

FRAGMENTATION AND NUCLEON STRUCTURE IN
SEMI-INCLUSIVE DEEP-INELASTIC SCATTERING AT THE
HERMES EXPERIMENT

BY

SYLVESTER JOHANNES JOOSTEN

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Physics
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2013

Urbana, Illinois

Doctoral Committee:

Professor Jen-Chieh Peng, Chair
Professor Naomi C.R. Makins, Director of Research
Professor Lance Cooper
Professor Gregory MacDougall

Abstract

Multiplicities for the semi-inclusive production of each charge state of π^\pm and K^\pm mesons in deep-inelastic scattering are presented as a function of the kinematic quantities x , Q^2 , z and $P_{h\perp}$. The multiplicities were extracted from data collected by the HERMES experiment at the HERA storage ring using 27.6 GeV electron and positron beams on a hydrogen or deuterium gas target. These results for identified hadrons constitute the most precise measurement to date, and will significantly enhance our understanding of the proton structure, as well as the fragmentation process in deep-inelastic scattering. Furthermore, the 3D binning at an unprecedented level of precision provides a handle to help disentangle the transverse momentum structure of both.

The high level of precision coupled with an intermediate energy regime requires a careful study of the complex interaction between the experimental systematics, theoretical uncertainties, and the applicability of the factorization theorem within the standard framework of leading-twist collinear QCD. This is illustrated by the extraction of the valence quark ratio d_v/u_v at leading-order in α_s . These results show a strong z -dependence below $z \approx 0.30$, which could be interpreted as evidence for factorization breaking. This evidence weakens somewhat when isospin invariance of the fragmentation functions is assumed to be broken.

Additionally, the multiplicities for the semi-inclusive production of π^0 mesons in deep-inelastic scattering are presented as a function of z . These multiplicities were extracted from the same data sample as used for the charged meson results. The neutral pion multiplicity is the same as the average charged pion multiplicity, up to $z \approx 0.70$. This is consistent with isospin invariance below $z \approx 0.70$. The results at high values of z show strong signs of isospin symmetry breaking.

The work for this thesis was supported by the National Science Foundation (NSF).

Contents

List of Figures	vi
List of Abbreviations	viii
List of Symbols	ix
1 Introduction	1
2 Deep Inelastic Scattering	3
2.1 Kinematic Variables	3
2.2 Structure Functions and the DIS Cross Section	5
2.3 Factorization and the Quark-Parton Model	6
2.4 Factorization beyond the Quark-Parton Model	8
2.5 Semi-inclusive DIS and the Fragmentation Process	12
2.5.1 The SIDIS Cross Section	13
2.5.2 Current Knowledge of FFs	13
3 The HERMES Experiment	18
3.1 The HERA ep -collider	19
3.2 The HERMES Target	19
3.3 The HERMES Spectrometer	21
3.3.1 Track Reconstruction	22
3.3.2 Particle Identification	22
3.3.3 Trigger and Data Acquisition	24
4 Multiplicity Analysis	27
4.1 Data Selection	28
4.2 Experimental Cuts	29
4.3 Particle Identification	29
4.3.1 Lepton-Hadron Separation	30
4.3.2 Determining Hadron Flavor	32
4.4 DIS Event Selection	34
4.4.1 Charge Symmetric Background	35
4.4.2 Trigger Inefficiencies	35
4.5 Kinematic Coverage and Binning	36
4.6 Multidimensional Smearing-Unfolding	36
4.6.1 Formalism	36
4.6.2 Unfolding in Practice	41
4.6.3 Importance of a Multi-Dimensional Approach	42

4.6.4	The HERMES Effect	45
4.7	Contamination by Exclusive Channels	47
4.8	Data Combination and Projections	55
4.9	Systematic Uncertainties	56
4.9.1	RICH Unfolding	58
4.9.2	Azimuthal Asymmetries in the Unpolarized Cross Section	63
4.9.3	Finite Bin Width and Monte-Carlo Model-Dependence	64
4.9.4	Time Stability	64
4.9.5	Minor Contributions	71
4.9.6	Comparison with the Statistical Uncertainties	71
5	Hadron Multiplicity Results	76
5.1	One-Dimensional Projection versus z	76
5.2	The Multiplicity Target Asymmetry	78
5.3	Two-Dimensional Projections	80
5.4	The Multiplicity Website	80
6	Interpretation in Leading-Order Collinear DIS	84
6.1	Calculating the Multiplicities in Leading-Order	84
6.2	Comparison with Leading-Order Calculations	85
6.3	Strangeness Suppression	89
7	Transverse-Momentum Structure	95
7.1	Theoretical Background	95
7.2	Results for Transverse-Momentum Dependence	96
8	Precocious Scaling and the Limits of Factorization	102
8.1	Limits of Factorization	102
8.1.1	Theoretical Interpretation	102
8.1.2	Target Remnant Influence in SIDIS Measurements	103
8.1.3	Factorization at Intermediate Energies and Precocious Scaling	109
8.2	The Applicability of QPM-like Factorization and d_v/u_v	112
8.2.1	Fragmentation Function Symmetries	113
8.2.2	The Pion Yield Difference Ratio and d_v/u_v	114
8.2.3	Results: d_v/u_v and Factorization Breaking	115
9	Neutral Pion Multiplicities	122
9.1	Measurement of the π^0 Yield	124
9.2	Corrections	126
9.3	Systematic Uncertainties	126
9.3.1	Background Subtraction	128
9.3.2	Calorimeter Calibration and Cluster Position	128
9.3.3	Azimuthal Asymmetries in the Unpolarized Cross Section	129
9.3.4	Monte-Carlo Model Dependence	129
9.3.5	Time Stability	129
9.4	Neutral Pion Results	129

List of Figures

2.1	Illustration of deep-inelastic scattering	4
2.2	Factorization in SIDIS	7
2.3	Global measurements for F_2^p	9
2.4	The factorization theorem in Feynman diagram language.	10
2.5	MSTW2008 PDFs at NNLO	11
2.6	Comparison of NLO fragmentation functions	14
2.7	Fragmentation Functions from e^+e^- annihilation	15
2.8	Charged-hadron production from pp and $p\bar{p}$	16
2.9	Previously available SIDIS multiplicities	17
3.1	The HERA ep -collider	18
3.2	The HERMES storage cell	20
3.3	The HERMES spectrometer	21
3.4	The calorimeter and pre-shower hodoscope	23
3.5	The RICH	24
3.6	Hadron identification with the RICH	25
4.1	Lepton-hadron separation	31
4.2	RICH efficiencies for single-track events	33
4.3	xQ^2 correlation and DIS event selection	35
4.4	3D binning design	38
4.5	Importance of a multi-dimensional approach	43
4.6	HERMES effect inflation factor	46
4.7	Elastic and quasi-elastic fractions	47
4.8	Impact of the HERMES effect correction	48
4.9	Exclusive vector meson fractions: inclusive	50
4.10	Exclusive vector meson fractions: semi-inclusive	51
4.11	Impact of the vector meson correction: x, Q^2	53
4.12	Impact of the vector meson correction: $z, P_{h\perp}$	54
4.13	Fractional impact of the vector meson correction	55
4.14	Systematic uncertainty breakdown: z	57
4.15	Systematic uncertainty breakdown: $zP_{h\perp}$	59
4.16	Systematic uncertainty breakdown: zx	60
4.17	Systematic uncertainty breakdown: zQ^2	61
4.18	RICH efficiencies for the systematic uncertainty	62
4.19	Monte-Carlo model dependence: z	65
4.20	Monte-Carlo model dependence: $zP_{h\perp}$ (pions)	66
4.21	Monte-Carlo model dependence: $zP_{h\perp}$ (kaons)	67
4.22	Comparing the 2000 and 2005 results: zx (pions)	69

