ GeV is imposed.

6.3.4 *Technical description of DJANGOH*

The computational procedures applied in DJANGOH are based on the methods used in AXO [98] library for Monte-Carlo integration and event generation. AXO relies on the Monte-Carlo integration algorithm VEGAS [99]. The computation is made in this order:

- Integration of the different contributions: partial cross-sections are determined according to the defined phase-space region. They give the relative weight of the corresponding contribution in the final step of event sampling. Moreover, the integration procedure supplies information for the construction of the distribution function applied for event generation.
- Estimation of the local maxima of the distribution function in a predefined number of hypercubes.
- According to the partial cross-sections that were calculated, events are generated randomly from the individual contributions. HERACLES is only taking care of the scattered lepton and the potential radiative photon. DJANGO is simulating the QCD effect and generates the hadronic part of the event.

6.3.5 *Consistency checks*

In order to test the self-consistency of DJANGOH, I generated a certain number of events of μp scattering with an incoming muon energy of 160 GeV. In Fig. 53), the energy of the radiated photon (if one is present), the outgoing muon and the struck quark are shown. The cutoff at low energy is given by the specified kinematic cuts in DJANGOH. Here the kinematic conditions are E_{beam} = 160 GeV, $0.004 < x < 0.4$, $0.1 < y < 0.9$, $Q^2 > 1$ $(Gev/c)^2$ and $W > 4$ GeV/c^2.

We expect naively a peak around $E_\gamma = 0$ GeV as soft photons (low energy photons) are more likely to be emitted than hard photons (high energy photons). The rest of the energy of the incoming muon is distributing accordingly between outgoing muon and struck quark.

Fig. 54 shows the θ_γ and $\theta_{\mu'}$ distributions. As discussed before in Chapter 2, the radiated photon has two privileged directions of emission, namely the s-peak and the p-peak, collinear

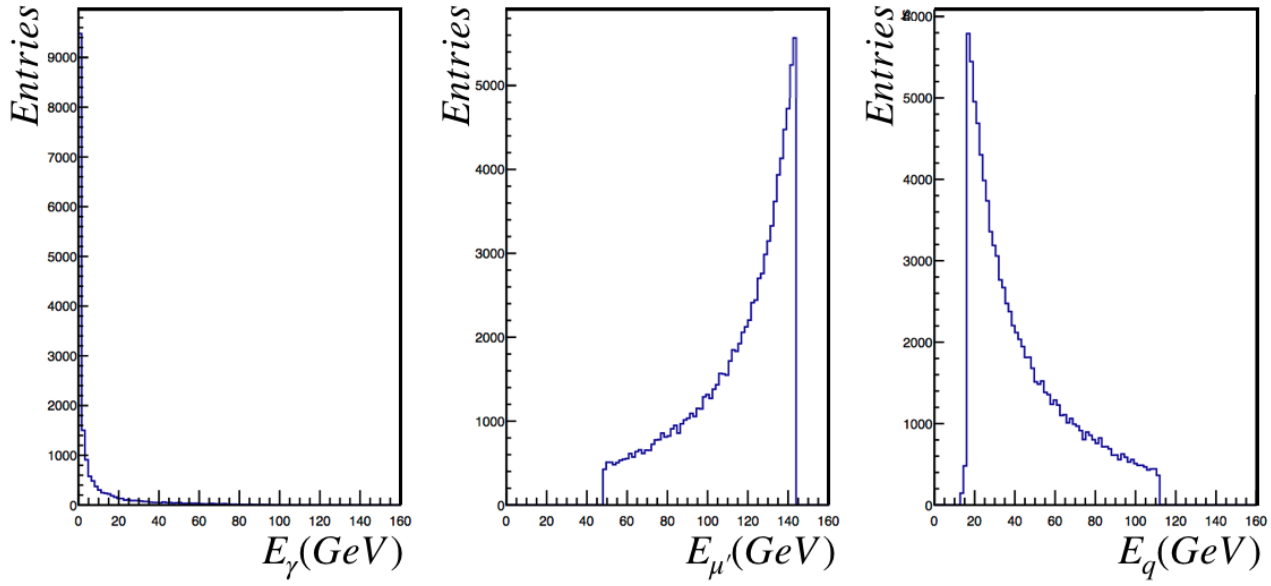


Figure 53: Energy distribution for, from left to right, radiative photon, outgoing lepton and struck quark. The emission of low energy radiative photon is privileged by DJANGOH, while the rest of the energy is distributed between the outgoing muon and the struck quark.

to the direction of propagation of the incoming and outgoing muons. In Fig. 54, the two distributions are plotted next to each other, enabling to see if the p-peak is matching with the position of the peak in the scattering angle of the outgoing muon, which is the case. The s-peak is around 0, which is also expected.

A last exercise that I have done is to see whether DJANGOH, when we put a $Q^2_{lep} > 1$ $(Gev/c)^2$ constraint on event generation, is producing radiative events with $Q^2_{had} < 1$ $(Gev/c)^2$, which would be the case if a radiative photon was emitted by the incoming lepton. A quick proof of this can be made starting from the relation between Q^2_{lep} and Q^2_{had}:

$$Q^2_{had} = Q^2_{lep} + 2E_\gamma(\nu_{lep} - \sqrt{\nu^2_{lep} + Q^2_{lep}}\cos\theta_\gamma)$$

When a real photon is emitted by the incoming lepton, $\cos\theta_\gamma \simeq 1$ then $\nu_{lep} - \sqrt{\nu^2_{lep} + Q^2_{lep}}$ $\cos\theta_\gamma \leqslant 0$ leading to $Q^2_{had} \leqslant Q^2_{lep}$. With an analoguous reasoning, if a real photon is emitted by the outgoing lepton, then $Q^2_{had} \geqslant Q^2_{lep}$. In Fig. 55, Q^2_{had} is shown as a function of Q^2_{lep}, for $Q^2_{lep} = 1$ $(GeV/c)^2$. Values of Q^2_{had} can be found both above and below $Q^2 = 1$, which was the point to be verified. Another information that is given by this plot is that as the distribution around $Q^2_{had} = Q^2_{lep}$ is narrow, most radiative photons are soft, ie. low energetic, as noted previously.

6.3.6 *Improvement of DJANGOH*

An upgrade to the original DJANGOH is the possibility to use different input energies for the incoming lepton for multiple event generation. Before only one input energy could be specified at the launch of the program. DJANGOH is now capable to take into account a new beam energy at each new event. Nonetheless, using different input energies for event generation is causing a problem: the cross-sections that are needed for event generation are depending on this input energy. The naive way would be to recompute the cross-section for each event, but this solution takes much time due to the computation of the cross-section being the slowest part of the generation.

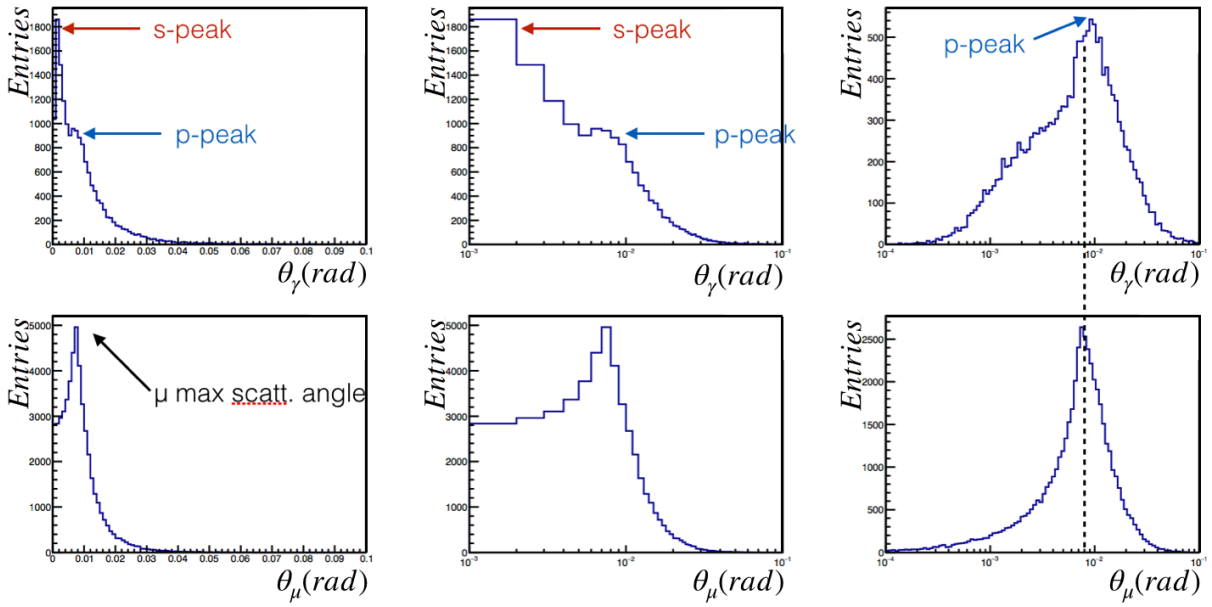


Figure 54: θ angle distribution for the radiative photon (up) and the outgoing muon (down). From left to right the scaling of the x-axis is changed (normal, logarithmic, logarithmic with constant bin size). One can see two peaks in the theta distribution of the radiative photon: one around zero (s-peak) and one a little bit further (p-peak). When compared to the θ distribution of the outgoing muon, especially with the last scaling, one can see the two peaks match.

A solution to circumvent this problem is the use of a grid of cross-sections. This grid is initialized after a rough specification of the type of dispersion in energy of the considered beam. Basically, the grid needs to have a mean energy and the standard deviation of energy to this mean energy, as well as the number of bins in the grid. Then for each bin, the energy of the center of the bin is taken and the cross-section corresponding to this energy is computed. The narrower the bins are the more accurate the cross-section is for the considered bin. As shown in Fig. 56, with a mean energy of 160 GeV, a distribution width of 20 GeV and 20 bins, the grid is giving an accurate map of the cross-sections. Though the difference of cross-section is not very large (5% of the total cross-section), it has to be taken into account for a proper event generation.

6.3.7 *TDJANGOH Interface*

In order to create a C++ class (called TDJANGOH) that plays the role of an interface, I had to modify some FORTRAN parts of DJANGOH, especially the input method. DJANGOH is working with an input file, where codewords with set values are specified in order to configure the generator. This was not convenient for the idea of an interface. Thus I have drawn correspondences between Common Blocks in FORTRAN and structures in C++ so that I can specify values in the C++ structure and the change is repercuted in the FORTRAN code and vice-versa. It is useful to specify the values for the input but also to recover the results of the hadronization that are located in the LUJETS Common Block. The idea is that within TGEANT the user specifies the input for DJANGOH, the interface pass it to the generator and the interface recovers the results of the generator and pass it to TGEANT (see Chapter 9).

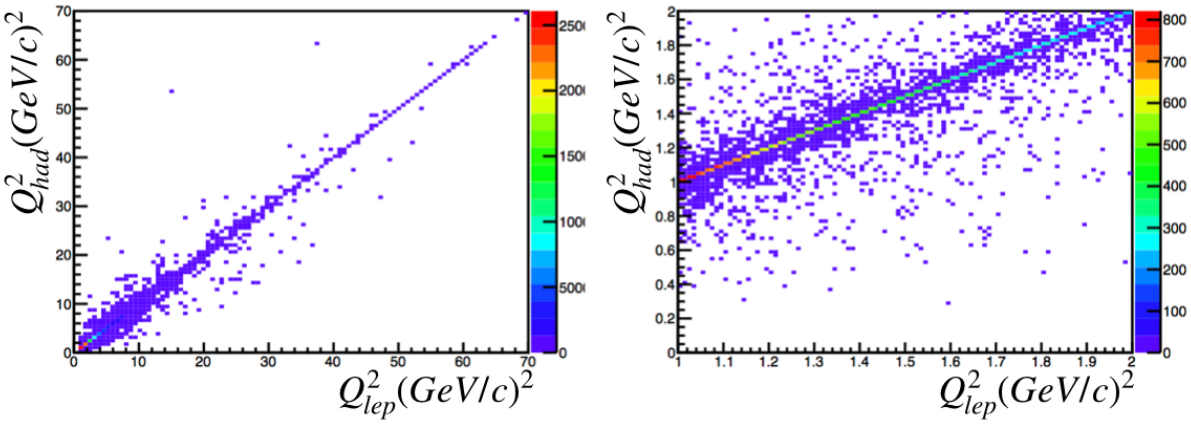


Figure 55: Q^2 correlation plot $Q^2_{had} = f(Q^2_{lep})$. On the left is the plot for the complete range of Q^2 given by the kinematic constraints. The scattering around the $Q^2_{had} = Q^2_{lep}$ line is small, indicating that most of the radiative photons are soft. On the right is the same plot but restricted to the $Q^2_{lep} \in [1, 2]$ $(GeV/c)^2$ and $Q^2_{had} \in [0, 2]$ $(GeV/c)^2$ region. The fact is that for $Q^2_{lep} = 1$ $(GeV/c)^2$, Q^2_{had} takes values above and below $Q^2 = 1$ $(GeV/c)^2$, as expected.

6.4 SUMMARY

The DJANGOH event generator including radiative events has been modified to simulate μp interactions at the COMPASS experiment. After this modification, several consistency checks were performed. All were conclusive. Some further improvements were brought to the generator. The generator was designed to only work with one input energy and now can work with multiple input energies thanks to a cross-section grid binned in beam energy. As DJANGOH is a FORTRAN framework, a C++ shell was built as an C++ interface to DJANGOH named TDJANGOH to be able to use it with the COMPASS MC tool.

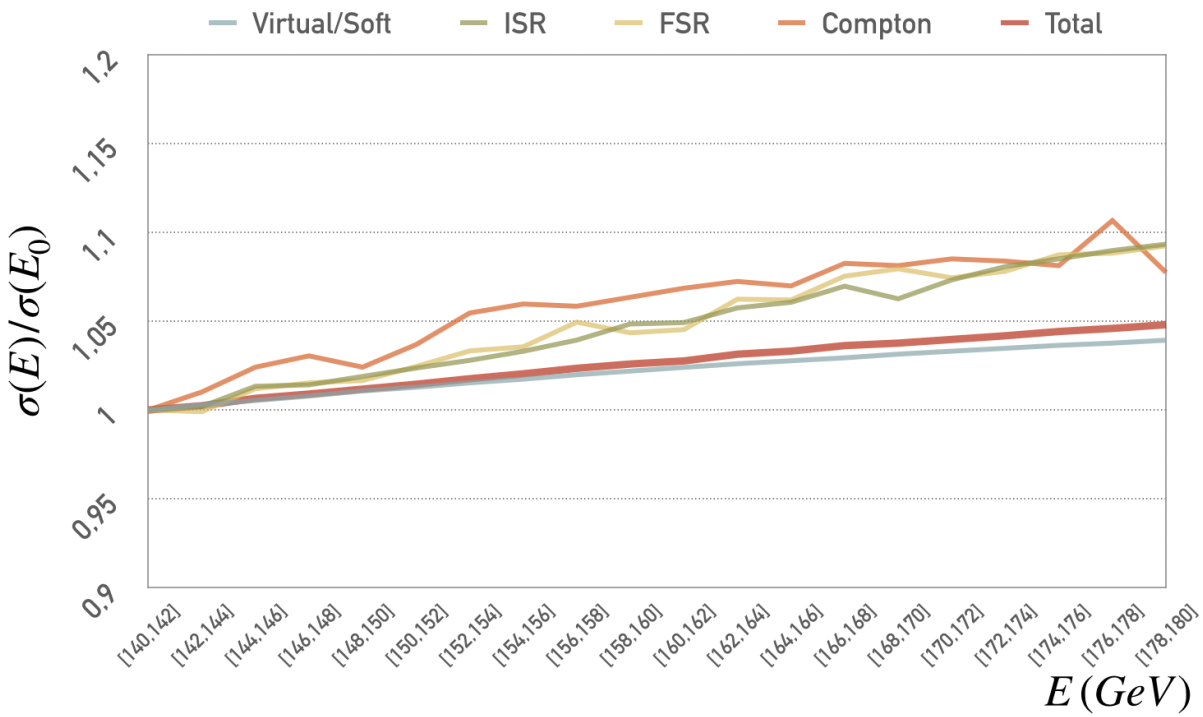


Figure 56: Ratios of cross-section in a given energy bin over the cross-section value at 140 GeV for virtual/soft cross-section (gray blue), initial state radiation cross-section (green), final state radiation cross-section (yellow), compton contribution cross-section (orange) and the total cross-section (red). The variation of the total cross-section with the energy goes up to 5%.

7

ISSUES IN RADIATIVE CORRECTION CALCULATION

7.1 RADIATIVE CORRECTIONS RELATED ISSUES

In the past, COMPASS has been using two programs for radiative corrections estimation: one is TERAD, a program that does analytic calculations of the (x,y)-dependent radiative correction factors, the other is RADGEN, which in addition to the (x,y)-dependent radiative correction factors allows one to take into account a kinematic smearing caused by the radiated photon. Despite being based on different QED calculation schemes, the two programs give compatible results for inclusive corrections.

Before going further in this discussion, some RADGEN formalism must be explained [87]. The relations of the observed cross-section σ_{meas} with the 1-photon exchange cross-section $\sigma_{1\gamma}$ are, for the inclusive case:

$$\sigma_{meas} = \delta_R(\Delta)(1+\delta_{vert}+\delta_{vac}+\delta_{sm})\sigma_{1\gamma}+\sigma_{el}+\sigma_{qel}+\sigma_{in}(\Delta), \tag{82}$$

and for the semi-inclusive case:

$$\sigma_{meas} = \delta_R(\Delta)(1+\delta_{vert}+\delta_{vac}+\delta_{sm})\sigma_{1\gamma}+\sigma_{in}(\Delta), \tag{83}$$

where $\sigma_{1\gamma}$ is the one photon exchange Born cross-section and δ_{vac}, δ_{vert} and δ_{sm} are corrections due to vacuum polarization by electron and muon pairs, vertex corrections and residuum of the cancellation of infrared divergent terms independent of the cut-off parameter Δ:

$$\begin{aligned} \delta_{vac} = \frac{\alpha}{\pi}\left[-\frac{20}{9}+\frac{2}{3}\ln\frac{Q^2}{m_e^2}+\frac{2}{3}\ln\frac{Q^2}{m_\mu^2}\right], \quad \delta_{vert} = \frac{\alpha}{\pi}\left[-2+\frac{3}{2}\ln\frac{Q^2}{m^2}\right], \\ \delta_{sm} = \frac{\alpha}{\pi}\left[-\frac{\pi^2}{6}+\mathrm{Li}_2\left(\cos^2\frac{\theta}{2}\right)-\frac{1}{2}\ln^2(1-y)\right], \end{aligned} \tag{84}$$

where m is the lepton mass and $\mathrm{Li}_2(x) = -\int_0^x \ln(1-y)/y\,dy$ is the dilogarithm. The cross-sections σ_{el}, σ_q and σ_{in} are the contribution from radiative processes for elastic, quasielastic and deep inelastic scattering in case of nuclear targets. They can be calculated in terms of the radiative tail from j^{th} mass level σ_j:

$$\sigma_{el} = \sigma_{el}(M_A,1), \quad \sigma_q = \sigma_q(M,1), \quad \sigma_{in} = \int_{M+m_\pi}^{W} \frac{d\sigma_{in}(M_h,\theta_{max})}{dM_h}, \tag{85}$$

where $M_h = \sqrt{W_{had}^2}$ and σ_j (j = el, q, in) has the form of an integral over θ_γ, the angle between the real and the virtual photons:

$$\sigma_j(M,\theta_{max}) = \int_0^{\theta_{max}} T_0(W_2^j(T_1+T_2+T_3+T_4+T_5)+W_1^j(T_6+T_7))d\theta_\gamma. \tag{86}$$

The structure functions $W_{1,2}^j$ have a different expressions for the different types of the tail. The termes T_i are kinematical factors [91]. θ_{max} is given as $\theta_{max} = min(\pi,\theta_\Delta)$ where θ_Δ is obtained via the relation:

$$W_{had}^2 = W^2 - 2\Delta(\nu+M-\sqrt{\nu^2+Q^2}\cos(\theta_\Delta)). \tag{87}$$

The cut-off parameter Δ is introduced to divide the integration region over the photon energy in the soft and hard energy region. We already saw that the hard energy region $\sigma_{in}(\Delta)$ can be computed without any approximations. The correction factor for the soft energy part is given by:

$$\delta_R(\Delta) = exp\left[-\frac{\alpha}{\pi}\left(\ln\frac{E}{\Delta} + \ln\frac{E'}{\Delta}\right)\left(\ln\frac{Q^2}{m^2} - 1\right)\right]. \tag{88}$$

In RADGEN the difference between the inclusive and semi-inclusive cross-section is only the elastic σ_{el} and quasielactic σ_{qel} cross-sections. The core of the problem is however in the calculation of σ_{in}. The semi-inclusive radiative corrections are given for the events with a hadron observed in the final state. The quantity $\sigma_{in}(M_h, \theta_{max})$ is known and is not a problem but the lower limit of the integral is. In fact:

$$W \geqslant M + M_\pi, \quad E_\gamma \geqslant \frac{Q^2_{had}}{2M} + M_\pi \xrightarrow{Q^2_{had}\to 0} M_\pi \tag{89}$$

As there must be energy conservation, photons with $\nu_{had} < 11$ GeV interacting with a proton at rest cannot produce a hadron with 12 to 40 GeV. This is the reason why we cannot use up to now TERAD for semi-inclusive correction calculation as it yields too large corrections. The same reasoning holds with RADGEN when it is not used as a generator. But used as a generator, it should be working and thus obviously there is a problem somewhere in the code of the generator.

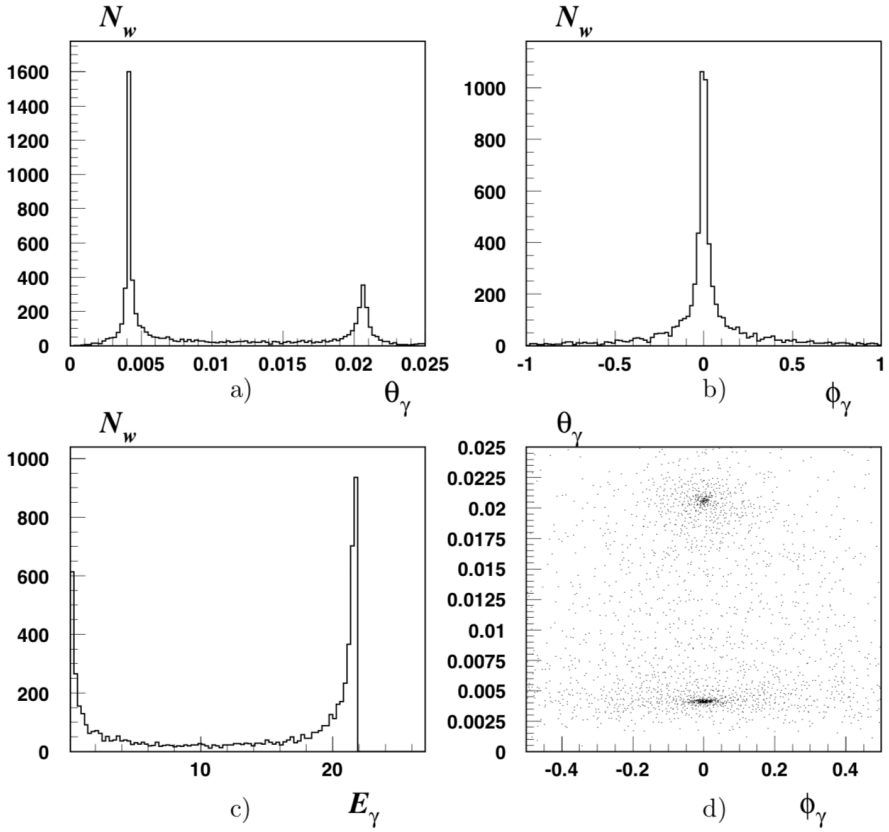


Figure 57: Distribution calculated by RADGEN for one point in HERMES kinematics ($x = 0.1$, $y = 0.8$, $E \approx 27.5$ GeV, $\nu_{obs} \approx 22$ GeV). The distribution of the radiation angles θ_γ a), ϕ_γ b) and of the energy c) of the radiated photon for $x = 0.1$ and $y = 0.8$. The two-dimensional distribution d) shows θ_γ vs ϕ_γ. In panel c), one can note that the hard photon emissions are sizeable. Figure taken from [87]

When looking at the radiated photon energy distribution calculated by RADGEN for one point in HERMES kinematics ($x = 0.1$, $y = 0.8$, $E \approx 27.5$ GeV, $\nu_{obs} \approx 22$ GeV) as represented

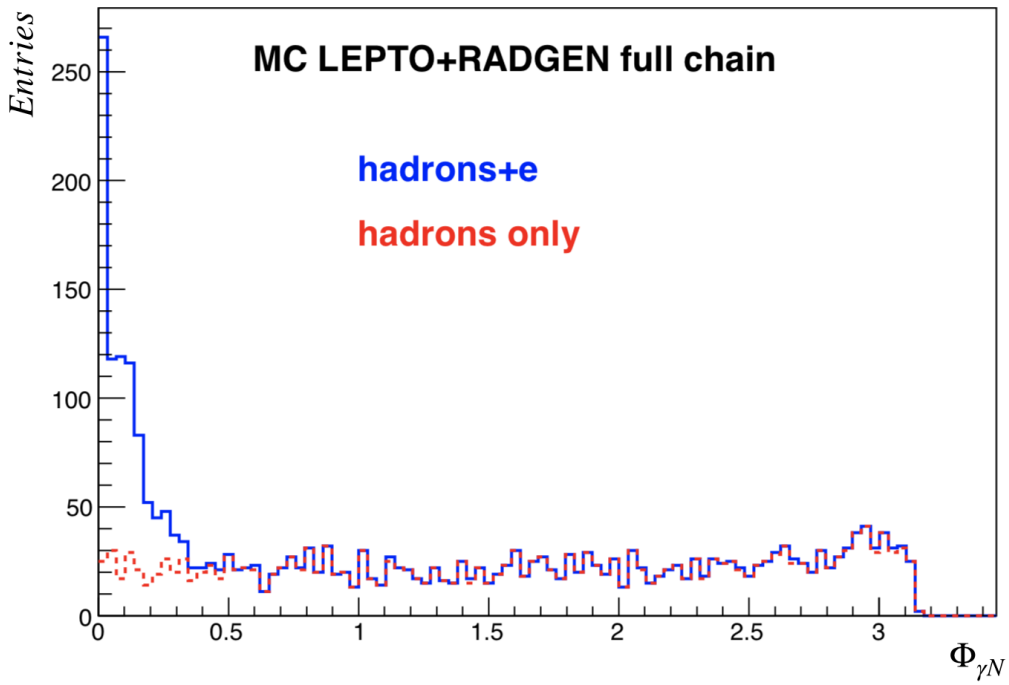


Figure 58: Hadron (red) and hadron+electron (blue) Monte-Carlo distributions versus $\Phi_{\gamma N}$ in the γ-nucleon reference plane. Figure taken from [100].

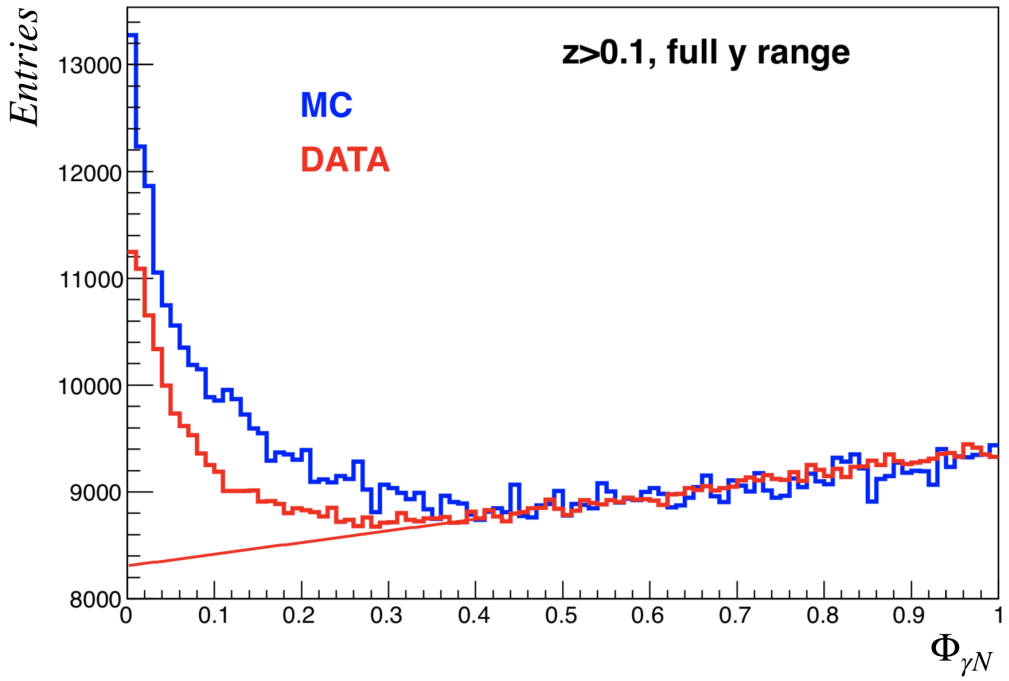


Figure 59: Electron distribution versus $\Phi_{\gamma N}$ in the γ-nucleon reference plane for $3 < p_e < 8$ GeV (region where RICH can discriminate electron in real data). Real Data are in red, Monte-Carlo with RADGEN in blue. Figure taken from [100].

in the lower-left panel of Fig. 57, hard photon emissions are sizeable. The consequences of having such radiated photon energy distribution can be investigated by looking at actual COMPASS data, e.g. 2006 data. As the ^{6}LiD target has a non-negligible radiation length part of photons will create e^+e^- pairs and electrons from the conversion should be seen in the spectrometer. When performing Monte-Carlo generation with COMGEANT+RADGEN,

a large amount of electrons is produced. Fig. 58 shows the distribution of electrons and hadrons as a function of $\Phi_{\gamma N}$. The electrons are concentrated around $\Phi_{\gamma N} \sim 0$. By comparing the Monte-Carlo generation with the real data for electrons with $z > 0.1$ and full y range (Fig. 59), one can note that on average about 1.8 more electrons are produced in Monte-Carlo than in real data.

7.2 COMPARISON BETWEEN RADGEN AND DJANGOH

As one saw in the previous section, a rather problematic result is happening with RADGEN. The generator is producing a high number of hard photons leading to great discrepancies with the real data for the electron distribution. It is expected that DJANGOH will yield a better description of the data. As a first step the results of RADGEN and DJANGOH for the distribution of radiative photons are compared. In Fig. 60, the real photon energy E_γ distributions are displayed for both generators for $0.8 \leqslant y \leqslant 0.9$ and $1 \leqslant Q^2 \leqslant 2(\mathrm{GeV/c})^2$. One can see that overall, without comparing any number, DJANGOH is producing less hard photons than soft photons in contrast to RADGEN.

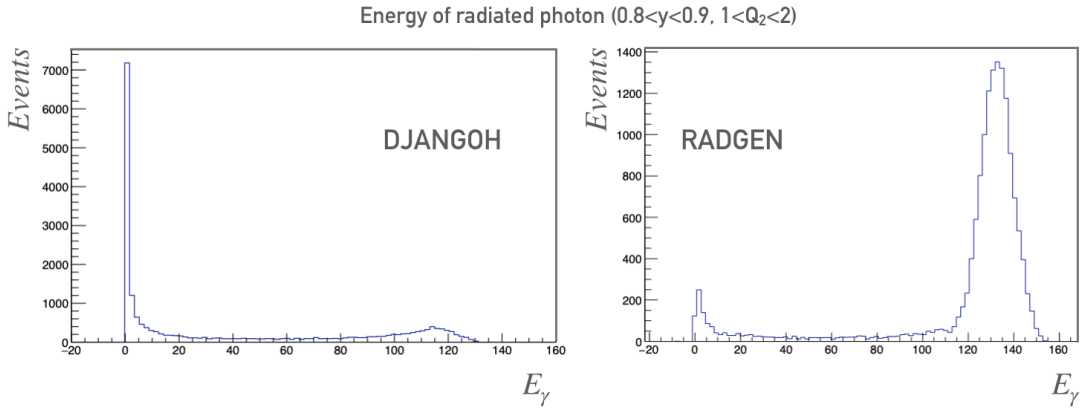


Figure 60: Left is the radiative photon energy distribution for $0.8 \leqslant y \leqslant 0.9$ and $1 \leqslant Q^2 \leqslant 2(\mathrm{GeV/c})^2$ for DJANGOH, right is the same distribution for RADGEN. DJANGOH is producing overall more soft photons than hard ones, unlike RADGEN. Nevertheless, the plot does not allow to conclude whether DJANGOH is producing less hard photons than RADGEN.

By comparing more thoroughly, computing the proportion of radiative events in the range between 20 and 160 GeV over the total number of DIS events (Fig. 61), RADGEN has a total of 18.6% of event in this range when DJANGOH reaches a total of only 8.6%. We can then definitely claim that DJANGOH is producing less hard photons than RADGEN, which is encouraging for the comparison to real data.

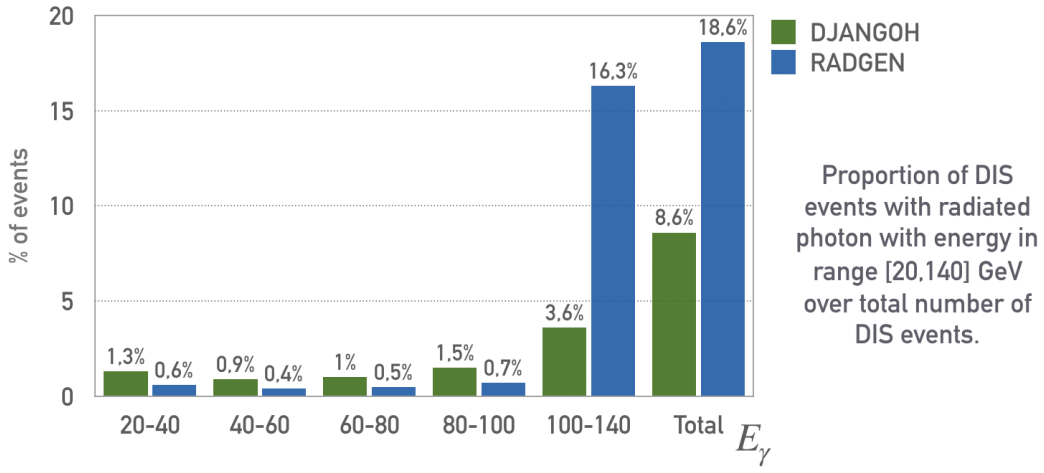


Figure 61: Comparison of proportion of event in high energy radiative photon range over total number of DIS event with the same kinematical restrictions as in Fig. 60. This comparison allows to conclude that DJANGOH is indeed producing less hard photons than RADGEN with more than a factor 2 between the two generators. Thus, RADGEN is producing much less soft photons than DJANGOH, when referring to Fig. 60. Figure taken from [101]

7.3 SUMMARY

RADGEN, when coupled to our MC simulation, was found not to be reproducing COMPASS data on the electron distribution. In fact it is producing a factor 1.8 more electron than what is seen in real data. This is due to the fact that RADGEN produces more hard photons than soft photon, thus leading to a high electron production due to photon conversion in the target. A comparison was done between RADGEN and DJANGOH. It shows that DJANGOH is producing more soft photons than hard photons, but that it also produces less hard photons than RADGEN, leading to a smaller electron production. This gave us incentive to implement DJANGOH in our MC simulation and compare the results of MC to real data.

8

RESULTS ON RADIATIVE CORRECTIONS

The DJANGOH generator can be used in two ways: as a event generator inside a MC simulation or in a standalone way to compute radiative corrections. Here, DJANGOH is able to compute radiative correction factors in bins of (x,y,z), which was not possible with the previously used TERAD code.

8.1 INCLUSIVE RADIATIVE CORRECTION FACTORS

The calculation of the inclusive radiative correction factors is done by computing σ_{Born} and $\sigma_{Born+\mathcal{O}(\alpha)}$. A correct way to obtain these factors is the following:

$$\eta(x,y) = \frac{\sigma_{Born}(x,y)}{\sigma_{Born+\mathcal{O}(\alpha)}(x,y)} = \frac{\frac{\sigma_{Born,tot}\cdot N_{Born}(x,y)}{N_{Born,tot}}}{\frac{\sigma_{Born+\mathcal{O}(\alpha),tot}\cdot N_{Born+\mathcal{O}(\alpha)}(x,y)}{N_{Born+\mathcal{O}(\alpha),tot}}}, \tag{90}$$

where $\sigma_{Born+\mathcal{O}(\alpha),tot}$ and $\sigma_{Born,tot}$ are the integrated Born+$\mathcal{O}(\alpha)$ and Born DIS cross-section over the imposed kinematic boundaries respectively, $N_{Born+\mathcal{O}(\alpha),tot}$ and $N_{Born,tot}$ are the total number of DIS events generated with Born+$\mathcal{O}(\alpha)$ and Born DIS cross-section respectively, and $N_{Born+\mathcal{O}(\alpha),tot}(x,y)$ and $N_{Born,tot}(x,y)$ are the DIS events generated in a given (x,y) bin with Born+$\mathcal{O}(\alpha)$ and Born DIS cross-section respectively. The results that are presented below are obtained with the TERAD F_2 and R parametrizations, which describes accurately the behaviour of F_2 at low Q^2 [102]:

- $F_2^p(x,Q2)$ for $Q^2 > 0.2$ GeV^2 and $0.000035 < x < 0.85$, as obtained from a fit to the world proton (and deuteron) data made by the SMC [103].
- For $Q^2 < 0.2$ GeV^2, a phenomenological model of Badelek and Kwiecinski [104], valid at $10^{-5} < x < 0.1$ and $0 < Q^2 < 1000$ GeV^2.
- $R(x,Q^2)$ as parameterised by SLAC (newer version, called R1998 [105]), valid for $Q^2 > 0.5$ GeV^2, extended to lower values of Q^2, including the $R \simeq Q^2$ behaviour at $Q^2 = 0$.

All first order QED corrections are included except for quark line radiation. The reason why these corrections are not included is that these corrections are negligible except at large $x \geqslant 0.5$ and $Q^2 \geqslant 10^3$ $(GeV/c)^2$, where the corrections reach the magnitude of barely one percent (see Fig. 62). In addition, these corrections are often not subtracted inside the parametrization, thus they are already taken into account in the parametrization [106]. In contrast they are considered in TERAD in a QPM-like approach.

8.2 COMPARISON BETWEEN DJANGOH AND TERAD

DJANGOH results for inclusive radiative corrections are compared with TERAD [107]. The two programs are using the same set of F_2^p and R. This are the only inputs (apart from the process input ie. μp scattering at 160 GeV muon energy) that need to be identical so that the comparison is relevant. One thing to be noted is that TERAD is using in addition $\mathcal{O}(\alpha^2)$ corrections. They should have an impact on the cross-section of TERAD, but are negligible.

The first check for consistency done in Fig. 63 is to compare the σ_{Born} of both programs. Using the same input information on $F_2^p(x,Q2)$ and $R(x,Q^2)$ does not guarantee that σ_{Born}

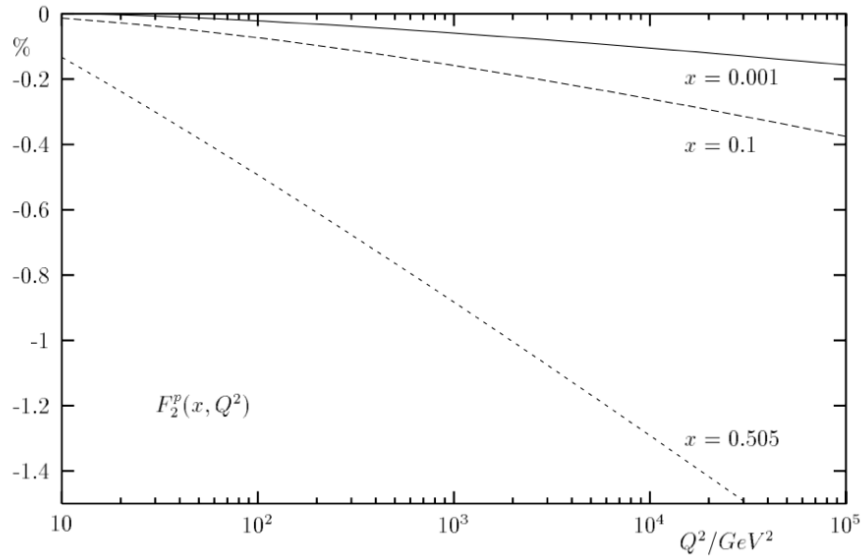


Figure 62: Q^2 dependence of the quarkonic QED corrections (in percent) to the structure function F_2^p for deep inelastic lepton-proton scattering at $x = 0.001$, $x = 0.1$ and $x = 0.505$. Figure taken from [106].

for both program, called σ^D_{Born} and σ^T_{Born}, is the same, even if the structure functions should define the Born cross-section unambiguously. The reason is that in TERAD, σ^T_{Born} is computed without constants like π, M_{proton}, α, etc. and its functional form is unknown, unlike in DJANGOH. This means that the ratio $r = \frac{\sigma^D_{Born}}{\sigma^T_{Born}}$ may slightly differ from 1 but must be constant as a function of x and y.

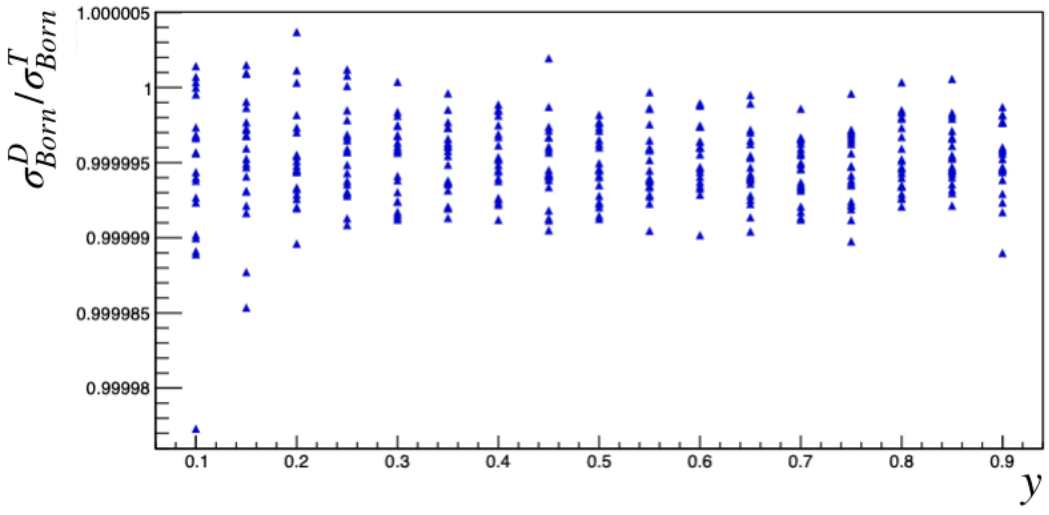


Figure 63: Ratio of the Born cross-sections calculated with DJANGOH and TERAD, for the same F_2 and R parametrizations, as a function of y at different values of x (staggered points at fixed y)

The radiative correction factors $\eta(x, y)$ for DJANGOH (η_D) and TERAD (η_T) are compared in Fig. 64. The relative difference shows that the two programs differ at most 3% in the region of lowest x and highest y. This results is extremely good, knowing that DJANGOH and TERAD are not using the same renormalization scheme.

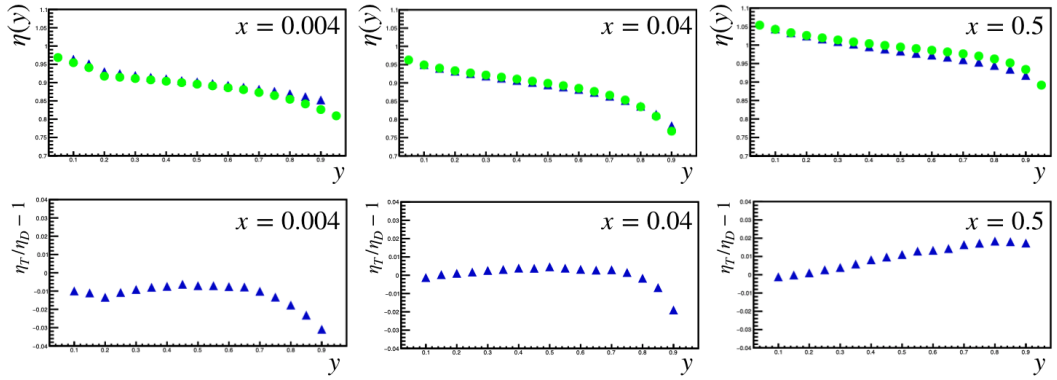


Figure 64: In top panels, comparison of radiative corrections factor $\eta(y)$ for fixed values of x, computed for proton target at 160 GeV and with the same F_2^p and R parametrizations. Green dots mark results of TERAD, blue triangles the results of DJANGOH. In bottom panels, relative difference of radiative corrections factors $(\eta_T/\eta_D) - 1$ as a function of y for fixed values of x

8.3 RADIATIVE CORRECTION FACTORS: EFFECT ON MULTIPLICITIES

The calculation of the radiative correction factors for the multiplicities can be done by computing M^h_{Born}, multiplicities obtained without radiative corrections, and $M^h_{Born+\mathcal{O}(\alpha)}$, multiplicities obtained with radiative corrections. The factors are given as:

$$\begin{aligned} \eta^h(x,y,z) &= \frac{M^h_{Born}(x,y,z)}{M^h_{Born+\mathcal{O}(\alpha)}(x,y,z)} \\ &= \frac{N^h_{Born}(x,y,z)/N^{DIS}_{Born}(x,y)}{N^h_{Born+\mathcal{O}(\alpha)}(x,y,z)/N^{DIS}_{Born+\mathcal{O}(\alpha)}(x,y)} \end{aligned} \tag{91}$$

where N^h is the number of hadrons and N^{DIS} the number of DIS events.

For the calculation of the radiative correction factors the cuts from the SIDIS analysis for the selection of DIS events and hadrons are used (see Chapter 10 for further details). The kinematical cuts used are:

- $0.004 \leqslant x \leqslant 0.4, x \in \{.004, .01, .02, .03, .04, .06, .1, .14, .18, .4\}$
- $0.1 \leqslant y \leqslant 0.7, y \in \{.1, .15, .2, .3, .5, .\}$
- $12 \leqslant p_h \leqslant 40$ GeV/c

Fig. 65 exhibits the semi-inclusive radiative correction factor $\eta^h(x,y,z)$. This factor goes from 2% correction at high x to 20% correction at high z and high y. This dependence on y and z is expected (for instance if a hadron has a high z in a non-radiative event, consider the same event but with the radiation of a real photon, ν_{lep} will remain the same but the hadron will have in reality less energy available from the virtual photon, thus having $z_{had} \leqslant z_{lep}$, leading to less events in the high z region for the multiplicities obtained with radiative correction). The results are obtained at generator level but they should be the same if one uses reconstructed MC to compute them. However as a huge number of events is needed to obtain such results (about 1 billion events), computing them would require an amount of reconstructed MC that largely overshoot our biggest MC productions.

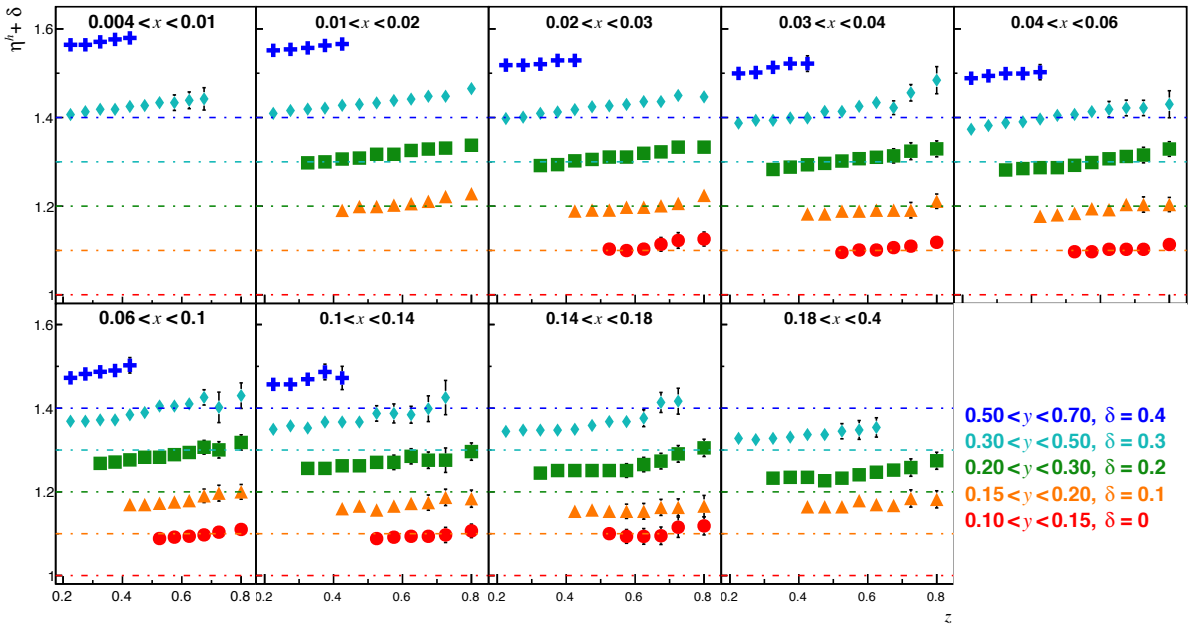


Figure 65: $\eta^h(x, y, z)$, positive hadrons in full points, negative in open points, in bins of x, staggered with y and versus z. The corrections go from 2% at high x to 20% at high z and high y.

8.4 SUMMARY

The DJANGOH event generator with radiative events is a way to access radiative correction factor in a three dimensional (x,y,z) binning, like in the multiplicity analysis. This is the first time that such corrections are available for the COMPASS data. The size of the correction, going from 2 to 20% is within the expectations. Moreover, this correction can directly be applied on the multiplicities. When compared with TERAD computation on the inclusive corrections, the results of DJANGOH are shown to be compatible within 3%, a really good result knowing that DJANGOH and TERAD are not using the same renormalization scheme. A four-dimensional (x,y,z,p_T) binning could also be done but it would need way more statistics that what was used for the three dimensional (x,y,z) binning.

9 INTEGRATION OF DJANGOH IN THE MONTE-CARLO CHAIN

As DJANGOH was giving encouraging results, we decided to implement the generator in our Monte-Carlo simulation named TGEANT. In this chapter we present TGEANT, what it is and how it works. The integration of the DJANGOH generator is then discussed. Lastly result of Monte-Carlo with DJANGOH for electron production is compared with real data.

9.1 TGEANT

TGEANT [108] is the COMPASS MC tool that was developed in object-oriented C++ and based on the Geant4 toolkit. TGEANT has extended the basic framework of Geant4 in order to simulate all aspects of the COMPASS experiment. A short insight into the event simulation is given in this section. The concept of application programming interfaces ensures a highly flexible design. Up to now TGEANT could be used with a number of generator as LEPTO, HEPGEN, PYTHIA etc.

9.1.1 *Software package*

The measured data is a convolution of physics signals with effects related to the experimental acceptance. These effects are introduced by the geometry of the experimental apparatus, by the reconstruction algorithms, by efficicies and by resolution of detectors. The extraction of physics signals requires an extensive knowledge of experimental acceptance, which can only be obtained by Monte Carlo. To perform a full Monte Carlo simulation, several aspects have to be taken into account. For the simulation of the primary physics interaction a dedicated event generator is used, the transport of all secondary particles through the experimental apparatus and the simulation of their interactions with material and of the detector response is carried out by the Monte Carlo software. The goal of TGEANT is to simulate the response of the experimental setup for a physics process under study. Basic requirement is an accurate geometry description of the experimental apparatus. The implementation of particles and physics models is necessary to simulate the physics interactions of particles with matter and the particle transportation. Geant4 offers data bases, which comprise a huge field of applications for all kinds of particles and processes. In TGEANT, the list of physics processes and particles is optimized to the kinematic range of the COMPASS experiment. TGEANT is delivered in a software package with four other sub-packages, namely the graphical user interface and the Toolbox as well as the two libraries libSettings and libEvent, which are needed for the data exchange between the different packages. The whole project is maintained in a Git repository on a CERN server. The interplay of the different software packages is presented in Fig. 66. The TGEANT output files can be either used in CORAL or in the Toolbox. The former case is the standard way of the Monte Carlo reconstruction, while the Toolbox is used particularly with regard to detector studies and tuning.

9.1.2 *Event simulation*

The event loop in TGEANT is the major part of the simulation software. One or more so-called primary particles are placed with a given momentum vector in the world volume. After the initialization phase, the event loop is started and primary particles are tracked by the

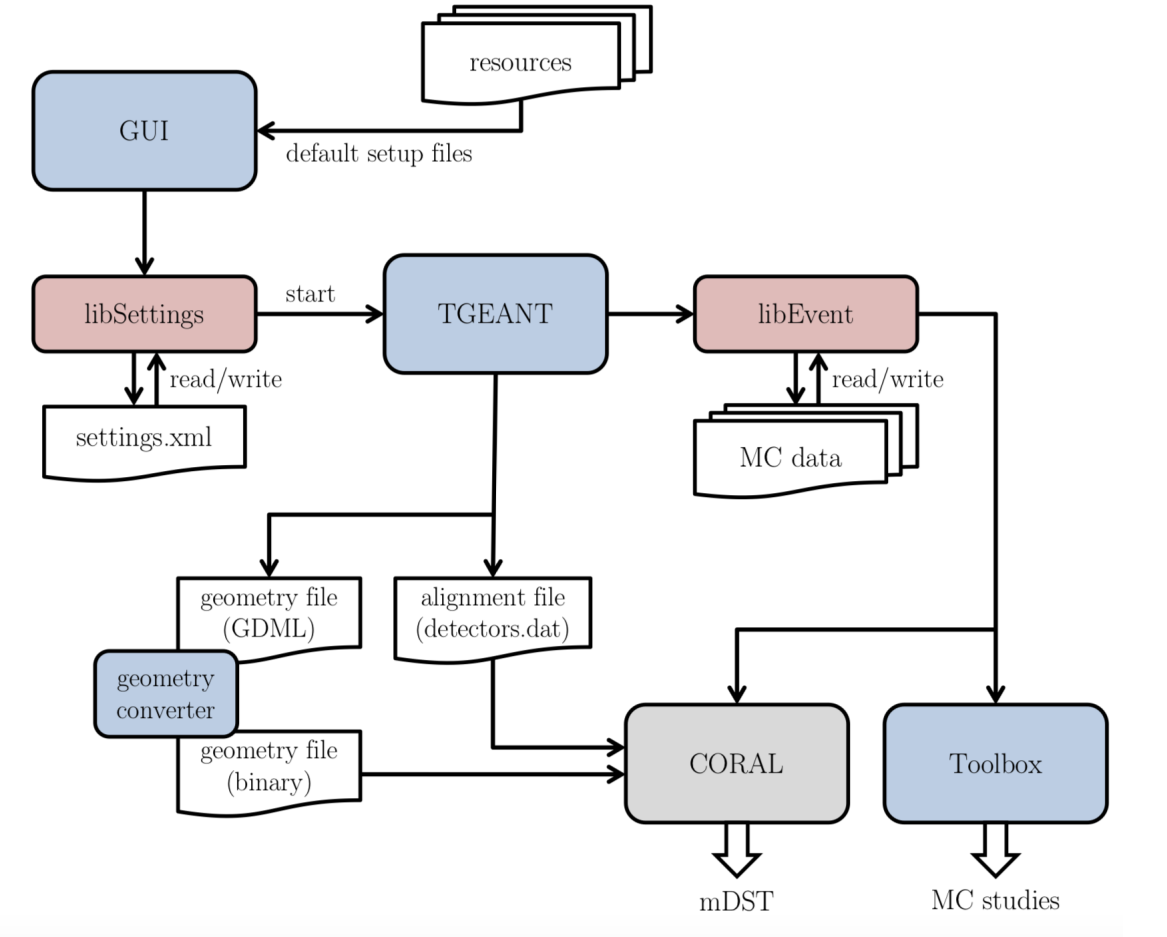


Figure 66: Flow chart of the TGEANT software package. The simulation software TGEANT is controlled by a setup file, which is easily created with the graphical user interface. The GUI can draw on different default setup files from the resources folder. The output files can either be analyzed with the Toolbox for the purpose of Monte Carlo studies or processed by CORAL in order to produce mDST files. For the latter case, TGEANT also provides the alignment and geometry files. Figure taken from [108].

Geant4 algorithm through the experimental setup. During the event loop, new particles can only be created by implemented physics processes, applied according to their cross sections. Once the event loop has ended, the output of simulated detector responses is processed. The flow chart of the event loop in TGEANT is illustrated in Fig. 67.

9.1.3 *Primary vertex generation*

The algorithm to generate a primary vertex in TGEANT is responsible for stopping the primary beam particle and for calling an event generator. The *target extrapolation* algorithm is used to trigger the event generator.

The goal of the target extrapolation method is to stop the movement of the primary beam particle at a random position inside the target volume. A flow chart of the method is presented in Fig. 68. It is a multi-purpose method to generate vertices within the target volume. A realistic vertex distribution can easily be simulated by using a beam file and a precise target alignment. The event generator is triggered exactly after the beam particle has traversed a random distance inside the target volumes.

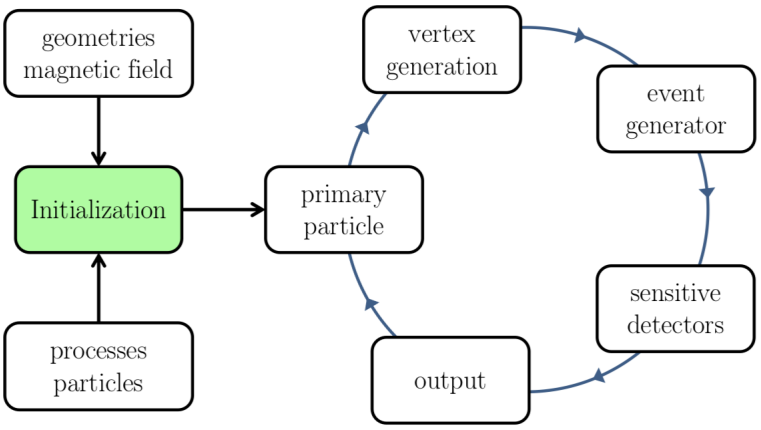


Figure 67: Flow chart of the event loop in TGEANT. Figure taken from [108].

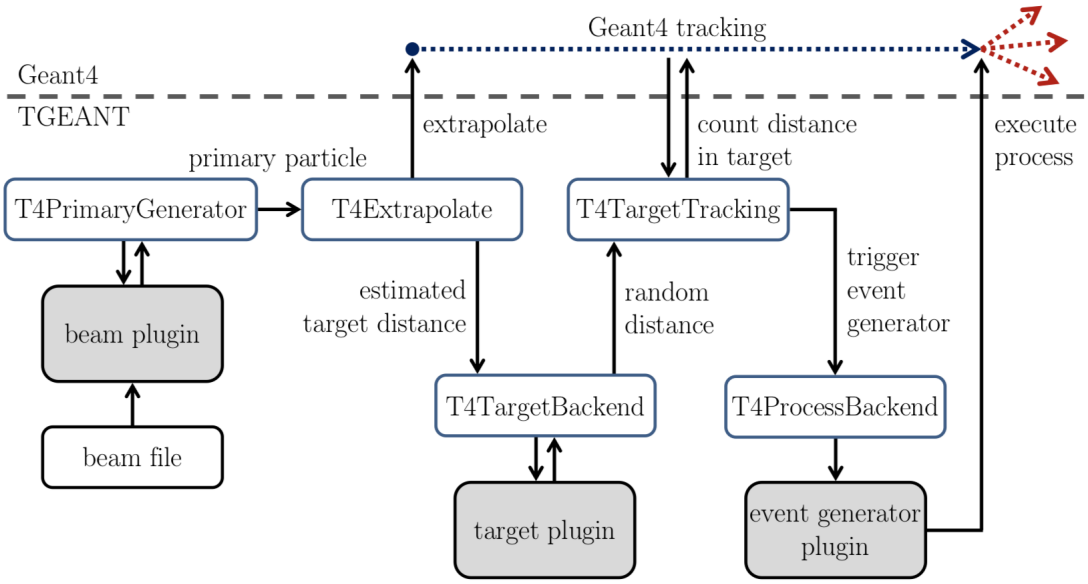


Figure 68: Flow chart of the target extrapolation method: The T4Extrapolate class extrapolates the primary particle to the desired starting z position. After traversing a random distance inside the target volumes, the beam particle is stopped by the T4TargetTracking class. At this point, the event generator is applied. This random distance is dictated by the T4TargetBackend derived target class. An estimated distance, which the beam particle is able to traverse in the target volume, is provided by the T4Extrapolate class and can be used optionally. Figure taken from [108].

9.1.4 *Event generators*

Several interfaces for different event generators are already installed in TGEANT and ready to use, see Fig. 69. The event generators are implemented as discrete Geant4 processes using the abstract T4ProcessBackend base class, which handles the interface to TGEANT. This involves the call of the event generator function and the forwarding of the 4-momentum of the incoming beam particle at the vertex position.

The simplified procedure of an event generator can be described as follows. The four-momentum of the incoming beam particle serves as input parameter and the target nucleon is at rest. During the simulated interaction, one or more outgoing particles are generated. The momentum distribution between these final state particles may be a complex procedure and needs to be randomized by the event generator according to the cross section of the in-

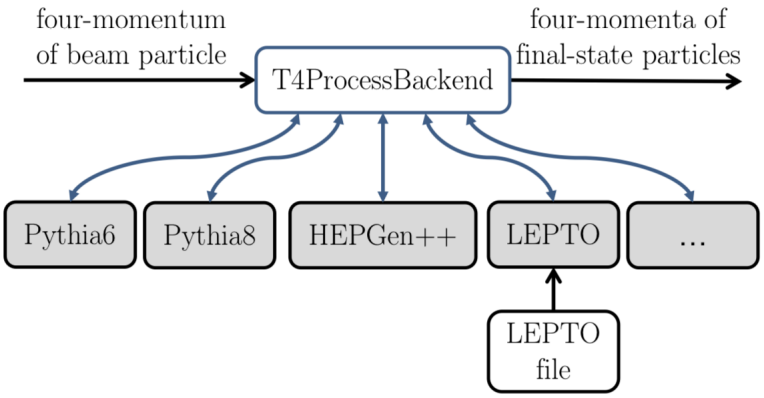


Figure 69: Inheritance diagram for the T4ProcessBackend base class. The four- momentum of the incoming beam particle is forwarded to the used event generator. The four-momenta of the final state particles are received in return. Figure taken from [108].

teraction. At the end, however, the energy and momentum conservation needs to be ensured and TGEANT has to exchange the initial state beam particle with all final state particles.

9.2 DJANGOH AS A PHYSICS GENERATOR FOR TGEANT

There are two ways to implement a generator in TGEANT:

- As an internal generator (Pythia and HEPGEN++ way): a C++ interface to the generator is needed, the conservation of $P(p, E)$ is perfect and it only needs a beamfile for the primary generator.
- As an external generator (Lepto): the beamfile is read by the standalone generator, the primary generation and process infos are stored inside a file and TGEANT has to do an extrapolation from the beamfile position (z=0 m) to the beam starting position (z=-9 m) when the primary generation of the event was done by LEPTO outside of the TGEANT simulation, occasioning $P(p, E)$ being not perfectly conserved.

The external implementation is quite complicated as three classes are needed in order to use the generator in TGEANT. First, the beamfile has to be read by the generator which then outputs a file containing the beamfile infos but in its own format. Then a first class has to be dedicated to the reading of the beamfile, recovering the information about the selected muon in the file. A second class takes care of passing information between the generator and TGEANT. The last class is the class of the generator itself (Fig. 70).

If you apply this method to DJANGOH, then you encounter several problems:

- A new file for the converted beamfile has to be created.
- DJANGOH has to be modified to allow backward propagation of the incoming muon.
- The result of fragmentation (LUJETS) has to be recovered in a file.

This results in many file accesses and consequently it is a non-efficient way to implement DJANGOH inside TGEANT.

The internal implementation recquires perhaps more work on the generator itself however this solution is much more efficient. Here, only two class are needed. One is the interface class that creates instances of DJANGOH that can be manipulated in any C++ environment. This class is a C++ interface that is handling the FORTRAN part of DJANGOH. The other class is taking care of passing the information between the interface to the generator and TGEANT (Fig. 71). This second method was used to implement DJANGOH inside TGEANT.

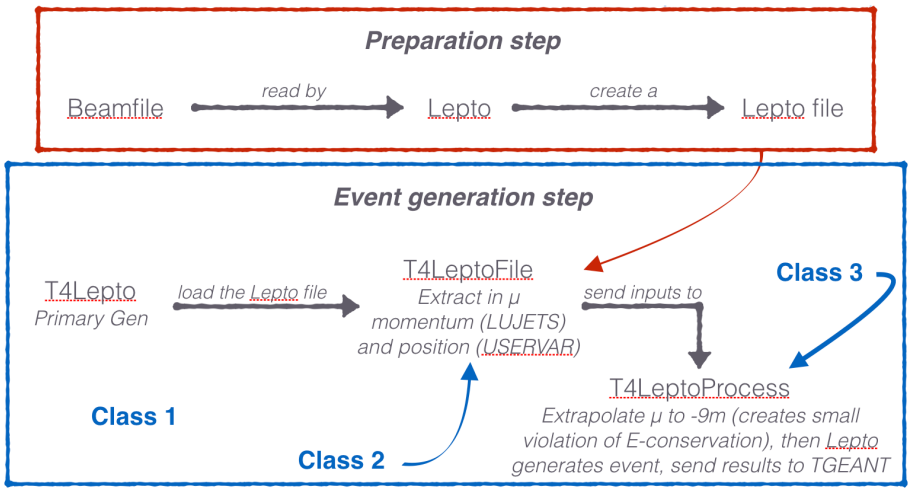


Figure 70: Diagram explaining the philosophy of the external generator implementation, taking the implementation of LEPTO as an example. First and foremost, a specific beamfile is created after the initial one. The external generator will read this pregenerated file to extract infos about the incoming particle. As these informations are at the interaction point, a backward propagation extrapolation in the target material has to be made, inducing some small violations of energy conservation. Then the generator is producing the event and sends the results to TGEANT.

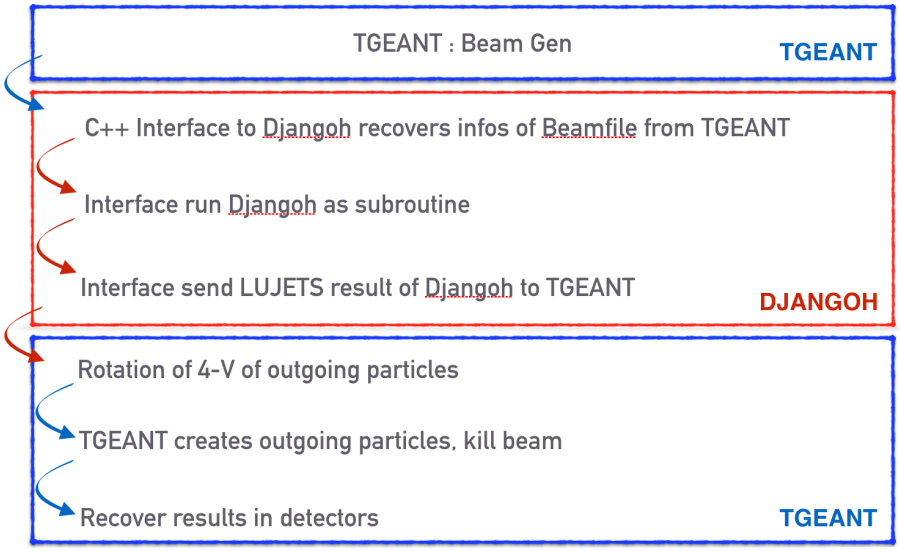


Figure 71: Diagram explaining the philosophy of the internal generator implementation. The beamfile is read by TGEANT and the C++ interface to the generator is recovering the informations of the incoming particle. This interface then runs the generator as a subroutine and sends back the results of the generation to TGEANT, which then creates the outgoing particles accordingly.

9.3 RESULTS ON ELECTROPRODUCTION FROM PHOTON CONVERSION

At this point, with DJANGOH fully integrated as an event generator for TGEANT, we can go back to our original goal: see if DJANGOH describes better the electroproduction from photon conversion at COMPASS than RADGEN did. Comparing the distribution of the absolute value of the Φ angle in the γ-nucleon reference frame from DJANGOH reconstructed MC with CORAL with real data as displayed in Fig. 72, we see an improvement compared to the factor 1.8 difference seen with RADGEN (Fig. 59). The shape is well described and the comparison is much improved. The discrepancy is only of the order of 10% at $|\Phi| \sim 0$ and

less than 5% elsewhere; showing that DJANGOH is able to reproduce with a good fidelity the electroproduction from photon conversion observed in data.

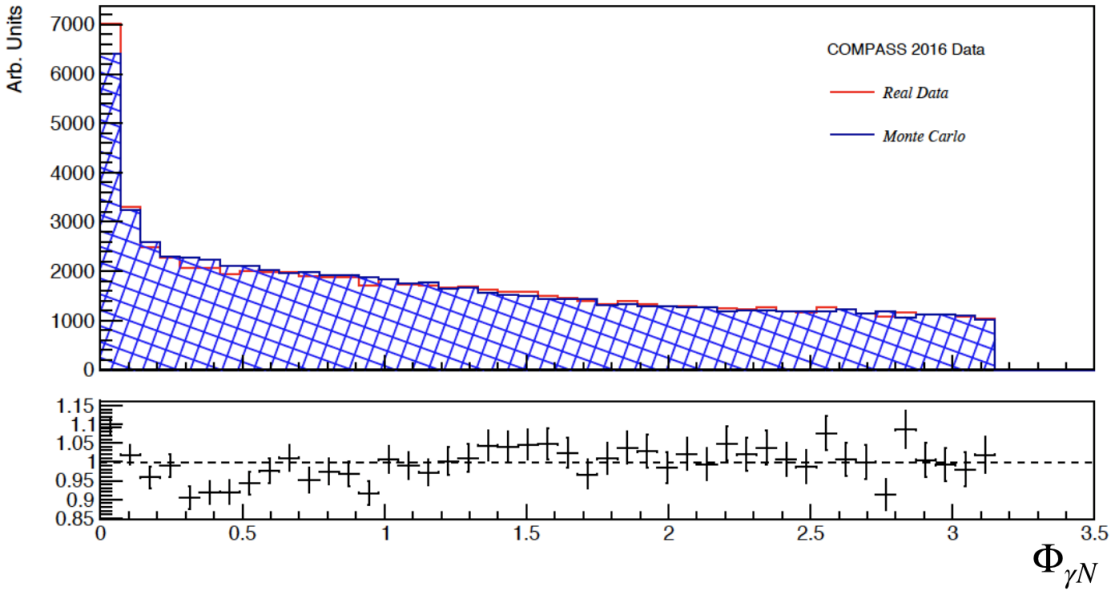


Figure 72: Electron distribution versus Φ in the γ-nucleon reference plane for $3 < p_e < 8$ GeV (region where RICH can discriminate electron in real data). Real Data are in red, Monte-Carlo with DJANGOH in blue, with the ratio data over MC on the bottom panel.

9.4 SUMMARY

TGEANT is a flexible Monte Carlo simulation, which allows a precise reproduction of the experimental setup. This flexibility allows one to use the event generator one sees fit to his needs, choosing between the internal or external implementation of this generator to the simulation. DJANGOH, after being integrated as an internal generator inside TGEANT, shows good performances in the data versus MC comparison, especially when comparing the electroproduction from photon conversion, being more accurate than previously used generator RADGEN.

Part IV

DATA ANALYSIS OF SIDIS CHARGED HADRON MULTIPLICITY (2016 DATA)

10 ANALYSIS OF 2016 RAW MULTIPLICITIES

The 2006 SIDIS COMPASS hadron multiplicity results, based on data taken with an isoscalar target (^{6}LiD) [45], do not constrain firmly the strange quark fragmentation function. With the analysis of new data taken on pure proton target (lH_2), the results will provide an independent new set of equations linking the multiplicities with the fragmentation functions but still involving the same quark fragmentation functions we are interested in. Fitting proton and deuteron data together will add constrains to the fragmentation function extraction. In order to perform this kind of study, one need a precision of 5 to 7% on the multiplicities should be obtained.

The analysis is performed on COMPASS data recorded in 2016 using a 160 GeV muon beam incident on a proton target (lH_2). Five weeks of the 2016 data are analyzed (named P07, P08, P09, P10 and P11).

10.1 METHOD OF EXTRACTION

The method of extraction of the multiplicities follows several steps. For each selection step, a number of cuts is applied on both geometrical and kinematic quantities. First, DIS events are selected and then SIDIS events (hadrons) are selected. For the DIS event selection, a study of the target radius was done in order to determine the optimal value for the target cut. After the event selection, the hadron candidates have to be identified as pions, kaons or protons and the identified hadron count has to be corrected using the RICH detection efficiency and purity by so-called unfolding. The obtained raw multiplicities are then binned. The unfolding is also done in bins, only in other variables than for the raw multiplicities. For the analysis the common event reconstruction codes from COMPASS are used. An individual analysis code is developed to study the SIDIS channel and select pion, kaon or protons production.

Input to the described analysis are samples of pre-selected events which fulfill the following requirements: an incoming muon with measured momentum, a reconstructed outgoing muon, an interaction vertex (called Best Primary Vertex in the COMPASS nomenclature) and $Q^2 > 0.8\ (\mathrm{GeV}/c)^2$. This explains why in the cut flow for the selection of DIS events, there is no effect for these two cuts.

10.2 DIS EVENT SELECTION

In the Table 10, the effect of the cuts for DIS events is summarized, showing the number of DIS events and the absolute percentage of the sample remaining. The last selection will be described later.

Table 10: List and effects of the cuts for DIS events. The percentage corresponds to the absolute percentage of the sample remaining.

Cut	# of events after cut	Absolute % of events after cut
Events with Best Primary Vertex	47.5 M	100%
Events with reconstruted scattered muon	47.5 M	100%
Events with primary interaction in the target material, target radius cut (explained in Section 10.3)	25.6 M	53.8%
Events with energy of beam muon energy in range [140 GeV, 180 GeV]	25.6 M	53.8%
Events with a well measured momentum (so-called 'BMS cut')	24.2 M	50.9%
Events with $\chi^2/\text{ndf} < 10$ for a well reconstructed beam track	24.2 M	50.9%
Events with muon beam trajectory extrapolation crossing entirely the target cell	23.4 M	49.2%
Events with $\chi^2/\text{ndf} < 10$ for a well reconstructed scattered muon track	23.4 M	49.2%
Events with Z coordinate of the first measured hit of scattered muon < 350 cm (Z_{SM1})	23.3 M	49.1%
Events with Middle, Ladder, Outer or LAST trigger	23.3 M	49.1%
Events with $Q^2 > 1$ (GeV/c)^2	18.5 M	38.9%
Events with $0.1 < y < 0.7$	8.39 M	17.7%
Events with $5 < W < 17$ GeV/c^2	8.34 M	17.6%
Events with $0.004 < x < 0.4$	8.32 M	17.5%
Events in the specified ν range (See Sec. 10.4)	-	-

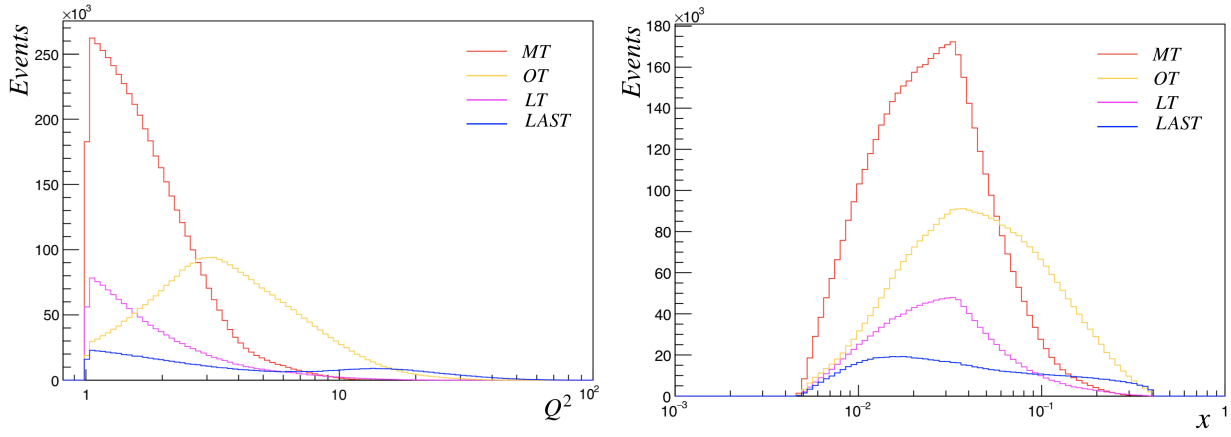
The cut $Q^2 > 1$ (GeV/c)^2 and the lower limit $W > 5$ GeV/c^2 select the deep inelastic scattering regime and the upper limit $W < 17$ GeV/c^2 to avoid low statistics regions. The lower limit $y > 0.1$ removes events with bad reconstruction of scattered muon (by ν-resolution deterioration as ν becomes really small) and the misidentification of halo muons as scattered muons. The upper limit $y < 0.7$ eliminates events where large radiative corrections have to be applied (corrections greater than 20%).

Four triggers are used in this analysis: the middle trigger (MT), the ladder trigger (LT), the outer trigger (OT) and the LAS trigger (LAST). They are all inclusive triggers, ie. only a scattered μ is required to fire the triggers. The region covered by triggers as a function of x and Q^2 is shown in Fig. 73. The middle trigger covers the low Q^2 region, the ladder trigger covers the middle Q^2 region while the outer and LAS cover the high Q^2 region. For x, the outer covers the low and high x, while middle and ladder cover the middle x region and the LAS the high x region.

In 2016, central slabs of the outer trigger were inefficient up to P07. Thus, events in P07 where the scattered μ track goes through the inefficient slabs are rejected. The situation improved from P08 onwards.

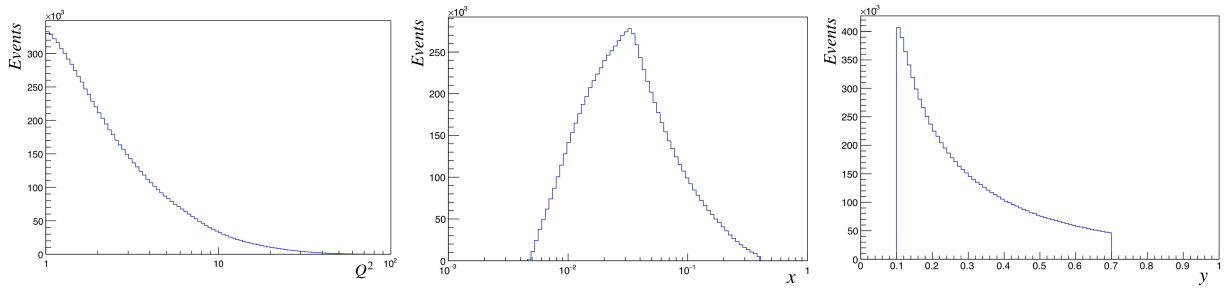
The Q^2, x and y distributions are illustrated in Fig. 74 for the DIS sample after event selections. The Q^2-x correlation is also shown as well as the x-y one. It can be noted that

most of the statistics is located in the low Q^2-x and low x-y values and Q^2 values reach up to 90 $(GeV/c)^2$.

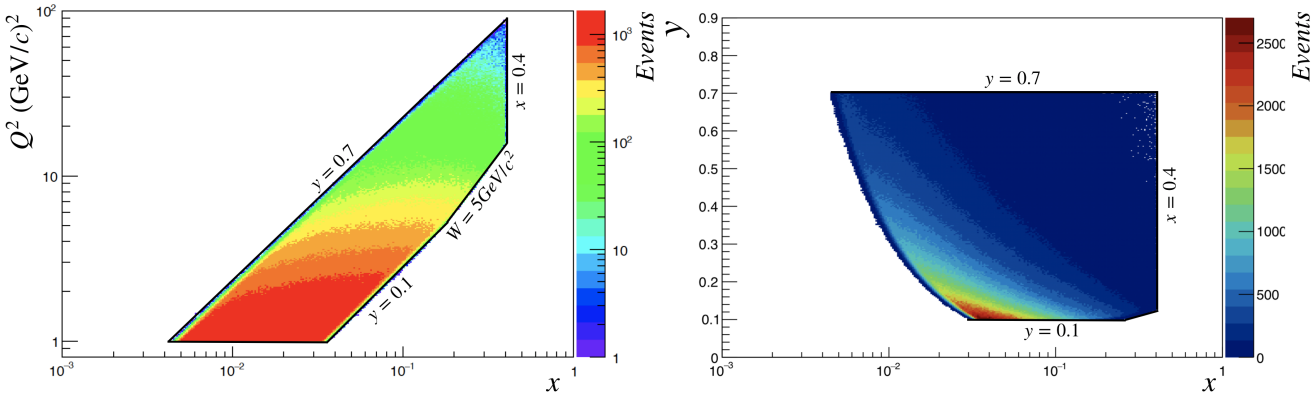


(a) (b)

Figure 73: Event distribution for the middle, ladder, outer and LAS triggers as a function of Q^2 (a) and x (b).



(a) Q^2 (b) x (c) y



(d) x-Q^2 (e) x-y

Figure 74: On the top panels, Q^2, x and y distributions. On the bottom, x-Q^2 and x-y correlations. All distributions are for the final DIS sample.

10.3 TARGET CUT EVALUATION

The 2.5 m long lH_2 target in 2016 is not perfectly straight (slight 'banana shaped', Fig. 75) due to the mylar tube. In the Monte-Carlo simulation we use a 2.5 m long cylinder tilted

with respect to the beam by average angles. This angle corresponds to the angle between the upstream and downstream ends of the target. Of course due to its 'banana shape' in reality, the description of the target in Monte-Carlo is not reaching 100% fidelity.

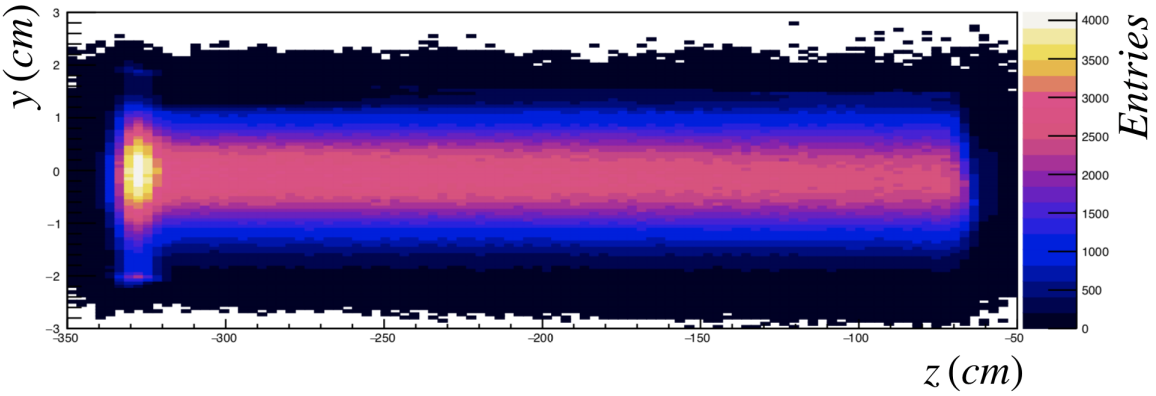


Figure 75: Vertex distribution in the target in the $z-y$ plane. On the upstream part of the target, one can see at y above 1 cm a deficit of vertices (lack of symmetry with respect to $y \sim 0$ cm in the color gradient) where the bubbles are.

After the radial cut on the real data target (1.9 cm radius) to get rid of the mylar and a cut along y (y = 1.2 cm) to get rid of the bubbles in the upstream part of the target due to its tilt, the volume of the real data target not intersecting the Monte-Carlo target is of 0.5% (Fig. 76). This brings a systematic error on the multiplicities that we wanted to avoid. To this end, we devised three different solutions:

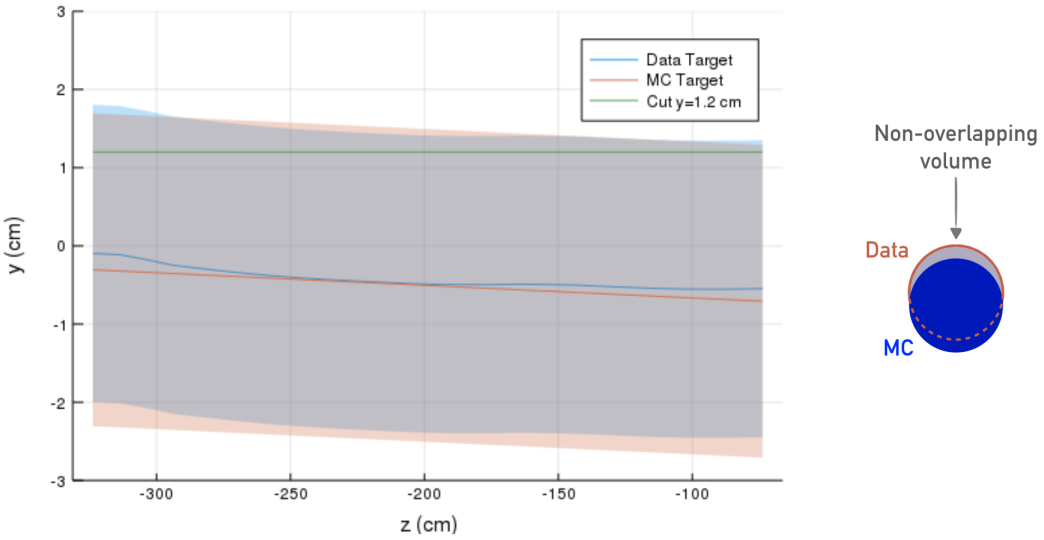


Figure 76: Left is a (y,z) view of the real data (blue) and the Monte-Carlo target (red), z being the direction of propagation of the beam. The actual cut used in the analysis corresponds to the intersection of both real data target (red) and Monte-Carlo target (blue) volumes. The green line shows the y = 1.2 cm cut. Right is a sketch showing the approximate overlap between the two volumes.

1. Cut more severely on the radius of the real data target (1.7 cm radial cut) to reduce the non-overlapping volume to zero.

2. Do a simultaneous cut on the real data target and on the Monte-Carlo target to only keep the overlap between the two volumes.

3. Do a better description of the target in the Monte-Carlo by using several cylinder volumes and fuse them together to increase the fidelity of the Monte-Carlo target description with respect to the real data one.

The last solution may be used in the future to maximise the efficiency of the analysis from a statistics point of view, but it is heavy and time consuming for a marginal gain. This gain is at most 0.5% with respect to the method chosen, while one can argue that as the events gained are on the edge of the target, they might be cut out by checking for the muon beam trajectory crossing entirely the target cell, thus the gain could even be less than this. The first two solutions were in competition and had the same spirit: cut more in the data target to avoid any systematic bias between real data target and Monte-Carlo target. We chose the second one, a simultaneous cut on both targets, as it was the one that was discarding less target volume, thus maximising statistics. Fig. 76 displays the volume that survives such cut. The number of events lost by cutting on both target volumes instead of only the data target volume is negligible.

10.4 HADRON SELECTION

In the Table 10, the effect of the cuts for hadrons is summarized, showing the number of hadrons and the absolute percentage of the sample remaining:

Table 11: List and effects of the cuts for hadrons. The percentage corresponds to the absolute percentage of the sample remaining.

Cut	# of events after cut	Absolute % of events after cut
Particle is not a scattered muon	37.0 M	100%
Maximum radiation length cumulated along all the trajectory < 15 radiation lengths	28.3 M	76.6%
$\chi^2/\mathrm{ndf} < 10$ for the hadron track	27.9 M	75.4%
Z coordinate of the first measured hit < 350 cm	27.9 M	75.3%
Z coordinate of the last measured hit > 350 cm	19.1 M	51.5%
$0.01 < \theta_{RICH} < 0.12$ (at RICH entrance)	12.6 M	34.1%
$x^2_{RICH} + y^2_{RICH} > 25$ cm^2 (rejection of RICH pipe)	12.5 M	33.7%
$12 < p_h < 40$ GeV/c	3.37 M	9.11%
$0.2 < z < 0.85$	2.67 M	7.21%

In addition, a cut on the kinematic variable ν was performed to reject actual DIS events, where the actual z bin cannot be reached for any hadron with momenta between 12 and 40 GeV/c with ν of the DIS event. The criteria is defined by:

$$\nu_{max} = \frac{\sqrt{(p^2_{max} + m^2_h)}}{z_{max}}, \tag{92}$$

$$\nu_{min} = \frac{\sqrt{(p^2_{min} + m^2_h)}}{z_{min}}, \tag{93}$$

where p_{max} (p_{min}) is the hadron momentum limit of 40 GeV/c (12 GeV/c), z_{max} (z_{min}) is the upper (lower) value of the z-bin and m_h is the mass of the considered hadron. As the

cut depends on the mass of the considered hadron, the number of DIS events is different for each identified hadron.

10.5 DOWNSTREAM TARGET VERTEX DISTRIBUTION

When looking at vertex distribution for hadrons and comparing data with Monte-Carlo a deficit of about 6% of vertices is observed at the downstream part of the target (between -100 and -70 cm as shown in Fig. 78). The same observation is made in Monte-Carlo, however it is more pronounced in data than in Monte-Carlo. After investigation, I discovered that in both data and Monte-Carlo there were hadrons that have their track not attached to the best primary vertex passing in a 2 mm-radius circle around the best primary vertex (Fig. 77). This contribution is only found in the downstream part of the target, while in the upstream part it is non-existant.

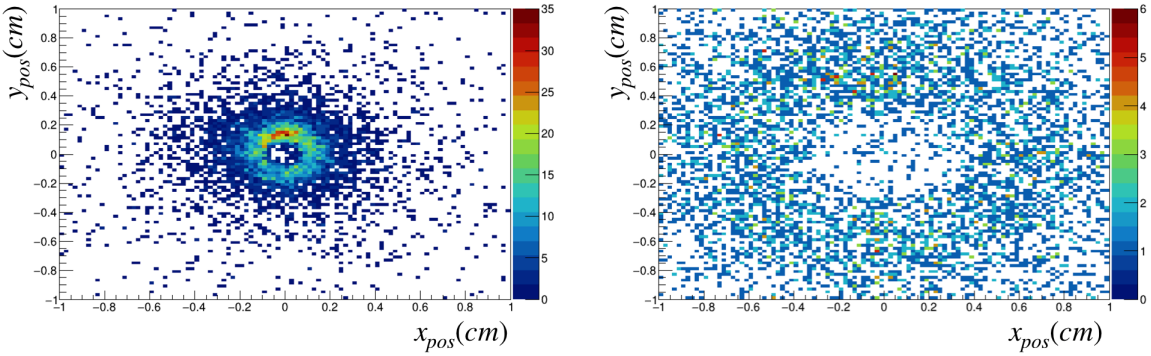


Figure 77: The distance to the best primary vertex of the extrapolated position of the unattached hadrons to the vertex position perpendicular to the beam direction. On the left, the plot corresponds to the downstream part of the target and on the right, to the upstream part.

This problem is introduced by the reconstruction software and could not be solved in time for this thesis. However we found a rescue procedure to reattach these hadrons to the best primary vertex. All the hadrons in a circle of radius 2 mm around the best primary vertex are used in the further analysis as if they were attached to it. The same quality cuts were applied to these hadrons as for the attached ones. With this procedure we were able to recover for the loss of hadron in the downstream part of the target as seen in Fig. 78.

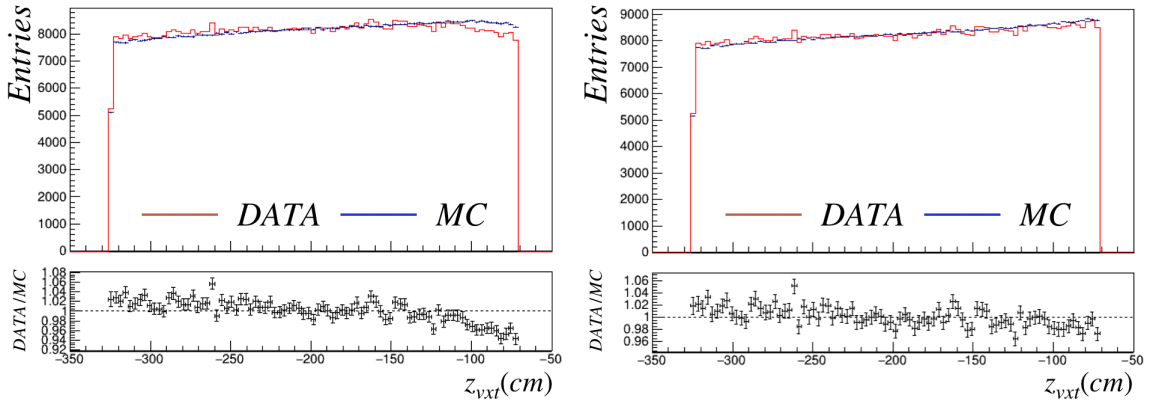


Figure 78: Comparison of the vertex distribution along the target of data and MC hadrons left) before and right) after the rescue procedure.