4.23	Comparing the 2000 and 2005 results: zx (kaons)	70
4.24	Systematic versus statistical uncertainty: z	73
4.25	Systematic versus statistical uncertainty: zx (pions)	74
4.26	Systematic versus statistical uncertainty: zx (kaons)	75
5.1	Multiplicity results: z	77
5.2	Multiplicity target asymmetry	79
5.3	Multiplicity results: $zP_{h\perp}$	81
5.4	Multiplicity results: zx	82
5.5	Multiplicity results: zQ^2	83
6.1	Comparison with LO calculations: z	86
6.2	Comparison with LO calculations: target asymmetry	88
6.3	Comparison with LO calculations: zx	90
6.4	Comparison with LO calculations: zQ^2	91
6.5	Kaon-to-pion ratio and strangeness suppression: z	92
6.6	Kaon-to-pion ratio: 2D	94
7.1	Multiplicity results: $zP_{h\perp}$	97
7.2	Average $P_{h\perp}$: z	98
7.3	Average $P_{h\perp}$: zx	99
7.4	Average $P_{h\perp}$: zQ^2	100
7.5	Multiplicity hadron charge asymmetry: $P_{h\perp}$	101
8.1	E665 rapidity spectrum for positive hadrons	105
8.2	E665 hadronic net charge versus rapidity	106
8.3	EMC hadronic net charge versus rapidity	107
8.4	E665 correlations between the hemispheres	108
8.5	$(\bar{d} - \bar{u})/(u - d)$ as a function of z from HERMES (1998)	110
8.6	LO d_v/u_v : z	116
8.7	LO d_v/u_v : z in x -slices	118
8.8	LO d_v/u_v : x in z -slices	119
8.9	LO d_v/u_v : z , comparison with Lund MC	121
9.1	HERMES π^0 results from 2001	123
9.2	$M_{\gamma\gamma}$ spectrum and π^0 fit	127
9.3	Neutral pion results: z (proton)	130
9.4	Neutral pion results: z (deuteron)	131
9.5	Comparison with the old results	132

List of Abbreviations

DIS	Deep-inelastic scattering.
FF	Fragmentation function.
ISV	Isospin-symmetry violation.
LO	Leading order.
NLO	Next-to-leading order.
NNLO	Next-to-next-to-leading order.
PDF	Parton distribution function.
pQCD	Perturbative QCD.
QED	Quantum electrodynamics.
QCD	Quantum chromodynamics.
RICH	Ring-imaging Čerenkov detector.
SIDIS	Semi-inclusive deep-inelastic scattering.
TRD	Transition-radiation detector.

List of Symbols

k, k'	Four-momenta (E, \vec{k}) of incident and scattered lepton l and l' .
P	Four-momentum of the target nucleon; $(M, \vec{0})$ in the lab frame.
s	Squared center-of-mass energy $(k + P)^2$.
q	Four-momentum $(k - k')$ of the virtual photon γ^* .
Q^2	Negative squared four-momentum transfer $-q^2$; approximately $4EE' \sin^2(\frac{\theta}{2})$ in the lab frame.
W^2	Squared invariant mass of the photon-nucleon system $(P + q)^2$.
ν	Energy transfer $\frac{P \cdot q}{M}$ to the target; $E - E'$ in the lab frame.
y	Fractional energy of the virtual photon $\frac{P \cdot q}{p \cdot k}$; $\frac{\nu}{E}$ in the lab frame.
x	Bjorken scaling variable $\frac{Q^2}{2P \cdot q}$.
z	Fractional energy of the hadron h , $\frac{P \cdot P_h}{P \cdot q}$; $\frac{E_h}{\nu}$ in the lab frame.
$P_{h\perp}$	Component of the hadron momentum P_h transverse to q ; $\frac{ \vec{q} \times \vec{P}_h }{ q }$ in the lab frame.
ϕ_h	Azimuthal angle between the lepton scattering plane and the hadron production plane.
$P_{\text{CM}}^{\parallel}$	Hadron momentum parallel to q in the photon-nucleon system.
η_{CM}	Center-of-mass rapidity in the photon-nucleon system $\frac{1}{2} \ln \frac{E_{\text{CM}}^h + P_{\text{CM}}^{\parallel}}{E_{\text{CM}}^h - P_{\text{CM}}^{\parallel}}$.
x_F	Feynman scaling variable $\frac{2P_{\text{CM}}^{\parallel}}{W}$.

1 Introduction

The semi-inclusive production of pseudoscalar meson in deep-inelastic scattering is used in this work to study various properties of the strong interaction. Quantum chromodynamics (QCD), the quantum field theory that describes the strong interaction, is only perturbative at short distance scales, Hard reactions can be described in perturbative QCD (pQCD) because of the existence of **factorization** theorems, which allow for the experimental cross section to be split in parts that are process dependent but perturbative, and process independent, **universal** parts that describe the long-distance physics. These non-perturbative parts are described by parameterizations that are constrained by experimental measurements. Examples of the physics described by these parameterizations are the nucleon structure, and the functions that described the decay of a free quark into a jet of final state hadrons.

They are described in various frameworks that are derived from pQCD, in the limit of specific kinematic regimes. A high-precision, multi-dimensional measurement, as presented in this work, allows for the non-perturbative parameterizations to be better constrained, while simultaneously probing the limits of the aforementioned frameworks. A pedagogical introduction and motivation, assuming only basic knowledge of particle physics, is provided in Chapter 2. The rest of the work provides an explanation of the more intricate theoretical details when needed. This allows most of the discussion to be read as stand-alone works by someone familiar with the field.

This work presents multiplicities for the semi-inclusive production of π^+ , π^- , K^+ and K^- mesons in deep-inelastic scattering at the HERMES experiment, in various three-dimensional binnings. The experimental setup of the HERMES experiment is discussed in Chapter 3. These results constitute the most precise measurement in semi-inclusive deep-inelastic scattering to date. The high statistical precision of the measurement requires a sophisticated extraction procedure to minimize the systematic uncertainties. This is explained in detail in Chapter 4.

The main results are presented and discussed in Chapter 5. Chapter 6 illustrates the degree to which these results will enhance the precision of the parameterizations of the fragmentation functions in the standard framework of leading-twist collinear QCD. Through the inclusion of the dependence on the transverse hadron momentum $P_{h\perp}$, these results reach beyond this collinear framework, and provide

a handle to help disentangle the transverse momentum structure of the nucleon and the fragmentation process. This is briefly discussed in Chapter 7. Next, the applicability of the standard framework of leading-twist collinear QCD for semi-inclusive deep-inelastic scattering at intermediate energies is tested in Chapter 8. Finally, the π^0 multiplicities are extracted and compared to those for the charged pions as a test for isospin symmetry in Chapter 9.

2 Deep Inelastic Scattering

This chapter provides a brief pedagogical overview of the theoretical concepts necessary to situate and understand the rest of this work. The more intricate details are saved for later chapters, and introduced only at the specific points where they are needed. This approach makes Chapters 5 to 9 relatively self-contained, allowing someone with a basic background in the field of hadronic physics to follow the discussion without having to read the rest of this work.

2.1 Kinematic Variables

The deep-inelastic scattering (DIS) from a lepton l with four-momentum k on a hadron N with four-momentum P , is the hard scattering reaction,

$$l + N \rightarrow l' + X, \quad (2.1)$$

where the nucleon is broken up into several hadrons. In the final state, the scattered lepton l' is detected with four-momentum k' , as illustrated in Fig. 2.1. The case where one or more of the produced hadrons are detected in coincidence with the scattered lepton is referred to as semi-inclusive DIS (SIDIS).