10.6 PARTICLE IDENTIFICATION WITH RICH DETECTOR

The π, K and p identification (PID) is performed by the RICH detector.

The method used for the RICH particle identification is described in Chapter 5. The idea is the following: when a particle is detected, six likelihood (LH) functions are calculated (π, K, p, e, μ and the background) and are then compared to perform particle identification. The evaluation is done separately for pions, kaons and protons. The largest likelihood is then selected. The method is improved by looking further to $LH(2^{nd})$, which is the second largest likelihood of the four compared likelihood values (π, K, p and the background). The electron and muon likelihoods are not considered in the assignment of $LH(2^{nd})$ as in the chosen momentum range (12 to 40 GeV/c) the RICH detector can not be used to efficiently distinguish electrons from π.

All π, K and p probabilities are used for the unfolding. The likelihood cuts of Table 8 of Chapter 5 are applied with some additional conditions as the present sample is not a clean one:

1. Pion selection
   - $LH(\pi) > 0$
   - $LH(\pi) > LH(K)$, $LH(p)$ and $LH(bg)$.
2. Kaon selection
   - $LH(K) > 0$
   - $LH(K) > LH(\pi)$, $LH(p)$ and $LH(bg)$.
3. Proton selection Three cases are considered depending on the momentum p_h of the particle and are distinguished by the kaon threshold ($\simeq$ 8.9 GeV/c) and proton threshold ($\simeq$ 17.95 GeV/c):
   (a) Kaon threshold $< p_h \leqslant$ proton threshold - 5 GeV/c
      - All LH = 0
   (b) $p_h >$ proton threshold + 5 GeV/c
      - $LH(p) > 0$
      - $LH(p) > LH(\pi)$, $LH(K)$ and $LH(bg)$.
   (c) Proton threshold - 5 GeV/c $< p_h <$ proton threshold + 5 GeV/c
      - Using (a) and (b) simultaneously.

10.7 RICH UNFOLDING BASED ON EFFICIENCY MATRICES

With the unfolding procedure the hadron identification is corrected on a hadron by hadron basis for the limited RICH efficiency and misidentification. In order to perform this correction, the RICH actual performance was evaluated from real data as described in Chapter 5. The result of this evaluation is presented through RICH performance matrices, M_{RICH}, binned in momentum and angle:

- p_h {12,13,15,17,19,22,25,27,30,35,40} GeV/c
- θ_h {0.01,0.04,0.12} rad

The 3-by-3 matrices M_{RICH} give a relation between the vector of counts for true hadron T_h and the vector for identified hadron I_h:

$$\begin{bmatrix} I_\pi \\ I_K \\ I_p \end{bmatrix} = \begin{bmatrix} \epsilon(\pi \to \pi) & \epsilon(K \to \pi) & \epsilon(p \to \pi) \\ \epsilon(\pi \to K) & \epsilon(K \to K) & \epsilon(p \to K) \\ \epsilon(\pi \to p) & \epsilon(K \to p) & \epsilon(p \to p) \end{bmatrix} \begin{bmatrix} T_\pi \\ T_K \\ T_p \end{bmatrix}. \tag{94}$$

The coefficients of the M_{RICH}, $\epsilon(t \to i)$, are the probabilities that a true hadron of type t is identified as a hadron of type i.

The number of true hadrons are obtained by inverting the performance matrices (Eq. 95):

$$\overrightarrow{T_h} = M^{-1}_{RICH} \overrightarrow{I_h}. \tag{95}$$

The numbers of identified and unfolded hadrons are displayed in Table. 12.

Table 12: Number of identified pions, kaons, and protons for the five analyzed periods before and after unfolding.

	π^+	π^-	K^+	K^-	p	$\bar{p}$
Identified	953970	789480	253045	153440	131066	60705
Unfolded	976213	814685	255132	150775	124221	52014

10.8 KINEMATIC BINNING

The multiplicities are evaluated in bins of the Bjorken variable x, the muon energy fraction carried by the virtual photon y and the virtual photon energy fraction carried by final state hadron z. They are calculated with the following formula:

$$\frac{dM^h(x,y,z)}{dz} = \frac{1}{N^{DIS}_{Events}(x,y)} \frac{dN^{DIS}_h(x,y,z)}{dz}, \tag{96}$$

where N^{DIS}_{Events} is the number of DIS events and N^{DIS}_h is the number of hadrons after RICH unfolding. As in practise, the multiplicities are measured in bins of x (9 bins), y (5 bins) and z (12 bins), the raw multiplicities (multiplicities without corrections) can be expressed as:

$$M^h_{raw}(x,y,z) = \frac{N^{DIS}_h(x,y,z)/\delta z}{N^{DIS}_{Events}(x,y)}, \tag{97}$$

where δz is the width of the z bin. For the multiplicity extraction, the binning in x, y and z is the following:

Table 13: Bin limits for the multidimensional binning

Variable	Binning
x	$\{0.004, 0.01, 0.02, 0.03, 0.04, 0.06, 0.1, 0.14, 0.18, 0.4\}$
y	$\{0.1, 0.15, 0.2, 0.3, 0.5, 0.7\}$
z	$\{0.2, 0.25, 0.3, 0.35, 0.4, 0.45, 0.5, 0.55, 0.6, 0.65, 0.7, 0.75, 0.85\}$

10.9 STATISTICAL ERROR PROPAGATION

The statistical error propagation used in the multiplicity calculation will be explained in the following. All further calculations are done in bins of (x,y,z). All the DIS events enter with the same weight of 1 in the error calculation:

$$E^2_{DIS} = N_{DIS}. \tag{98}$$

Same for unidentified hadrons:

$$E^2_{Had} = N_{Had}. \tag{99}$$

For identified hadrons, the squared error includes the RICH statistical error:

$$E^2_{Had} = \sum_{i=1}^{N_{Had}} E^2_{RICH,i}, \tag{100}$$

where

$$E^2_{RICH}[0 < h < 3] = cov(M^{-1}_{hr}, M^{-1}_{hr}) + (M^{-1}_{hr})^2, \tag{101}$$

with M_{hr} the element of the RICH unfolding matrix for hadron h from RICH identified hadron r and

$$cov(M^{-1}_{hr}, M^{-1}_{hr}) = \sum_{0<i,j,k,l<3} M^{-1}_{hi} M^{-1}_{jr} M^{-1}_{hk} M^{-1}_{lr} cov(M_{ij}, M_{kl}). \tag{102}$$

For the raw multiplicities the error takes into account the correlation between hadrons and DIS events:

$$E^2_{raw} = \left[\frac{E^2_{Had}}{N^2_{DIS}} - \left(\frac{N^2_{Had}}{N^2_{DIS}} \right)^2 E^2_{DIS} \right] / z^2_{width}. \tag{103}$$

10.10 RESULTS FOR RAW MULTIPLICITIES ($h^\pm$, $\pi^\pm$, $K^\pm$ AND $p/\bar{p}$)

The raw multiplicity results shown in this section are without any correction except for the RICH unfolding correction for identified hadrons. The unidentified hadron multiplicities are displayed as a function of z in bins of x and staggered vertically with y in Figs. 79 to 80 (see Appendix B for identified hadrons). The charged hadron multiplicities strongly depend on z as expected with a small dependence with x also.

The raw multiplicities for positive and negative hadrons, pions and protons are very similar but with a small asymmetry at high x, explained by the fact that at high x in the valence region, the u quark is dominant in the target. For kaons, $M^{K^+}_{raw} > M^{K^-}_{raw}$ as K^- ($\bar{u}s$) can only be produced by sea quarks or subleading particles.

In total each charged hadron multiplicities yield more than 300 data points.

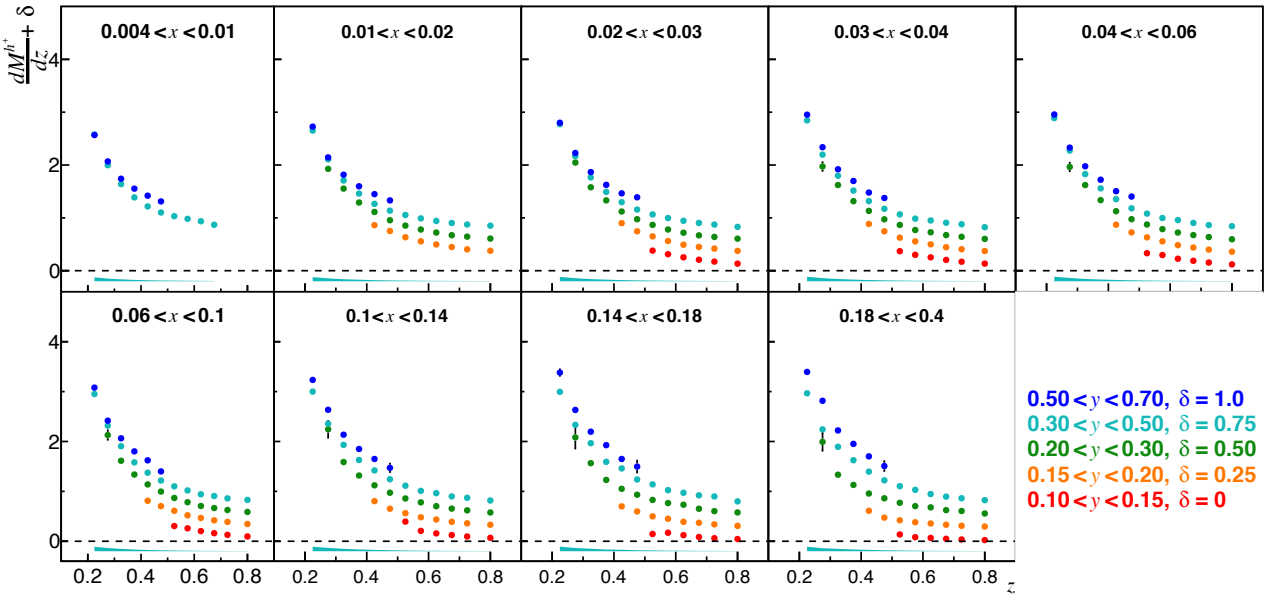


Figure 79: Unidentified positive hadron raw multiplicities as a function of z in bins of x and scattered vertically with y. Statistical error is shown but is small in most of the bins.

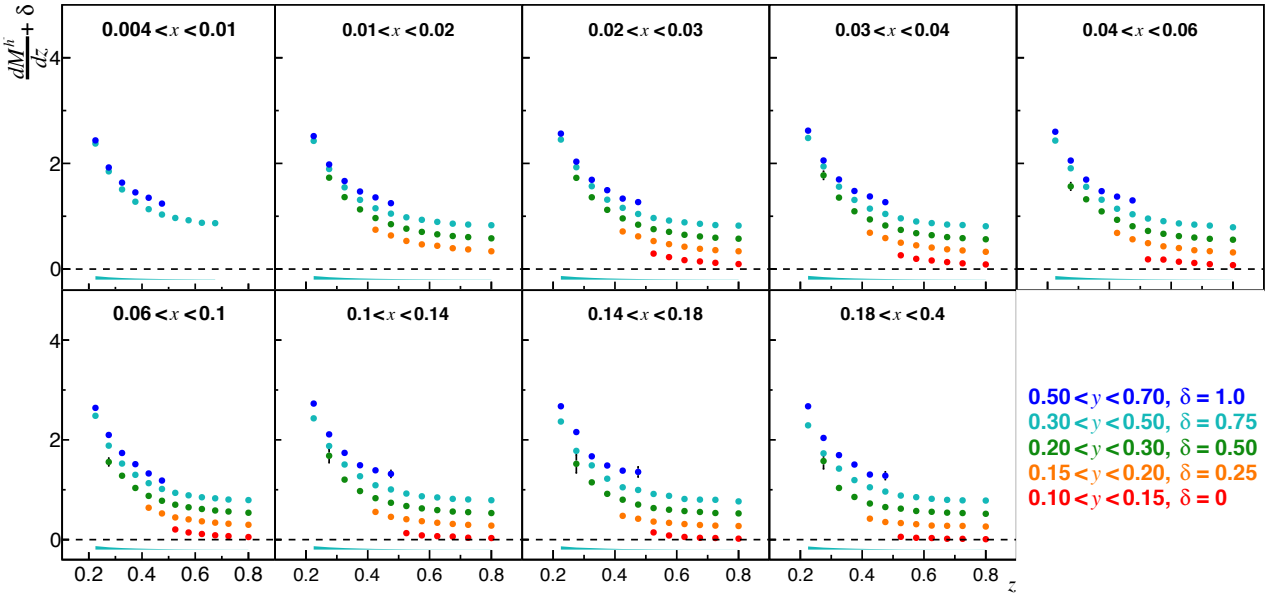


Figure 80: Same as Fig. 79 but for unidentified negative hadrons.

10.11 SUMMARY

From deep inelastic muon scattering on a pure proton target (lH_2) from the 2016 COMPASS data, raw multiplicities for unidentified hadrons, pions, kaons and protons were extracted in a three dimensional (x,y,z) binning. The data cover a wide kinematic domain defined by $Q^2 > 1$ $(GeV/c)^2$, $y \in [0.1, 0.7]$, $x \in [0.004, 0.4]$, $W \in [5, 17]$ GeV and $z \in [0.2, 0.85]$. The hadron momentum is taken in the range $[12, 40]$ GeV/c.

The raw multiplicities for identified hadrons are corrected by the RICH identification and misidentification efficiencies. The dominant uncertainty, before the application of correction factors, is the statistical error but for most of the bins $\sigma/M^h_{raw} < 1\%$.

11 CORRECTION FACTORS TO THE MULTIPLICITIES

The *raw* charged multiplicities presented in the previous chapter have to be corrected. The corrections discussed here are the acceptance correction, i.e. the correction due to the geometrical limitation of the spectrometer and the data reconstruction efficiency, as well as the diffractive vector meson correction and the radiative correction. The electron contamination correction is included in the acceptance correction.

11.1 DETERMINATION OF THE SPECTROMETER ACCEPTANCE

11.1.1 *Monte Carlo sample from DJANGOH*

The COMPASS detector does not cover the full phase-space, thus the measured multiplicities have to be corrected for the finite detector acceptance, which is of the order of 70%. The correction is done using a Monte-Carlo dataset containing about 400 million events generated in the kinematic region $Q^2 > 0.8$ $(GeV/c)^2$, $x \in [10^{-4}, 0.9]$, $y \in [0.01, 0.95]$, thanks to the Blue Waters facility computing power [109, 110]. The kinematic region is larger than the one of interest in the analysis. This is done in order to be able to take into account bin-to-bin migration of events. Two different MC samples with a different beam charge are providing 200 millions events each to take into account any asymmetry in the spectrometer related to the beam charge. As there is a difference of two slabs in the OT in the experimental setup between the periods P07 and P08 onwards, two different acceptance corrections have been determined. The two different MC samples were both reconstructed twice. One reconstruction is done without the two inefficient OT slabs and will give P07 acceptance while the other is done with all the slabs and will give $P08^+$ (P08 and following periods) acceptances. Eventually from two samples, four different acceptances are computed (P07-μ^+, P07-μ^-, $P08^+$-μ^+ and $P08^+$-μ^+).

The events are generated with the DJANGOH generator with a parametrization of the parton distribution functions (MSTW08 [111]). In addition, the use of JETSET inside DJANGOH allows the hadronization of quarks q to final-state hadrons h according to the Lund model. The COMPASS high p_T tuning was used for the JETSET parameters [112]. The events of DJANGOH are then propagated into the experimental setup simulated by TGEANT. The beam and beam reconstruction is not simulated, but reconstructed beam tracks from data are being used. The output of this chain is referred to as *generated* sample. These events are then reconstructed with the same CORAL code as used to reconstruct real data. This new sample is called *reconstructed* sample. The same DIS event and unidentified hadron selection that are used on real data (except cuts related to beam and beam reconstruction) are applied to the MC data sample for reconstructed MC events and particles.

The acceptance is calculated as the ratio of reconstructed over generated particles. In both cases, the particle ID is taken from the MC truth. The following selection is made on the generated particles:

1. Energy of the beam muon in range [140,180] GeV
2. Z coordinate of the generated event vertex (z_{vtx}) within the target region $\in$ [-325 cm, -71 cm]

3. Primary interaction in the target material (PHAST routine PaAlgo:InTarget() for both data and MC (Section 10.3) target positions to have a complete overlap of coverage)
4. Beam track crossing the entire target (PHAST routine PaAlgo:CrossCells())
5. $Q^2 > 1$ (GeV/c)^2
6. $0.1 < y < 0.7$
7. $5\ \text{GeV}/c^2 < W < 17\ \text{GeV}/c^2$
8. $0.004 < x < 0.4$
9. ν range used in data
10. $0.2 < z < 0.85$

The inclusive kinematic variables Q^2, x and y and semi-inclusive variables z and p_h are shown in Figs. 81 and 82 for real data (red) and reconstructed MC data (blue). The ratio between real data and reconstructed MC data is shown at the bottom of each panel. A relative good agreement is reached, except in the low statistics high x and high Q^2 regions and for θ_h. It was shown in previous work that the COMPASS acceptance nearly factorizes in $A_{muon} \cdot A_{hadron}$. Thus for multiplicities A_{muon} drops out to a large extent and the difference of data and MC in edges of the muon kinematics does not introduce a bias. The description of θ_h should be improved in future analyses.

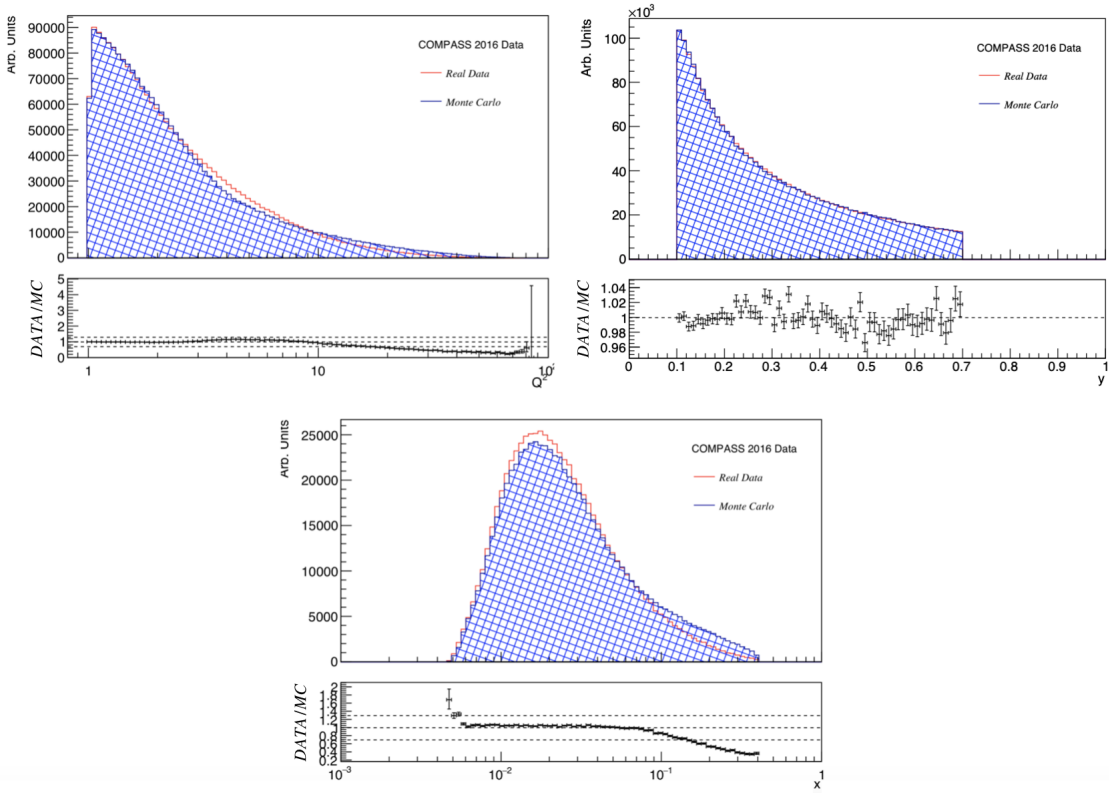


Figure 81: Kinematical variables for DIS events (Q^2, y and x) for Data (red) and Monte-Carlo (blue), as well as the ratio Data/Monte-Carlo.

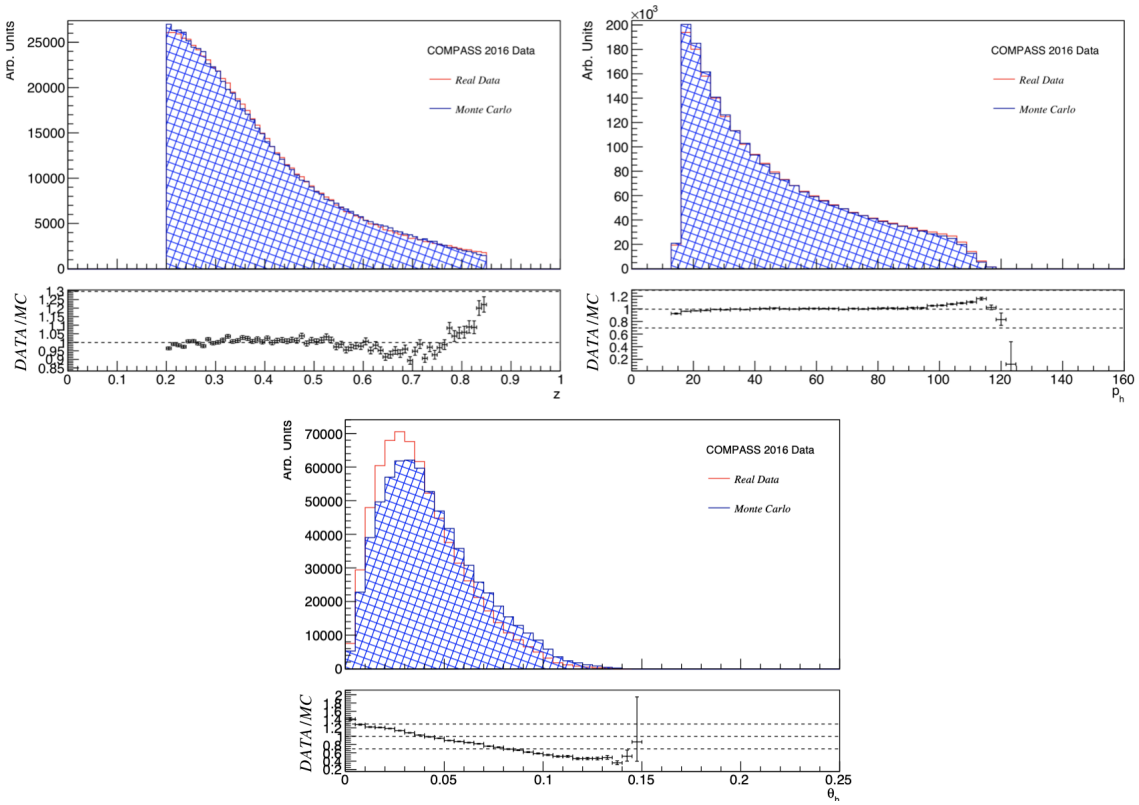


Figure 82: Kinematical variables for charged hadrons (z, p_h and θ_h) for Data (red) and Monte-Carlo (blue), as well as the ratio Data/Monte-Carlo.

11.1.2 *Acceptance calculation*

In the following, r and g refers to 'reconstructed' and 'generated' quantities. The lepton variables are used for both reconstructed and generated samples.

The acceptance is determined as the ratio of reconstructed multiplicities M_r^h over the generated multiplicities M_g^h and is binned as for the data in x, y and z:

$$A^h(x,y,z) = \frac{M_r^h(x_r,y_r,z_r)}{M_g^h(x_g,y_g,z_g)} = \frac{N_r^h(x_r,y_r,z_r)/N_r^{DIS}(x_r,y_r,z_r)}{N_g^h(x_g,y_g,z_g)/N_g^{DIS}(x_g,y_g,z_g)} \tag{104}$$

where x_g, y_g and z_g are the generated kinematic values and x_r, y_r and z_r are the reconstructed kinematic values. The acceptance being calculated in this fashion, the kinematic bin migration due to reconstruction limitations is accounted for. A more rigorous bin migration correction would involve an unfolding procedure but is not done in this analysis.

For this method, the error estimation is difficult to rigorously calculate as the numbers of evaluated hadrons and DIS events, as in both the reconstructed and generated case, they are not independent. In the following, all quantities are binned in (x,y,z). An estimate is made by assuming on the basis that there are much more DIS events than hadrons. Due to the z kinematic bin migration effects, there exist particles in N_r, which does not belong to N_g. Decomposing N_r into two contributions namely N_{r^0}, which are contained in N_g and $N_{r'}$, which are not, the final acceptance error yields:

$$\begin{aligned} E_{acc}^2 = &\left(\frac{G_D}{R_D+R'_D}\right)^2\left[\frac{(R_h+A)(G_h-R_h+1)}{(G_h+2)^2(G_h+3)}+\frac{R'_h}{G_h^2}+\frac{R'^2_h}{G_h^3}\right] \\ &+\left(\frac{G_D}{R_D+R'_D}\right)^4\left(\frac{R_h+R'_h}{G_h}\right)^2\left[\frac{(R_D+1)(G_D-R_D+1)}{(G_D+2)^2(G_D+3)}+\frac{R'_D}{G_D^2}+\frac{R'^2_D}{G_D^3}\right], \end{aligned} \tag{105}$$

where G_h (G_D) are the generated hadrons (DIS events) in a given x, y, z bin, R_h (R_D) the reconstructed hadrons (DIS events) and R'_h (R'_D) all other particles (events) that are reconstructed as hadrons (DIS events) in a given x, y, z bin.

The acceptance correction factors $A^h(x,y,z)$ for unidentified hadrons, pions, kaons and protons are shown in Figs. 83 to 86 for the P07 sample. The acceptance results are displayed versus z and each pad is a (x,y) bin, left to right for increasing x and top to bottom for increasing y. Note that the acceptance is very similar for positive and negative particles and for μ^+ and μ^- beams.

In Fig. 87 the acceptance correction factors are compared for the P07 and P09 samples. The difference visible in the highest x bins is expected and is due to the change in the trigger acceptance between P07 and P08.

The acceptance correction is then applied to the raw multiplcities

$$M^h(x,y,z) = \frac{M^h_{raw}(x,y,z)}{A^h(x,y,z)}. \tag{106}$$

The acceptance correction presented here was calculated with preliminary efficiencies, which should be improved in the future thus improving the correction. In addition, going back to the data over MC comparison of θ_h, two leads could help to improve the description. One would be to have a more detailed look at the model used inside the MC, an other would be to do acceptance in bins of θ_h, which would require to generate more MC events.

11.2 DIFFRACTIVE VECTOR MESON CORRECTION

It is usually assumed that hadrons produced in SIDIS originate from lepton-parton scattering. But the scattering of a lepton off a nucleon can also result in the diffractive production of vector mesons. These particles decay into lighter mesons that cannot be distinguished from the ones resulting from the hadronization of a quark originating from the target nucleon. The fragmentation functions extracted from multiplicities describe single quark fragmentation only. Thus diffractive processes have to be subtracted from multiplicities.

In order to study these diffractive processes, the HEPGEN generator is used. HEPGEN [113] is a generator of Monte Carlo events, which is dedicated to studies of hard exclusive single photon or meson production processes at the COMPASS experiment kinematic domain. In addition, generation of single photon or meson production accompanied by the diffractive dissociation of the nucleon is allowed. Five processes are implemented in the generator: single photon production (DVCS+BH), exclusive π^0 production, exclusive ρ^0 production, exclusive ρ^+ production and exclusive Φ production. HEPGEN has been ported to C++ as HEPGEN++.

For pions and kaons, the dominant vector meson contribution comes from the diffractive production of ρ^0 and Φ (Fig. 88), respectively:

$$\begin{aligned} \gamma * p &\to \rho^0 p \to p\pi^+\pi^- \\ \gamma * p &\to \Phi p \to pK^+K^- \end{aligned} \tag{107}$$

These processes are mainly exclusive but in 20% of cases a diffractive dissociation of the target nucleon occurs. Other channels (excited ρ, ω, etc.) are expected to contribute much less and are not taken into account. As pions and kaons stemming from diffractive vector meson decay cannot be separated from the ones resulting from SIDIS, the evaluation of their contribution to the multiplicities is based on a Monte Carlo study. Three Monte Carlo samples are produced based on different generators (SIDIS using DJANGOH, diffractive Φ and ρ^0 using HEPGEN++). All this samples are reconstructed with the same event reconstruction

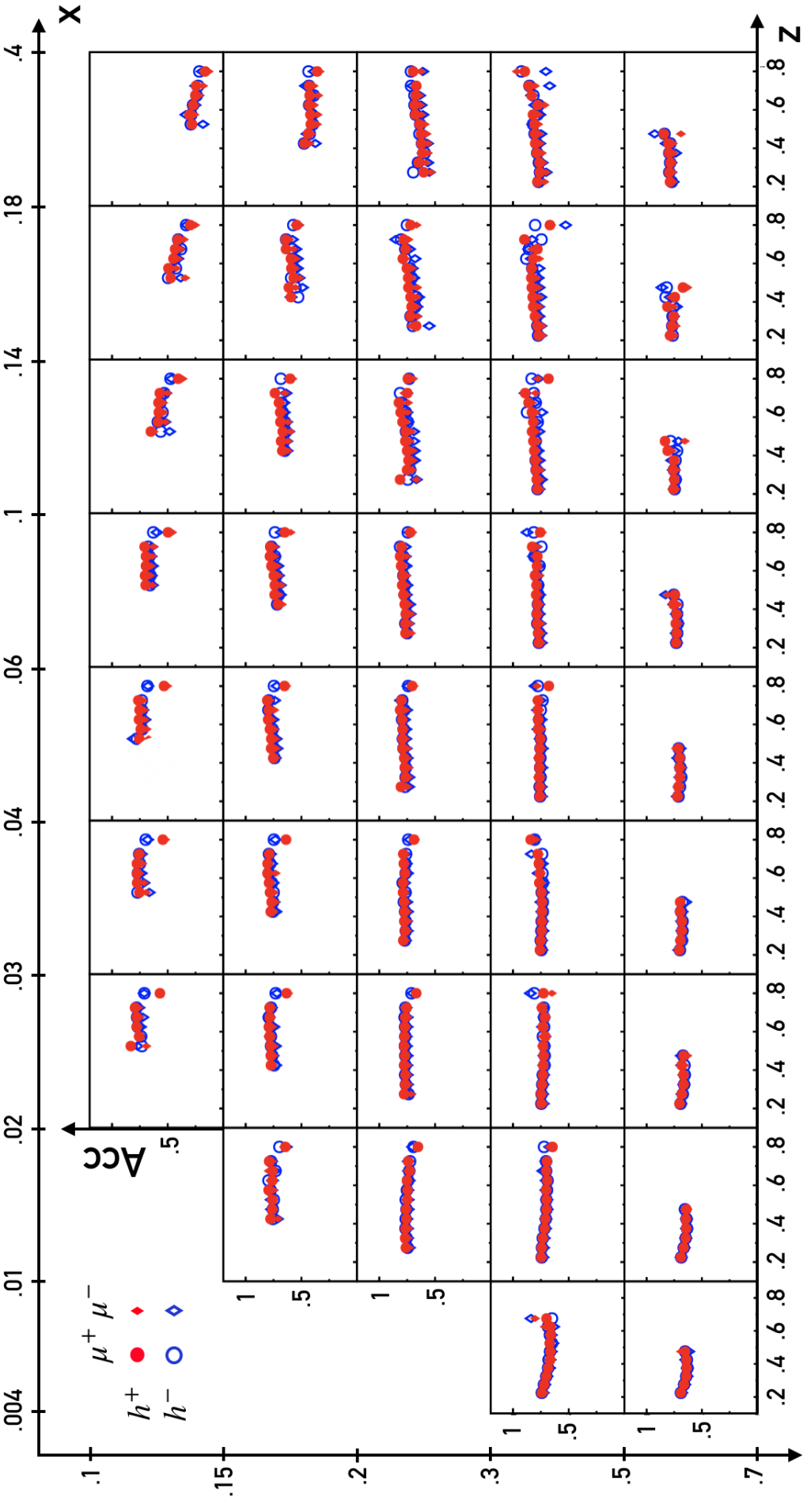


Figure 83: Charged hadron acceptance in x, y and z bins. The red full markers correspond to positive hadrons, blue open markers to negative hadrons, circles markers correspond to hadrons obtained with μ^+ beam and diamonds markers to hadrons obtained with μ^- beam.

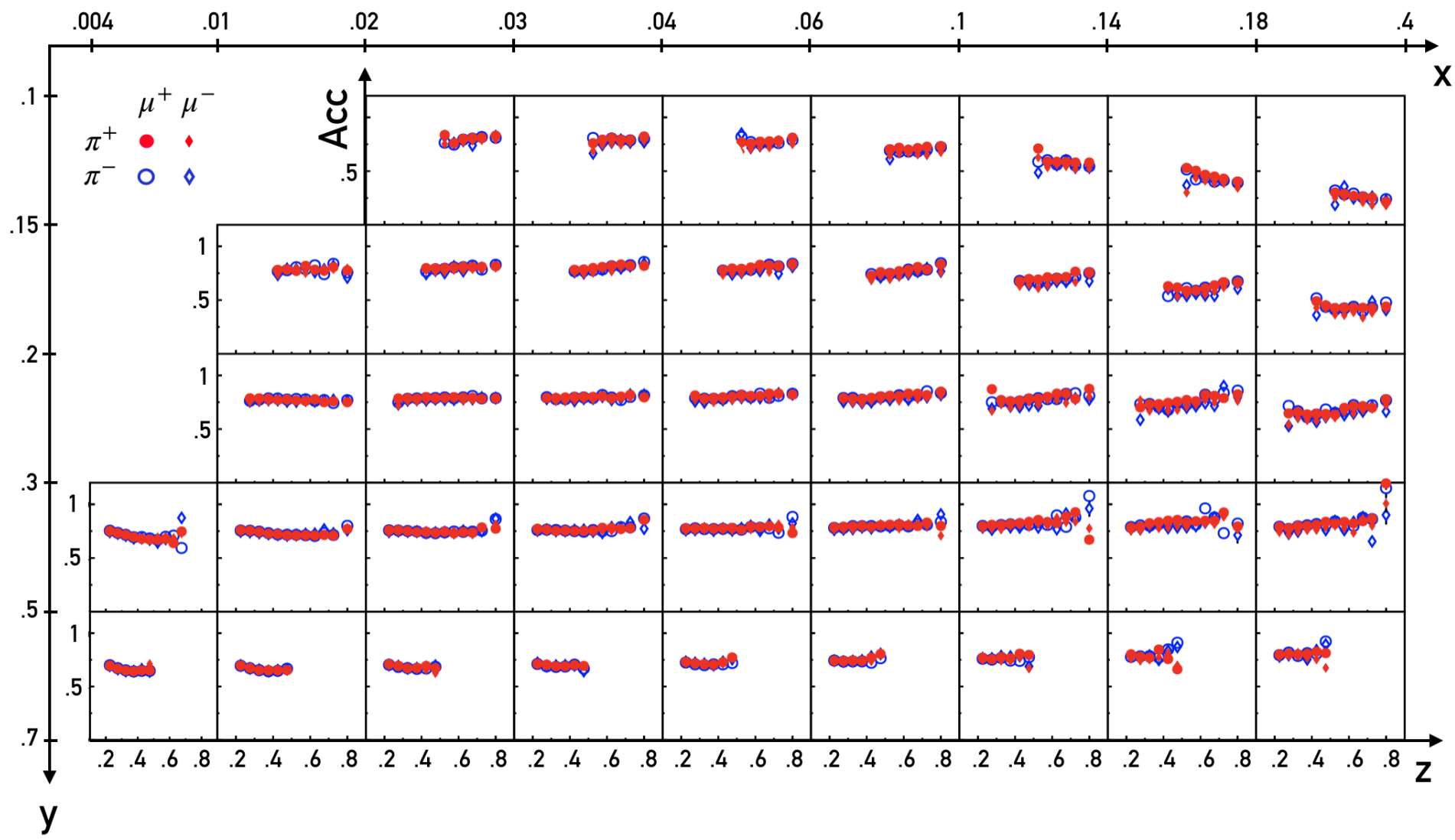


Figure 84: Charged pion acceptance in x, y and z bins. The red full markers correspond to positive hadrons, blue open markers to negative hadrons, circles markers correspond to pions obtained with μ^+ beam and diamonds markers to hadrons obtained with μ^- beam.

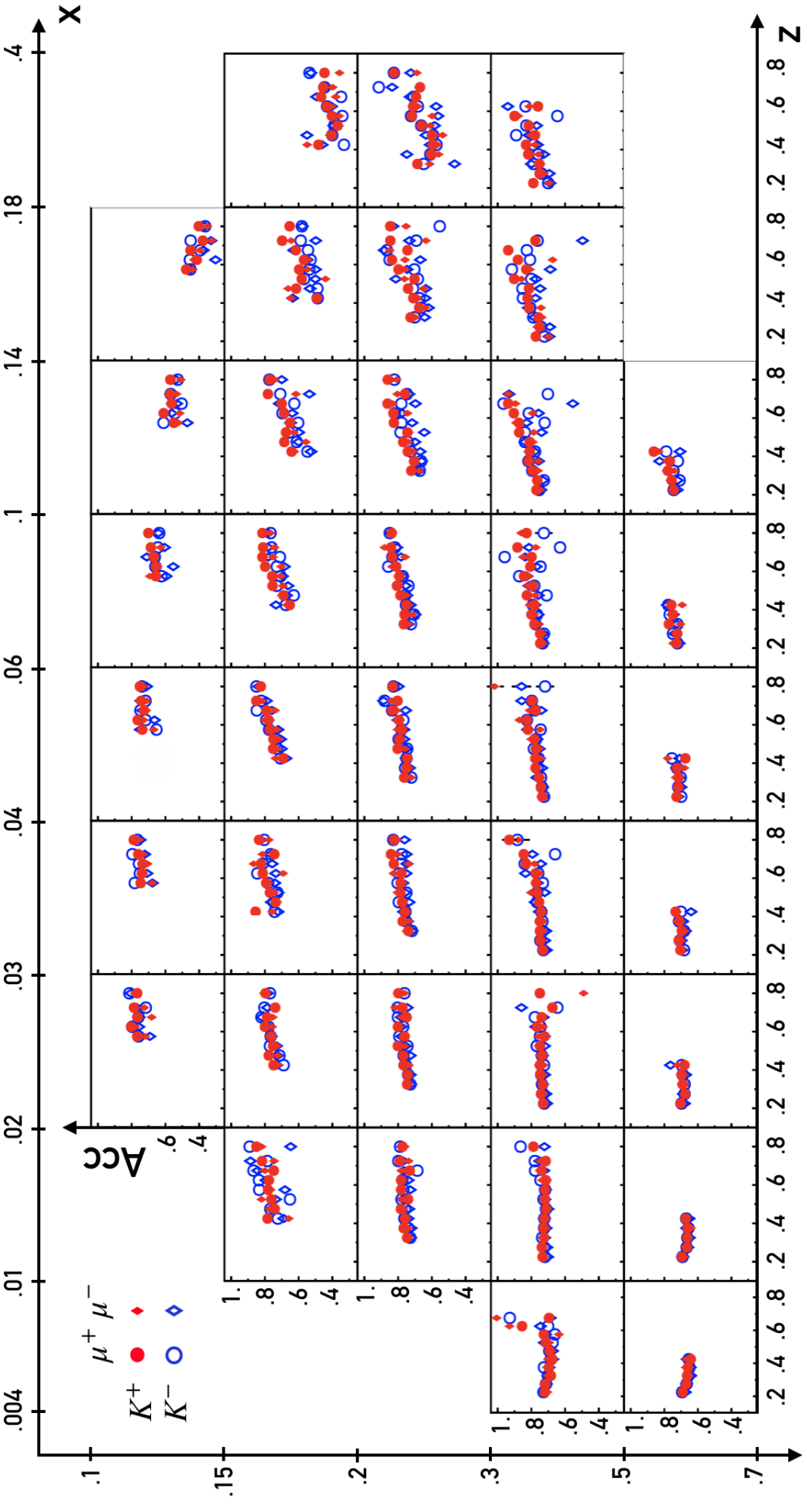


Figure 85: Charged kaon acceptance in x, y and z bins. The red full markers correspond to positive hadrons, blue open markers to negative hadrons, circles markers correspond to kaons obtained with μ^+ beam and diamonds markers to hadrons obtained with μ^- beam.

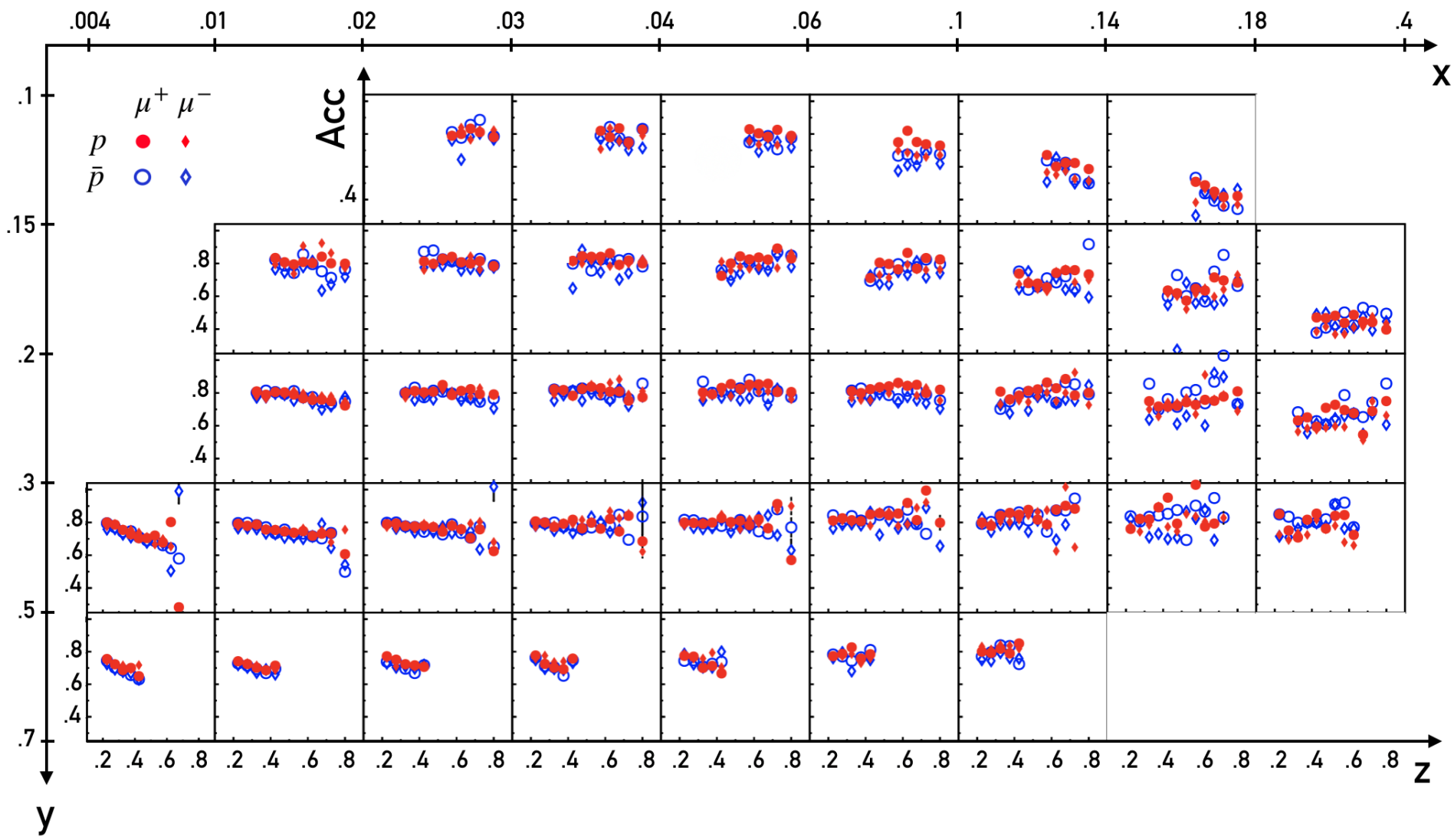


Figure 86: Charged proton acceptance in x, y and z bins. The red full markers correspond to positive hadrons, blue open markers to negative hadrons, circles markers correspond to kaons obtained with μ^+ beam and diamonds markers to hadrons obtained with μ^- beam.

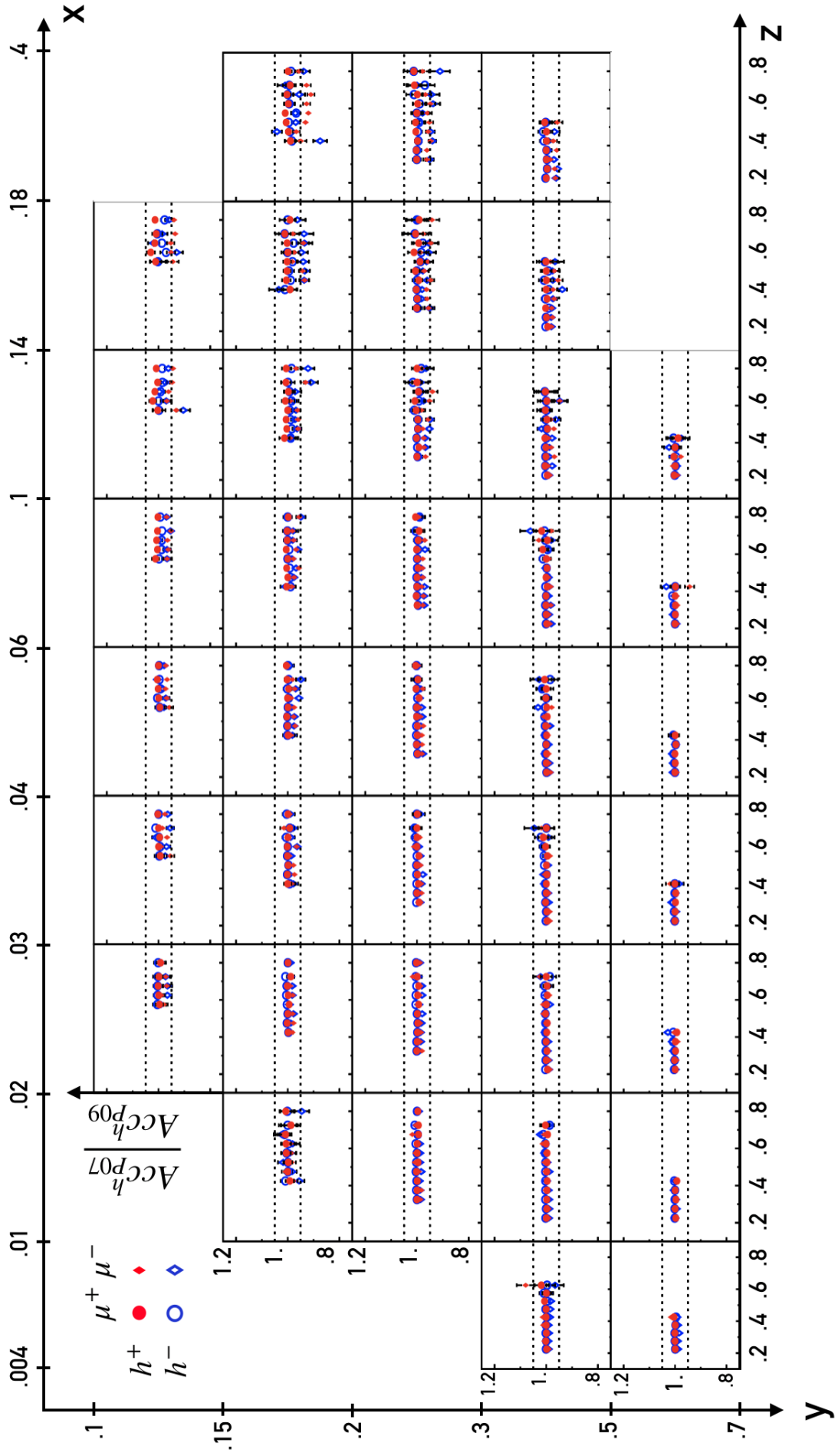


Figure 87: Ratio of P07 over P09 acceptance in x, y and z bins. The dashed lines represent a 5% discrepancy.

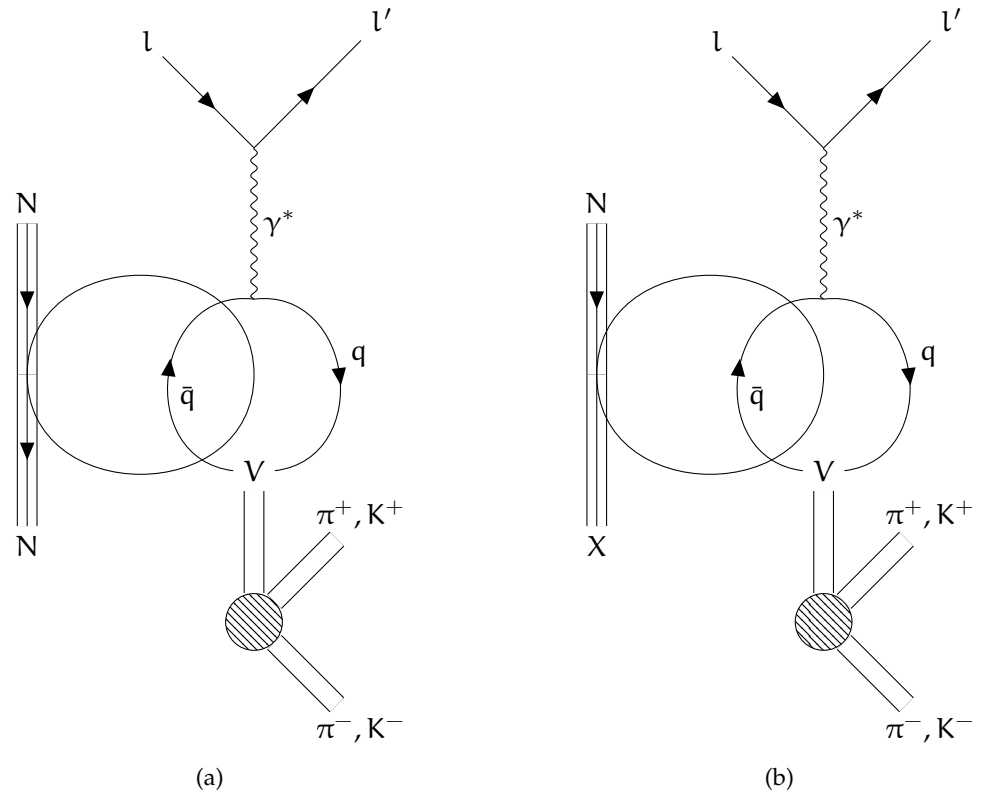


Figure 88: Vector meson diffractive production (V in the figure, being ρ^0, Φ, etc..). In the VMD model [114], the γ^* virtual photon creates a $q\bar{q}$ pair with compatible quantum numbers. Two cases can be encountered: vector meson exclusive production (where the same nucleon is found in the final state) (a) and vector meson production with nucleon diffractive dissociation (b).

chain as data. For the diffractive vector meson samples, both exclusive events and events with diffractive dissociation of the proton are simulated.

The fraction of pions (resp. kaons) resulting from a diffractive ρ^0 (resp. Φ) is calculated in the same binning as the raw multiplicities as:

$$\begin{aligned} f^{\pi}_{\rho^0}(x,y,z) &= \frac{N^{\pi}_{\text{HEPGEN++}}(x,y,z)}{N^{\pi}_{\text{DJANGOH}}(x,y,z) + N^{\pi}_{\text{HEPGEN++}}(x,y,z)} \\ f^{K}_{\Phi}(x,y,z) &= \frac{N^{K}_{\text{HEPGEN++}}(x,y,z)}{N^{K}_{\text{DJANGOH}}(x,y,z) + N^{K}_{\text{HEPGEN++}}(x,y,z)}, \end{aligned} \tag{108}$$

where $N^{\pi}_{\text{HEPGEN++}}$, N^{π}_{DJANGOH}, $N^{K}_{\text{HEPGEN++}}$ and N^{K}_{DJANGOH} are the number of kaons reconstructed from the HEPGEN++ and DJANGOH MC samples normalized by the corresponding MC luminosity (L_{MC}). The luminosity depends on the event weighting and the process cross-section σ_{int} (DIS for DJANGOH event and diffractive vector meson production for HEPGEN++ events):

$$\sum_{events} w_i = L_{MC} \cdot \sigma_{int}. \tag{109}$$

The final weighted number of DIS events and hadrons is summarized in Table 14.

The same MC samples are also used to determine diffractive vector meson contribution to the DIS sample. The diffractive vector meson events can also lead to a contamination in DIS events. Here, the two channels studied are diffractive ρ^0 and Φ with the fraction of the contamination expressed as:

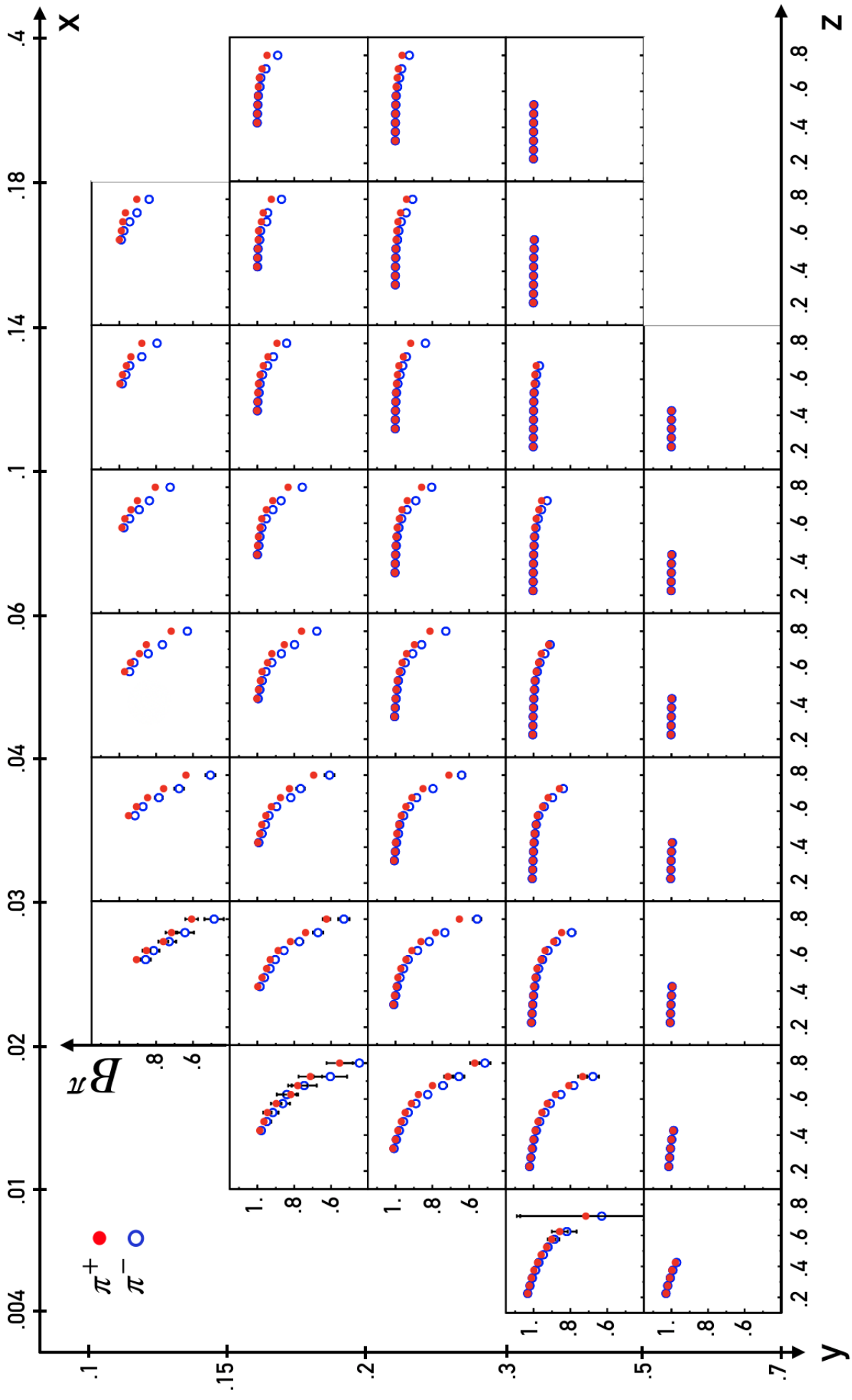


Figure 89: Correction factor B^π for diffractive vector meson contamination (ρ^0) as a function of z for (x,y) bins. The red markers correspond to positive pions and blue markers for negative pions.

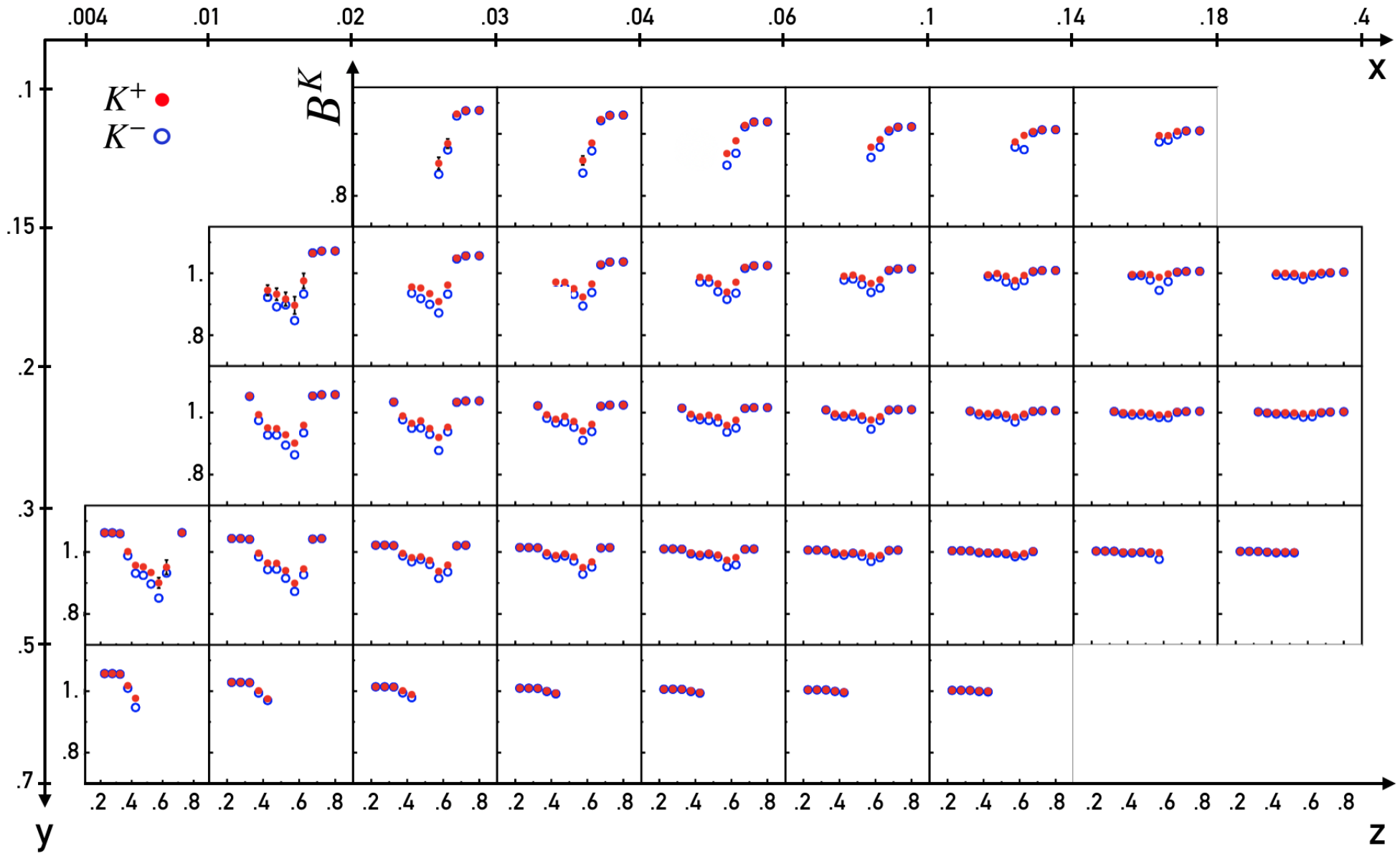


Figure 90: Correction factor B^K for diffractive vector meson contamination (Φ) as a function of z for (x,y) bins. The red markers correspond to positive kaons and blue markers for negative kaons.

	DJANGOH	ρ^0	Φ
Generated Events	8.4M	19.1M	19.2M
Weighted Generated Events	8.4M	280.2M	576.9M
Integrated Cross-Section [pb]	227010	12200	2500
Monte-Carlo Luminosity [pb^{-}1]	36.8	22966.6	23075.2
DIS Events [pb]	3.4M	96209	18083
h^+ [pb]	630880	25301	3890
h^- [pb]	511014	25250	4033
π^+ [pb]	453794	25257	-
π^- [pb]	377335	25212	-
K^+ [pb]	102019	-	3872
K^- [pb]	75158	-	4015

Table 14: Weighted number of DIS events and hadrons for the diffractive vector meson correction.

$$
\begin{aligned}
f^{\rho^0}_{DIS}(x,y,z) &= \frac{N^{DIS}_{\rho^0,HEPGEN++}(x,y,z)}{N^{DIS}_{DJANGOH}(x,y,z) + N^{DIS}_{\rho^0,HEPGEN++}(x,y,z) + N^{DIS}_{\Phi,HEPGEN++}(x,y,z)}, \\
f^{\Phi}_{DIS}(x,y,z) &= \frac{N^{DIS}_{\Phi,HEPGEN++}(x,y,z)}{N^{DIS}_{DJANGOH}(x,y,z) + N^{DIS}_{\rho^0,HEPGEN++}(x,y,z) + N^{DIS}_{\Phi,HEPGEN++}(x,y,z)}.
\end{aligned}
\tag{110}
$$

The total contribution from the diffractive vector-meson contribution to the DIS sample is $f^{VM}_{DIS} = f^{\rho^0}_{DIS} + f^{\Phi}_{DIS}$. The final corrections read as follows

$$
\begin{aligned}
B^h(x,y,z) &= \frac{\frac{N^\pi(x,y,z)}{N^h(x,y,z)}\left(1 - f^\pi_{\rho^0}(x,y,z)\right) + \frac{N^K(x,y,z)}{N^h(x,y,z)}\left(1 - f^K_\Phi(x,y,z)\right) + \frac{N^p(x,y,z)}{N^h(x,y,z)}}{1 - f^{VM}_{DIS}(x,y,z)} \\
B^\pi(x,y,z) &= \frac{1 - f^\pi_{\rho^0}(x,y,z)}{1 - f^{VM}_{DIS}(x,y,z)} \\
B^K(x,y,z) &= \frac{1 - f^K_\Phi(x,y,z)}{1 - f^{VM}_{DIS}(x,y,z)}.
\end{aligned}
\tag{111}
$$

and are displayed in Fig. 89 for the correction for pions and in Fig. 90 for the correction for kaons. The correction for pions has the strongest impact at high z ($z > 0.5$) and low x ($x < 0.02$) where it can reach 50%, while the correction for kaons has the strongest impact at middle z ($0.4 < z < 0.6$) and low x ($x < 0.02$), where it can reach 20%.

11.3 RADIATIVE CORRECTIONS

The experimental multiplicities include also QED radiative effects that should be corrected for the differential cross-section with respect to the 1γ cross-section. The correction factor taking into account these contributions is the radiative correction factor defined as:

$$
\eta(x,y,z) = \frac{d^2M_{1\gamma}/dxdydz}{d^2M_{measured}/dxdydz}, \tag{112}
$$

where $M_{1\gamma}$ denotes multiplicities obtained using the cross-section in the one photon exchange approximation and $M_{measured}$ denotes multiplicities obtained using the measured

cross-section which includes radiative effects. The bias on the μ kinematics upon real photon emission affects in turn the reconstruction of the kinematic variables x, y and z. This effect is now taken into account thanks to the use of DJANGOH. Previously in this chapter, DJANGOH was used as an event generator for our Monte Carlo simulation but it can also be used to compute radiative corrections. At generator level, one sample of multiplicities with radiative corrections and one without are generated and from these two samples $\eta(x, y, z)$ is calculated. Fig. 65 in Chapter 8, Section 8.3 shows the obtained correction factor. A statistical error is associated to this correction due to the non-analytical nature of the calculation. This is the first time that radiative corrections in bins of (x, y, z) are used in a COMPASS analysis.

11.4 ELECTRON CONTAMINATION

The pion (and thus hadron) sample is contaminated by electrons and positrons. With the DJANGOH event generator, we are able to describe almost correctly the electron production from radiative photons: in Fig. 72 one can see that the discrepancy of electroproduction in Φ in the hadron production plane is up to 10% at low Φ. Given the overall small size of the correction it was decided to use the MC sample for corrections at momenta, where the electron identification cannot be provided by the RICH detector. The fraction of electrons in the pion samples obtained in the range $12 < p_h < 40$ GeV in the MC sample is shown in Fig. 91. This contamination goes from 5% at low z to 1% at high z. The correction is taken into account in the acceptance correction, taking the electrons in the reconstructed sample (as in data) and not in the generated one. The correction looks reasonable and further scrutinizing of this analysis is on the way.

11.5 SUMMARY

The most important correction is the acceptance $A(x, y, z)$, which accounts for the geometrical limitations of the apparatus, the data reconstruction efficiency and the detector efficiencies. The acceptance for charged hadrons is mostly about 70% but can drop down to 30% in some bins. The electron contamination correction, which corrects for the inability of the RICH to distinguish electrons and pions above 8 GeV/c, is embedded inside the acceptance correction.

The correction factor for the vector meson production $B^h(x, y, z)$ contaminating the hadron sample varies from 0 to 40% at high z for pions and 0 to 20% at medium z for kaons. As these cross-sections are experimentally not known, model calculations are used to evaluate this correction.

The correction factor related to the radiative corrections, taking into account the different cross-section and the change of hadron and lepton kinematic variables due to the emission of a real photon, hence biaising the kinematic distributions, is going from 2 to 20%, the highest correction being located at high y, high z and low x.

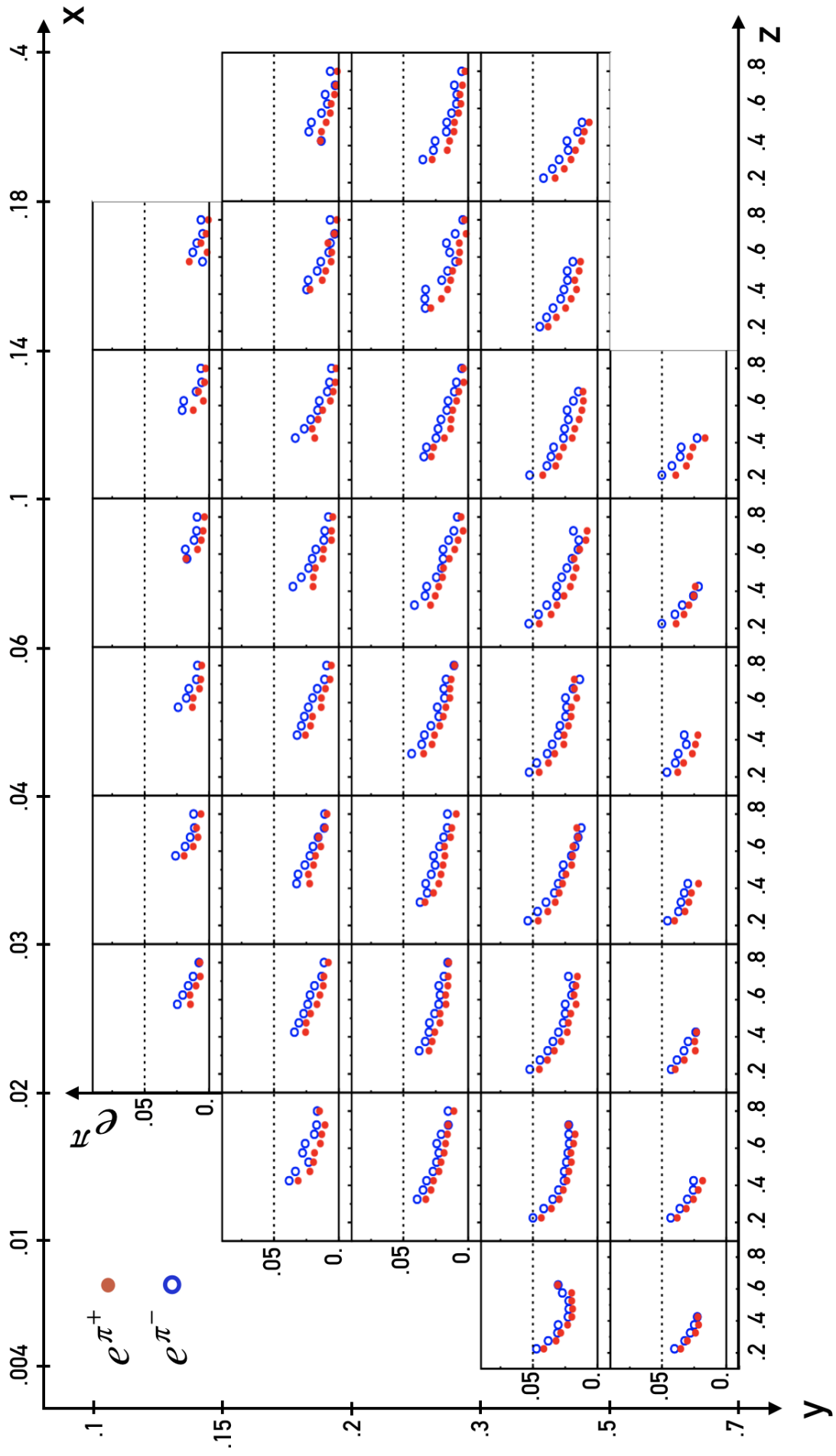


Figure 91: Fraction of electron contamination for pions as a function of z in bins of x (columns) and y (rows). Red markers are for π^+ and blue markers for π^-.

12 FINAL CHARGED HADRONS MULTIPLICITIES

This chapter is divided in two parts: one is focused on the discussion on the systematic uncertainties, the other presents the final multiplicities of unidentified charged hadrons, charged pions, charged kaons and protons/antiprotons extracted from SIDIS of 160 GeV muons off a pure proton target (lH_2). These results are obtained from the raw multiplicities of Chapter 10 and the correction factors of Chapter 11.

12.1 SUMMARY OF SYSTEMATIC STUDIES

The various systematic studies that were performed are summarized hereafter.

12.1.1 *Systematic uncertainty associated to the RICH unfolding*

The first stage of pion identification is based on the likelihood ratios: $LH(\pi)/LH(2^{nd})$ and $LH(\pi)/LH(bg)$. These cuts are optimized to minimize the pions misidentified as kaons. The systematic error associated to the selection of these cuts is performed varying the cuts around optimized values. Two sets of cuts *loose* and *severe* were used (see Table 15)

	Loose			Severe		
	π	K	$p(\bar{p})$	π	K	$p(\bar{p})$
$\frac{LH(\pi)}{LH(2^{nd})}$	> 1.00	–	–	> 1.06	–	–
$\frac{LH(\pi)}{LH(bg)}$	> 2.00	–	< 2.3 (2.2)	> 2.04	–	< 2.0 (1.9)
$\frac{LH(K)}{LH(2^{nd})}$	–	> 1.06	–	$> -$	> 1.10	–
$\frac{LH(K)}{LH(bg)}$	–	> 2.00	< 3.0 (2.9)	–	> 2.16	< 2.7 (2.6)

Table 15: Set of loose and severe cuts to evaluate the RICH systematic errors.

To evaluate the systematic error associated to the selection of the particle likelihood cuts, the particle identification is performed using the *loose* and *severe* sets of likelihood cuts and the corresponding RICH probability matrices and final multiplicities are extracted ($M_{raw}^{h^\pm,loose}$ and $M_{raw}^{h^\pm,severe}$ respectively). The largest difference between $M_{raw}^{h^\pm,loose}$ and $M_{raw}^{h^\pm,severe}$ with the nominal multiplicity $M_{raw}^{h^\pm}$ is taken as an estimate of the systematic error:

$$\sigma_{sys}^{RICH_{LH}} = MAX(|M_{raw}^{h^\pm,loose} - M_{raw}^{h^\pm}|, |M_{raw}^{h^\pm,severe} - M_{raw}^{h^\pm}|). \quad (113)$$

The difference between the altered RICH probability matrices and the optimal one are plotted in Figs. 92 and 93. For pions, the largest differences ($< 5\%$) are observed in the high momentum p_h region. For kaons and protons, the difference reaches 10% at low p_h ; small differences ($< 1\%$) are observed at the highest p_h value.

A second source of systematic error is that associated with the calculation of the RICH probability matrices M_{RICH}. This is estimated by generating two sets of altered RICH probability matrices. As represented in Eq. 114 the matrices are constructe using the statistical error associated to the original probability matrix elements.

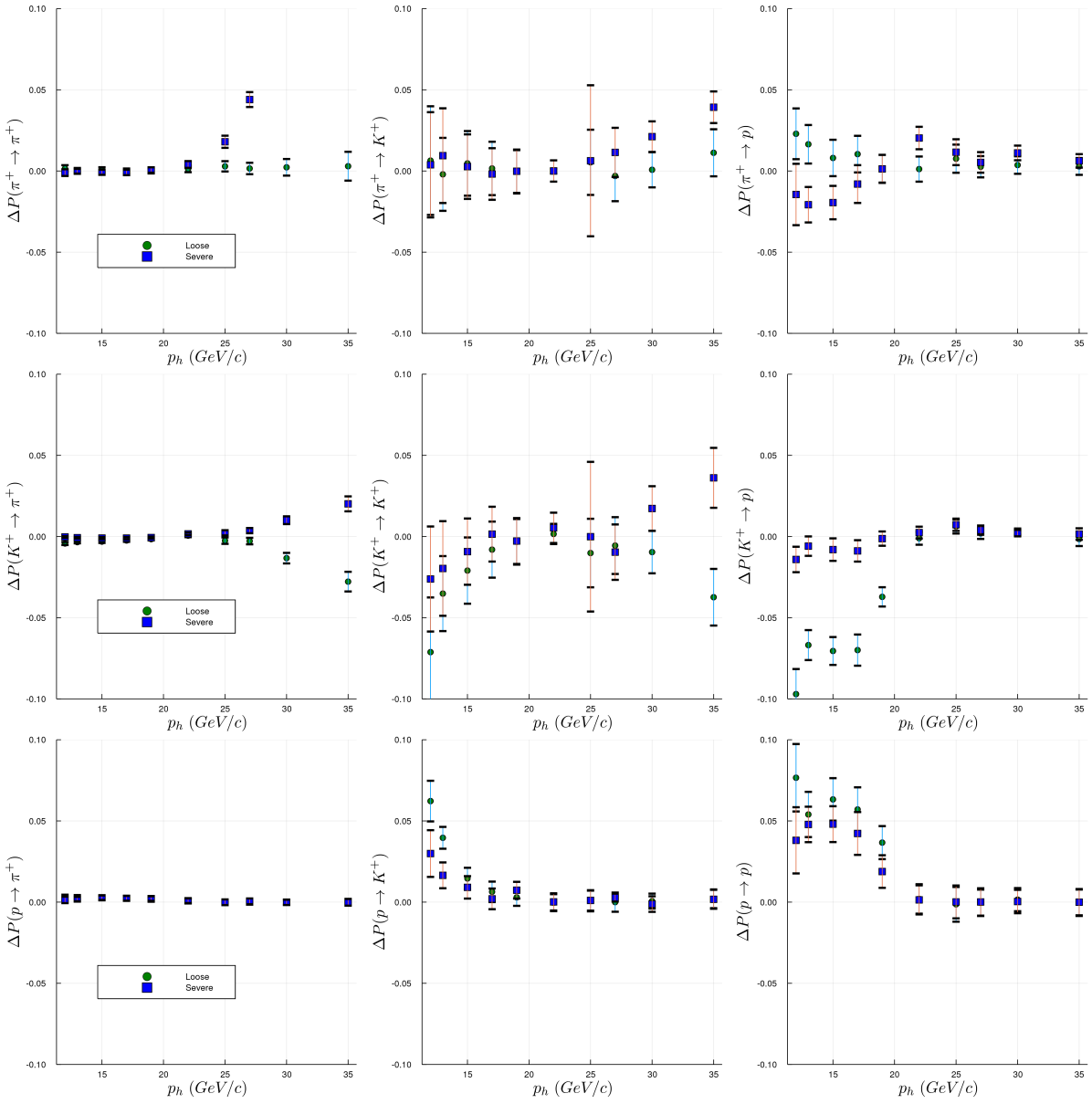


Figure 92: Difference between the identification and misidentification probabilities of loose and severe cuts with the optimal cuts for positive hadrons.

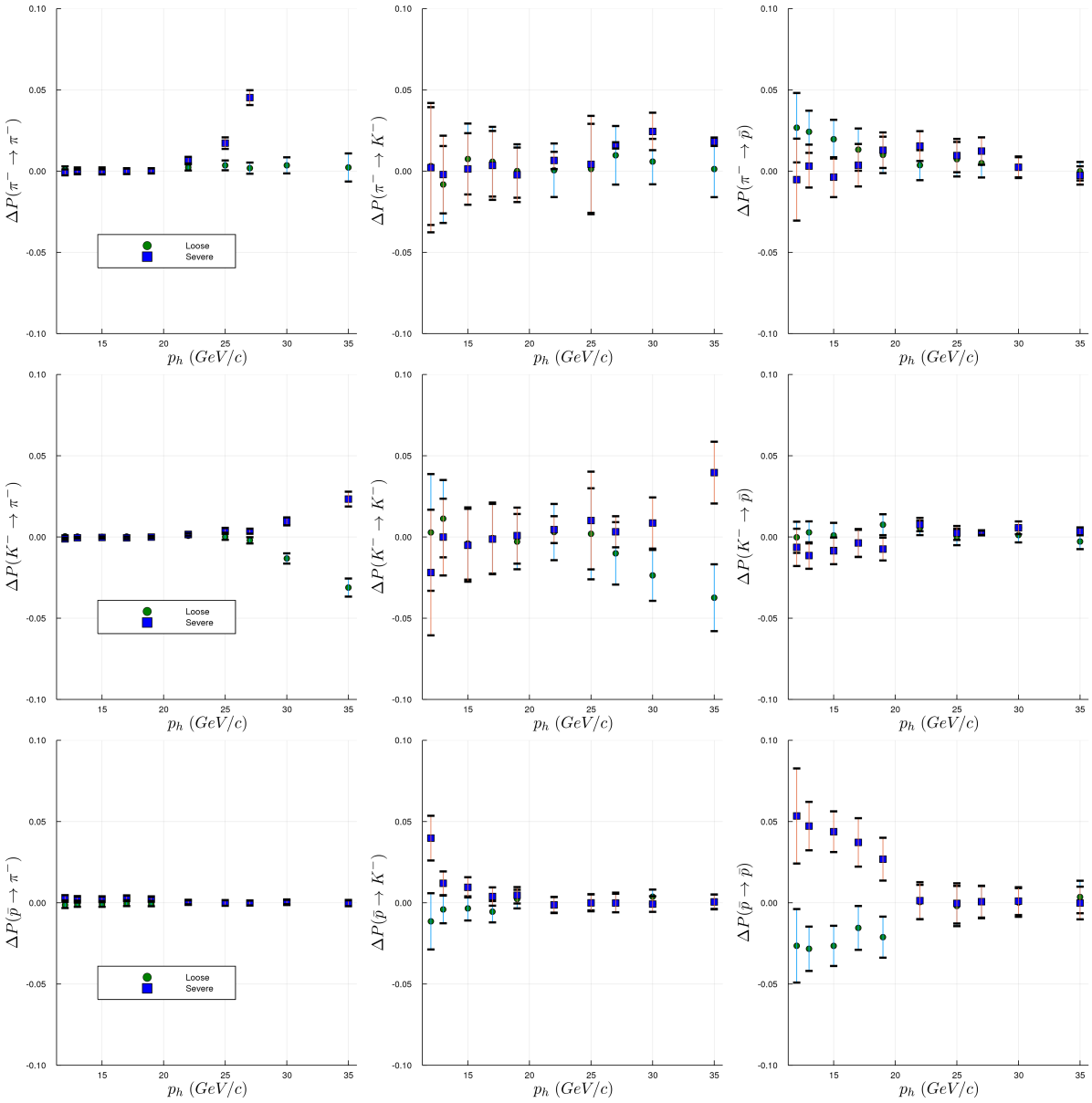


Figure 93: Same as Fig. 92 for negative hadrons.

$$M^{\pm}_{RICH} = \begin{bmatrix} P(\pi \to \pi) \pm \sigma_{P(\pi\to\pi)} & P(K \to \pi) \mp \sigma_{P(K\to\pi)} & P(p \to \pi) \mp \sigma_{P(p\to\pi)} \\ P(\pi \to K) \mp \sigma_{P(\pi\to K)} & P(K \to K) \pm \sigma_{P(K\to K)} & P(p \to K) \mp \sigma_{P(p\to K)} \\ P(\pi \to p) \mp \sigma_{P(\pi\to p)} & P(K \to p) \mp \sigma_{P(K\to p)} & P(p \to p) \pm \sigma_{P(p\to p)} \end{bmatrix} \quad (114)$$

The raw multiplicities $M^{h^{\pm},+}_{raw}$ and $M^{h^{\pm},-}_{raw}$ are then recalculated using the altered probability matrices $M^{\pm}_{RICH}$. The largest difference between $M^{h^{\pm},+}_{raw}$ and $M^{h^{\pm},-}_{raw}$ with $M^{h^{\pm}}_{raw}$ is taken as the sytematic error:

$$\sigma^{RICH_{stat}}_{sys} = MAX(|M^{h^{\pm},+}_{raw} - M^{h^{\pm}}_{raw}|, |M^{h^{\pm},-}_{raw} - M^{h^{\pm}}_{raw}|). \quad (115)$$

The final systematic uncertainty associated to the particle identification and unfolding correction (σ^{RICH}_{sys}) is the largest value of $\sigma^{RICH_{stat}}_{sys}$ and $\sigma^{RICH_{LH}}_{sys}$. The error goes from $< 0.2\%$ at low y for all x and z bins to $\sim 20\%$ for high y and high z, where multiplicities are low.

12.1.2 *Systematic uncertainty associated to the stability of data over time*

The data samples used in the analysis were recorded over a period of 5 weeks. As a quality check, the raw charged hadron multiplicities from period P07 and from all periods (averaged using the flux as weight) were compared. The results were found to be be compatible within statistical fluctuations for all unidentified hadrons, pions, kaons and protons (Fig. 94 for unidentified hadrons, other species in Appendix C). Consequently, no systematic error will be assigned for the data compatibility.

12.1.3 *Systematic uncertainty associated to the stability of data over beam charge*

The data samples used in the analysis were recorded with two different beam charges. Using the same method than in the previous subsection, a comparison was made for both beam charges and the results were compatible within statistical fluctuations. Consequently, no systematic error will be assigned for the beam charge change.

12.1.4 *Systematic uncertainty associated to the rescue procedure*

The rescue procedure might introduce background into the selected hadron sample. By applying quality cuts and the requirements at the RICH entrance, this contribution is negligible. Consequently, no systematic uncertainty is assigned for this rescue procedure.

12.1.5 *Systematic uncertainty associated to Monte Carlo sample: DJANGOH dependence*

To determine the acceptance dependence with the physical model chosen, different PDF sets were used to generate different MC samples. Moreover different JETSET parameters were also used [115]. For each sample the acceptance is calculated and the hadron multiplicities are corrected with each acceptance. The systematic uncertainty of the acceptance, considering two acceptances A^h and A'^h, is estimed in each kinematic bin (x,y,z):

$$\sigma^{A^h}_{syst} = \left| \frac{\left(\frac{A'^h}{A^h} - 1\right) M^h_{corr,acc}}{A^h} \right| \quad (116)$$

The value of $\sigma^{A^h}_{syst}$ was found to be of $\sim 5\%$.

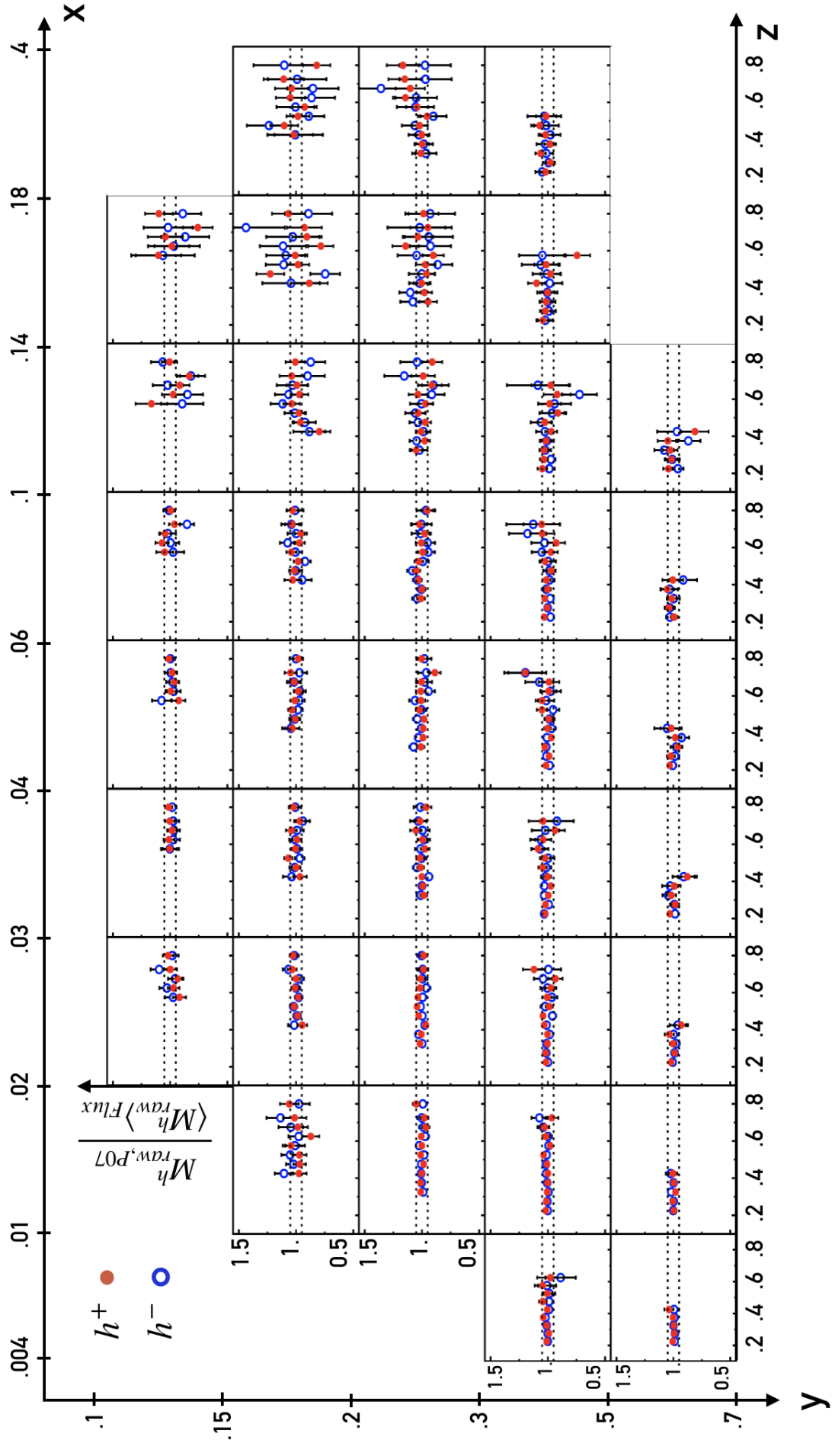
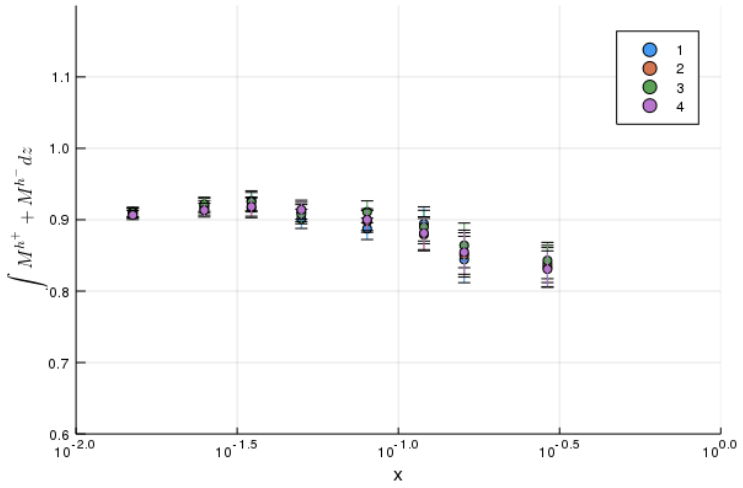
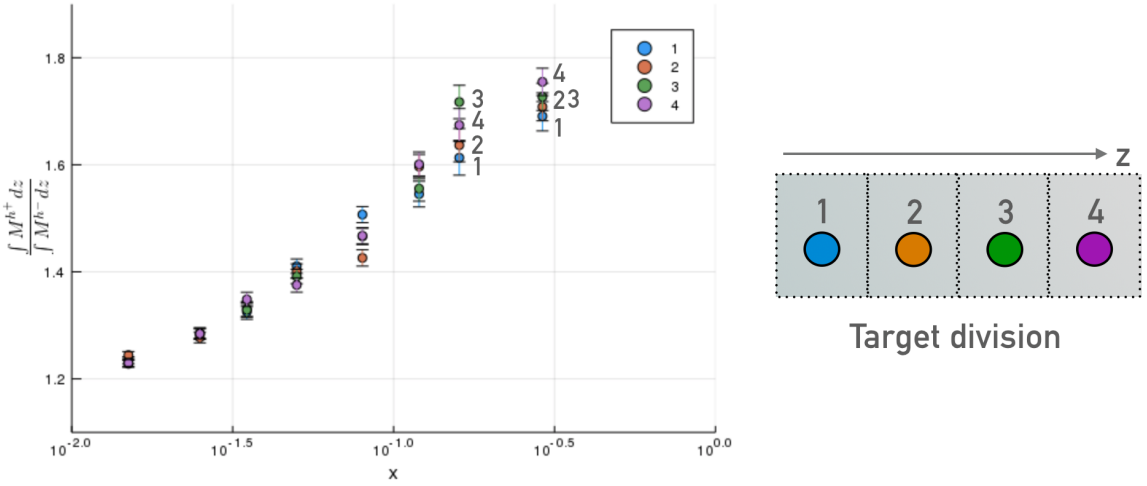


Figure 94: Ratio of P07 raw multiplicities over the raw multiplicities of all periods averaged over flux versus z. The red markers are for positive hadrons and the blue markers for negative hadrons. Each column corresponds to a given x bin and each row to a given y bin. The dashed lines are delimiting a 5% discrepancy.

To study the quality of the spectrometer description the target was split into four different parts. For the comparison the multiplicities were integrated over z and averaged over y and then compared for each part of the target (Fig. 95). From this comparison, a conservative systematic error of $\sim 5\%$ was derived.



(a)



(b)

Figure 95: Comparison of the multiplicity sum (a) and ratio (b) for unidentified hadrons for different target slices.

In the end, a total systematic uncertainty of $\sim 10\%$ is assumed for the acceptance.

12.1.6 *Systematic uncertainty associated to the diffractive vector meson correction*

In HEPGEN, the cross section for exclusive vector meson production is given by the GPD model of Goloskokov and Kroll. The theoretical uncertainty on the predicted cross section close to COMPASS kinematics is around 30% [116]. Propagating this uncertainty leads to a maximum relative uncertainty below 6%.

12.2 CHARGED HADRON MULTIPLICITIES ($h^{\pm}$,$\pi^{\pm}$,$K^{\pm}$,$p/\bar{p}$)

The final results for the multiplicities are obtained as:

$$M^h_{Final}(x,y,z) = M^h_{raw}(x,y,z)\frac{\eta^h(x,y,z)}{A^h(x,y,z)}B^h(x,y,z), \tag{117}$$

including all corrections described in the previous section. The x,y and z binning is given in Table 13. A^h corresponds to the acceptance correction (Section 11.1), η^h to the radiative correction factor (Section 11.3) and B^h to the diffractive vector meson correction factor (Section 11.2). The statistical error propagation is performed in all (x,y,z) bins assuming that all corrections are independent:

$$\begin{aligned} E^2_{Final} = &\left(\frac{\eta^h \cdot B^h}{A^h}\right)^2 E^2_{raw} + \left(\frac{\eta^h \cdot B^h \cdot M^h_{raw}}{Acc^2}\right)^2 E^2_{Acc} \\ &+\left(\frac{\eta^h \cdot M^h_{raw}}{Acc}\right)^2 E^2_{VM} + \left(\frac{B^h \cdot M^h_{raw}}{Acc}\right)^2 E^2_{RC}. \end{aligned} \tag{118}$$

The corresponding systematic uncertainties from the different sources are added quadratically. The largest contribution in most bins comes from the systematic uncertainty of the acceptance.