The four-momentum transfer q is carried by a virtual photon, or a virtual W - or Z -boson. The measurements for this work were performed at the HERMES experiment, where a 27.6 GeV positron or electron beam was scattered off a fixed target. This corresponds to a center-of-mass energy $\sqrt{s} = 7.26$ GeV far below the Z -boson mass. The variables P and q fully constrain the kinematic dependence of unpolarized DIS. In fact, for a given beam energy E and target mass M , there are only two independent variables that the cross section can depend on. These can be defined in a Lorentz-invariant way as,

$$Q^2 \equiv -q^2, \quad (2.2)$$

where the minus sign is included so that Q^2 is positive. The second scalar is the

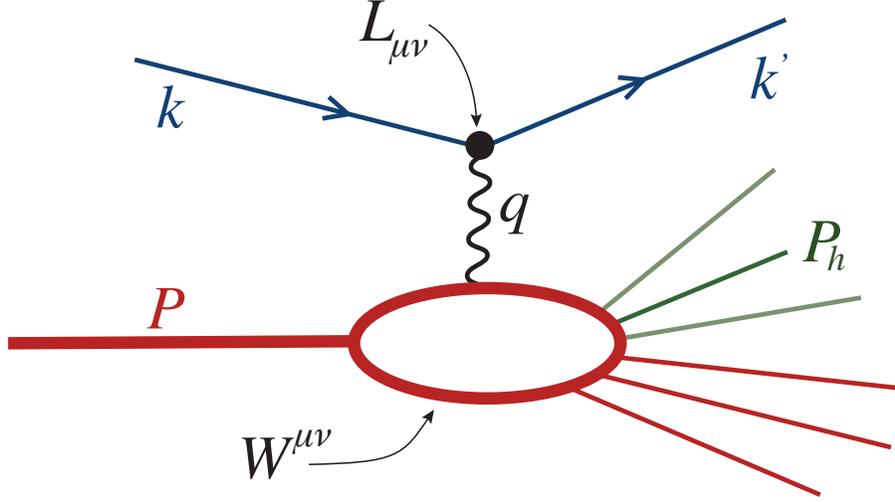


Figure 2.1: Diagram illustrating the deep-inelastic scattering of a lepton with four-momentum k off a proton with four-momentum P . In the case of semi-inclusive deep-inelastic scattering, a hadron with four-momentum P_h is detected in coincidence with the scattered lepton with momentum k' .

Bjorken scaling variable,

$$x \equiv \frac{Q^2}{2P \cdot q}. \quad (2.3)$$

Additional useful invariants are,

$$\nu \equiv \frac{P \cdot q}{M} \stackrel{\text{lab}}{=} E - E', \quad (2.4)$$

$$y \equiv \frac{P \cdot q}{p \cdot k} \stackrel{\text{lab}}{=} \frac{\nu}{E}, \quad (2.5)$$

$$s \equiv (P + k)^2, \quad (2.6)$$

$$W^2 \equiv (P + q)^2 = M^2 + 2M\nu - Q^2. \quad (2.7)$$

Due to the laboratory-frame identities, ν and y are generally referred to as respectively the energy transfer from the beam to the target, and the fractional beam energy transfer. s is the squared center-of-mass energy of the experiment, and W^2 the squared invariant mass of the final hadronic state. For elastic scattering, $W^2 = M^2$, and therefore $x = 1$, while for inelastic scattering, the Bjorken scaling variable ranges between $0 < x < 1$. An event is deep-inelastic if there is enough energy available for the formation of multiple final state hadrons,

$$W^2 = M^2 + \frac{1-x}{x} Q^2 \gg M^2. \quad (2.8)$$

A lower bound for the deep-inelastic region is therefore given by approximately $Q^2 > 1 \text{ GeV}^2$ and $W^2 > 4 \text{ GeV}^2$. In fact, much of the theoretical framework used to describe DIS and SIDIS is derived in the Bjorken limit, where $Q^2 \rightarrow \infty$ for fixed values of x .

The detection of a hadron h in coincidence with the scattered lepton in unpolarized DIS adds three more degrees of freedom to the scattering cross section, corresponding to the three degrees of freedom in the hadron four-momentum P_h for a given hadron with mass M_h ,

$$z \equiv \frac{P \cdot P_h}{P \cdot q} \stackrel{\text{lab}}{=} \frac{E_h}{\nu}, \quad (2.9)$$

$$P_{h\perp} \equiv \frac{|\vec{q} \times \vec{P}_h|}{|\vec{q}|}, \quad (2.10)$$

$$\phi_h \equiv \arccos \frac{(\hat{q} \times \vec{k}) \cdot (\hat{q} \times \vec{P}_h)}{|\hat{q} \times \vec{k}| |\hat{q} \times \vec{P}_h|}, \quad (2.11)$$

with $\hat{q} \equiv \vec{q}/|q|$. The invariant scaling variable z , ranging between $0 < z < 1$, is equal to the energy fraction transferred to the hadron h in the laboratory frame. The transverse momentum component $P_{h\perp}$, and the azimuthal angle ϕ_h between the lepton scattering plane and the hadron production plane, are both defined with respect to the virtual photon momentum in the laboratory frame.

A comprehensive list of the kinematic variables can be found in the front matter at Page [ix](#).

2.2 Structure Functions and the DIS Cross Section

As explained in the previous section, the inclusive process in Fig. 2.1 can be expressed as a function x and Q^2 , for a given beam energy E and target mass M . The general structure of the inclusive DIS cross section can be derived using only quantum electrodynamics (QED) and general symmetry arguments regarding quantum-chromodynamics (QCD). Contraction of the two vertices in Fig. 2.1 gives,

$$d\sigma = \frac{4\alpha^2}{s} \frac{d^3\vec{k}'}{2|\vec{k}'|Q^4} L_{\mu\nu}(k, q) W^{\mu\nu}(p, q), \quad (2.12)$$

The leptonic tensor $L_{\mu\nu}$ describes the lepton-photon coupling, and can be expressed as,

$$L_{\mu\nu} = 2 \cdot \left[k'_\mu k'_\nu k'_\mu k'_\nu - g_{\mu\nu} (\vec{k}' \vec{k}' - m_l^2) \right], \quad (2.13)$$

with m_l the lepton mass and $g_{\mu\nu}$ the Minkowski metric.

The hadronic tensor $W^{\mu\nu}$ describes the coupling of the virtual photon to the hadronic system, and function of P and q . It is Lorentz invariant, and $W^{\mu\nu} = W^{\nu\mu*}$. Additionally, because of conservation of electromagnetic current, $W^{\mu\nu} = 0$. This only leaves two independent tensor structures for $W^{\mu\nu}$. Each of these tensors is multiplied by a scalar **structure function** F , that only depends on the invariants x and Q^2 ,

$$W^{\mu\nu} = -\left(g_{\mu\nu} - \frac{q_\mu q_\nu}{q^2}\right) F_1(x, Q^2) + \left(P_\mu - q_\mu \frac{P \cdot q}{q^2}\right) \left(P_\nu - q_\nu \frac{P \cdot q}{q^2}\right) \frac{1}{P \cdot q} F_2(x, Q^2). \quad (2.14)$$

Putting Eqs. (2.12) to (2.14) together then results in,

$$\frac{d\sigma}{dx dQ^2} = \frac{4\pi\alpha}{Q^4} \left[y^2 F_1(x, Q^2) + \left(\frac{1-y}{x} - \frac{My}{2E} \right) F_2(x, Q^2) \right]. \quad (2.15)$$

It is possible leverage the kinematic dependence of the inclusive cross section in Eq. (2.15) to measure these structure functions directly.

2.3 Factorization and the Quark-Parton Model

Together, the structure functions F_1 and F_2 constrain the cross section of the interaction of the virtual photon with the nucleon, and as such encode the nucleon structure in the unpolarized DIS process. To extract the nucleon structure in a process-independent way, one needs to be able to factorize out the process-dependent hard-scattering reaction. A first such attempt was made by Bjorken [1] and Feynman [2] in their parton model, inspired by the results from the results from the DIS experiments at SLAC in the late sixties. This model provided much of the intuitive basis for modern hadronic physics, and factorization theorems can be thought of as field theoretic realizations of the parton model.

The quark-parton model (QPM) describes a hadron as an extended object, made up of partons, and held together by their mutual interactions. While it is impossible to explicitly compute the structure of the virtual partonic states that make up the hadron, the scattering of a free parton, for example by an electron, is assumed to be calculable. This dichotomy mirrors the features of QCD, where it is impossible to perturbatively calculate long distance behavior (confinement), while asymptotic freedom allows perturbation theory to be used at short distances.

Consider the example of DIS by a virtual photon exchange at high energies, depicted in Fig. 2.1. When boosting to the electron-nucleon center-of-mass frame, two important things happen to the nucleon. It is Lorentz-contracted in the direc-

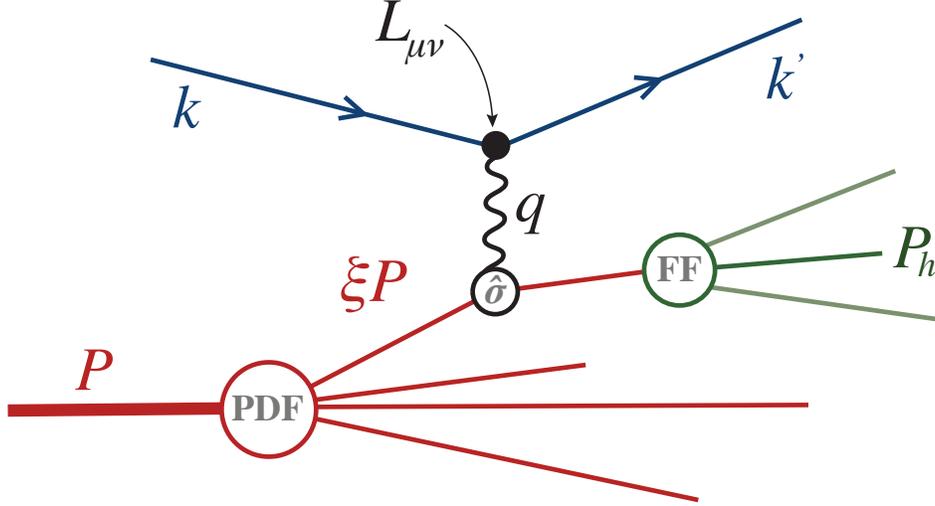


Figure 2.2: Diagrammatic illustration of factorization in semi-inclusive deep-inelastic scattering.

tion of the collision, and its internal interactions are time-dilated. As the center-of-mass energy increases, the lifetime of any virtual partonic state is therefore lengthened, while the time it takes the electron to traverse the hadron is shortened. Hence, the nucleon will be in a single virtual state, characterized by a definite number of partons, during the interaction. The partons do not interact during this time frame, and may be thought of as carrying a momentum fraction ξ of the nucleon, in the center-of-mass frame. These arguments motivate a picture where the electrons interact with a parton of definite momentum, rather than the hadron as a whole, as long as the momentum transfer is very high. Additionally, the scattering reaction is assumed to occur instantaneously, without the interference of final state interactions that happen on a much longer timescale.

The DIS process at high energies can therefore be treated as the incoherent scattering off individual partons, as depicted in Fig. 2.2. This allows the cross section to be written as the sum of probabilities, rather than amplitudes. The parton distribution function (PDF) $f_q(\xi)$ is defined as the probability that a noninteracting parton of species q with momentum fraction ξ is encountered. The DIS cross section can therefore be written as,

$$\frac{d\sigma}{dx dQ^2} \propto \int_0^1 d\xi \sum_q f_q(\xi) \frac{d\hat{\sigma}_q(x/\xi, Q^2)}{dx dQ^2}, \quad (2.16)$$

with $\hat{\sigma}$ the electron-parton cross section. After the interaction, the parton is on-

shell,

$$(\xi P + q)^2 = \xi^2 M^2 + 2\xi P \cdot q - Q^2 \approx 0, \quad (2.17)$$

and therefore, as $Q \gg M$, $\xi \approx Q^2/(2P \cdot q) = x$, relating the Bjorken scaling variable x to the momentum fraction carried by the struck parton. At leading order, the structure functions F_1 and F_2 can then be written as,

$$F_1(x) = \frac{1}{2} \sum_q \int_0^1 d\xi e_q^2 f_q(\xi) \delta(\xi - x) = \frac{1}{2} \sum_q e_q^2 f_q(x), \quad (2.18)$$

$$F_2(x) = \sum_q \int_0^1 d\xi e_q^2 \xi f_q(\xi) \delta(\xi - x) = \sum_q x e_q^2 f_q(x), \quad (2.19)$$

where e_q refers to the parton charge. The relation $2xF_1 = F_2$, known as the Callan-Gross relation [3], follows directly from incoherent scattering off spin-1/2 partons.

In the QPM, the PDFs are simple number densities that describe the proton structure. They are **universal** (i.e., process-independent) quantities, and therefore have a **predictive power**. For example, PDFs measured in DIS can be used to calculate the cross section for the Drell-Yan process ($p+p \rightarrow l+\bar{l}+X$). Factorization into terms that are perturbative, and terms that are non-perturbative but universal, allows for high-energetic lepton-hadron processes to be described and calculated in a consistent way.

2.4 Factorization beyond the Quark-Parton Model

Figure 2.3 shows a compilation of various measurements of F_2^p as a function of Q^2 in various x -bins. The QPM prediction that the structure functions be independent of Q^2 is clearly violated, except for a narrow region of $0.1 < x < 0.15$. This is not entirely unexpected, as the previous discussion only used qualitative properties of QCD.

The problems with the QPM picture can be made apparent when considering the one-loop QCD corrections to the diagram in Fig. 2.2, which give rise to collinear singularities in the cross section¹, caused by QCD-radiation with very low transverse momentum k_T . Of course, these collinear internal lines should not belong to the hard graph in the first place. It is therefore natural to introduce a cutoff μ_F such that radiative corrections with a transverse momentum $k_T < \mu_F$ are

¹ Besides the collinear divergences, there are also complications due to ultra-violet and soft divergences. The former can be removed by standard counter terms, while the latter automatically cancel in the sum over all graphs. This implies that factorization does not hold for on the level of an individual graph. A full treatment is beyond the scope of this work, a pedagogical explanation can be found in Ref. [10], and a full formal treatment in, e.g., Ref. [11].