12.2.1 *Final charged hadron multiplicities*

The 300 data points for each of the charged hadron multiplicities $M^{h^{\pm}}$ are shown in Figs. 96 to 103 as a function of z, in bins of x and staggered vertically with y. A strong z dependence is observed for all (x,y) bins as well as a small dependence with x. The Q^2 values are in the range 1 to 30 $(GeV/c)^2$. The statistical uncertainties are too small to be visible in almost all kinematic bins. The bands at the bottom of each x bin panel are the systematic errors for the bin $0.3 < y < 0.5$ (bin that covers the largest z range). They are very similar but not shown for the other y bins.

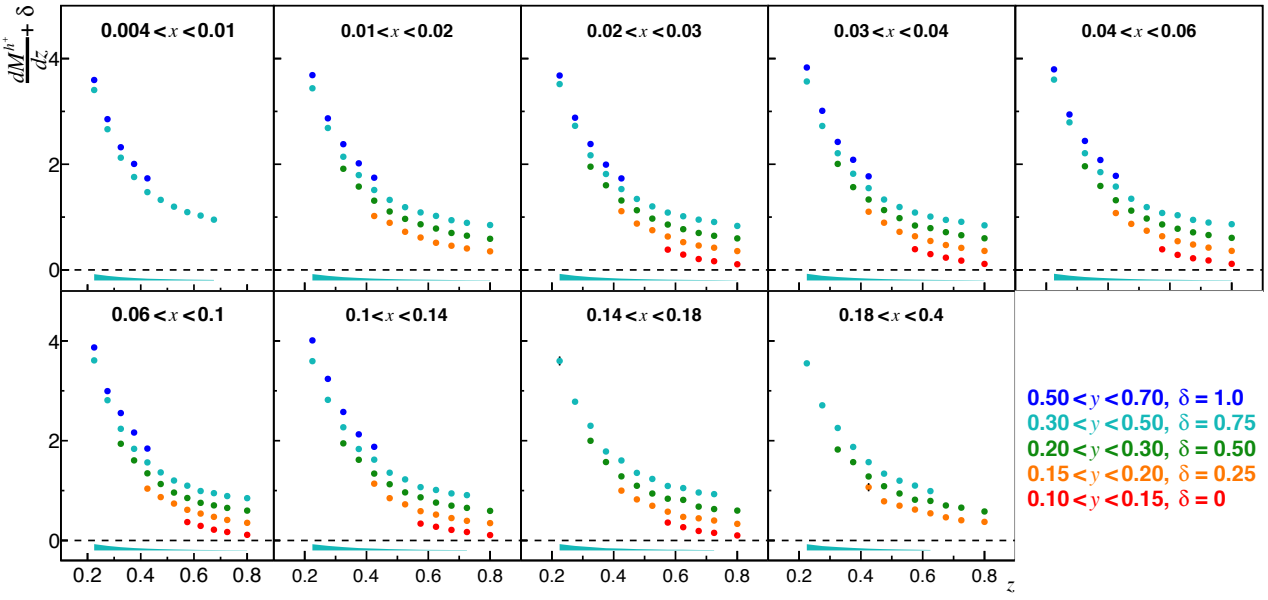


Figure 96: Unidentified positive hadron multiplicities (with all corrections) as a function of z in bins of x staggered vertically with y.

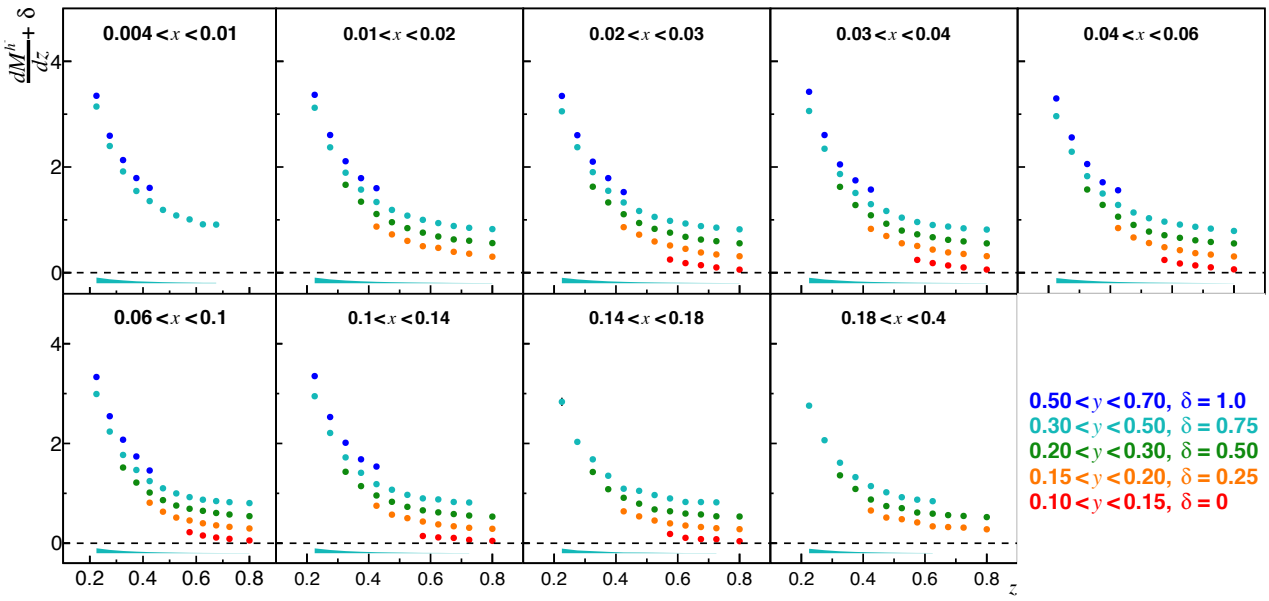


Figure 97: Same as Fig.96 for unidentified negative hadrons.

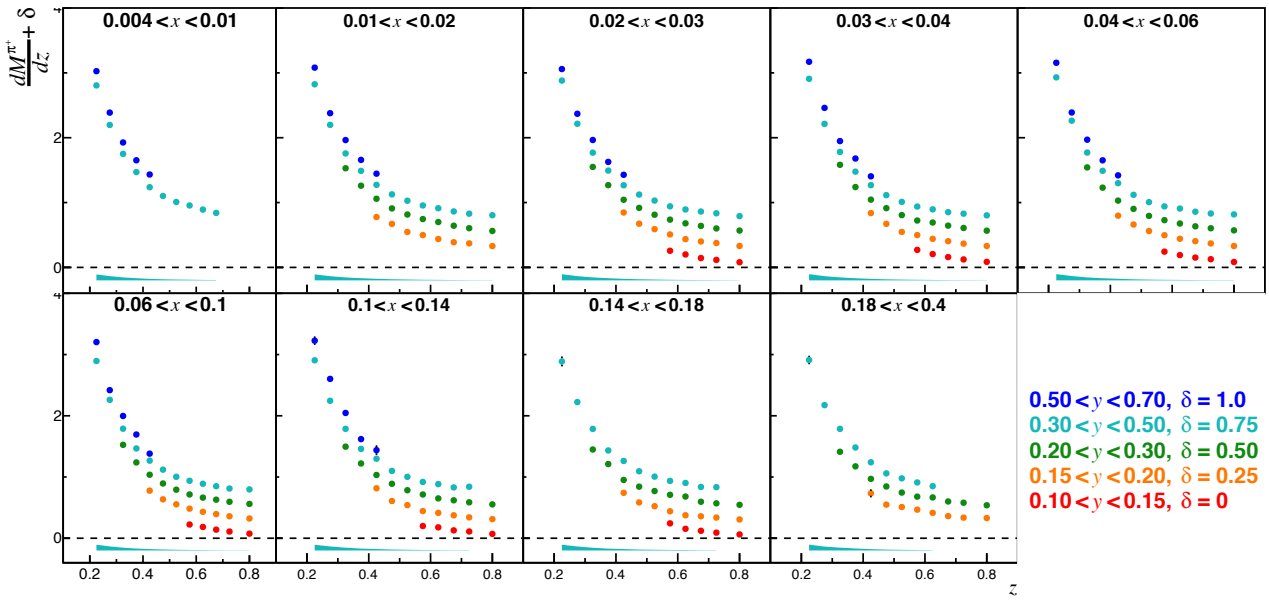


Figure 98: Same as Fig.96 for positive pion.

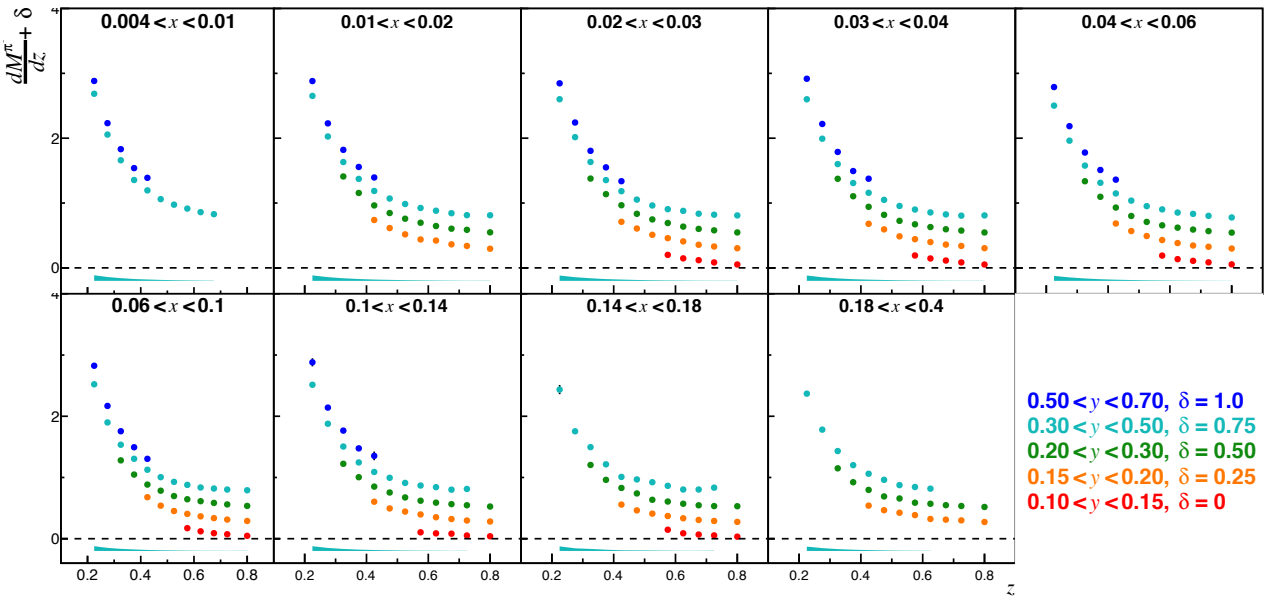


Figure 99: Same as Fig.96 for negative pions.

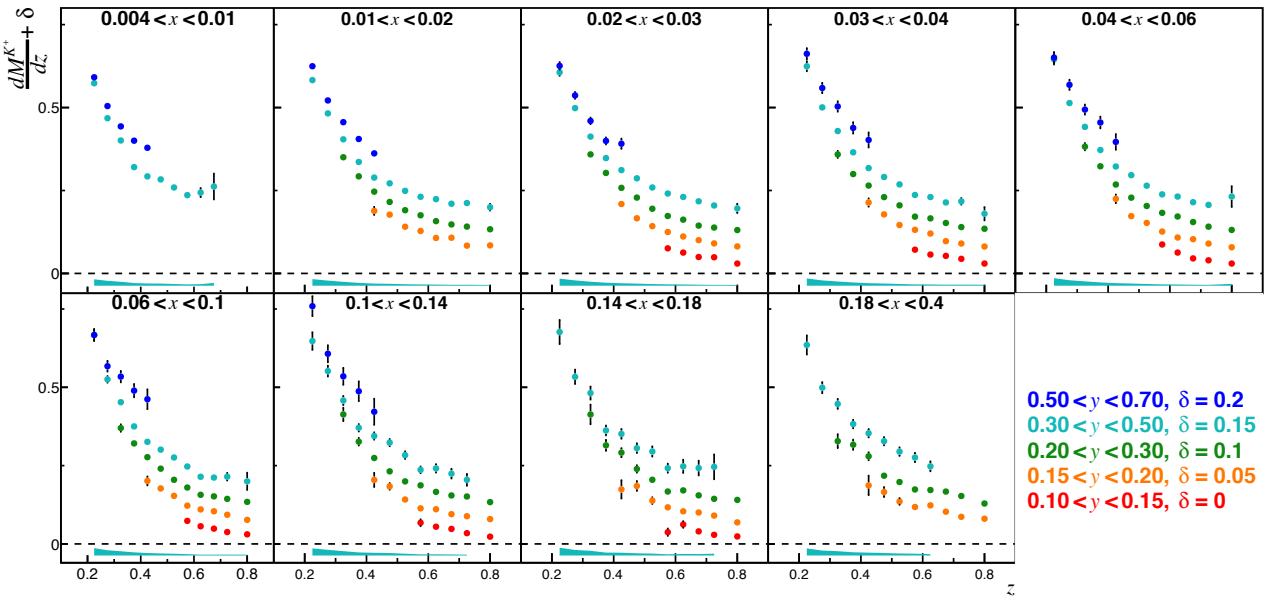


Figure 100: Same as Fig.96 for positive kaons.

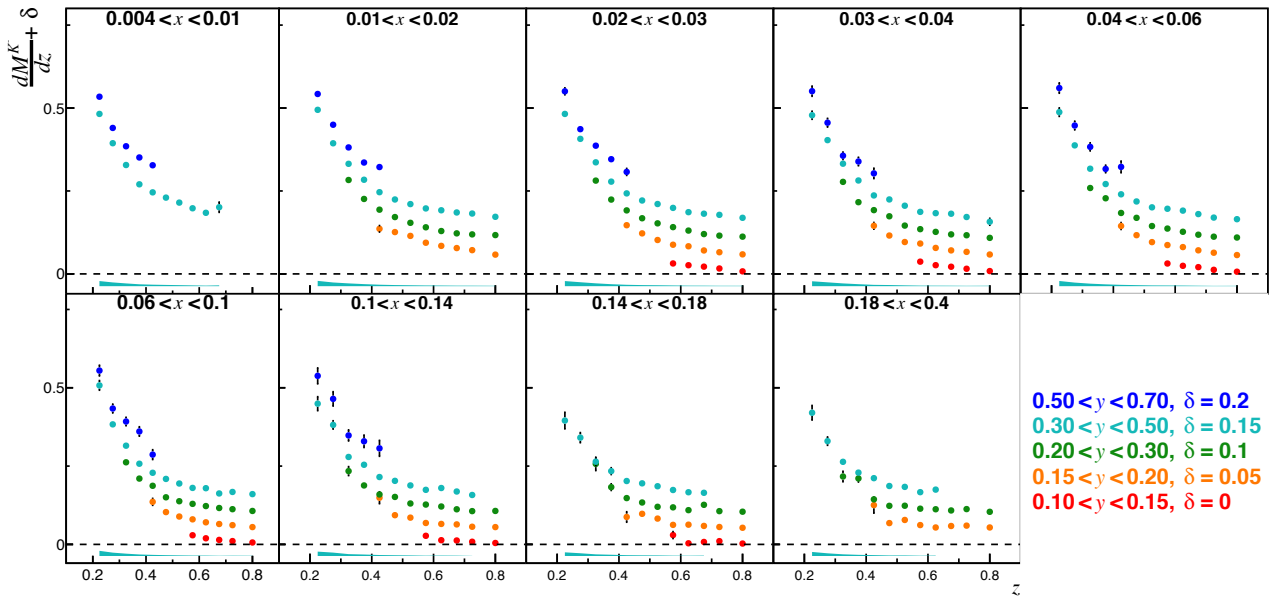


Figure 101: Same as Fig.96 for negative kaons.

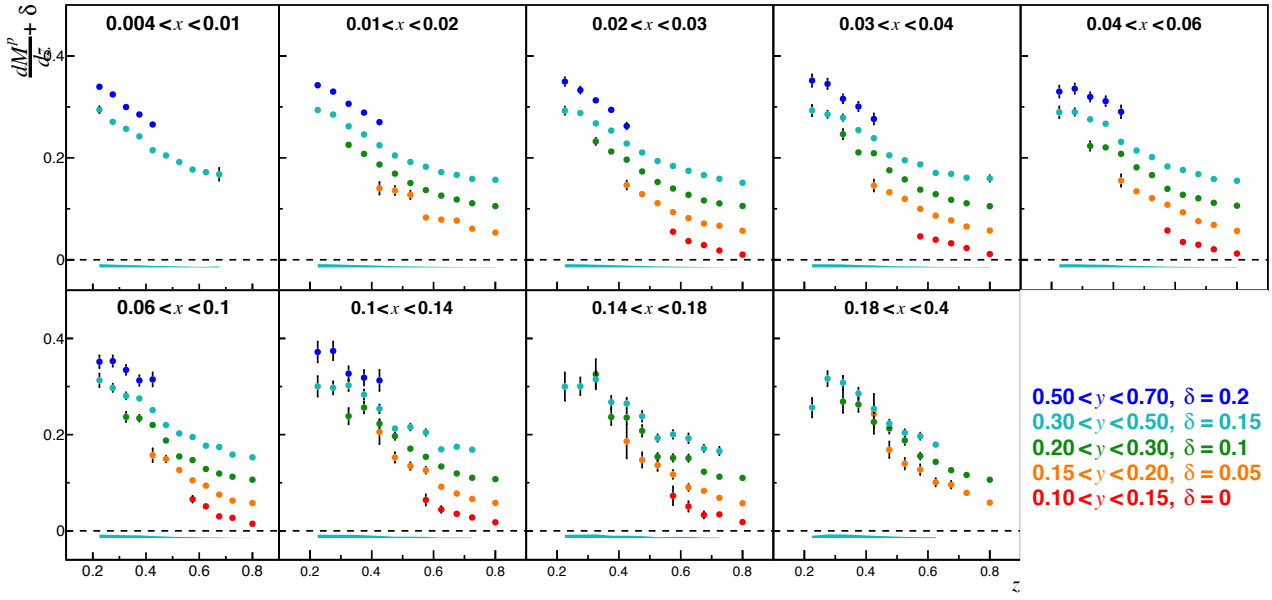


Figure 102: Same as Fig.96 for protons.

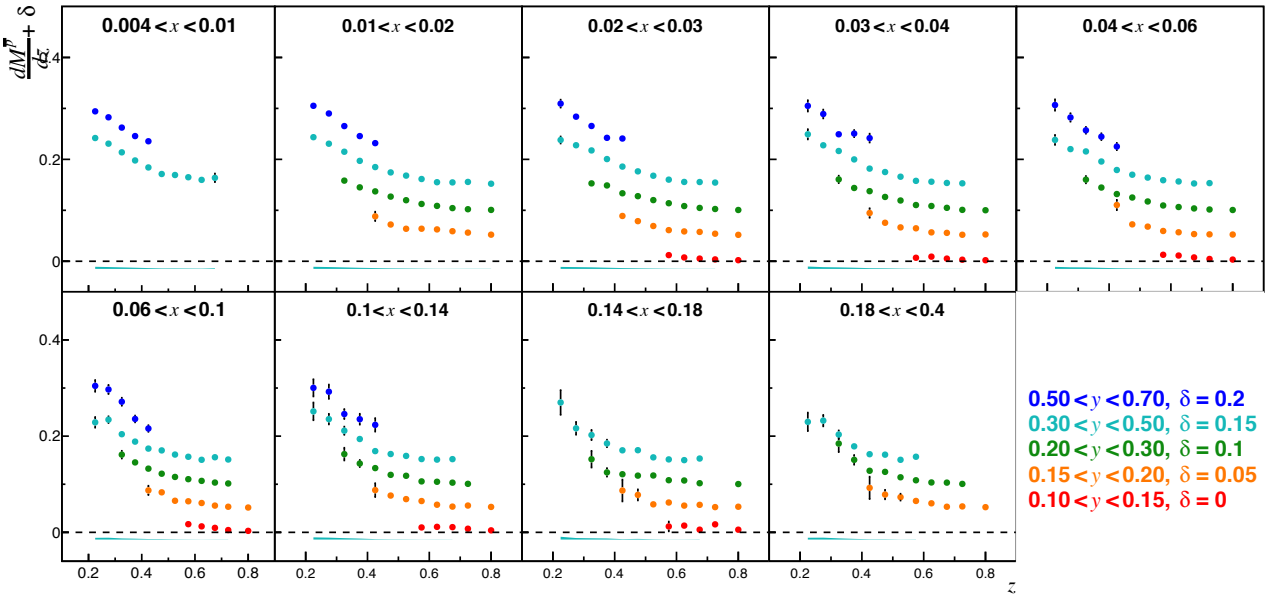


Figure 103: Same as Fig.96 for antiprotons.

When removing the y staggering on the previous results, one can see that the results from different y bins agree very well in the overlap region (Fig. 104). That means that our multiplicities results have no y dependence and can be averaged over y, using the square of the statistical error as weight.

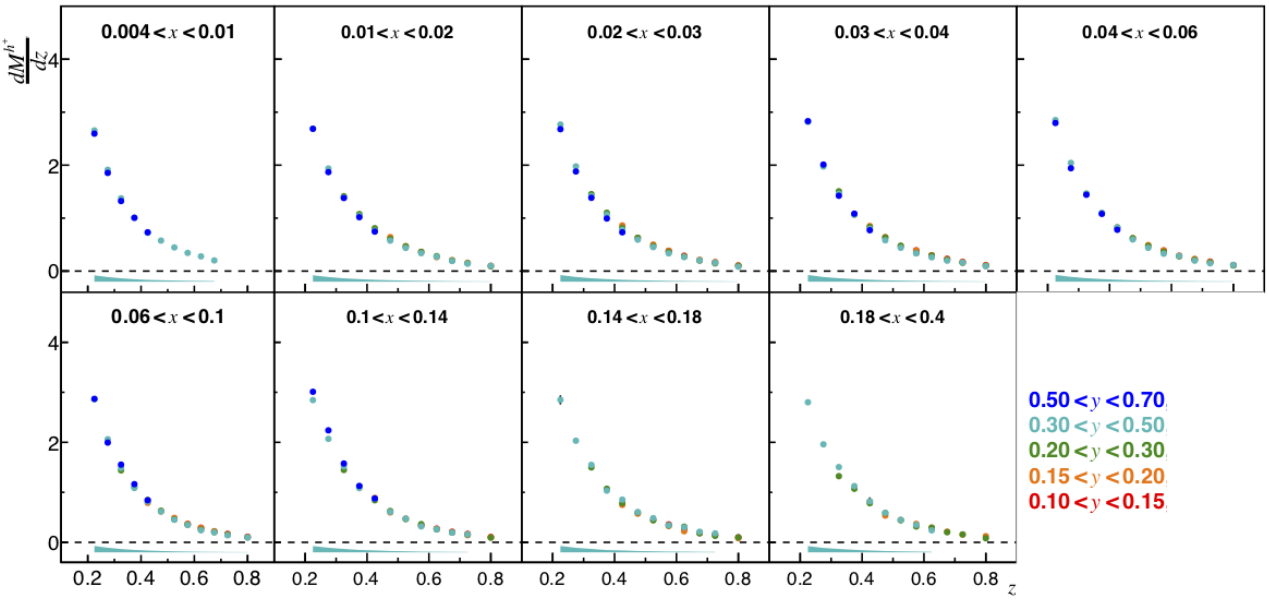


Figure 104: Unidentified positive hadron multiplicities (with all corrections) as a function of z in bins of x. The vertical staggering with y has been suppressed showing that the different y bins do overlap.

The multiplicities are shown as a function of z and in bins of x in Figs. 105 to 108. An asymmetry between all positive and negative charged hadrons is observed, increasing with x. The size of the asymmetry depends on the hadron species. Having more π^+ than π^- is due to the fact there is a dominant u quark distribution in the target but the asymmetry is smaller than for hadrons. The strong asymmetry between K^+ and K^- is due to the fact there

is a dominant valence u quark distribution in the target, whereas producing leading K^- is only possible with sea quarks. The same statement can be made for proton and antiproton.

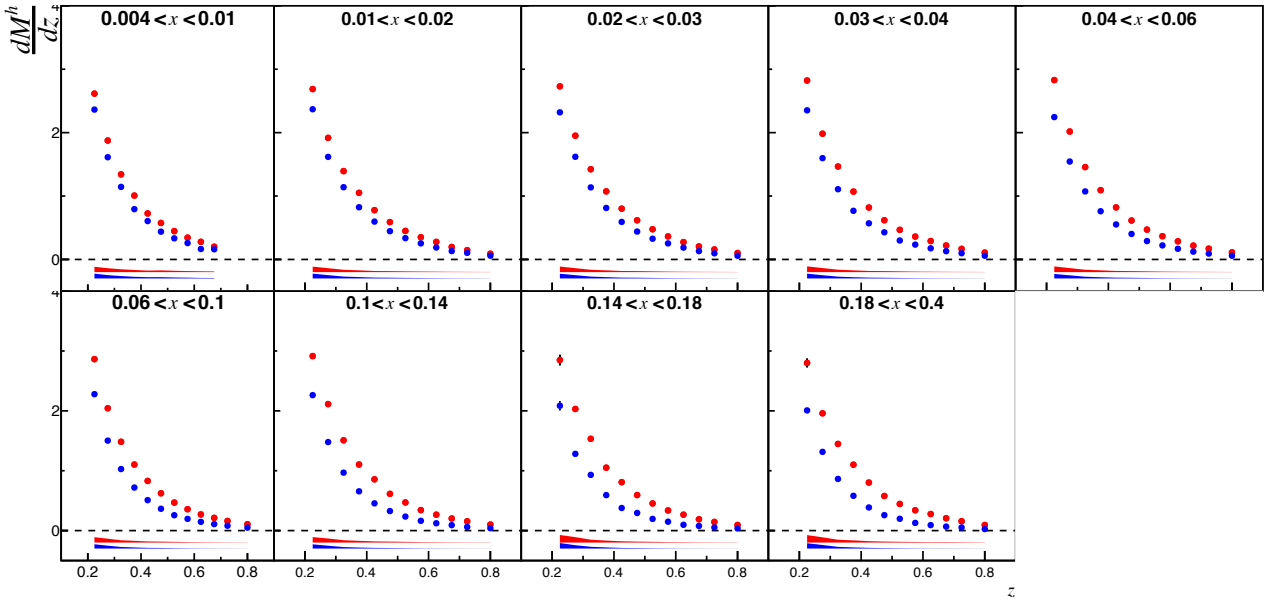


Figure 105: Unidentified positive (red) and negative (blue) hadron multiplicities (with all corrections) averaged over y as a function of z in bins of x.

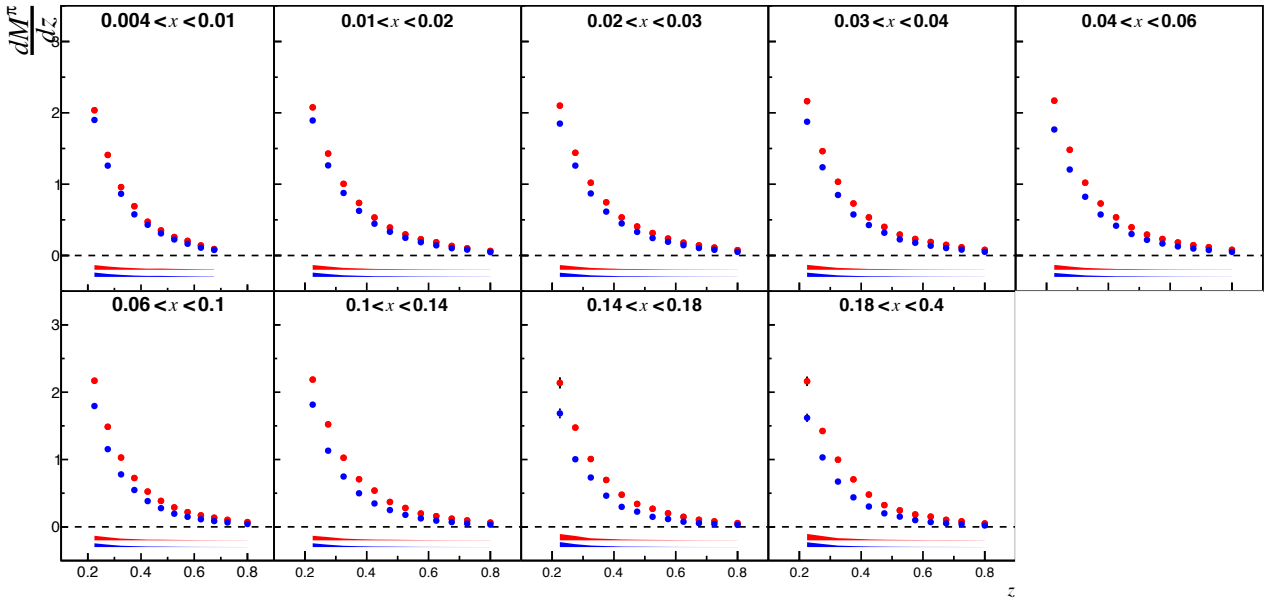


Figure 106: Same as Fig.105 for pions.

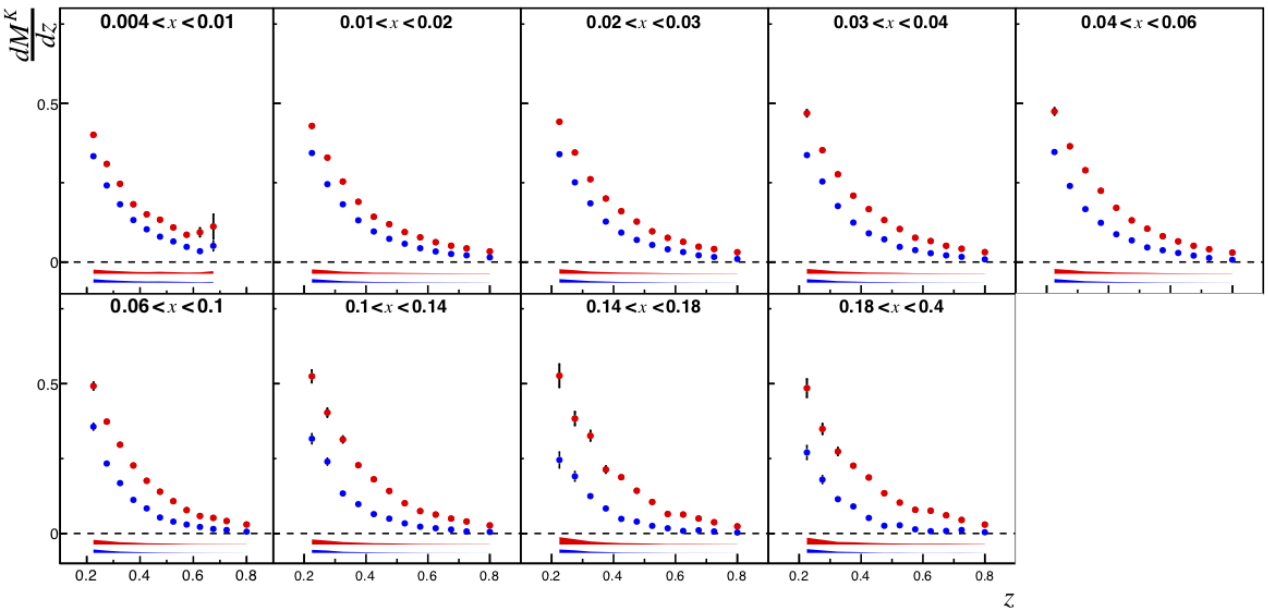


Figure 107: Same as Fig.105 for kaons.

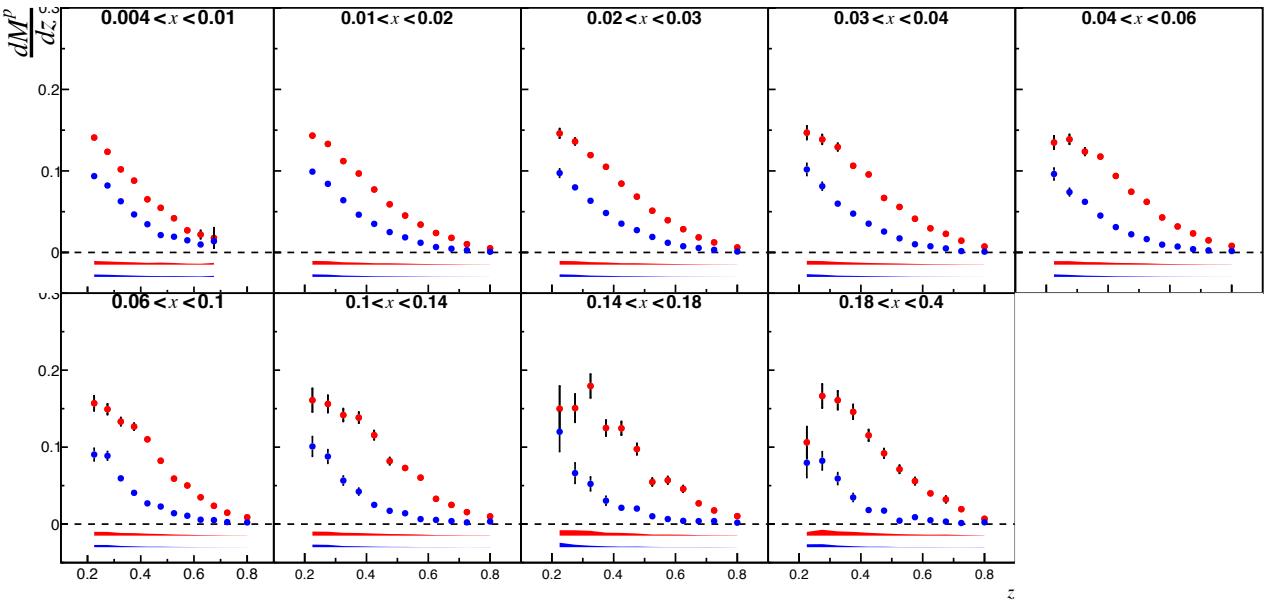


Figure 108: Same as Fig.105 for protons/antiprotons.

12.3 RATIO AND SUM OF CHARGED HADRON MULTIPLICITIES

From the y-averaged multiplicities, we can still integrate on z. This allows us to have an idea of the underlying physics as some limiting cases for the x dependence, when integrating over z, can be investigated. In addition it also allows us to compare our results with other experiments that have different binning in (x, Q^2, z). From the integration over z only the 8 x bins that have a sufficient z coverage subsist. One interesting quantity to look at is the ratio of charged hadron multiplicities as with this quantity most of the systematic uncertainties cancel. The ratio is calculated as following:

$$\frac{\mathscr{M}^{h^+}}{\mathscr{M}^{h^-}} = \frac{\int_{0.2}^{0.85} \langle M^{h^+} \rangle_y \, dz}{\int_{0.2}^{0.85} \langle M^{h^-} \rangle_y \, dz}. \tag{119}$$

In some cases the multiplicities are only integrated over z between 0.4 and 0.6:

$$\frac{\mathscr{M}_r^{h^+}}{\mathscr{M}_r^{h^-}} = \frac{\int_{0.4}^{0.6} \langle M^{h^+} \rangle_y \, dz}{\int_{0.4}^{0.6} \langle M^{h^-} \rangle_y \, dz}. \tag{120}$$

An alternative approach is to look at the sum of charged hadron multiplicities $\mathscr{M}^{h^+} + \mathscr{M}^{h^-}$ in order to study multiplicities independent of the hadron charge. For some type of hadrons, these sums extracted on a proton target (lH_2, this analysis) and an isoscalar target (^{6}LiD, COMPASS published results [44, 45]) should be similar.

Hereafter, the results from this analysis will be compared to COMPASS published results and HERMES published results [47]. HERMES has performed measurements of electron and positron of 27.6 GeV scattering off proton and deuteron targets at DESY-HERA. The datasets selected for the comparison are the multiplicities of charged pions and charged kaons in bins of x and z. The range of z of these datasets is of $[0.1, 1.1]$ and the range in y of $[0.1, 0.85]$.

For the proton multiplicities the results will be compared to EMC published results [117]. EMC has performed measurement of muon-proton scattering at 120 and 280 GeV. The dataset selected is the multiplicities of proton and antiproton in bins of x and integrated over z between 0.4 and 0.6. The range of x of the dataset is [0.02,0.3] and the range in y is [0.07,0.85].

12.3.1 *Ratio of charged hadron multiplicities*

In Figs. 109 to 112, $\mathscr{M}^{h^+}/\mathscr{M}^{h^-}$ from COMPASS is depicted for a proton target (blue triangle) and for an isoscalar target (orange circles). The same ratio from HERMES is presented for a proton target (violet open squares) and a deuteron target (green open stars) for charged pions and charged kaons.

The ratio of unidentified charged hadron and pion multiplicity for a proton target should lie above COMPASS results on isoscalar target, the reason being the different quark mixture in the two targets (more u in proton target thus higher $h+/h-$ ratio in proton than in isoscalar target). The difference is expected to be $\sim 10-20\%$, as obtained here. The ratio of charged kaon multiplicity on proton target is also expected to be larger than COMPASS results for isoscalar target by $\sim 10-20\%$, as observed here. These expectations are obtained by evaluating the multiplicities for both targets with Eq. 29 taking DSS07 [55] LO fragmentation functions for hadrons and taking PDFs from MSTW08 [111].

The COMPASS results for proton and isoscalar targets are compared to HERMES results for proton and deuteron targets. In all the x range, the proton results for pions are compatible within error bars, as were the deuteron results, while for kaons discrepancies between COMPASS and HERMES results for both proton and deuteron/isoscalar targets are seen.

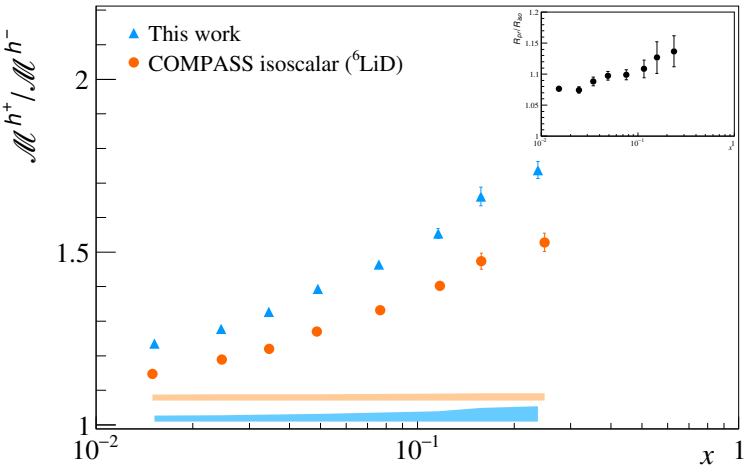


Figure 109: Ratio of $\frac{\mathcal{M}^{h^+}}{\mathcal{M}^{h^-}}$ for a proton target (blue closed points) and an isoscalar target (orange closed points) (COMPASS data). On the top right is displayed the ratio of proton target result over isoscalar target result.

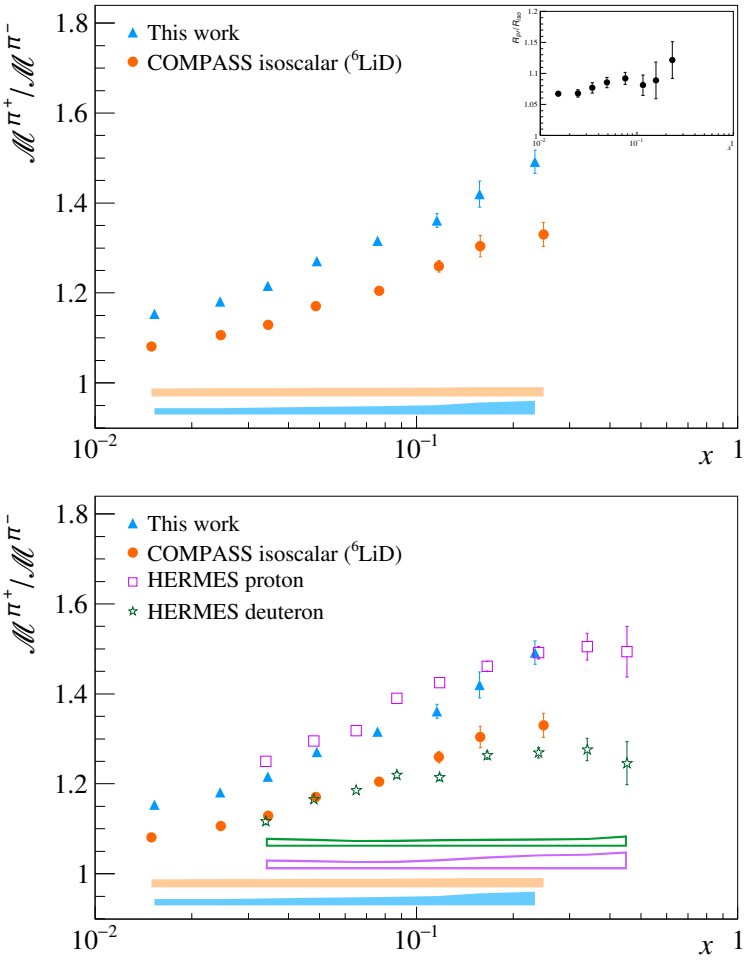


Figure 110: Ratio of $\frac{\mathcal{M}^{\pi^+}}{\mathcal{M}^{\pi^-}}$ from COMPASS for a proton target (blue closed points) and an isoscalar target (orange closed points) and from HERMES for a proton target (violet open points) and a deuteron target (green open points). On the top right is displayed the ratio of proton target result over isoscalar target result.

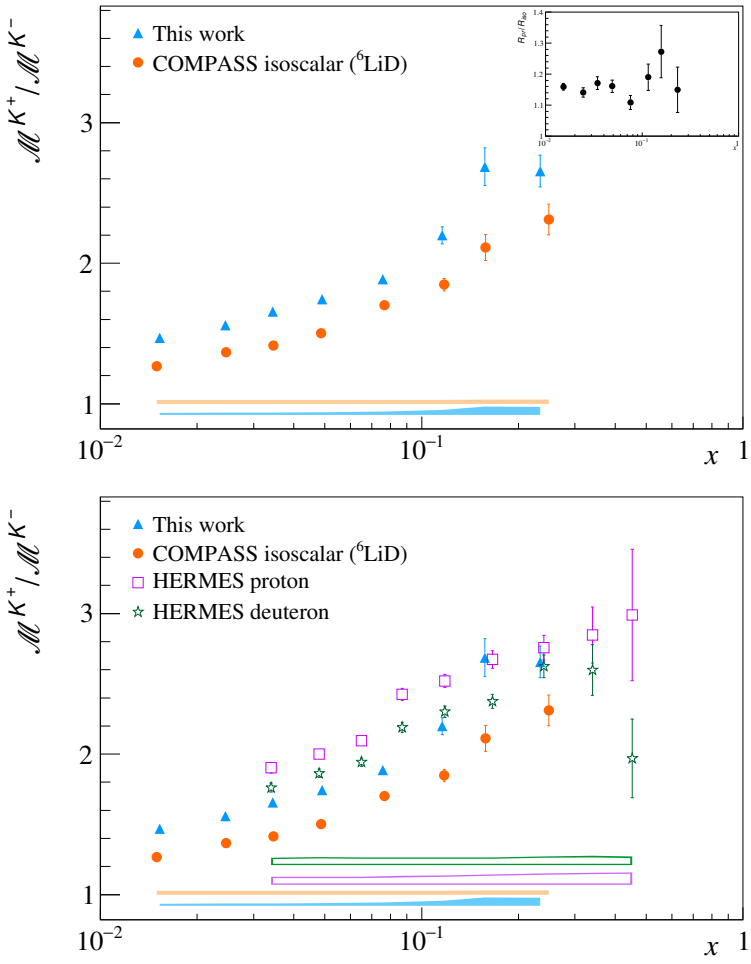


Figure 111: Same as Fig. 110 for $\frac{\mathcal{M}^{K^+}}{\mathcal{M}^{K^-}}$.

While the pion multiplicities were found to be well described both in LO and NLO pQCD, this was not the case for kaon multiplicities. The region of large z appears to be problematic for kaons. Investigations on these subjects were conducted by the COMPASS collaboration on the kaon ratio versus z for large z [118] and tensions with pQCD prediction were observed. Moreover a strong ν dependence of the ratio was found and this dependence may explain the discrepancy to HERMES. HERMES data points are generally obtained at lower $\langle\nu\rangle$ than in COMPASS. When comparing the few data points with the exact same kinematics, the results for kaon multiplicities ratio are agreeing. As for the tension with pQCD prediction, at high-z ($z > 0.75$) a dependence of the ratio on the missing mass parameter $M_X = \sqrt{M_p^2 + 2M_p\nu(1-z) - Q^2(1-z)^2}$ was found. Thus, the ratio is in fact a function of ν and z. This points to the need of a correction within the pQCD formalism to take into account the phase-space available for the hadronisation of the target remnants. For other experiments using lower beam energies than COMPASS deviations may even appear at lower z. The ratio of proton over antiproton is agreeing within uncertainties with the EMC results and shows a major improvement in the precision of the measurement of these multiplicities.

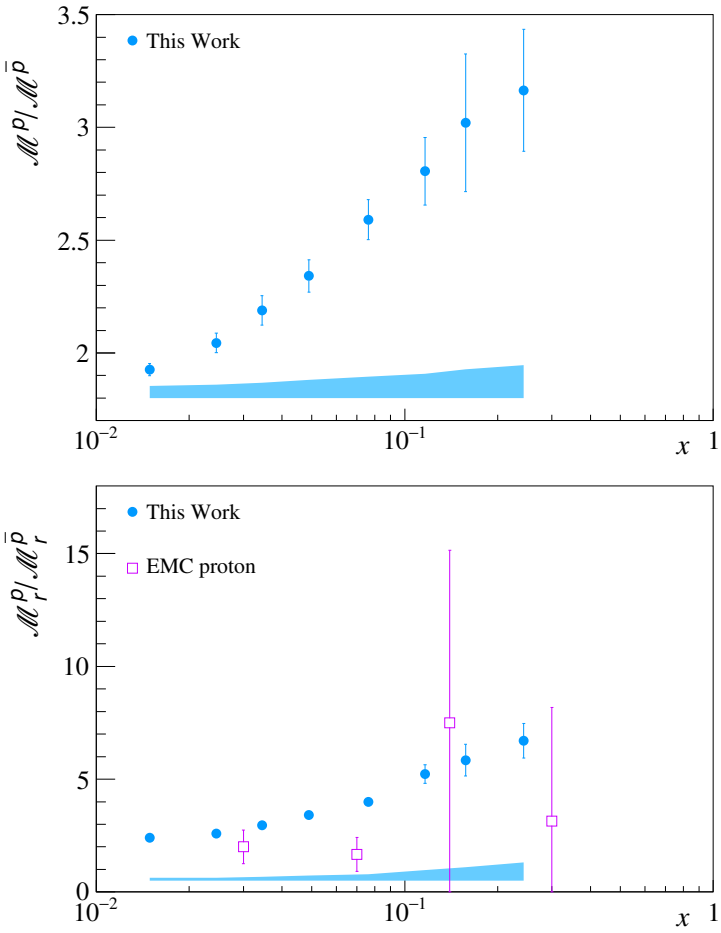


Figure 112: Ratio of $\frac{\mathscr{M}^p}{\mathscr{M}^{\bar{p}}}$ from COMPASS for a proton target (blue closed points) and from EMC for a proton target (violet open points).