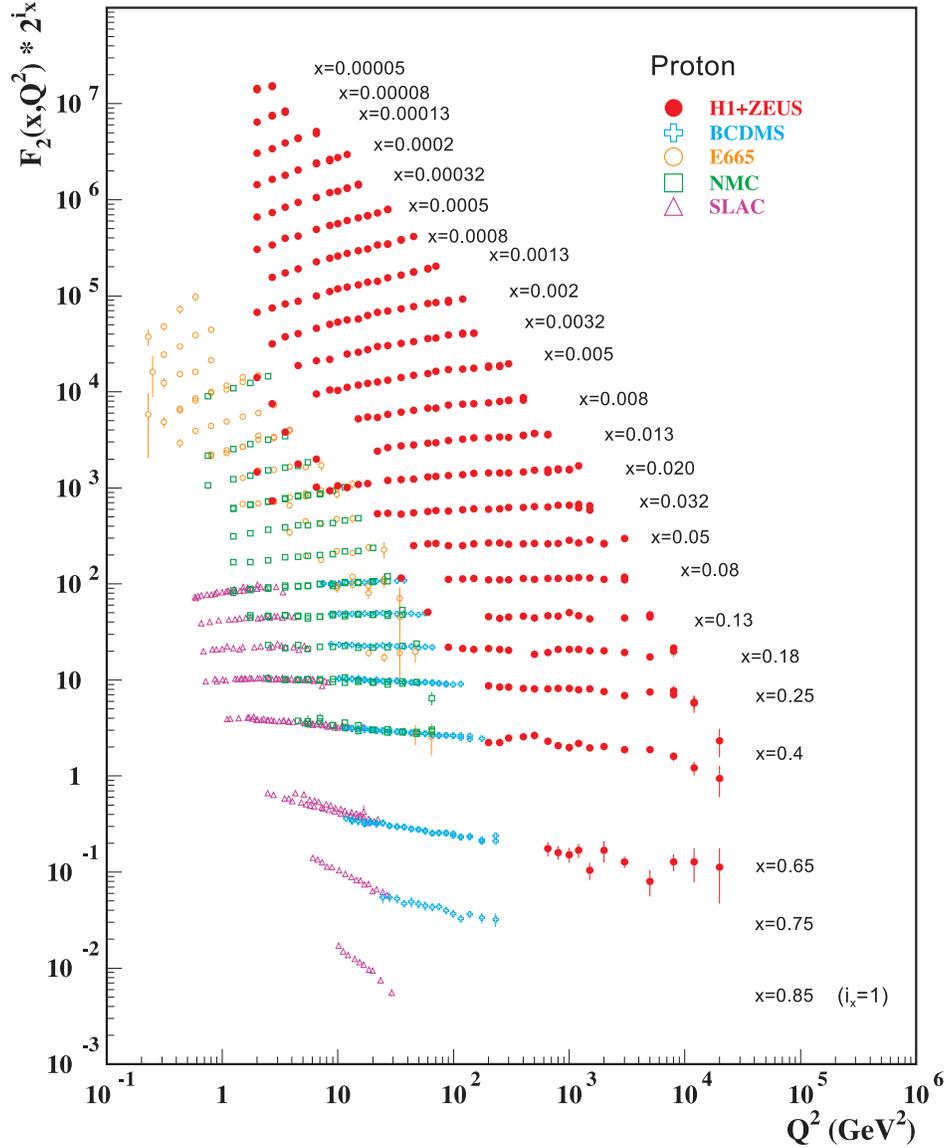


Figure 2.3: The proton structure function F_2^p , as a function of Q^2 in bins of fixed x , measured at the ep -collider experiments ZEUS and H1 [4], and fixed target experiments for ep -scattering at SLAC [5], and μp -scattering at BCDMS [6], E665 [7] and NMC [8]. The error bars are the statistical and systematic uncertainties added in quadrature. Some points have been slightly offset in Q^2 for clarity. The H1+ZEUS combined binning in x is used in this plot; all other data are rebinned to the x values of these data. For the purpose of plotting, F_2^p has been multiplied by 2^{i_x} , where i_x is the number of the x bin, ranging from $i_x = 1$ ($x = 0.85$) to $i_x = 24$ ($x = 0.00005$). Figure from [9].

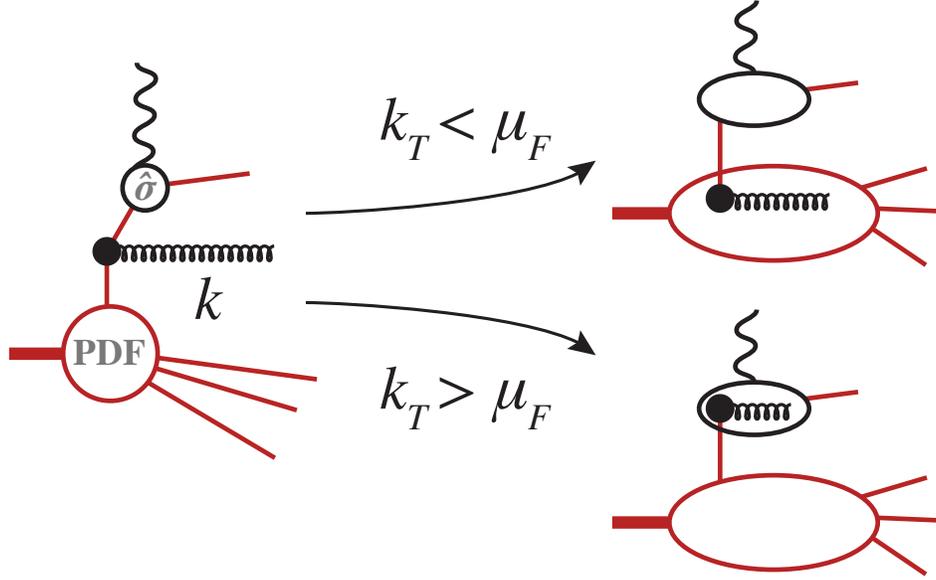


Figure 2.4: Illustration of the property of factorization in Feynman diagram language. An appropriate factorization scale μ_F exists, for which the perturbative expansion of the cross section can be rearranged in such a way that the collinear internal lines are absorbed into the PDFs.

absorbed in the PDF. This is represented in Feynman diagram language in Fig. 2.4. Hence, the property of factorization means that an appropriate **factorization scale** μ_F exists, for which the perturbative expansion of the cross section can be rearranged in such a way that the collinear internal lines are absorbed into $f_q(\xi)^2$. In the case of DIS, the photon virtuality Q constitutes a natural choice for this factorization scale.

The precise way in which the collinear divergences are absorbed in the PDFs, is referred to as the **factorization scheme**. The PDFs are therefore only universal if the same common factorization scheme is used for all processes. The standard choice is the modified minimal subtraction ($\overline{\text{MS}}$) scheme [11–14]. In this scheme, the PDFs retain their probabilistic interpretation when working in the light-cone gauge. The validity of factorization to all orders in α_s has been proved for a limited amount of lepton-hadron processes, including DIS, SIDIS and Drell-Yan (see, e.g., Refs. [11, 13]).

As a consequence of factorization, both $f_q(\xi, \mu_F)$ and $d\hat{\sigma}$ depend on μ_F . Analogous to the renormalization scale μ for α_s , the fact that the cross section cannot depend on the factorization scale gives rise to a set of renormalization group equa-

² In real calculations, the split is accomplished with the aid of dimensional regularization, which is more subtle than a simple division of the integral in two parts.

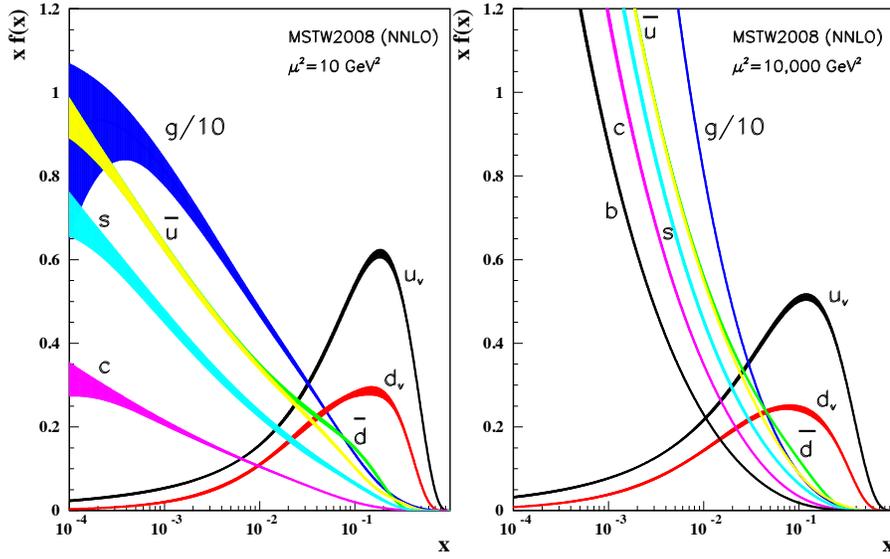


Figure 2.5: Distributions of x times the unpolarized parton distributions $f_q(x)$ (where $f = u_v, d_v, \bar{u}, \bar{d}, s, c, b, g$ and their associated uncertainties using the NNLO MSTW2008 [15] parameterization at scale $\mu_F^2 = 10 \text{ GeV}^2$ and $\mu_F^2 = 10,000 \text{ GeV}^2$. Figure from [9].

tions,

$$\frac{d}{d \ln \mu_F} f_q(x, \mu_F) = \sum_b \int_x^1 \frac{d\xi}{\xi} P_{qb}(x/\xi, \alpha_s(\mu_F)) f_b(\xi, \mu_F). \quad (2.20)$$

Equation (2.20) is known as the DGLAP (Dokshitzer-Gribov-Lipatov-Altarelli-Parisi) equation [16–19]. The splitting functions $P_{qb}(x/\xi)$ specify the probability that a quark or gluon of type b and momentum fraction ξ is the parent of q with momentum fraction x . Note the sum over parton flavor indices: the evolution of, for example, a u -quark ($q = u$) can involve a gluon ($b = g$) through the element P_{ug} of the kernel that describes the gluon splitting into a $\bar{u}u$ pair. The scale variation is a key prediction of pQCD, and it implies that a measurement of the PDFs at an input scale μ_0 automatically constrains their values for at all other scales $\mu_F > \mu_0$.

A common interpretation of the DGLAP evolution is that, with increasing resolution of the virtual photon, a larger number of quarks and gluons can be resolved. The fraction of quarks with low x therefore increases, while the number with high x decreases. Figure 2.5 shows the MSTW2008 PDFs [15], which are determined at the next-to-next-to leading order (NNLO) in α_s . The depletion at large x and increase at small x , are clearly visible in this figure.

2.5 Semi-inclusive DIS and the Fragmentation Process

Isolated quarks have never been observed in nature, due to the property of confinement in QCD. When a quark or antiquark is ejected from a bound state of quarks and gluons by the absorption of a high-energy photon, as it separates from the ensemble, a shower or “jet” of hadrons is produced. This process can be understood in terms of the Lund string fragmentation model [20–22], which is based on the original chain-fragmentation model by Field and Feynman [23]. In this model, the fragmentation process proceeds through the generation of additional quark-antiquark pairs from the color field, which eventually combine with the original quark or antiquark and with each other until a configuration of color-singlet multi-quark states is reached.

Because of the long distance scale of this process, it cannot be calculated perturbatively. The concept of factorization in distance scales can be readily extended to include this hadronization process. In the collinear case, the process can be described by the fragmentation functions (FFs) $D_q^h(z, Q^2)$, which can be interpreted as the number density of hadrons h produced by fragmentation of a struck quark or antiquark q^3 . The FFs follow from a QCD description of the fragmentation process, and as such must follow the same symmetry relations as QCD. This topic is discussed in Section 8.2.1 and Chapter 9.

The scale dependence of the FFs is calculable in pQCD, and governed by renormalization group equations very similar to those for PDFs [14, 24–27]. However, due to singular behavior of the evolution kernels at low z , as well as hadron mass effects and other power corrections of the order M/Q , generally referred to as “higher-twist effects”⁴, the range of applicability of FFs is intrinsically bound to $z > z_{\min}$. A sufficient restriction to avoid these issues is given by $z > 0.05$ for pions and $z > 0.1$ for kaons [29]. This unfortunately prevents the use conservation rules, such as momentum conservation of the fragmenting parton in the hadronization process,

$$\sum_h \int_0^1 dz z D_q^h(z, Q^2) = 1 \quad (2.21)$$

as a viable constraint in a global fit to determine the FFs.

³ See Chapter 7 for a definition that includes the transverse-momentum dependence.

⁴ Using the definition of “twist” from Ref. [28].

2.5.1 The SIDIS Cross Section

The SIDIS cross section for the production of a hadron h ,

$$l + N \rightarrow l' + h + X, \quad (2.22)$$

is proportional to certain combinations of both PDFs and FFs. It can be written in terms two structure functions in a way very similar to the DIS case in Eq. (2.15) [12, 29–32],

$$\frac{d\sigma^h}{dx dy dz} = \frac{2\pi\alpha^2}{Q^2} \left[\frac{1 + (1-y)^2}{y} 2F_1^h(x, z, Q^2) + \frac{2(1-y)}{y} F_L^h(x, z, Q^2) \right]. \quad (2.23)$$

It should be noted that Eq. (2.23) only applies to hadron production in the **current** fragmentation region, i.e., it only describes the fragmentation of the struck quark q . This topic is discussed in detail in Chapter 8.

The structure functions F_1^h and F_L^h are given at LO in α_s by,

$$2F_1^h(x, z, Q^2) = \sum_q e_q^2 f_q(x, Q^2) D_q^h(z, Q^2) \quad (2.24)$$

$$F_L^h(x, z, Q^2) = 0. \quad (2.25)$$

The diagram in Fig. 2.2 provides a graphical representation of this LO formalism. At NLO in α_s , the structure functions are given by,

$$2F_1^h(x, z, Q^2) = \sum_q e_q^2 \left(f_q(x, Q^2) D_q^h(z, Q^2) \right. \quad (2.26)$$

$$\left. + \frac{\alpha_s(Q^2)}{2\pi} \left[q \otimes C_{qq}^1 \otimes D_q^h \right. \right. \quad (2.27)$$

$$\left. + q \otimes C_{gq}^1 \otimes D_g^h + g \otimes C_{qg}^1 \otimes D_q^h \right] (x, z, Q^2), \quad (2.28)$$

$$F_L^h(x, z, Q^2) = \frac{\alpha_s(Q^2)}{2\pi} \sum_q e_q^2 \left[q \otimes C_{qq}^L \otimes D_q^h \right. \quad (2.29)$$

$$\left. + q \otimes C_{gq}^L \otimes D_g^h + g \otimes C_{qg}^L \otimes D_q^h \right] (x, z, Q^2), \quad (2.30)$$

with \otimes denoting a standard convolution. The NLO $\overline{\text{MS}}$ coefficient functions $C_{ab}^{1,L}$ can be found in, e.g., Refs. [12, 30–32].