12.3.2 *Sum of charged hadron multiplicities*

In Figs. 113 to 116, $\mathscr{M}^{h^+} + \mathscr{M}^{h^-}$ from COMPASS is depicted for a proton target (blue triangle) and for an isoscalar target (orange circles). The same sum from HERMES is presented for a proton target (violet open squares) and a deuteron target (green open stars) for charged pions and charged kaons.

For a proton target, the charged pion multiplicities integrated over z can be expressed at LO pQCD as [119]:

$$\left(\mathscr{M}^{\pi^+} + \mathscr{M}^{\pi^-}\right)_p = \mathscr{D}^{\pi}_{fav} + \mathscr{D}^{\pi}_{unf} - \frac{S}{U+D+S}\left(\mathscr{D}^{\pi}_{fav} - \mathscr{D}^{\pi}_{unf}\right), \tag{121}$$

where $U = 4u + 4\bar{u}$, $D = d + \bar{d}$, $S = us + \bar{s}$ and $\mathscr{D}^K(Q^2) = \int D^K(z, Q^2)dz$. As $\frac{S}{U+D+S}$ is small and the Q^2 dependence of $\mathscr{D}^{\pi}_{fav} + \mathscr{D}^{\pi}_{unf}$ is weak, the pion multiplicity sum is expected to be almost flat. The same reasoning can be done for the isoscalar case [119]:

$$\left(\mathscr{M}^{\pi^+} + \mathscr{M}^{\pi^-}\right)_d = \mathscr{D}^{\pi}_{fav} + \mathscr{D}^{\pi}_{unf} - \frac{2S}{5U' + 2S}\left(\mathscr{D}^{\pi}_{fav} - \mathscr{D}^{\pi}_{unf}\right), \tag{122}$$

where $U' = u + \bar{u} + d + \bar{d}$. The expression of the sum for protons is the same than for pions for both targets. Thus the same conclusions can be made.

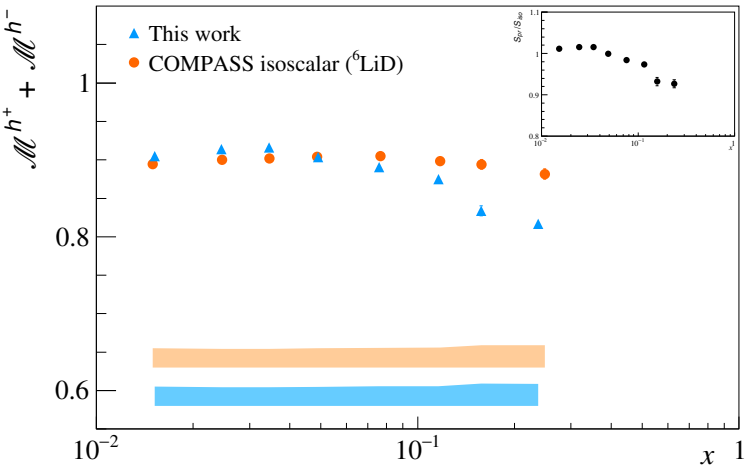


Figure 113: Sum of $\mathscr{M}^{h^+} + \mathscr{M}^{h^-}$ from COMPASS for a proton target (blue closed points) and an isoscalar target (orange closed points). On the top right is displayed the ratio of proton target result over isoscalar target result.

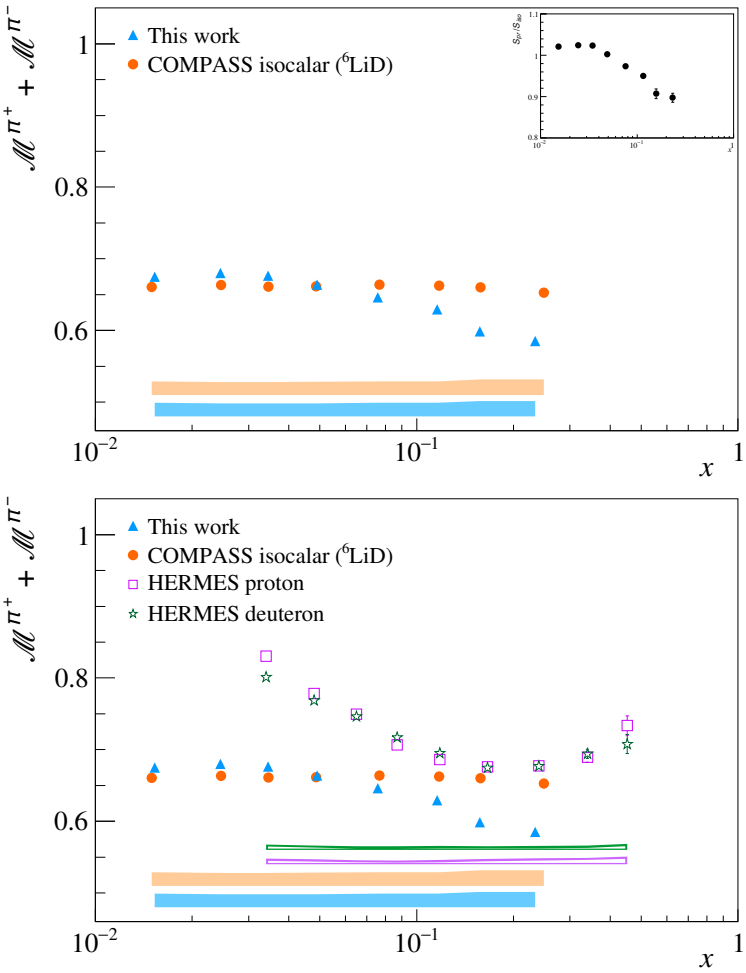


Figure 114: Sum of $\mathscr{M}^{\pi^+} + \mathscr{M}^{\pi^-}$ from COMPASS for a proton target (blue closed points) and an isoscalar target (orange closed points) and from HERMES for a proton target (violet open points) and a deuteron target (green open points). On the top right is displayed the ratio of proton target result over isoscalar target result.

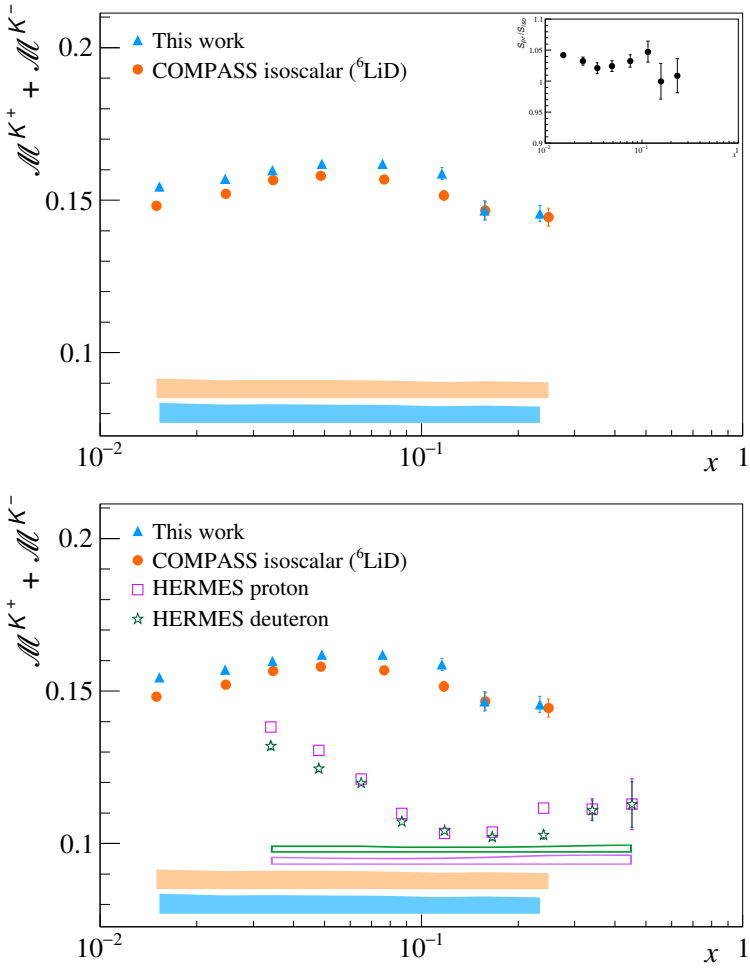


Figure 115: Same as Fig. 114 for $\mathscr{M}^{K^+} + \mathscr{M}^{K^-}$.

Similarly for kaons, for a proton target [119]:

$$\left(\mathscr{M}^{K^+} + \mathscr{M}^{K^-}\right)_p = \mathscr{D}^K_{fav} + \mathscr{D}^K_{unf} + \left[\frac{(s+\bar{s})\left(\mathscr{D}^K_{str} - \mathscr{D}^K_{fav}\right) + (d+\bar{d})\left(\mathscr{D}^K_{unf} - \mathscr{D}^K_{fav}\right)}{4(u+\bar{u}) + d + \bar{d} + s + \bar{s}}\right]. \tag{123}$$

At high value of x, the sea content of the nucleon can be neglected:

$$\left(\mathscr{M}^{K^+} + \mathscr{M}^{K^-}\right)_p = \mathscr{D}^K_{fav} + \mathscr{D}^K_{unf} + \left[\frac{d\left(\mathscr{D}^K_{unf} - \mathscr{D}^K_{fav}\right)}{4u + d}\right], \tag{124}$$

and taking as approximation $u = 2d$:

$$\left(\mathscr{M}^{K^+} + \mathscr{M}^{K^-}\right)_p = \frac{8\mathscr{D}^K_{fav} + 10\mathscr{D}^K_{unf}}{9}. \tag{125}$$

For the isoscalar case [119]:

$$\left(\mathscr{M}^{K^+} + \mathscr{M}^{K^-}\right)_d = \frac{U\mathscr{D}^K_U + S\mathscr{D}^K_S}{5U + 2S}, \tag{126}$$

At high value of x, the strange content of the nucleon can be neglected:

$$\left(\mathscr{M}^{K^+} + \mathscr{M}^{K^-}\right)_d = \frac{\mathscr{D}^K_U}{5}. \tag{127}$$

From the proton target result for $\mathscr{M}^{K^+} + \mathscr{M}^{K^-}$ at high x ($x = 0.25$) we extract $8\mathscr{D}^K_{fav} + 10\mathscr{D}^K_{unf} \approx 1.22 - 1.30$. This differs from the earlier DSS fit result at $Q^2 = 3\ (\mathrm{GeV}/c)^2$,

$8\mathscr{D}^K_{fav} + 10\mathscr{D}^K_{unf} \approx 0.35$ [55], which was mainly based on HERMES results. The same kind of result was already found for an isoscalar target, where $\mathscr{M}^{K^+} + \mathscr{M}^{K^-}$ at high x ($x = 0.25$) we extract $D^K_U \approx 0.65 - 0.70$ and differs from the earlier DSS fit result at $Q^2 = 3$ (GeV/c)^2, $D^K_U \approx 0.43$ [55].

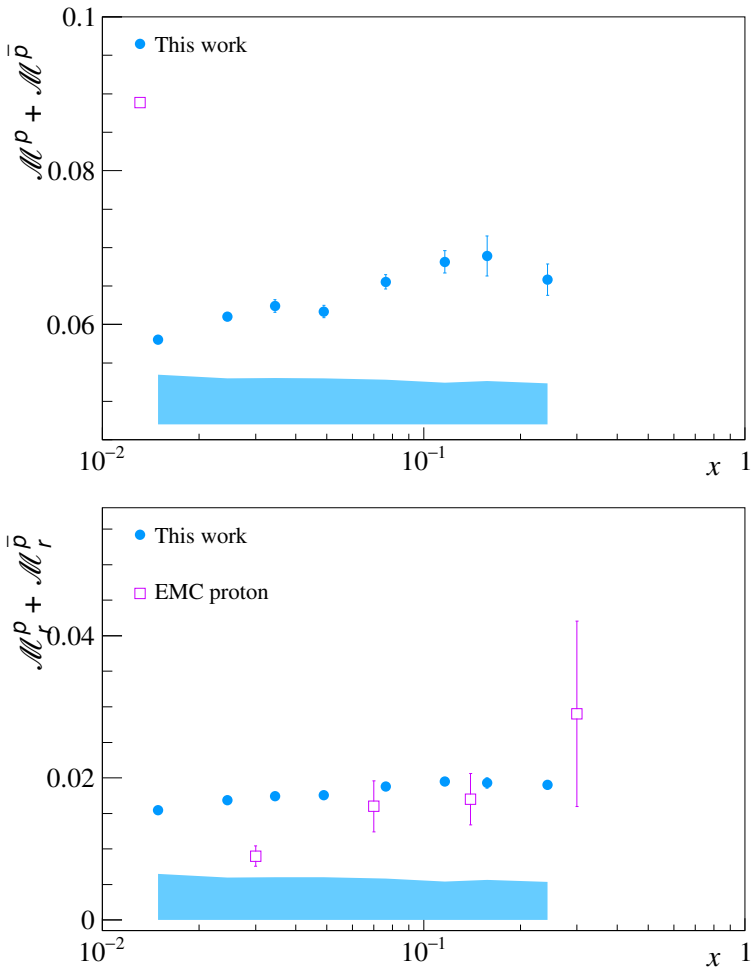


Figure 116: Sum of $\mathscr{M}^p + \mathscr{M}^{\bar{p}}$ from COMPASS for a proton target (blue closed points) and from EMC for a proton target (violet open points).

The sum of charged pion (hence charged hadron as hadrons are mostly pions) multiplicity for a proton target should lie at the same level than COMPASS results for an isoscalar target, while for charged kaon the result for a proton target should be slightly (~ 5%) above results on isoscalar target. These expectations are obtained by evaluating the multiplicities for both targets with Eqs. 121 to 127 taking DSS07 [55] LO fragmentation functions for hadrons and integrate them over z from 0.2 to 0.85 and taking PDFs from MSTW08 [111]. The COMPASS pion and hadron results for an isoscalar target have been moved up by 6% compared to the data from the publication in order to take into account an improvement on the knowledge of the radiative corrections since the publication of the results. At high x for the hadron and pion results on proton target, the sum should be flat as for the results on isoscalar target. As discussed in Chapter 11, the MC description of the data has to be improved to better simulate the θ_h dependence of the data. The discrepancy between COMPASS results and HERMES results for pions and kaons, both for a proton and deuteron targets, has to be noted. They lie well above the COMPASS points for pions and well below the COMPASS points for kaons and exhibit a different x behaviour. Towards low x, COMPASS data show a flat behaviour, unlike the rise that is suggested by the HERMES data. Though this discrepancy cannot be explained by the mean Q^2 of multiplicity sets (of $\sim 3-5$ (GeV/c)^2 for COMPASS data and

$\sim$ 1.5 $(\mathrm{GeV}/c)^2$ for HERMES) as there is only a very weak Q^2 evolution of the multiplicity sum. A possible explanation for kaons comes from the fact that HERMES was operating at a lower ν than in COMPASS and it was found that there might be phase space limitation effects which may be larger in the case of HERMES than in COMPASS, while it is already seen for $z > 0.75$ data for COMPASS [118]. Another explanation to reconcile the results is to look for hadron mass correction [120], which reduce the apparent large discrepancy between the measurements for kaons. HERMES points lie well below the COMPASS points and exhibit a different x behaviour. The sum of proton/antiproton is agreeing within uncertainties with the EMC results.

12.4 SUMMARY

The final unidentified hadron ($h^{\pm}$), pion ($\pi^{\pm}$), kaon ($K^{\pm}$) and proton/antiproton ($p/\bar{p}$) multiplicities extracted from 2016 COMPASS data of muon deep inelastic scattering on a pure proton (lH_2) target were presented as a function of z and in bins of x and y. Averaging these results over y, two dimensional projection are obtained. The subsequent integration over z allows the comparison of the sum and ratio of charged hadron multiplicities with other experiments results like HERMES [47]. Comparisons are also made with COMPASS results for an isoscalar target [44, 45]. For the ratio, the results are in agreement with the expectations. The discrepancy that was already observed for the kaon ratio between COMPASS and HERMES for a deuteron/isoscalar target is also present for a proton target. For the sum, the results for hadrons and pions differs from the expectations in the high x region. Overall, all the results for a proton target for the sum seem to suffer from a drop at high x, particularly visible on hadrons and pions, which should be flat. The kaons seem to be less harmed by this issue. The cause of this drop is still under investigation. A large discrepancy with HERMES results is observed for the sum in the shape for both pions and kaons.

13

DETERMINATION OF QUARK FRAGMENTATION FUNCTIONS INTO KAONS

In the previous chapter, we obtained charged kaon multiplicities $M^K(x, Q^2, z)$. As these multiplicities can be expressed as a combination of the PDFs $q(x, Q^2)$ and the FFs $D_q^K(z, Q^2)$, assuming that the PDFs are known, they can be used to extract the quark fragmentation functions into kaon.

13.1 LO QCD FIT OF KAON MULTIPLICITIES

D^K can be obtained from a pQCD fit to the existing multiplicities. While DSEHS or JAM are doing such studies in NLO, in COMPASS only a LO QCD code is available. The extraction of FFs from the measured $K^\pm$ multiplicities is performed in a LO pQCD fashion according to Eq. 129:

$$\frac{dM^h(x, Q^2, z)}{dz} = \frac{d^3\sigma^h(x, Q^2, z)/dxdQ^2dz}{d^3\sigma^{DIS}(x, Q^2, z)/dxdQ^2} = \frac{\sum_q e_q^2(x, Q^2)D_q^h(z, Q^2)}{\sum_q e_q^2 q(x, Q^2)}, \tag{128}$$

with $q(x, Q^2)$ are the quark PDFs for the flavour q. Using a standard χ^2 minimization over the data points with the minimization framework Minuit2 of the ROOT package:

$$\chi^2 = \sum_j \frac{\left[T_j\left(x_j, Q_j^2, z_j\right) - M_j\left(x_j, Q_j^2, z_j\right)\right]^2}{\sigma_j^2}, \tag{129}$$

where M_j are the measured multiplicities and σ_j^2 are the quadratic sum of the statistical errors. T_j are the multiplicities evaluated for a set of parameters for a given parametrisation. Due to the large Q^2 span covered by the data, it is mandatory to properly take into account the Q^2 dependence of FFs and PDFs. Thus the T_j are evaluated at $Q_0^2 = 1$ (GeV/c)^2 and evolved to their actual Q_j^2 using the LO DGLAP Q^2 evolution code provided by M. Hirai and S. Kumano [49].

As described in Chapter 1, four independent FFs (see Eq. 27) are extracted: D_{fav}^K, D_{unf}^K, D_s^K and D_g^K. For the current analysis, the following parametrization, similar to the one from DSEHS described in Chapter 1 is used:

$$\begin{aligned} zD_i(z, Q_0^2) &= \frac{N_i z^{\alpha_i}(1-z)^{\beta_i}(1+\gamma_i(1-z)^{\delta_i})}{\int_{0.2}^{0.85} z'^{\alpha_i}(1-z')^{\beta_i}(1+\gamma_i(1-z')^{\delta_i}dz'}, \; i = \{fav\} \\ zD_i(z, Q_0^2) &= \frac{N_i z^{\alpha_i}(1-z)^{\beta_i}}{\int_{0.2}^{0.85} z'^{\alpha_i}(1-z')^{\beta_i}dz'}, \; i = \{s, unf, g\}. \end{aligned} \tag{130}$$

The PDFs set from MSTW08 in LO pQCD [111], stored by the LHAPDF data group [121] was chosen.

13.2 UNCERTAINTIES CALCULATION

The statistical and systematic uncertainties of the extracted FFs are determined by using a bootstrap method [122]. Therefore a number of resamples of the original multiplicity data

are constructed. For each resample the data points are fluctuated proportional to their uncertainty, resulting in a set of data points:

$$M'_j = M_j + R \cdot \sigma_j, \tag{131}$$

where M_j and σ_j are the original value and the corresponding uncertainty of data point j, respectively. The factor R is a Gaussian randomly distributed value in the range $[-\infty, \infty]$. Its determination is differently treated for statistical and systematic uncertainties. The statistical uncertainties are uncorrelated and the deviation of each data point j is calculated with an individual random value R_j. Thus, for statistical uncertainties, Eq. 133 can be written as:

$$M'_{j,stat} = M_j + R_j \cdot \sigma_{j,stat}. \tag{132}$$

For systematic uncertainties, for each resample only one random value R' of the standard Gaussian distribution is generated:

$$M'_{j,sys} = M_j + R' \cdot \sigma_{j,sys}. \tag{133}$$

For both statistical and systematic uncertainties, a set of 100 resamples is created. Then for each resample the fit is performed as it is done for the original data. To calculate the uncertainty bands, for each z bin, the mean and the RMS values are calculated from the resulting FFs. The uncertainties are centered around the mean value and have the width of the corresponding RMS value.

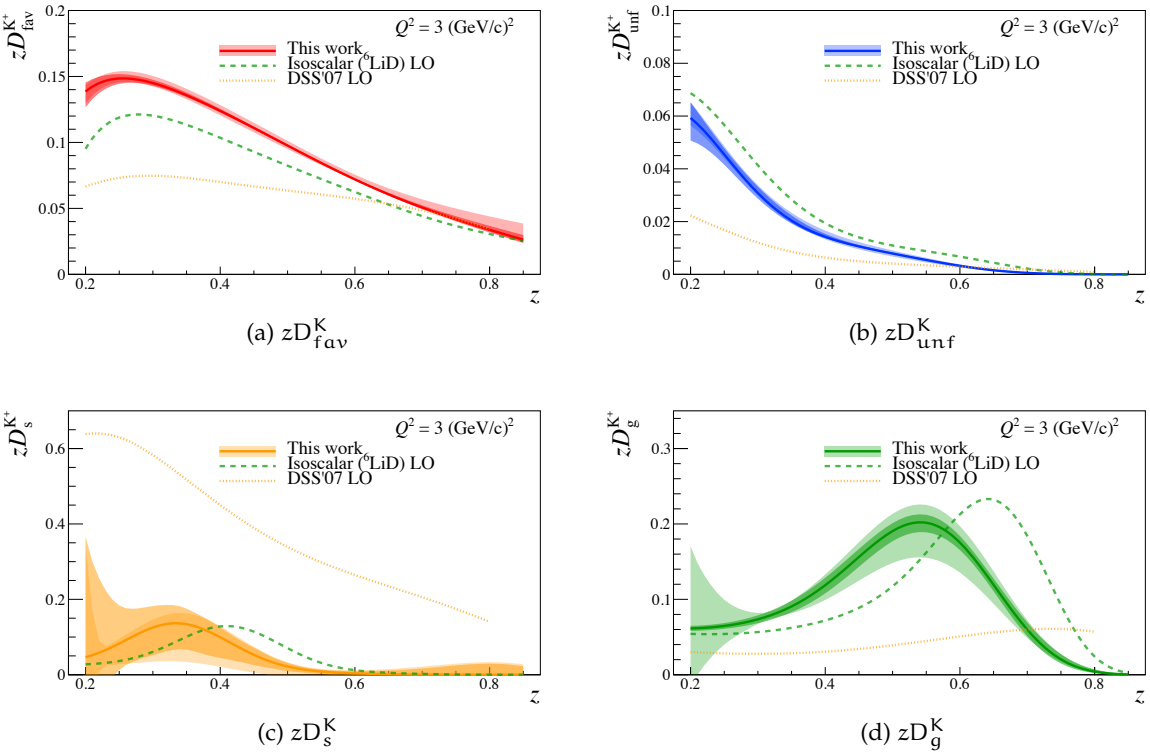
13.3 KAON FRAGMENTATION FUNCTIONS

The final parameters of the fit are displayed in Table 16. The χ^2 per degrees of freedom for the results is 3.5. A similar fit has been performed with COMPASS data on isoscalar target (see Appendix D). In Fig. 117, the four FFs are shown. The values obtained are significantly higher for D^K_{fav} and D^K_{unf} compared to DSS'07 LO results and in agreement with the LO fit of COMPASS data on isoscalar target, while for D^K_{str} the values are much smaller than DSS'07 LO for both sets. This needs a more detailed investigation and also a fit in NLO. The reason I chose the DSS07 fit and not a newer one is that this is the only fit of DSS/DSEHS that is done at LO and it makes more sense to compare LO order extraction than for example LO versus NLO. The fragmentation functions are calculted up to $z = 0.85$ and in the previous chapter, we saw that pQCD is not describing the ratio of charged kaons for z above 0.7. In this thesis it was assumed that this region is not contraining much the fit hence it should not be a problem but future work should perform a close examination.

These results point out that there is a bad sensitivity of these proton target measurements to the strange quark. In fact these measurement are u dominated and go down to $x \sim 10^{-2}$ only, thus is not probing far into the sea. There is not enough sensitivity to the three fragmentation functions to kaons favoured, unfavoured and strange (the gluon FF is not considered as in LO it only comes from the DGLAP evolution). A more refined analysis of COMPASS plus HERMES data taking into account our findings on the multiplicity ratio is needed. An extension of this study would be to do a K^0 analysis, which would bring a sample with an extended kinematic range and with independent equation, to study whether it improves the fit. For COMPASS data on isoscalar target such study was conducted but was not conclusive as the K^0 sample was not good enough due to reinteractions in the isoscalar target. In the pure proton target there are no such reinteractions, hence the K^0 sample needs smaller corrections.

	N	α	β	γ	δ
D^K_{fav}	0.0647 $\pm$ 0.0007	-1.2 $\pm$ 0.2	0.46 $\pm$ 0.07	-1.6 $\pm$ 0.7	3.4 $\pm$ 0.2
D^K_s	0.03 $\pm$ 0.01	14 $\pm$ 7	27 $\pm$ 10	-	-
D^K_{unf}	0.005 $\pm$ 2	4 $\pm$ 5	19 $\pm$ 10	-	-
D^K_{glu}	0.08 $\pm$ 0.1	16 (fixed)	11 (fixed)	-	-

Table 16: Fit parameters for $Q_0^2 = 1$ (GeV/c)^2. The associated χ^2 is of 3.5.



(a) zD^K_{fav} (b) zD^K_{unf}

(c) zD^K_s (d) zD^K_g

Figure 117: The favoured (top left), unfavoured (top right), strange (bottom left) and gluon (bottom right) quark FFs $zD(z)$ into kaons from the COMPASS LO fit. The fit is done based on both the statistical and systematic errors. The green dashed lines are from the same COMPASS LO fit but with COMPASS results for an isoscalar target. The orange dashed line is DSS07 LO fit.

14 SUMMARY AND CONCLUSION

In order to improve the knowledge on quark fragmentation into charged hadrons, the production of charged hadrons in deep inelastic scattering on a pure proton target (lH_2) was studied in a semi-inclusive measurement.

The multiplicities of charged unidentified hadrons, pions, kaons and protons have been determined in a binning of three kinematic variables: x, the fraction of momentum of the nucleon held by the struck quark, y, the fraction of energy of the incoming lepton held by the virtual photon and z, the fraction of energy of the virtual photon transferred to the observed hadron h. The measurement was performed using 5 periods (weeks) of 2016 COMPASS data ($\sim 8.3 \times 10^6$ DIS events and $\sim 2.6 \times 10^6$ hadrons). The kinematic domain covered by these data is the following: $Q^2 > 1$ (GeV/c)^2, $y \in [0.1, 0.7]$, $x \in [0.004, 0.4]$, $W \in [5, 17]$ GeV/c^2, and $z \in [0.2, 0.85]$.

The hadron identification as pion, kaon and proton was provided by the Ring Imaging CHerenkov (RICH) detector. The performance of this detector was determined using the same data used in the aforementioned analysis. Having a good knowledge of the RICH response is mandatory to calculate pion, kaon and proton multiplicities. Due to the usage of the RICH, the hadron momentum range is restricted from 12 GeV/c to 40 GeV/c and the polar angle in the range from 0.01 to 0.12 rad in order to operate in a region where the RICH can fully discriminate pions, kaons and protons. With the lower cut in hadron momentum it is ensured that the particle momenta are well above the kaon identification threshold (about 9.45 GeV), while with the upper limit regions are avoided, where effects arising from saturation ($\beta \to 1$) begin to appear, in order to ensure a good charged hadron separation. The polar angle range is chosen, so that the RICH efficiencies are generally high and precisely measured. In this phase-space, high identification efficiencies are found for pions ($> 97\%$), kaons ($> 95\%$) and protons ($> 90\%$) with low misidentifications probabilities ($< 5\%$ for pions, $< 7\%$ for kaons, $< 10\%$ for protons).

The hadrons multiplicities were also corrected for the geometric limitations of the spectrometer, detector performance and the data reconstruction efficiency. The global acceptance correction is estimated using a MC simulation. A new MC generator, DJANGOH [92, 101], was applied both for generating radiative events in the MC simulation and computing radiative correction to the multiplicities in a three dimensional binning in x, y and z. For the first time we have a solid determination of radiative effects dependent on the z variable. DJANGOH gives compatible results when compared to radiative corrections obtained with TERAD (analytic calculation) and gives solid results about the electron production from photons. The contribution to the hadron yield from the vector meson production was also estimated.

The sum and ratio of charged hadrons are of special interest as they are integrated quantities and can be compared to results from other experiments. For the ratios of multiplicities, the results from COMPASS for a proton target are as expected with respect to the results from COMPASS for an isoscalar target [44, 45]. A discrepancy is found with results from HERMES [47] for kaons, confirming the discrepancy already found for deuteron/isoscalar targets, but are explained by the different kinematics of the data points of the two experiments [118]. For the sums of multiplicities, in the x region not affected by a possible acceptance problem, the results from COMPASS for a proton target are compatible with the LO pQCD expections. Again, the COMPASS results differ with the HERMES ones on this quantity for both pions and kaons, which might be explained by the fact that HERMES operates at lower ν than

COMPASS and possible hadron mass corrections [118, 120]. The concept of FFs is then not applicable in part of the phase space [118].

The K^+ and K^- multiplicities were used to extract the favoured D^K_{fav}, unfavoured D^K_{unf} and strange D^K_s quark fragmentation with a fit at LO, assuming the PDFs known. The result of the fit points out that there is a bad sensitivity to the strange quark of these measurements. The fit gives too much contribution to the favoured and unfavoured fragmentation functions at the expense of the strange fragmentation function.

The pion, kaon and proton multiplicity sets obtained in this analysis, which represent in total more than 1800 data points, are a major input for the global fit of world data done at NLO. The kaon multiplicity set is particularly awaited as it will enlarge significantly the available data set [123]. The proton multiplicity set is also a novelty which will interest the fitters.

This analysis will be pursued inside COMPASS. The new RICH calibration will improve the identification efficiency of charged hadrons at high momenta, while the inclusion of more data from 2016 and data from 2017 will improve the statistics. Further study on the Monte Carlo and the comparison between data and Monte-Carlo will improve the systematics. As soon as the problem of the drop in x of the charged pion and charged hadron sums is solved, the extraction of the pion fragmentation functions can be done. In contrast to the kaons, the extraction of the pion fragmentation functions with a LO fit of the pion multiplicities is more likely to converge as the measurement have a good sensitivity of the u and d quark, hence the favoured and unfavoured pion fragmentation functions. For kaons, the extraction of K^0 multiplicities from data could help constrain the kaon fragmentation function fit. This analysis has already started. In addition to the data taken in 2016 and 2017, more data will be taken in 2021 on deuteron target. Eventually as it was shown that our kaon and proton multiplicity ratios at z above 0.7 are not described by pQCD [118], it would be interesting to take it into account in our analysis and further investigate this kinematic region.

Part V

APPENDIX

A

RICH EFFICIENCIES

In this appendix the results for the RICH particle identification efficiency are shown in Figs. 118 to 121 for π^-, K^-, p and $\bar{p}$ for the various particle types and charges. In each figure, the momentum dependence for the different angular bins is shown. The efficiencies are weakly dependent on the angle, while it is more strongly correlated with the momentum, especially in the region near the threshold.

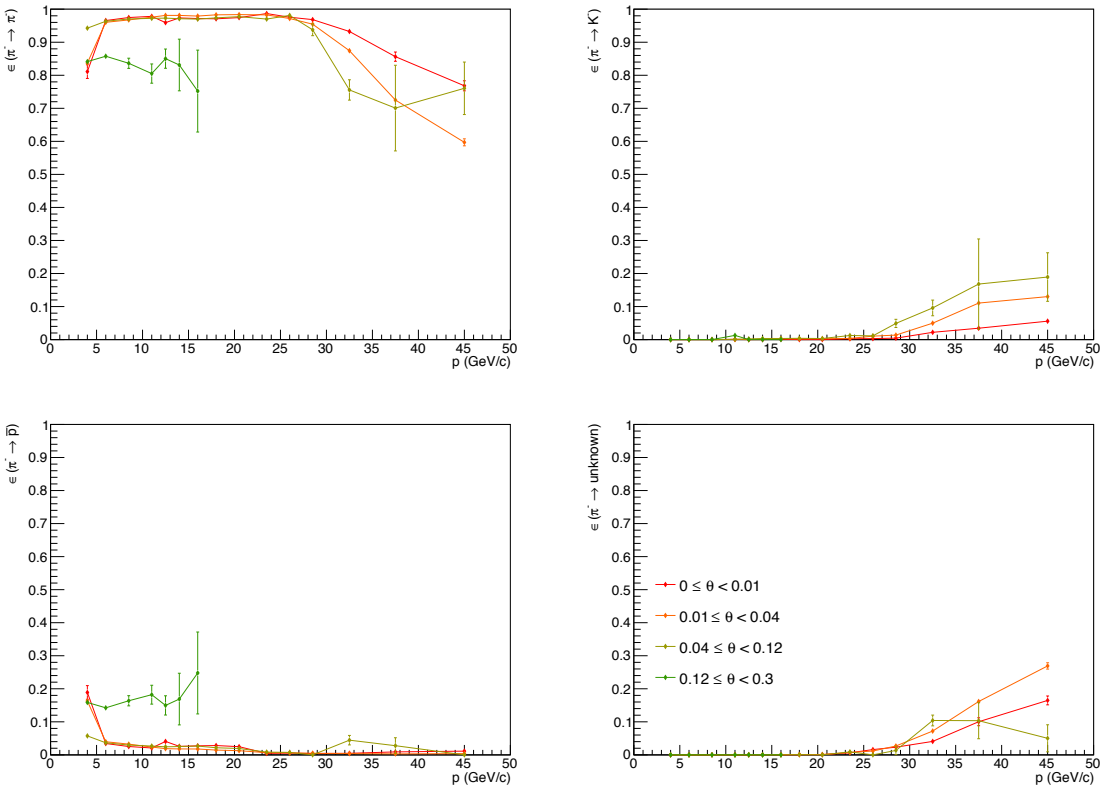


Figure 118: Identification probabilities $\epsilon(p \to j)$ for π^-.

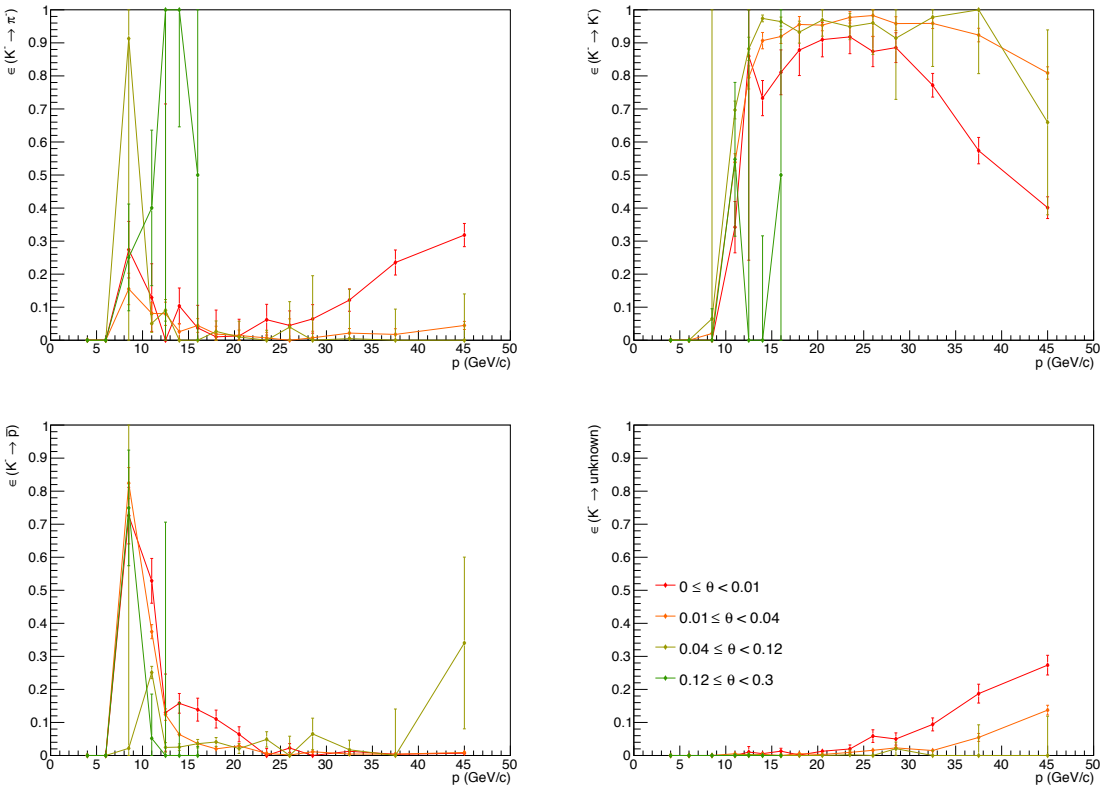


Figure 119: Identification probabilities $\epsilon(p \to j)$ for K^-.

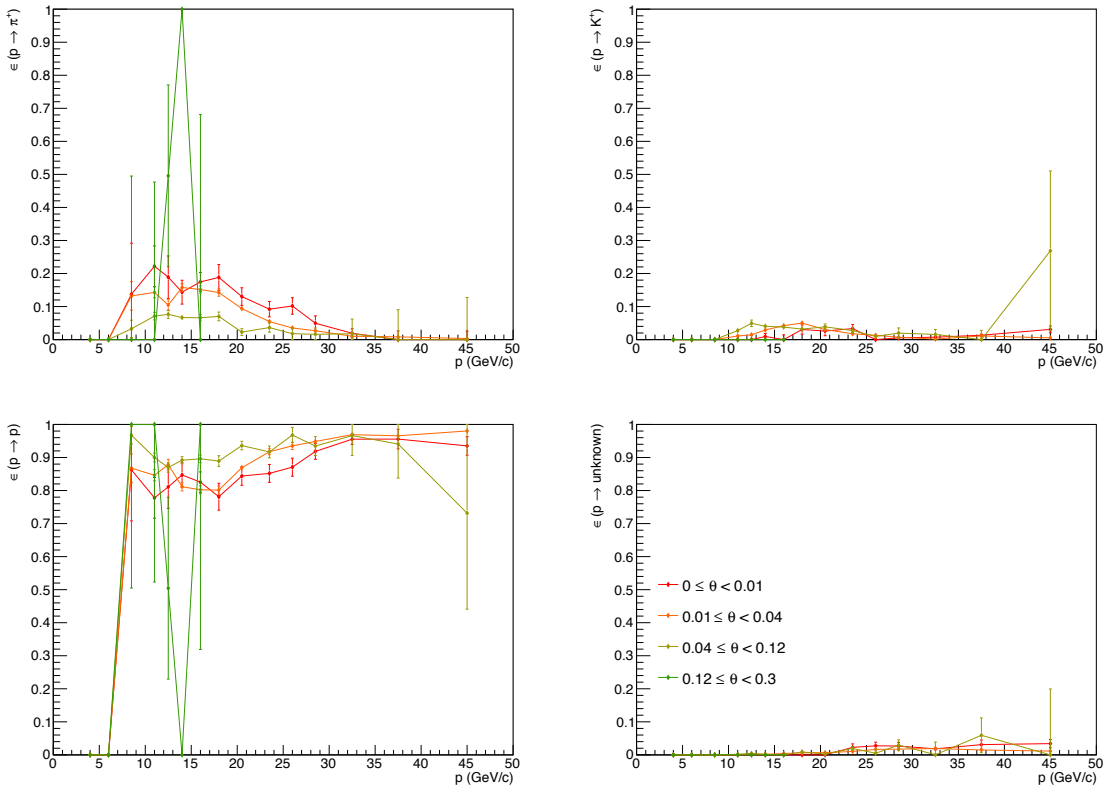


Figure 120: Identification probabilities $\epsilon(p \to j)$ for p.

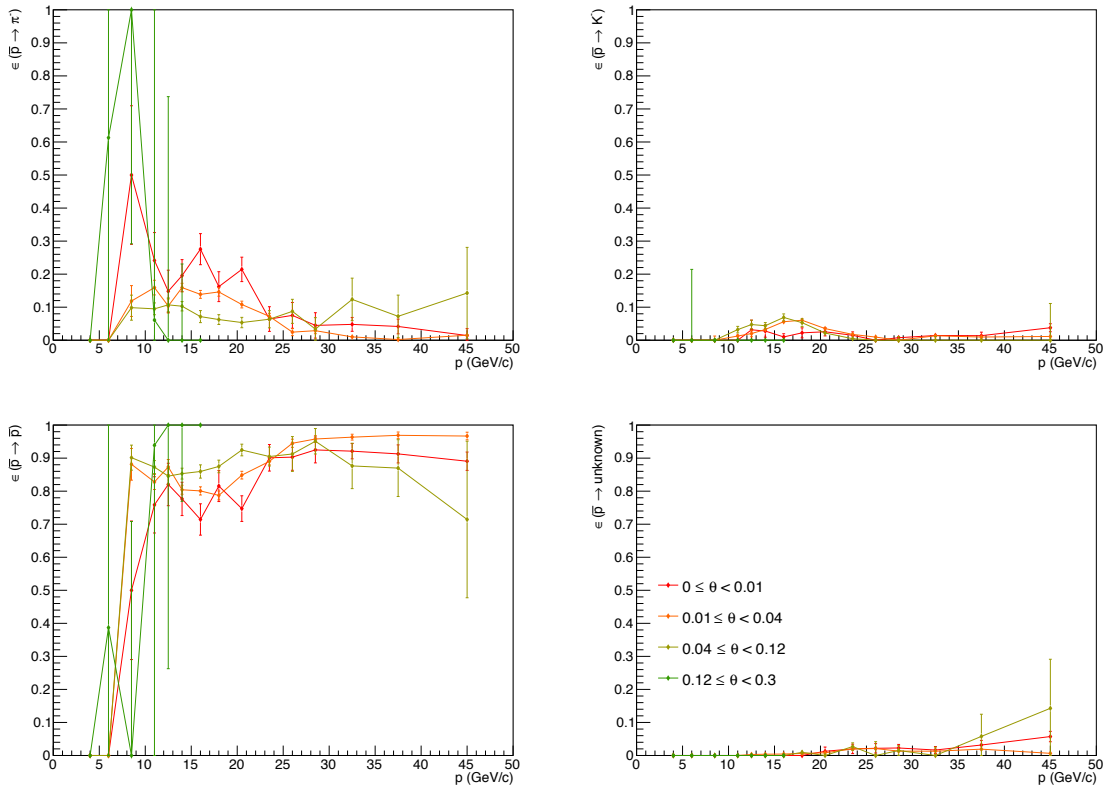


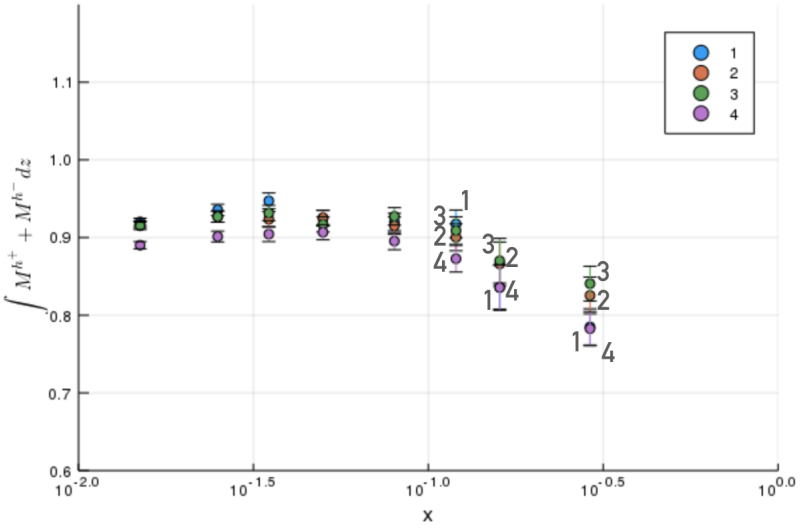
Figure 121: Identification probabilities $\epsilon(p \to j)$ for $\bar{p}$.

B

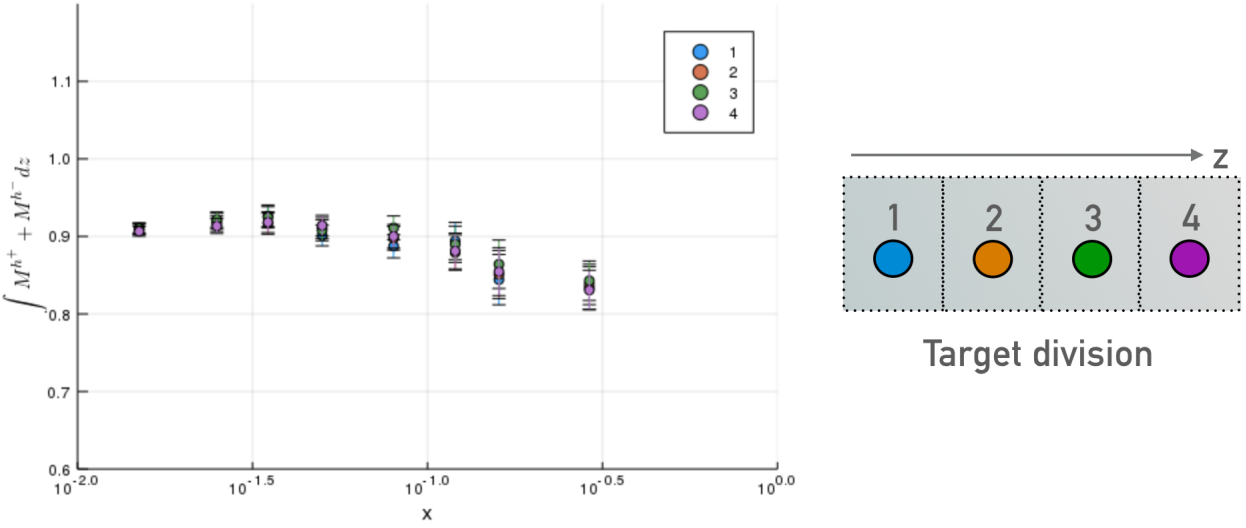
RAW MULTIPLICITY ANALYSIS

B.1 EFFECT OF THE RESCUE PROCEDURE ON THE MULTIPLICITY SUM

An other way to look at the effect of the rescue procedure is through the multiplicity sum. Dividing the multiplicities in four bins of target, the multiplicity sum should be the same in all bins of target. Before the rescue procedure one can see that this is not the case as the multiplicity sum for the last part of the target is consistently smaller as seen in Fig. 122. After the rescue procedure however, the result is as expected.