2.5.2 Current Knowledge of FFs

The available knowledge for the PDFs is increasingly precise, where the modern sets agree extremely well with each other, within their small estimated uncertainties [15, 37, 38]. While evolving rapidly, the parameterizations for the FFs are

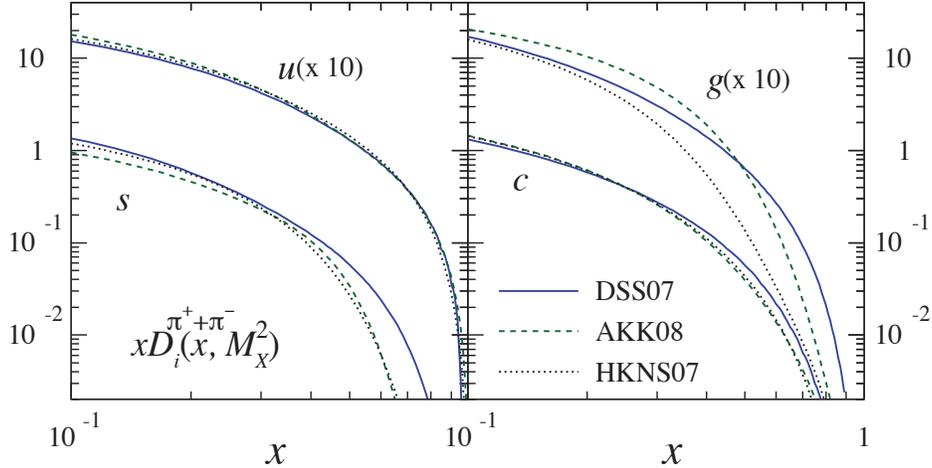


Figure 2.6: Comparison of u , s , c and g NLO fragmentation functions for $\pi^+ + \pi^-$ at the mass of the Z boson. The different lines correspond to the results of the most recent analyses performed by DSS07 [29, 33], AKK08 [34, 35] and HKNS07 [36]. Figure from [9].

much less precisely constrained. The discrepancy in Fig. 2.6 between the results of the most recent NLO extractions [29, 33–36] provides a first estimate of the uncertainties involved. The differences for other species of hadrons like kaons and protons are even larger.

Most extractions, e.g. in Refs. [34–36, 61, 62] are largely constrained by high statistic measurements from electron-positron annihilation into charged hadrons,

$$e^+e^- \rightarrow q + \bar{q} \rightarrow h + X \quad (2.31)$$

from LEP and SLAC [63–67]. A summary of the wealth of available information from this channel is summarized in Fig. 2.7. While they provide a very clean access in the sense that the cross sections have no dependence on PDFs, these data cannot disentangle quark and anti-quark fragmentation as they always refer to the charge-sum of certain hadron species (e.g., $\pi^+ + \pi^-$). Additionally, the vast majority of the electron-positron annihilation data were taken at the mass scale of the Z -boson, where all electroweak couplings become roughly equal. Therefore only flavor singlet combinations of FFs can be determined.

In recent years, the database has been expanded with the inclusion of results for inclusive single-hadron production with high-transverse momentum in proton-proton collisions at RHIC,

$$p + p \rightarrow h + X. \quad (2.32)$$

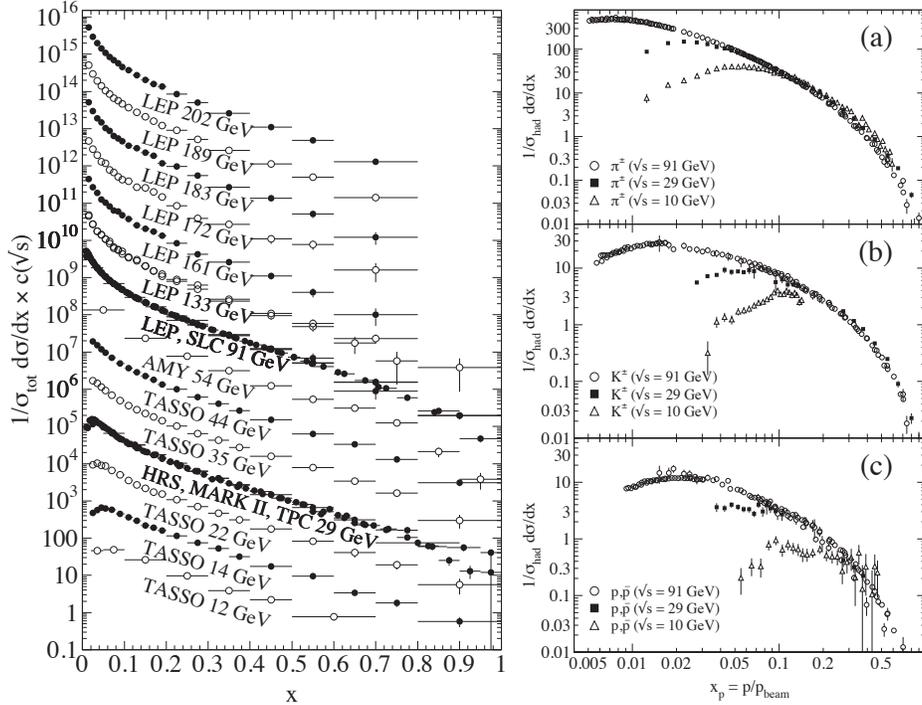


Figure 2.7: Fragmentation functions from e^+e^- annihilation as a function of the scaled momentum x . The left panel shows the distributions from Refs. [39–55] for all charged particles in various ranges of \sqrt{s} . The distributions were scaled by $c(\sqrt{s}) = 10^i$ with i ranging from $i = 0$ ($\sqrt{s} = 12$ GeV) to $i = 13$ ($\sqrt{s} = 202$ GeV). The right panels show the distributions from Refs. [51, 55–60] for (a) π^\pm , (b) K^\pm and (c) p, \bar{p} at $\sqrt{s} = 10, 29$ and 91 GeV. Figure from [9].

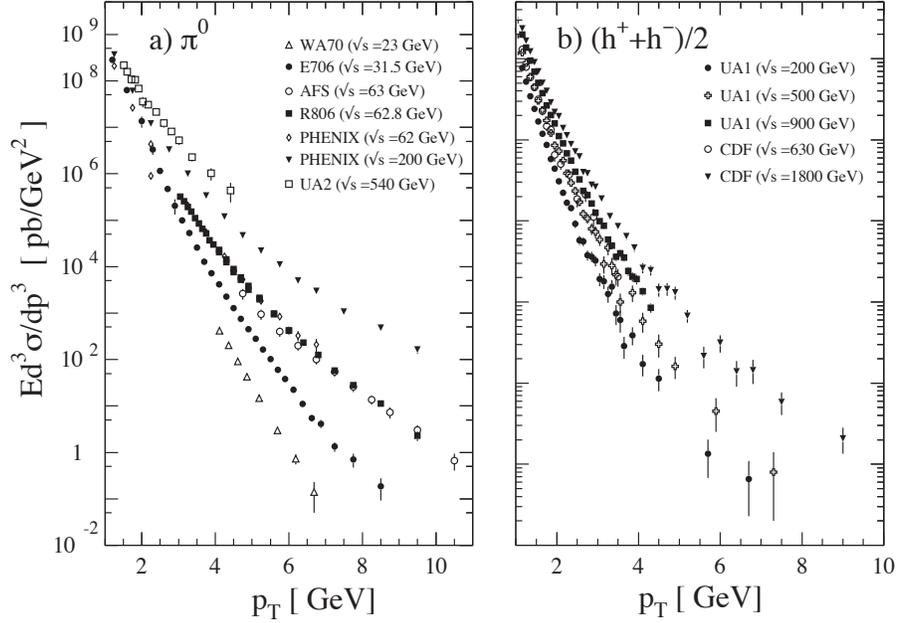


Figure 2.8: Selection of inclusive **(a)** π^0 and **(b)** charged-hadron production data from pp [68–71] and $p\bar{p}$ [72–74] collisions as a function of transverse momentum p_T . Figure from [9].