(a)



(b)

Figure 122: Comparison of the multiplicity sum before the application of the rescue procedure (a) and after (b) for different target slices.

B.2 RESULTS FOR RAW MULTIPLICITIES OF IDENTIFIED HADRONS

The raw multiplicity results shown in this section are without any correction except the RICH unfolding correction for identified hadrons. The unidentified hadron multiplicities are displayed as a function of z in bins of x and staggered vertically with y in Figs. 123 to 128. The charged hadron multiplicities strongly depends on z as expected with a small dependence with x also.

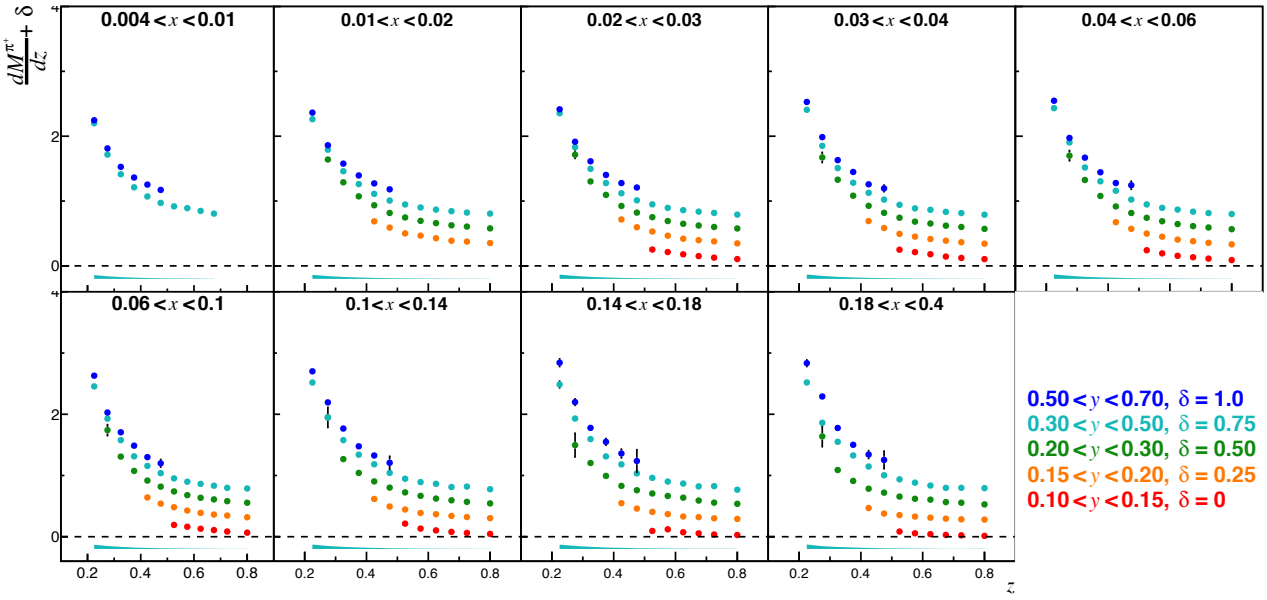


Figure 123: Same as Fig. 79 but for positive pions.

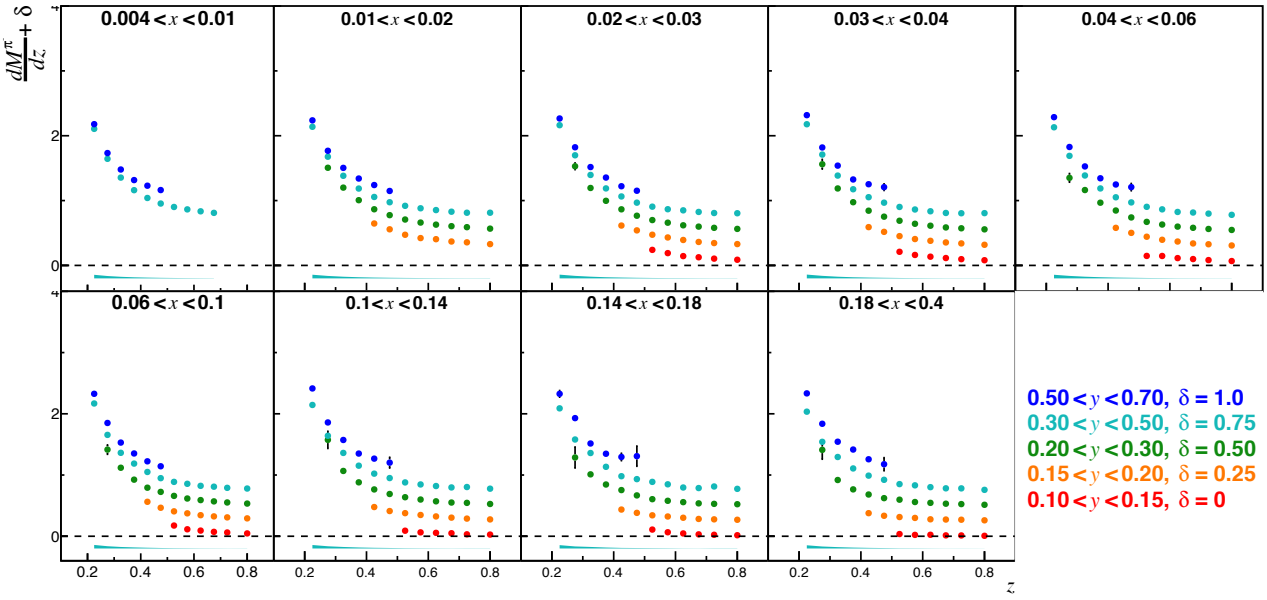


Figure 124: Same as Fig. 79 but for negative pions.

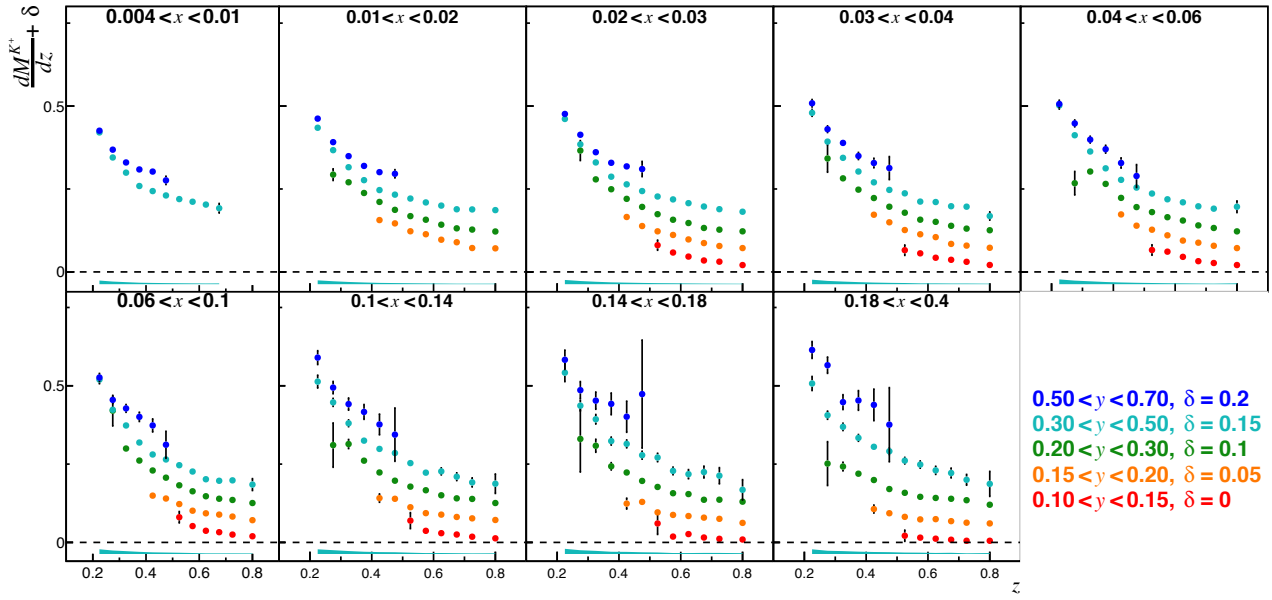


Figure 125: Same as Fig. 79 but for positive kaons.

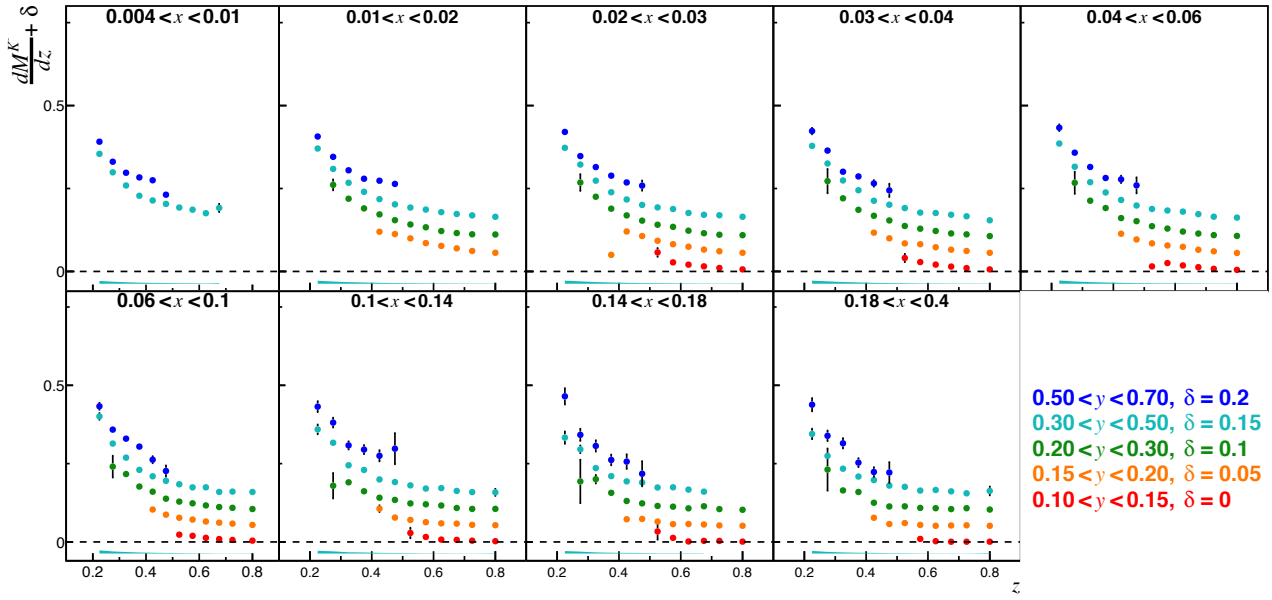


Figure 126: Same as Fig. 79 but for negative pions.

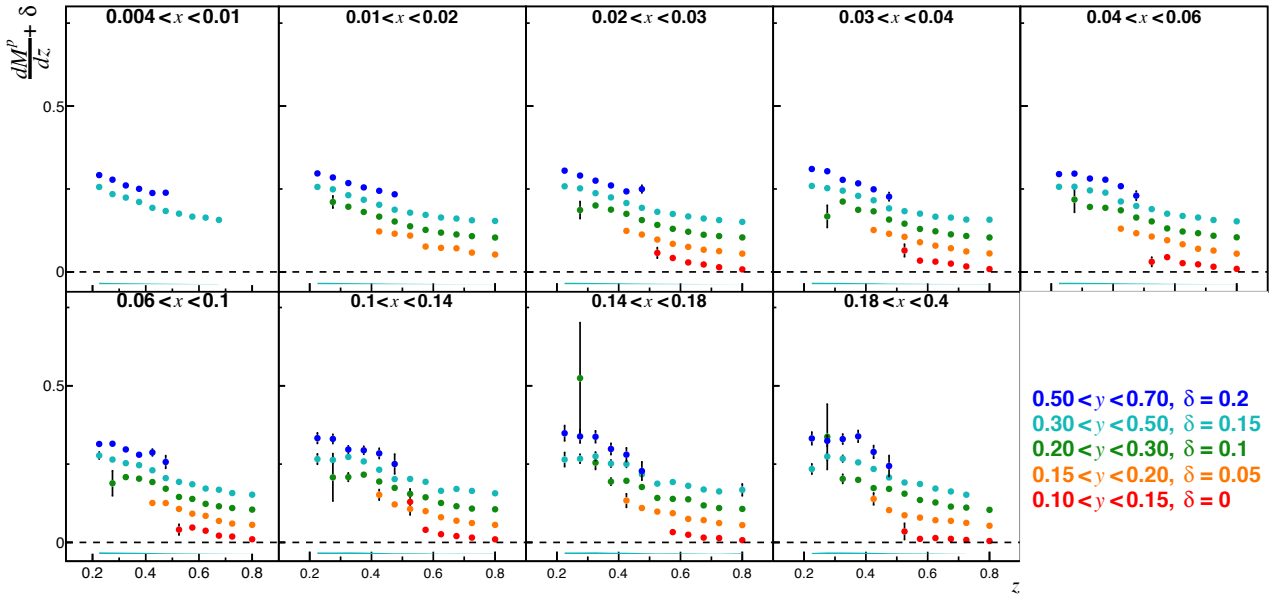


Figure 127: Same as Fig. 79 but for protons.

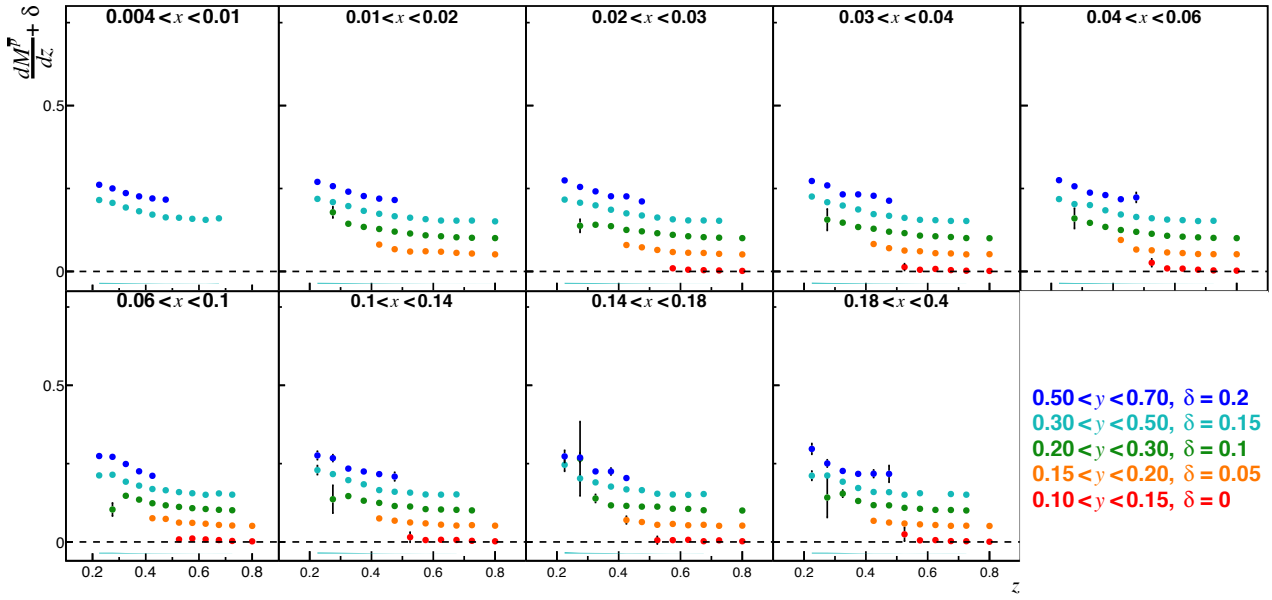


Figure 128: Same as Fig. 79 but for antiprotons.

C

SYSTEMATIC STUDIES

C.1 SYSTEMATIC UNCERTAINTY ASSOCIATED TO THE STABILITY OF DATA OVER TIME

The data samples used in the analysis were recorded over a period of 5 weeks. As a quality check, the raw charged hadron multiplicities from period P07 and from all periods (averaged using the flux as weight) were compared. The results of this study are presented in Figs. 129 to 131 for identified hadrons.

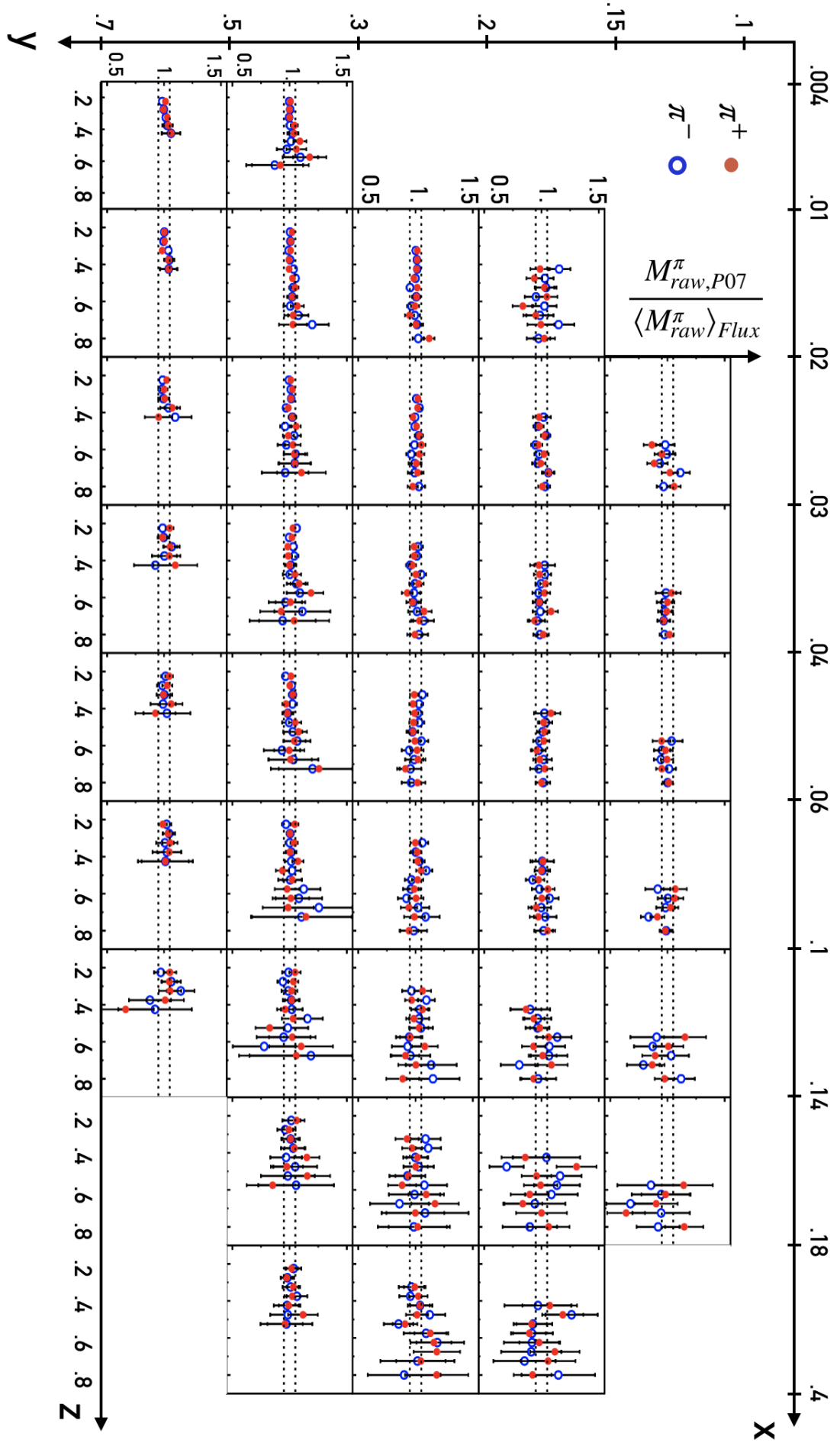


Figure 129: Same as Fig. 94 but for pion multiplicities.

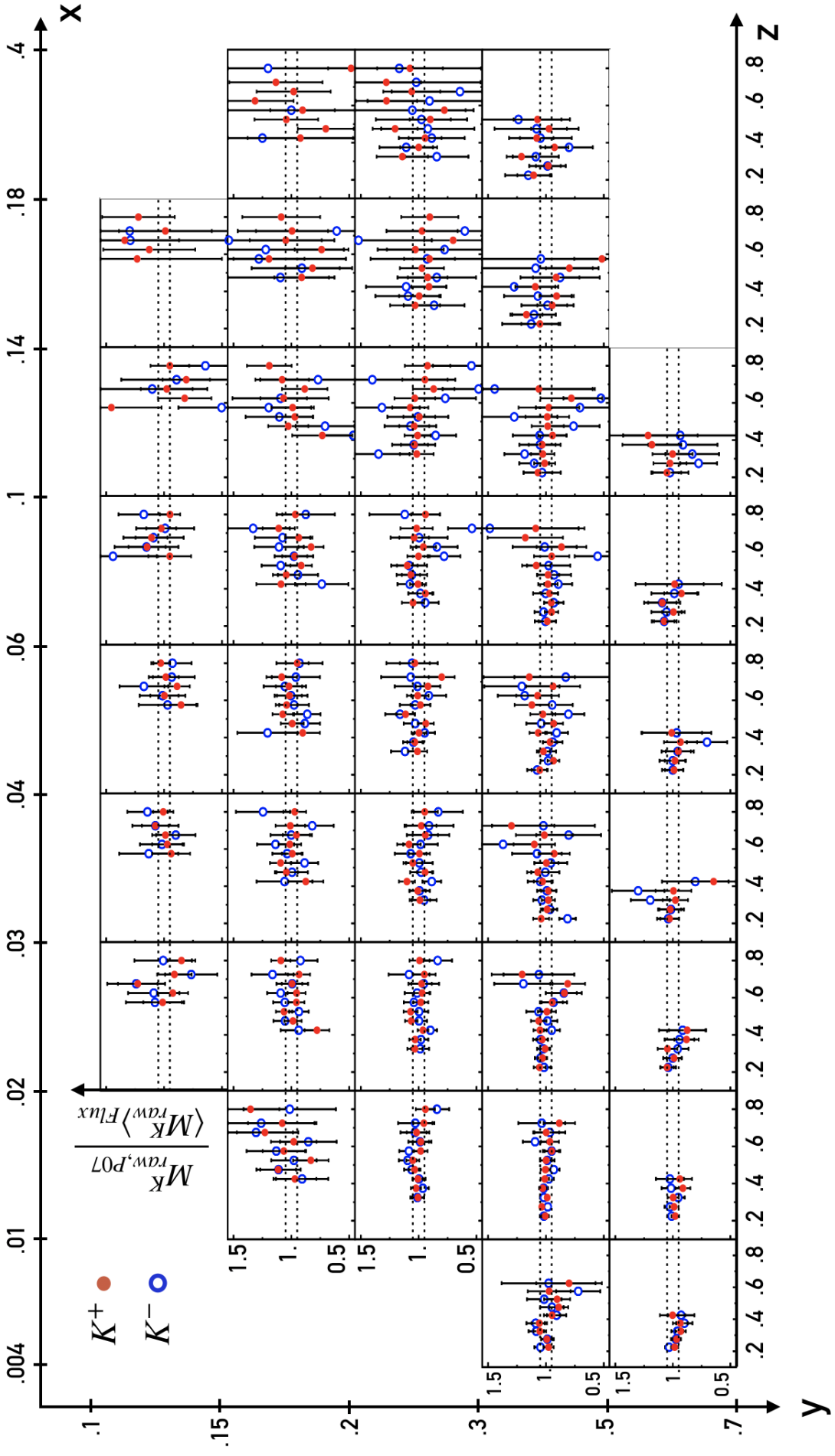


Figure 130: Same as Fig. 94 but for kaon multiplicities.

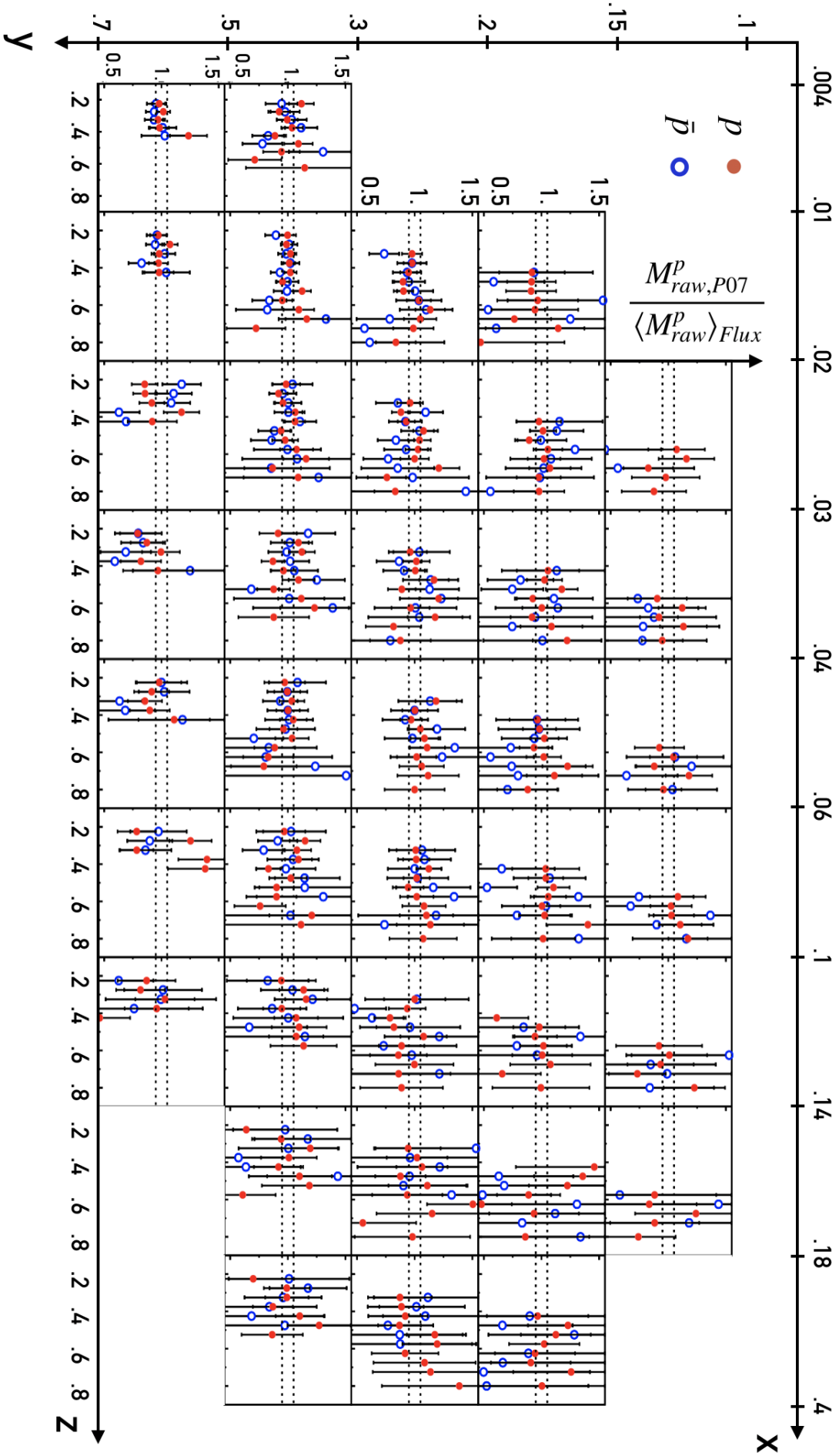


Figure 131: Same as Fig. 94 but for proton multiplicities.

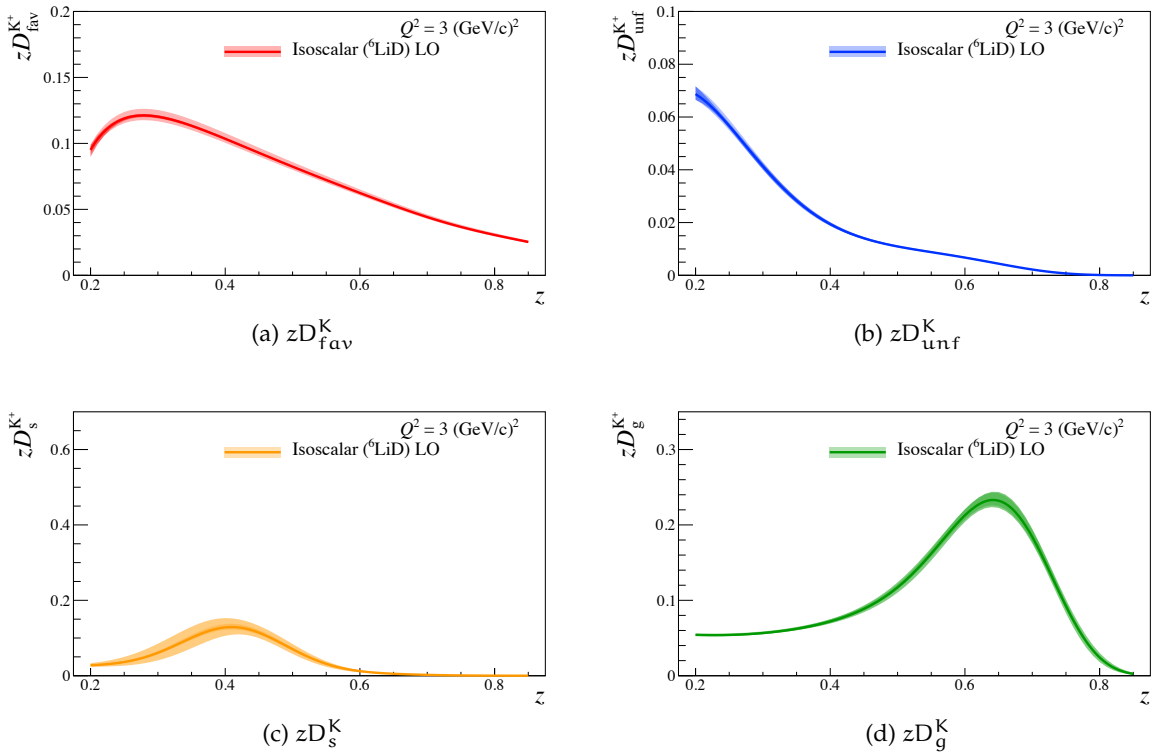
D

FURTHER LO QCD FIT OF KAON MULTIPLICITIES

A similar fit than the one perfomed in Chapter 13 was done with COMPASS data on isoscalar target. The result of the fit is presented in Fig. 132 and the parameters in Table 17.

	N	α	β	γ	δ
D^K_{fav}	0.05148 ± 0.0003	-1.4 ± 0.1	0.18 ± 0.05	-1.8 ± 0.4	3.6 ± 0.1
D^K_s	0.024 ± 0.001	19 ± 5	25 ± 2	-	-
D^K_{unf}	0.0081 ± 0.0002	2.8 ± 0.5	12 ± 1	-	-
D^K_{glu}	0.086 ± 0.006	30 ± 4	13 ± 5	-	-

Table 17: Fit parameters for $Q_0^2 = 1$ (GeV/c)^2. The associated χ^2 is of 3.2.



(a) zD^K_{fav} (b) zD^K_{unf}

(c) zD^K_s (d) zD^K_g

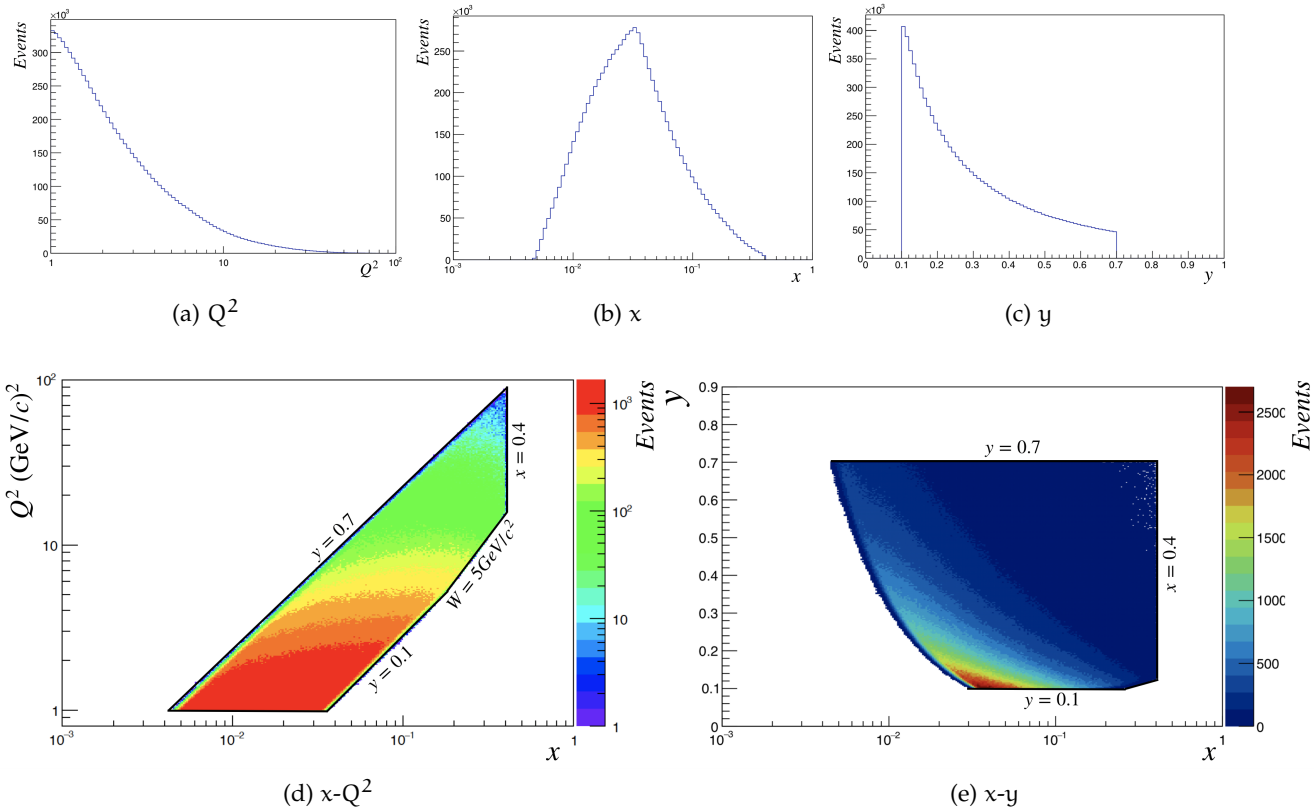
Figure 132: The favoured (top left), unfavoured (top right), strange (bottom left) and gluon (bottom right) quark FFs $zD(z)$ into kaons from the COMPASS LO fit. The fit is done based on both the statistical and systematic errors. The green dashed lines are from the same COMPASS LO fit but with COMPASS results for an isoscalar target. The orange dashed line is DSS07 LO fit.

E

RÉSUMÉ FRANÇAIS

Dans le but d'améliorer notre connaissance concernant la fragmentation des quarks en hadrons chargés, la production de hadrons chargés en diffusion inélastique profonde sur une cible de proton pure (lH_2) est étudiée dans le cadre d'une mesure semi-inclusive.

Les multiplicités de hadrons chargés non-identifiés, des pions chargés, des kaons chargés et des protons ont été déterminées dans un binning de trois variables cinématiques : x, la fraction de l'impulsion du nucleon portée par le quark touché, y, la fraction d'énergie du lepton incident portée par le photon virtuel et z, la fraction d'énergie du photon virtuel transférée au hadron h. Les résultats ont été obtenus en utilisant 5 périodes (semaines) des données 2016 de COMPASS ($\sim 8.3 \times 10^6$ événements DIS et $\sim 2.6 \times 10^6$ hadrons). Le domaine cinématique couvert est le suivant: $Q^2 > 1$ (GeV/c)^2, $y \in [0.1, 0.7]$, $x \in [0.004, 0.4]$, $W \in [5, 17]$ GeV/c^2, et $z \in [0.2, 0.85]$ (voir Fig. 133).



(a) Q^2 (b) x (c) y

(d) x-Q^2 (e) x-y

Figure 133: En haut, les distributions de Q^2, x et y. En bas, les corrélations x-Q^2 et x-y. Toutes les distributions sont pour la sélection finale d'événements DIS.

L'identification des hadrons en pions, kaons et proton est réalisée par le détecteur Ring Imaging CHerenkov (RICH). La performance de ce détecteur est déterminée en utilisant les mêmes données que pour l'analyse décrite supra. Avoir une bonne connaissance de la réponse du RICH est indispensable pour obtenir les multiplicités de pion, kaon et proton. Du fait de l'utilisation du RICH, l'intervalle d'impulsion des hadrons est restreinte de 12 GeV/c à 40 GeV/c et l'angle polaire à l'intervalle entre 0.01 et 0.12 rad dans le but d'opérer dans une région où le RICH peut faire pleinement la différence entre les pions, kaons et protons.

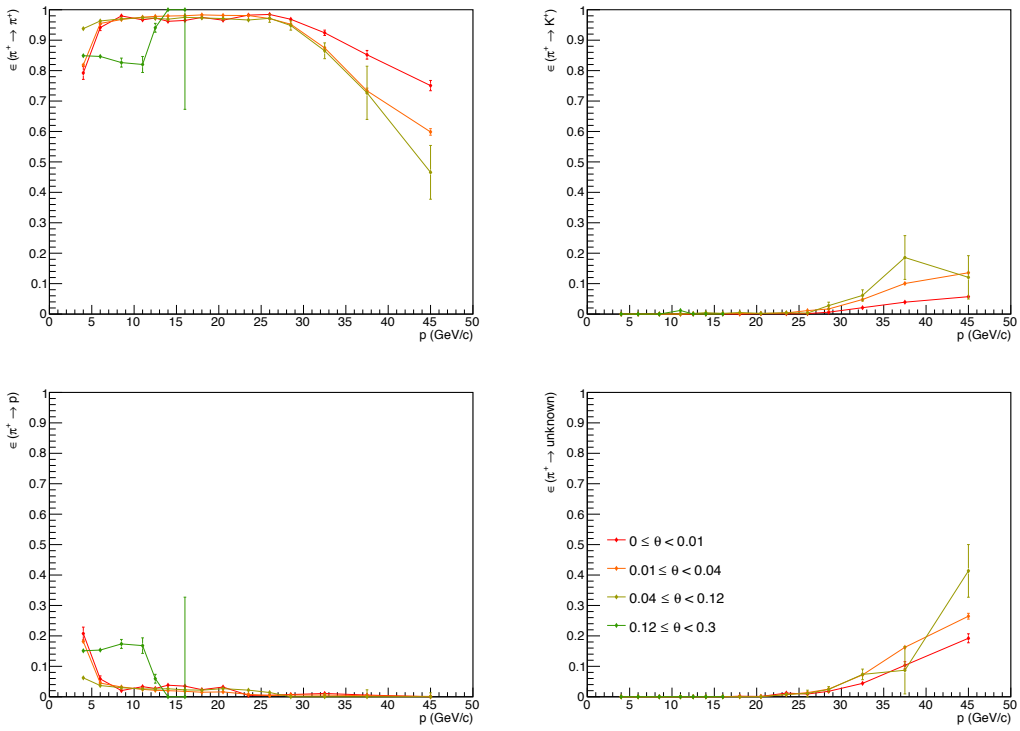


Figure 134: Probabilités d'identification $\epsilon(p \to j)$ pour π^+.

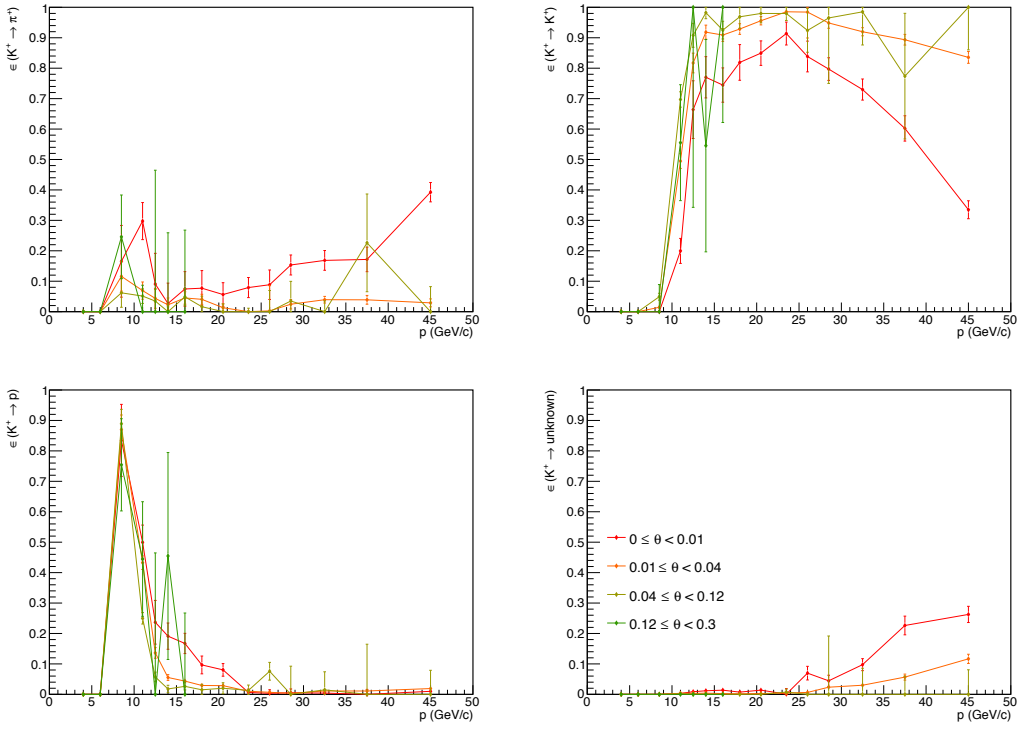


Figure 135: Probabilités d'identification $\epsilon(p \to j)$ pour K^+.

Avec une coupure basse sur l'impulsion des hadrons il est assuré que l'impulsion des particules sont bien au-dessus du seuil d'identification des kaons (environ 9.45 GeV), tandis que pour la limite haute permet d'éviter des effets de saturation ($\beta \to 1$) qui commencent à apparaitre. La bonne séparation des hadrons chargés est donc assurée. L'intervalle d'angle polaire est choisi de sorte que les efficacités du RICH sont généralement hautes et mesurées précisément. Dans cet espace des phases, les efficacités d'identification sont hautes pour les pions ($>$ 97%), kaons ($>$ 95%) et protons ($>$ 90%) avec des probabilités d'erreur d'identification faibles ($<$ 5% pour les pions, $<$ 7% pour les kaons, $<$ 10% pour les protons) comme le montrent les Fig. 134 et 135.

Les multiplicités de hadrons sont aussi corrigées pour les limitations géométriques du spectromètre, les performances des détecteurs et l'efficacité de reconstruction des données. La correction globale d'acceptance est estimée grâce à une simulation Monte-Carlo (MC). Un nouveau générateur d'événement MC, DJANGOH [92, 101], a été utilisé à la fois pour la génération d'événements radiatifs dans la simulation MC et pour le calcul des corrections radiatives aux multiplicités dans un binning tri-dimensionnel en x, y et z. Pour la première fois une solide détermination des effets radiatifs dépendents de la variable z a été faite (Fig. 136). DJANGOH donne des résultats compatibles avec ceux obtenus pour les corrections radiatives avec TERAD (calculs analytiques) et donne de très bons résultats concernant la production d'électrons par conversion de photons. La contribution de la production des mésons vecteurs à la production de hadrons est aussi estimée.

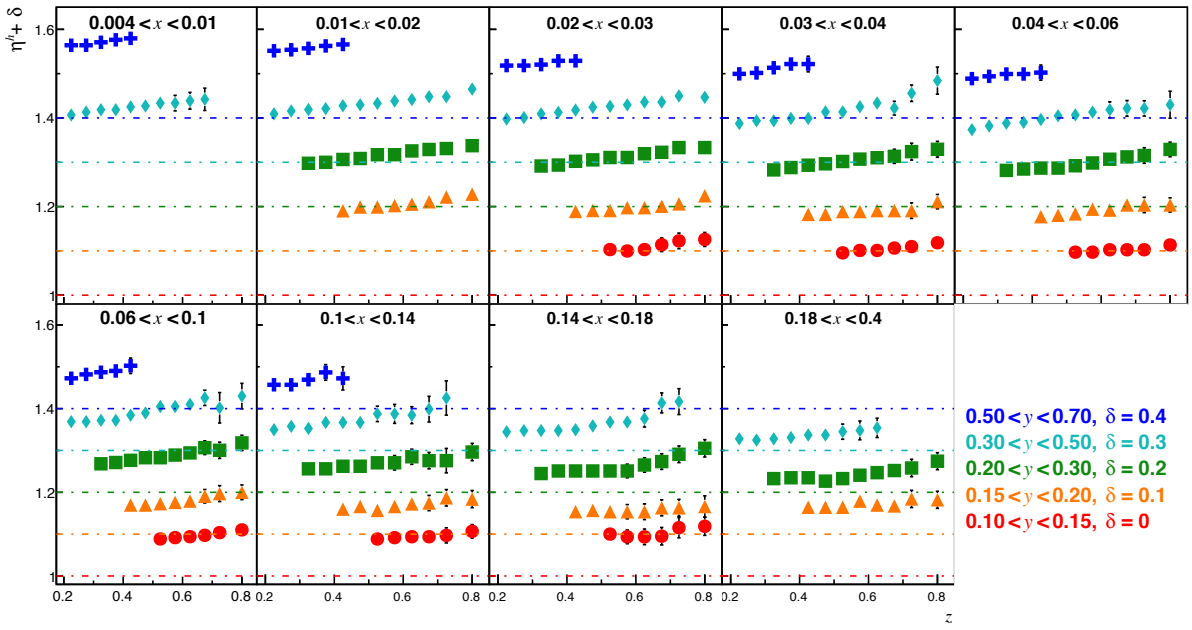


Figure 136: $\eta^h(x, y, z)$, hadrons positifs en points pleins, negatifs en points ouverts, en bins de x, décalés verticalement avec y et en fonction de z. La correction va de 2% à haut x à 20% à haut z et haut y.

La somme (Figs. 139 et 140) et le ratio (Figs. 137 et 138) des hadrons chargés a un intérêt spécial car ce sont des quantités intégrées et peuvent donc être comparées aux résultats d'autres expériences. Pour le ratio de multiplicités, les résultats de COMPASS sur une cible de proton sont comme attendus en comparaison des résultats de COMPASS sur une cible isoscalaire [44, 45]. Une difference est trouvée avec les résultats d'HERMES [47] pour les kaons, confirmant la différence déjà vue pour les cibles deutons/isoscalaires. Néanmoins, cette différence peut être dans les deux cas expliquée par les cinématiques différentes des données entre les deux expériences [118]. Pour la somme des multiplicités, dans le domaine

des x non-affectés par un probable problème d'acceptance, les résultats de COMPASS pour une cible de protons sont compatibles avec nos attentes provenant de calculs de LO pQCD. Ici aussi les résultats de COMPASS diffèrent de ceux d'HERMES pour les pions et kaons, ce qui peut être expliqué par le fait qu'HERMES opère à plus petit ν que COMPASS et ainsi qu'une correction de la masse des hadrons [118, 120]. Le concept de fonctions de fragmentations n'est donc pas applicable dans une partie de l'espace des phases [118].

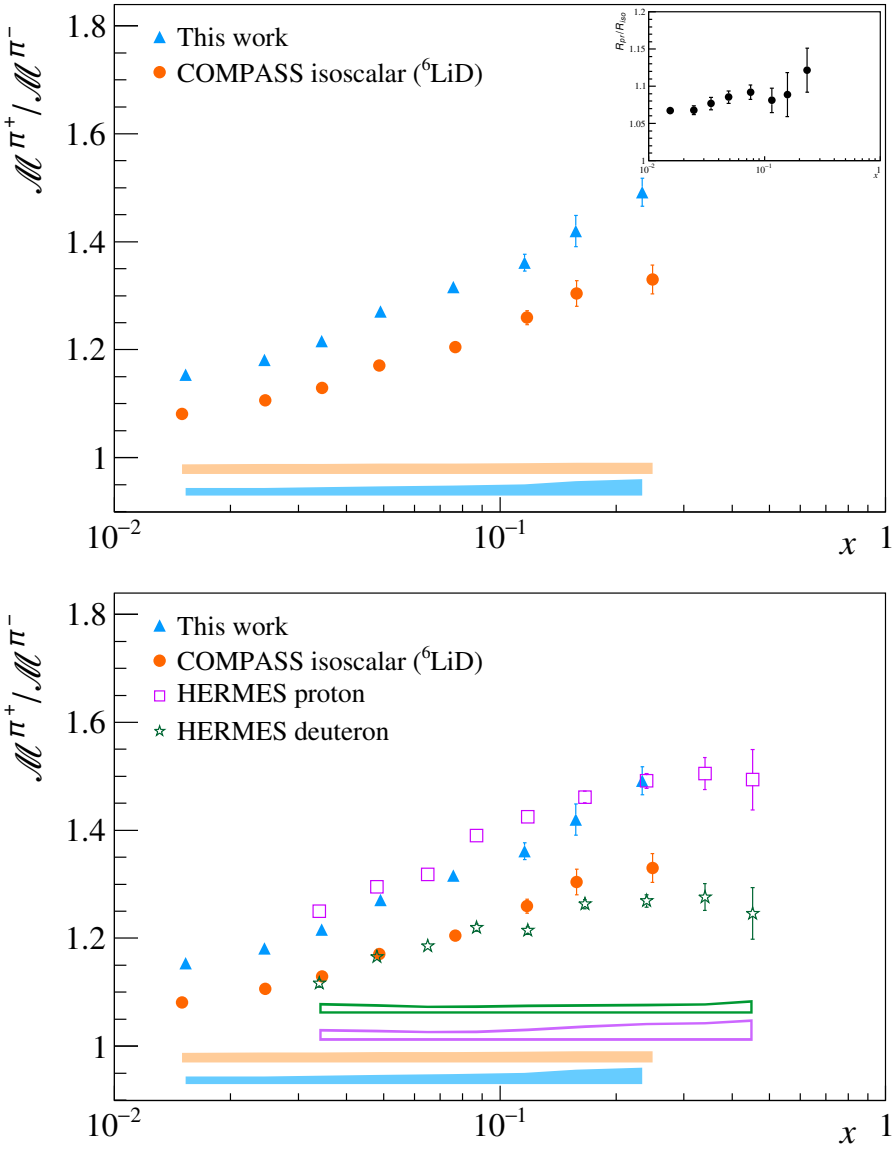


Figure 137: Ratio de $\frac{\mathscr{M}^{\pi^+}}{\mathscr{M}^{\pi^-}}$ de COMPASS pour une cible de protons (points fermés bleus) et pour une cible isoscalaire (points fermés oranges) et de HERMES pour une cible de protons (points ouverts violets) et pour une cible deuton (points ouverts verts). En haut à droite, le ratio des résultats avec la cible de protons sur ceux avec la cible isoscalaire est tracé.

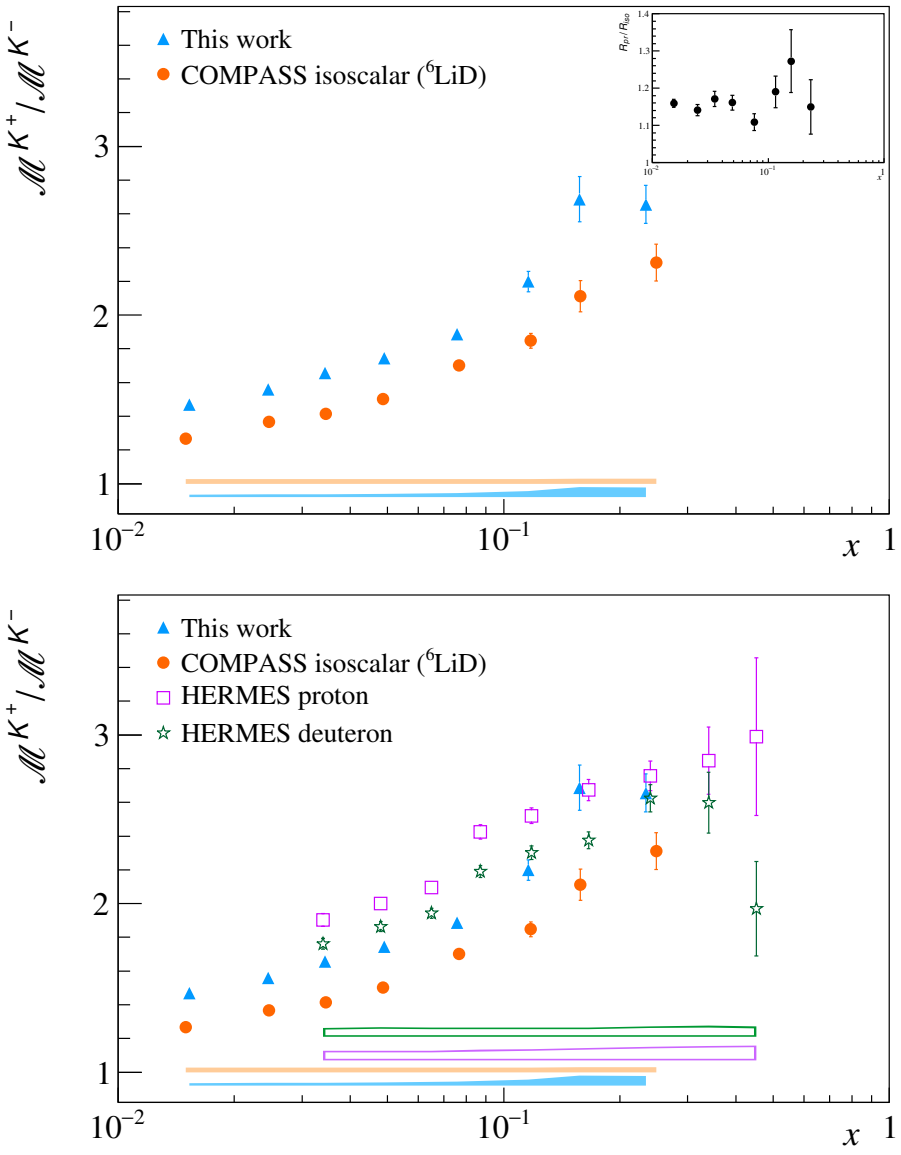


Figure 138: Même figure que Fig. 137 pour $\frac{\mathscr{M}^{K^+}}{\mathscr{M}^{K^-}}$.

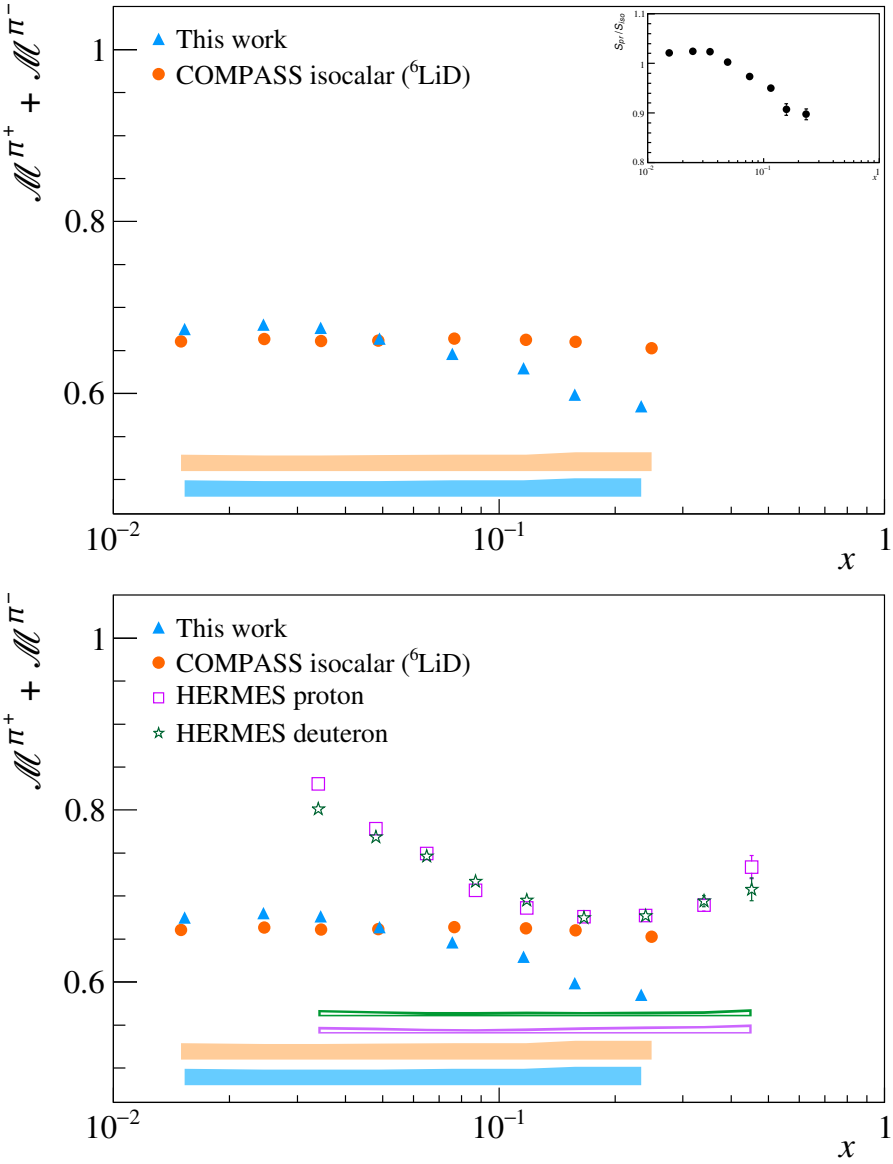


Figure 139: Somme de $\mathscr{M}^{\pi^+} + \mathscr{M}^{\pi^-}$ de COMPASS pour une cible de protons (points fermés bleus) et pour une cible isoscalaire (points fermés oranges) et de HERMES pour une cible de protons (points ouverts violets) et pour une cible deuton (points ouverts verts). En haut à droite, le ratio des résultats avec la cible de protons sur ceux avec la cible isoscalaire est tracé.

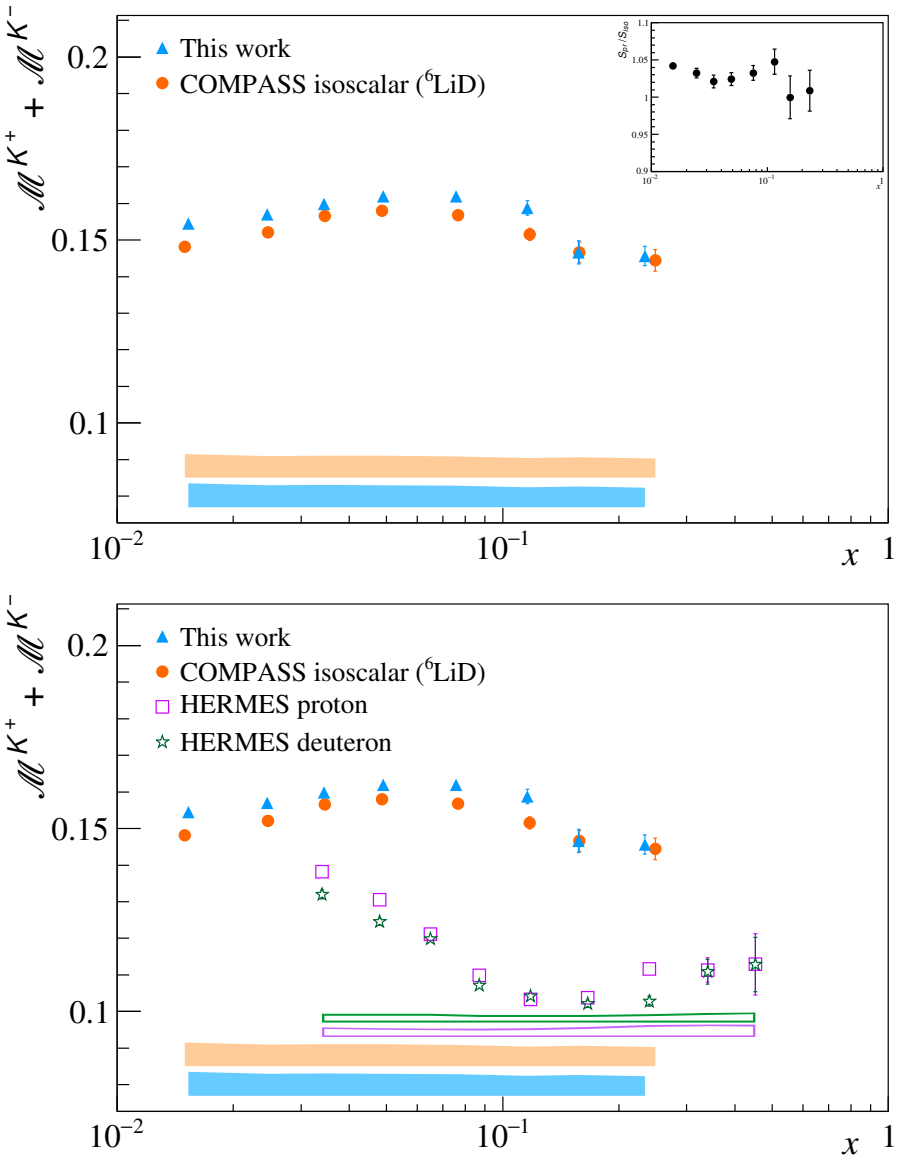
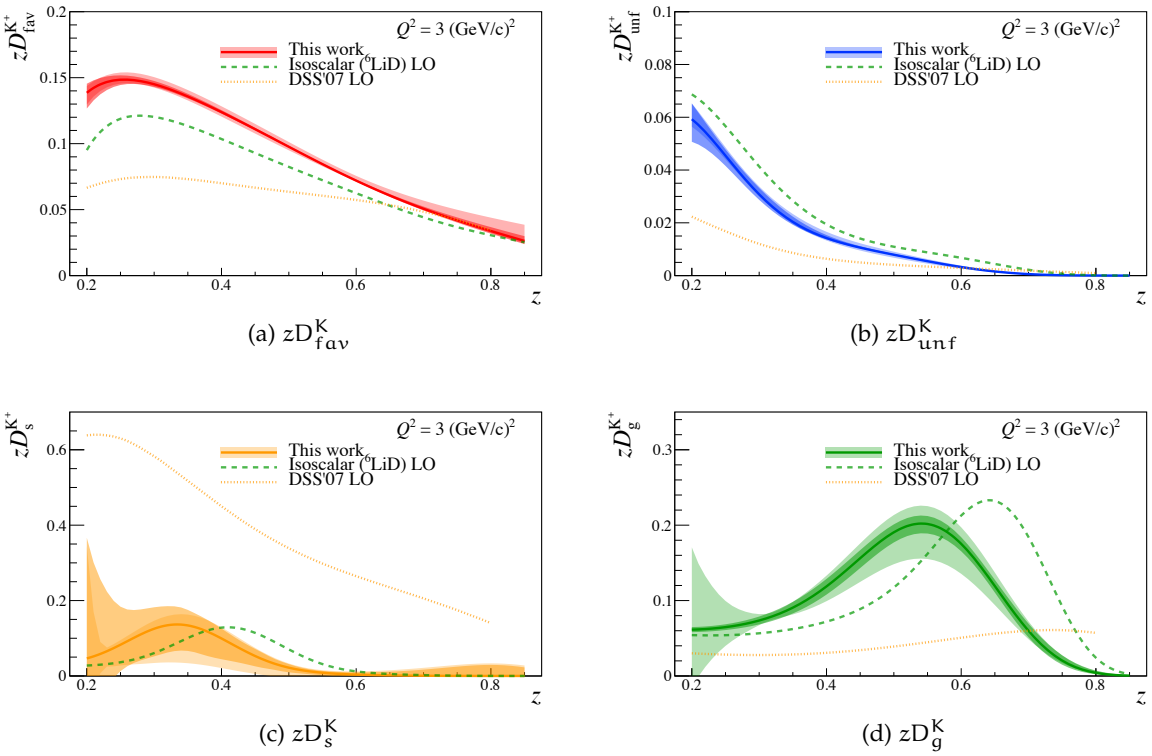


Figure 140: Même figure que Fig. 139 pour $\mathscr{M}^{K^+} + \mathscr{M}^{K^-}$.

Les multiplicités de K^+ et K^- sont utilisées pour extraire les fonctions de fragmentation de quark favorisées D^K_{fav}, défavorisées D^K_{unf} et étrange D^K_s avec un fit à LO, supposant les PDFs connues (Fig. 141). Le résultat de ce fit indique qu'il y a une mauvaise sensibilité des mesures au quark étrange. Le fit donne beaucoup trop de contribution aux fonction de fragmentation favorisées et défavorisées au détriment des fonctions de fragmentation étranges.



(a) zD^K_{fav} (b) zD^K_{unf} (c) zD^K_s (d) zD^K_g

Figure 141: Les FFs de quark $zD(z)$ en kaons favorisées (en haut à gauche), défavorisées (en haut à droite), étranges (en bas à gauche) et de gluons (en bas à droite) obtenues avec le fit LO de COMPASS. Le fit est réalisé avec les erreurs statistiques et systématiques. Les lignes vertes pointillées sont le même fit LO de COMPASS mais avec les résultats de COMPASS sur cible isoscalaire. Les lignes oranges pointillées sont le fit LO DSS07.

Les ensembles de multiplicités de pions, kaons et protons obtenus dans cette analyse, qui représentent au total plus de 1800 points de données, sont une contribution majeure pour les fits globaux des données mondiales réalisés à NLO. L'ensemble des multiplicités de kaons est particulièrement attendu car il va élargir significativement l'ensemble de données disponible [123]. L'ensemble des multiplicités de protons est aussi une nouveauté qui va intéresser les fitters.

Cette analyse va être poursuivie à COMPASS. Les nouvelles calibrations du RICH vont améliorer l'efficacité d'identification des hadrons chargés à haute impulsion, tandis que l'inclusion de l'ensemble des données de 2016 et 2017 vont améliorer la statistique. Des études plus approfondies sur le Monte-Carlo et la comparaison des données et du Monte-Carlo vont améliorer les systématiques. Dès que le problème de la chute des sommes de pions chargés et des hadrons chargés sera réglée, l'extraction des fonction de fragmentation de pion pourra être faite. Au contraire des kaons, l'extraction des fonctions de fragmentation de pion avec un fit LO des multiplicités de pions est plus à même de converger car les mesures ont une bonne sensibilité aux quarks u et d et donc aux fonctions de fragmentation favorisées et défavorisées. Pour les kaons, l'extraction des multiplicités de K^0 ds données pourrait aider à contraindre le fit des fonctions de fragmentation de kaon. Cette étude est en cours. En plus des données de 2016 and 2017, plus de données vont être prises en 2021 sur

cible de deuton. Finalement, comme il a été montré que nos ratios de multiplicités de kaons et de protons pour des z au-dessus de 0.7 ne sont pas décrites par pQCD [118], il serait intéressant de le prendre en compte dans notre analyse et examiner plus en détail cette région cinématique.

BIBLIOGRAPHY

[1] M. Breidenbach et al. "Observed Behavior of Highly Inelastic Electron-Proton Scattering." In: *Phys. Rev. Lett.* **23** (1969), p. 935.

[2] M. Gell-Mann. "A Schematic Model of Baryons and Mesons." In: *Phys. Lett.* **8** (1964), p. 214.

[3] G. Zweig. *DEVELOPMENTS IN THE QUARK THEORY OF HADRONS. VOL. 1. 1964 - 1978*. Hadronic Press, 1964.

[4] R. P. Feynman. "Very High-Energy Collisions of Hadrons." In: *Phys. Rev. Lett.* **23** (1964), p. 1415.

[5] M.G. Alexseev et al. (COMPASS Collaboration). "Quark helicity distribution from longitudinal spin asymmetries in muon-proton and muon-deuteron scattering." In: *Phys. Lett. B* **693** (2010), pp. 227–235.

[6] F. Halzen and A.D. Martin. *An Introductory Course in Modern Particle Physics*. John Wiley & Sons, Inc., 1984.

[7] A. Efremov M. Anselmino and E. Leader. "The theory and phenomenology of polarized deep inelastic scattering." In: *Physics Reports* 261(1-2) (1995), pp. 1–124.

[8] IOS Press, ed. *Proceedings of the International School of Physics Enrico Fermi*. 2008.

[9] J.D. Bjorken and E.A. Paschos. "Inelastic electron-proton and γ-proton scattering and the structure of the nucleon." In: *Physical Review* **185(54)** (1969), pp. 1975–1982.

[10] R.P. Feynman. *Photon-Hadron Interactions*. Advanced Book Program Series. Addison Wesley, 1998.

[11] C.G. Callan and D.J. Gross. "High-energy electroproduction and the constitution of the electric current." In: *Phys. Rev. Lett.* **22** (1969), pp. 156–159.

[12] E.L. Berger. "Semi-Inclusive Inelastic Electron Scattering from Nuclei." In: *ANL-HEP-CP-87-45* (1987).

[13] M. Tanabashi et al. (Particle Data Group). "The Review of Particle Physics (2018)." In: *Phys. Rev. D* **98** (2018), p. 030001.

[14] A. Pich. "Quantum chromodynamics." arXiv:hep-ph/9505231. 1995.

[15] D.E. Soper G. Sterman J.C. Collins. "Factorization of hard processes in QCD." arXiv:hep-ph/0409313. 2004.

[16] Y.L. Dokshitser. "Calculation of the structure functions for deep inelastic scattering and e^+e^- annihilation by perturbation theory in quantum chromodynamics." In: *Journal of Experimental and Theoretical Physics* **46(4)** (1977), pp. 641–652.

[17] V.N. Gribov and L.N. Lipatov. "Deep inelastic ep scattering in perturbation theory." In: *Soviet Journal of Nuclear Physics* **15** (1972), p. 438.

[18] V.N. Gribov and L.N. Lipatov. "e^+e^- pair annihilation and deep inelastic scattering in perturbation theory." In: *Soviet Journal of Nuclear Physics* **15** (1972), p. 675.

[19] G. Altarelli and G. Parisi. "Asymptotic freedom in parton language." In: *Nuclear Physics B* **126(2)** (1977), pp. 298–318.

[20] S.J. Joosten. "Fragmentation and nucleon structure in semi-inclusive deep-inelastic scattering at the HERMES experiment." PhD thesis. University of Illinois, 2013.

[21] Motylinski P. et al. Harland-Lang L.A. Martin A.D. "Parton distributions in the LHC era: MMHT 2014 PDFs." In: *Eur. Phys. J. C* **75** (2015), p. 204.

[22] J. Levelt and P. J. Mulders. "Quark correlation functions in deep-inelastic semi-inclusive processes." In: *Phys. Rev. D* **49(1)** (1994), pp. 96–113.

[23] R. Panknin. "Extraction of fragmentation functions from charged kaon and pion production at COMPASS." PhD thesis. University of Bonn, 2012.

[24] G. Kramer B. A. Kniehl and B. Pötter. "Testing the universality of fragmentation functions." In: *Nuclear Physics B* **597(1-3)** (2001), pp. 337–369.

[25] B. Andersson et al. "Parton fragmentation and string dynamics." In: *Physics Reports* **97(2-3)** (1983), pp. 31–145.

[26] J. M. Niczyporuk. "Quark flavor tagging in polarized semi-inclusive deep inelastic scattering." PhD thesis. Massachussetts Institute of Technology, 1997.

[27] T. Uematsu. "Q^2 dependence of quark and gluon fragmentation functions in a parton picture based on QCD." In: *Phys. Lett. B* **79(1-2)** (1978), pp. 97–102.

[28] ALEPH Collaboration. "Measurement of α_s from scaling violations in fragmentation functions in e^+e^- annihilation." In: *Phys. Lett. B* **364(4)** (1995), pp. 247–248.

[29] OPAL Collaboration. "Measurement of the longitudinal, transverse and asymmetry fragmentation functions at LEP." In: *Z. Phys. C* **68(2)** (1995), pp. 203–213.

[30] DELPHI Collaboration. "Measurement of the quark and gluon fragmentation functions in z^0 hadronic decays." In: *Eur. Phys. J. C* **6(1)** (1999), pp. 19–33.

[31] TPC/Two-Gamma Collaboration. "Charged-hadron inclusive cross sections and fractions in e^+e^- annihilation at $\sqrt{s} = 29$ GeV." In: *Phys. Rev. Lett.* **61** (1988), pp. 1263–1266.

[32] BABAR Collaboration. "Production of charged pions, kaons and protons in e^+e^- annihilations into hadrons at $\sqrt{s} = 10.54$ GeV." In: *Phys. Rev. D* **88** (2013), p. 032011.

[33] Belle Collaboration. "Precision measurement of charged pion and kaon differential cross sections in e^+e^- annihilation $\sqrt{s} = 10.54$ GeV." In: *Phys. Rev. Lett.* **111** (2013), p. 062002.

[34] UA5 Collaboration. "Charged particle multiplicity distributions at 200 and 900 GeV c.m. energy." In: *Z. Phys. C* **43(3)** (1989), pp. 357–374.

[35] UA1 Collaboration. "A study of the general characteristics of proton-antiproton collisions at $\sqrt{s} = 0.2$ to 0.9 TeV." In: *Nuclear Physics B* **335(2)** (1989), p. 261.

[36] ALICE Collaboration. "Charged-particle multiplicity measurement in proton-proton collisions at $\sqrt{s} = 7$ TeV with ALICE at LHC." In: *Journal of High Energy Physics* **2010(2)** (2010).

[37] CMS Collaboration. "Transverse-momentum and pseudorapidity distributions of charged hadrons in pp collisions at $\sqrt{s} = 0.9$ and 2.36 TeV." In: *Journal of High Energy Physics* **2010(2)** (2010).

[38] CMS Collaboration. "Transverse-momentum and pseudorapidity distributions of charged hadrons in pp collisions at $\sqrt{s} = 7$ TeV." In: *Phys. Rev. Lett.* **105** (2010), p. 022002.

[39] ATLAS Collaboration. "Measurement of inclusive jet charged-particle fragmentation functions in $pb + pb$ collisions at $\sqrt{s_{NN}} = 2.76$ TeV with the ATLAS detector." In: *Phys. Lett. B* **739** (2014), pp. 340–342.

[40] PHENIX Collaboration. "Midrapidity neutral-pion production in proton-proton collisions at $\sqrt{s} = 200$ GeV." In: *Phys. Rev. Lett.* **91** (2003), p. 241803.

[41] STAR Collaboration. "Forward neutral pion production in $p+p$ and $d+Au$ collisions at $\sqrt{s_{NN}} = 200$ GeV." In: *Phys. Rev. Lett.* **97** (2006), p. 152302.

[42] BRAHMS Collaboration. "Prodcution of mesons and baryons at high rapidity and high p_T in proton-proton collisions at $\sqrt{s} = 200$ GeV." In: *Phys. Rev. Lett.* **98** (2007), p. 252001.

[43] ATLAS Collaboration. "A determination of the fragmentation functions of u quarks into charged pions." In: *Phys. Lett. B* **160** (1985), pp. 417–420.

[44] C. Adolph et al. "Multiplicities of charged pions and charged hadrons from deep-inelastic muon scattering off an isoscalar target." In: *Phys. Lett. B* **764** (2017), pp. 1–10.

[45] C. Adolph et al. "Multiplicities of charged kaons from deep-inelastic muon scattering off an isoscalar target." In: *Phys. Lett. B* **767** (2017), pp. 133–141.

[46] Eoo-108 Collaboration. "Semi-inclusive charged-pion electroproduction off protons and deuterons : Cross sections, ratios and acces to the quark-parton model at low energies." In: *Phys. Rev. C* **85** (2012), p. 015202.

[47] A. Airapetian et al. (HERMES Collaboration). "Multiplicities of charged pions and kaons from semi-inclusive deep-inelastic scattering by the proton and the deuteron." In: *Phys. Rev. D* **87** (2013), p. 074029.

[48] S. Kretzer. "Fragmentation functions from flavour inclusive and flavour tagged e^+e^- annihilations." In: *Phys. Rev. D* **62** (2000), p. 054001.

[49] T.-H. Nagai M. Hirai S. Kumano and K. Sudoh. "Determination of fragmentation functions and their uncertainties." In: *Phys. Rev. D* **75** (2007), p. 094009.

[50] B.A. Kniehl S. Albino and G. Kramer. "Improvements from new theoretical input and experimental data." In: *Nuclear Physics B* **803(1-2)** (2008), pp. 42–104.

[51] E. Leader A.V. Sidorov and D.B. Stamenov. "Importance of semi-inclusive DIS processes in determining fragmentation functions." arXiv:1312.5200v3. 2013.

[52] Manuel Epele Roger J. Hernández-Pinto D. de Florian R. Sassot and Marco Stratmann. "Parton-to-pion fragmentation reloaded." In: *Phys. Rev. D* **91** (2015), p. 014035.

[53] Manuel Epele Roger J. Hernández-Pinto D. de Florian R. Sassot and Marco Stratmann. "Parton-to-kaon fragmentation revisited." In: *Phys. Rev. D* **95** (2017), p. 0194019.

[54] N. Sato J.J. Ethier and W. Melnitchouk. "First simultaneous extraction of spin-dependent parton distributions and fragmentation functions from a global QCD analysis." In: *Phys. Rev. Lett.* **119** (2017), p. 132001.

[55] R. Sassot D. de Florian and Marco Stratmann. "Global analysis of fragmentation functions for pions and kaons and their uncertainties." In: *Phys. Rev. D* **75** (2007), p. 114010.

[56] S. E. Kuhn J. J. Ethier Nobuo Sato W. Melnitchouk and A. Accardi. "Iterative Monte Carlo analysis of spin-dependent parton distributions." In: *Phys. Rev. D* **93** (2016), p. 074005.

[57] C. Itzykson and J.-B. Zuber. *Quantum Field Theory*. Dover Publications, Inc., 2005.

[58] H. Spiesberger. http://compassvm03.cern.ch/compass/software/analysis/transparencies/2016/am_160901/talks/DJANGOH-COMPASS.pdf. Seminar : Radiative Corrections in SIDIS with DJANGOH. 2016.

[59] K. Kurek B. Badelek D. Bardin and C. Scholz. "Radiative correction schemes in deep inelastic muon scattering." In: *Z. Phys. C* **66** (1995), p. 591.

[60] http://compassvm03.cern.ch/compass/run/run2017/.

[61] P. Abbon et al. "The COMPASS setup for physics with hadron beams." In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **779** (2015), pp. 69–115.

[62] L. Gatignon. http://sba.web.cern.ch/sba/BeamsAndAreas/M2/M2-OperatorCourse.pdf. 2006.

[63] P. Abbon et al. "The COMPASS experiment at CERN." In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **577** (2007), pp. 455–518.

[64] F. Thibaud. "Développement de détecteurs Micromegas pixellisés pour les hauts flux de particules et évaluation de la contribution diffractive à la leptoproduction de hadrons à COMPASS." PhD thesis. Université Paris-Sud, 2014.

[65] H. Pereira Da Costa. "Développement de chambres à dérive pour les hauts flux dans l'expérience COMPASS." PhD thesis. Université Paris-Sud, 2001.

[66] A. Zvyagin. "D-meson production by muons in the COMPASS experiment at CERN." PhD thesis. Ludwig-Maximilians-Universität München, 2010.

[67] The COMPASS collaboration. "The COMPASS trigger system for muon scattering." In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **550(1-2)** (2005), pp. 217–240.

[68] T. Bellunato et al. "Refractive index dispersion law of silica aerogel." In: *European Physical Journal C* **52** (2007), pp. 759–764.

[69] E. Albrecht et al. "The radiator gas and the gas system of COMPASS RICH-1." In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **502(1)** (2003), pp. 266–269.

[70] E. Albrecht et al. "The mirror system of COMPASS RICH-1." In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **502(1)** (2003), pp. 236–240.

[71] P. Abbon et al. "Particle identification with COMPASS RICH-1." In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **631(1)** (2011), pp. 26–39.

[72] F. Piuz. "Ring Imaging CHerenkov systems based on gaseous photo-detectors : trends and limits around particle accelerators." In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **502(1)** (2003), pp. 76–90.

[73] P. Abbon et al. "Design and construction of the fast photon detection system for COMPASS RICH-1." In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **616(1)** (2010), pp. 21–37.

[74] P. Abbon et al. "Read-out electronics for fast proton detection system for COMPASS RICH-1." In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **587(2-3)** (2008), pp. 371–387.

[75] T. Ypsilantis and J. Seguinot. "Theory of ring imaging cherenkov counters." In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **343(1)** (1994), pp. 30–51.

[76] R. J. Barlow. *Statistics: A Guide to the Use of Statistical Methods in the Physical Sciences*. Wiley, 1998.

[77] V. Blobel and E. Lohrmann. *Statistische und numerische Methoden der Datenanalyse*. Teubner, 1998.

[78] P. Schiavon. "Studies of the Cerenkov emission angle reconstruction. Part 1." In: *COMPASS Note* (1997).

[79] P. Schiavon. "Studies of the Cerenkov emission angle reconstruction. Part 2." In: *COMPASS Note* (1997).

[80] P. Schiavon. "Patter recognition for COMPASS RICH 1." In: *COMPASS Note* (1999).

[81] Q. M. Curiel Garcia. "Measurement of Hadron Multiplicities in Deep Inelastic Scattering and Extraction of Quark Fragmentation Functions." PhD thesis. Université Paris-Sud, 2014.

[82] M. Stolarski. http://wwwcompass.cern.ch/compass/software/analysis/transparencies/2016/am_161013/talks/stolarski_161013.pdf. talk at the October analysis meeting. 2016.

[83] M. Wilfert and J. Giarra. "RICH performance in 2011/12." In: *COMPASS Note* (2017).

[84] M. Stolarski. http://compassvm03.cern.ch/compass/software/analysis/transparencies/2019/am_190509/talks/stolarski_190509_1.pdf. talk at the May analysis meeting. 2019.

[85] A.A. Akhundov et al. "Model independent QED corrections to the process $ep \to eX$." In: *Fortsch. Phys.* **44** (1996), p. 373.

[86] B. Badelek. "TERAD15 user guide." In: *COMPASS Note* **2015-6** (2015).

[87] I. Akushevich et al. "RADGEN 1.0: Monte Carlo generator for radiative events in DIS on polarized and unpolarized targets." In: *Hamburg 1998/1999, Monte Carlo generators for HERA physics* (), pp. 554–565.

[88] N.Shumeiko A.Soroko I.Akushevich A.Ilyichev and A.Tolkachev. "POLRAD 2.0. FORTRAN code for the radiative corrections calculation to deep inelastic scattering of polarized particles." In: *Comp. Phys. Comm.* **104** (1997), p. 201.

[89] J. Drees and M. Dueren. "FERRAD 3.5." This code was first created by J. Drees for EMC and further developed by M. Dueren.

[90] D. Bardin and N. Shumeiko. "On an exact calculation of the lowest-order electromagnetic correction to the point particle elastic scattering." In: *Nucl. Phys. B* **127** (1977), p. 242.

[91] L.W.Mo and Y.S.Tsai. "Radiative Corrections to Elastic and Inelastic ep and μp Scattering." In: *Rev. Mod. Phys.* **41** (1969), p. 205.

[92] G. A. Schuler K.Charchula and H.Spiesberger. "Combined QED and QCD radiative effects in deep inelastic lepton-proton scattering: the Monte Carlo generator DJANGO6." In: *Comput. Phys. Commun.* **81** (1994), p. 381.

[93] J. Rathsman G. Ingelman A. Edin. "LEPTO 6.5: A Monte Carlo generator for deep inelastic lepton - nucleon scattering." In: *Comput.Phys.Commun.* **101** (1997), pp. 108–134.

[94] L. Lönnblad. "Ariadne version 4 — A program for simulation of QDC cascades implementing the colour dipole model." In: *Comput. Phys. Commun.* **71** (1992), p. 15.

[95] S. Mrenna T. Sjöstrand and P. Z. Skands. "PYTHIA 6.4 Physics and Manual." In: *JHEP* **05** (2006), p. 026.

[96] T. Sjöstrand and M. Bengtsson. "The Lund Monte Carlo for jet fragmentation and e^+ e^- physics - jetset version 6.3 - an update." In: *Comp. Phys. Commun.* **43** (1987), p. 367.

[97] J. P. Rachen R. J. Protheroe A. Mücke R. Engel and T. Stanev. "Monte-Carlo simulations of photohadronic processes in astrophysics." In: *Comput. Phys. Commun.* **124** (2000), p. 290.

[98] J. Vermaseren S. de Jong. "AXO User Manual." In: *NIKHEF-H Report* (1987).

[99] G. P. Lepage. "A new algorithm for adaptive multidimensional integration." In: *J. Comp. Physics* **27** (1978), p. 192.

[100] M. Stolarski. http://compassvm03.cern.ch/compass/software/analysis/transparencies/2015/am_151208/talks/stolarski_151209.pdf. Radiative Corrections for Semi-Inclusive events. 2015.

[101] N. Pierre. "DJANGOH : a Monte-Carlo Generator with Radiative Corrections." In: *COMPASS Note* 2018-3 (2018).

[102] B. Badelek N. Pierre. "Comparison of radiative processes calculations in TERAD and DJANGOH." In: *COMPASS note 2017-3* (2017).

[103] B. Adeva et al. "SM Collaboration." In: *Phys. Rev. D* **58** (1998), 112001 (Appendix).

[104] B. Badelek and J. Kwiecinski. "Electroproduction structure function F_2 in the low Q^2, low x region." In: *Phys. Lett. B* 295 (1992), p. 263.

[105] K. Abe et al. "E143 Collaboration." In: *Phys. Lett. B* 452 (1999), p. 194.

[106] H. Spiesberger. "QED Radiative Corrections for Parton Distributions." In: *Phys. Rev. D* **52** (1995), pp. 4936–4940.

[107] A.A. Akhundov et al. "Model independent QED corrections to the process $ep \to eX$." In: *Fortsch. Phys.* **44** (1996), p. 373.

[108] T. Szameitat. "New Geant4-based Monte Carlo software for the COMPASS-II experiment at CERN." PhD thesis. Physikalisches Institut Albert-Ludwigs-Universität Freiburg, 2016.

[109] Brett Bode, Michelle Butler, Thom Dunning, Torsten Hoefler, William Kramer, William Gropp, and Wen-mei Hwu. "The Blue Waters Super-System for Super-Science." In: *Contemporary High Performance Computing*. Chapman & Hall/CRC Computational Science. Chapman and Hall/CRC, 2013, pp. 339–366. ISBN: 978-1-4665-6834-1.

[110] William Kramer, Michelle Butler, Gregory Bauer, Kalyana Chadalavada, and Celso Mendes. "Blue Waters Parallel I/O Storage Sub-system." In: *High Performance Parallel I/O*. Ed. by Prabhat and Quincey Koziol. CRC Publications, Taylor and Francis Group, 2015, pp. 17–32. ISBN: 978-1-4665-8234-7.

[111] R. S. Thorne A. D. Martin W. J. Stirling and G. Watt. "Parton distributions for the LHC." In: *Eur. Phys. J. C* **63** (2009), pp. 189–285.

[112] The COMPASS Collaboration. "Results on longitudinal spin physics at COMPASS." In: *Phys. Lett. B* **718** (2013), p. 922.

[113] A. Sandacz and P. Sznajder. "HEPGEN - generator for hard exclusive leptoproduction." arXiv:1207.0333v1. 2012.

[114] J. J. Sakurai. "Vector-meson dominance and high-energy electron-proton inelastic scattering." In: *Phys. Rev. Lett.* **22** (1969), pp. 981–984.

[115] L. Silva. "Measurement of the Gluon Polarisation Through High p_T Hadron Production in COMPASS." PhD thesis. Universidade Técnica de Lisboa, 2011.

[116] S. V. Goloskokov and P. Kroll. "The role of the quark and gluon in GPDs in hard vector-meson electroproduction." In: *European Physical Journal C* **53** (2008), p. 367.

[117] The European Muon Collaboration. "Forward produced protons and antiprotons in deep inelastic muon-proton scattering." In: *Phys. Lett. B* **103** (1981), p. 388.

[118] The COMPASS collaboration. "K^- over K^+ multiplicity ratio for kaons produced in DIS with a large fraction of the virtual-photon energy." In: *Phys. Lett. B* **786** (2018), pp. 390–398.

[119] Private communications with M. Stolarski.

[120] A. Accardi S. W. Casper J. V. Guerrero J. J. Ethier and W. Melnitchouk. "Hadron mass corrections in semi-inclusive deep-inelastic scattering." In: *JHEP* **1509** (2015), p. 169.

[121] R. S. Thorne A. D. Martin W. J. Stirling and G. Watt. "LHAPDF6: parton density access in the LHC precision era." In: *Eur. Phys. J. C* **75** (2015), p. 132.

[122] W. H. Press et al. *Numerical Recipes in C (2^{n}d Ed.)* The Art of Scientific Computing, 1992.

[123] N. Pierre. "Kaon multiplicities in SIDIS from COMPASS." In: *XXVII International Workshop on Deep Inelastic Scattering and Related Subjects*. 2019.

université PARIS-SACLAY | ÉCOLE DOCTORALE PHENIICS

Titre : Multiplicités de hadrons en diffusion inélastique profonde de muons sur nucléons à COMPASS

Mots clés : Multiplicités, nucléon, quark, fragmentation, étrangeté

Résumé : Un des buts de la collaboration COMPASS est l'étude de la structure de spin du nucléon. La question de la polarisation des quarks de la mer est un sujet capital en physique hadronique, en particulier pour la polarisation du quark étrange. En vue de mieux contraindre la polarisation des quarks, une connaissance précise des fonctions de fragmentation (FFs), qui expriment l'hadronisation d'un quark q en un hadron h dans l'état final, est nécessaire. Les FFs peuvent être extraites depuis les multiplicités de hadrons produites en diffusion inélastique profonde semi-inclusive (SIDIS). Les données ont été prises à COMPASS avec un faisceau de muons de 160 GeV/c diffusant sur une cible de protons pure (lH_2). La présente thèse présente les mesures des multiplicités de hadrons chargés (pions, kaons et protons) faites à partir des données SIDIS collectées en 2016. Elle détaille aussi les améliorations apportées au générateur d'événement DJANGOH dans le but d'améliorer la description des corrections radiative inclusive et semi-inclusive qui sont ensuite utilisées comme facteurs de corrections aux multiplicités. Les données couvrent un large spectre cinématique : $Q^2 > 1$ $(GeV/c)^2$, $y \in [0.1,0.7]$, $x \in [0.004,0.4]$, $W \in [5,17]$ GeV et $z \in [0.2,0.85]$. Ces multiplicités, qui représentent un total d'environ 1800 points de données, apportent une contribution importante aux fit QCD globaux des données mondiales à NLO, visant à la détermination des FFs. Les FFs de quarks en kaons sont particulièrement attendues car elles pourront mieux contraindre la polarization du quark étrange.

Title : Multiplicities of hadrons in deep-inelastic scattering of muons on nucleons at COMPASS

Keywords : Multiplicities, nucleon, quark, fragmentation, strangeness

Abstract : One of the goals of the COMPASS collaboration is the study of the nucleon spin structure. The question of the polarization of the sea quark is a burning issue in the hadronic physics, especially for the strange quark polarization. In order to better constrain the quark polarization, a precise knowledge of the quark Fragmentation Functions (FFs) into hadrons, which are the final state hadronisation of quark q into hadron h, is mandatory. The FFs can be extracted from hadron multiplicities produced in Semi-Inclusive Deep Inelastic Scattering (SIDIS). Data were taken at COMPASS from a 160 GeV/c muon beam scattering off a pure proton target (lH_2). This thesis presents the measurement of charged hadrons (pions, kaons and protons) multiplicities from SIDIS data collected in 2016. It also details the improvements brought to the DJANGOH event generator to better describe the inclusive and semi-inclusive radiative corrections in DIS that are then used as correction factors to the multiplicities. The data cover a large kinematical range : $Q^2 > 1$ $(GeV/c)^2$, $y \in [0.1,0.7]$, $x \in [0.004,0.4]$, $W \in [5,17]$ GeV and $z \in [0.2,0.85]$. These multiplicities, which represent about 1800 data points in total, provide an important input for global QCD fit of world data at NLO, aiming at the FFs determination. The quark FFs into kaons are particularly awaited as they can better constrain the strange quark polarization.

Université Paris-Saclay
Espace Technologique / Immeuble Discovery
Route de l'Orme aux Merisiers RD 128 / 91190 Saint-Aubin, France