These include measurements of the transverse-momentum spectra of neutral pions at central rapidities at PHENIX [68], and at forward rapidities with STAR [75], as well as similar measurements for identified pions and kaons for forward rapidities at BRAHMS [76]. A selection these results is shown in Fig. 2.8. These data are particularly sensitive to the gluon fragmentation function through the dominant channel $g + g \rightarrow g + X$, while providing additional information on the flavor separation of the fragmentation functions.

Factorization of the SIDIS cross section in terms of PDFs and FFs has been proved to all orders of α_s (see, e.g., Refs. [11, 13]). The SIDIS process can therefore be used to help constrain FFs. The multi-dimensional multiplicities for pseudoscalar mesons,

$$M_n^h(x, Q^2, z, P_{h\perp}) \equiv \frac{1}{\frac{d^2\sigma_{\text{DIS}}(x, Q^2)}{dx dQ^2}} \int_0^{2\pi} d\phi_h \frac{d^5\sigma^h(x, Q^2, z, P_{h\perp}, \phi_h)}{dx dQ^2 dP_{h\perp} d\phi_h}, \quad (2.33)$$

presented in this work, were extracted using the 27.6 GeV lepton beam of the HERA storage ring at DESY, which operated with electrons or positrons. The extraction of multiplicities of pions and kaons separately for positive and negative charge provides sensitivity to the individual quark and antiquark flavors in the fragmentation process, (cf. the expression for the SIDIS cross section in Section 2.5.1).

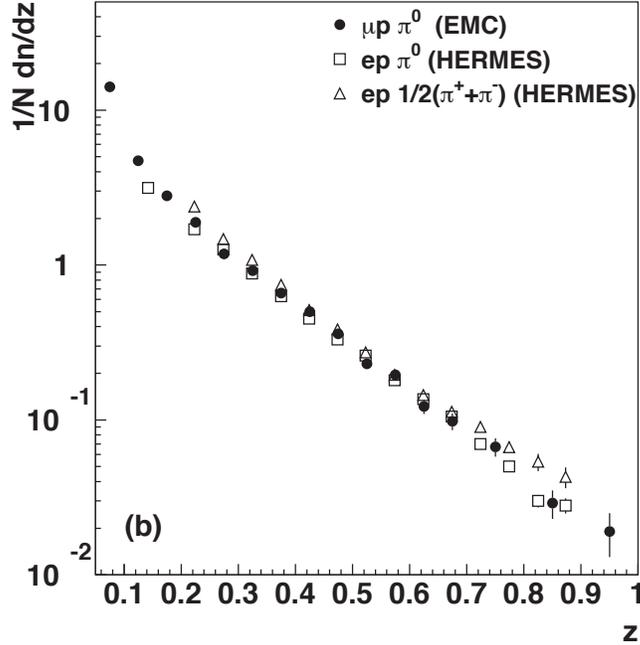


Figure 2.9: SIDIS multiplicities for pions as a function of z from the fixed-target DIS experiments EMC [77] and HERMES [78]. The HERMES points will be superseded by the higher-precision results in this work. Figure from [9].

The data presented here for proton and deuteron targets are the most precise results for multiplicities currently available at this energy scale. They greatly expand the global database SIDIS multiplicities, of which a selection is shown in Fig. 2.9. In particular, the data extend those of an earlier HERMES publication [78]). A very preliminary version of a subset of these data has already been used in a global analysis of FFs [29], in addition the available global database discussed in this section. The impact these preliminary results had, as well as their difference with the official HERMES results is discussed in Chapter 6.

It should be noted that, through the inclusion of the dependence on $P_{h\perp}$, these results reach beyond the standard framework of collinear factorization. The topic of transverse-momentum dependence of the fragmentation process is discussed in Chapter 7.

3 The HERMES Experiment

The HERMES experiment was located at the DESY research center in Hamburg, Germany. It shared the longitudinally polarized electron or positron beam beam of the HERA electron-proton collider with the H1 and ZEUS experiments.

It was originally conceived to measure the spin-dependent structure functions of the proton and neutron, triggered by the observations that only a small fraction of the nucleon spin can be attributed to the valence quarks by the EMC experiment [79, 80]. This required an experimental setup with a longitudinal beam and target polarizations, two novel technological achievements at HERMES. Since the commissioning in 1995, the scope of the experiment was expanded to include many topics in the field of hadron structure, hadron formation and hadronic interactions.

The following chapter provides a brief overview of the experimental setup, while highlighting those parts that are necessary to understand the results in this work. A detailed discussion of the HERMES spectrometer can be found in Ref. [81].

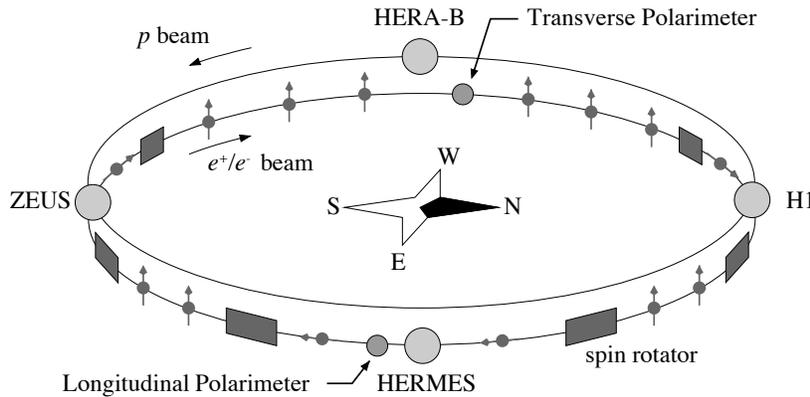


Figure 3.1: Schematic view of the HERA ep -collider.

3.1 The HERA ep -collider

The HERA facility consists of two independent storage rings with a circumference of 6.3 km, for leptons¹ and protons. Figure 3.1 shows an overview of the facility, which was decommissioned in 2007. The H1 and ZEUS experiments are located at the interaction points of both beam lines, while HERMES used only the lepton beam.

The leptons (protons) were pre-accelerated to 0.45 GeV (0.05 GeV) in the linear accelerator LINAC II (III), and accelerated further to 14 GeV (40 GeV) in the DESY II (III) and PETRA storage rings. The final energy of 27.7 GeV (920 GeV) is reached in the HERA storage ring.

Initial lepton currents reached values up to 50 mA. Interactions with residual gas in the beam line, as well as with the experiments leads to an exponential decrease in the beam current. To optimize the available luminosity, the lepton beam is therefore when it reaches a current of approximately 10 mA. This corresponds to a life time of about 12 h to 14 h. In practice, the beam dump was often accomplished by injecting the HERMES target cell with a high-density unpolarized gas.

The lepton beam is self-polarizing due to the Sokolov-Ternov effect [82]. Small differences in the spin flip amplitude for the emission of synchrotron radiation leads to a large transverse spin polarization of the leptons. Depending on the HERA operational conditions, polarization values up to 50 % can be achieved. Spin rotators [83, 84] positioned at either side of the HERMES spectrometer are used to obtain a longitudinal beam polarization. During the course of the experiment, both polarization states are provided by changing the field direction of the spin rotators.

3.2 The HERMES Target

The HERMES experiment uses a gas target in order to minimize the impact on the lifetime of the HERA lepton beam. It is able to operate with polarized hydrogen and deuteron with a polarization of about 80 %. In addition, it can use a variety of unpolarized gasses, H₂, D₂, He, N₂, Ne, Kr and Xe.

Gas targets have the benefit of a very low dilution due to undesired materials, as well as a very high polarization than can be flipped within milliseconds, when compared to solid targets. The main disadvantage lies in the very low target areal density of the order of 2×10^{11} atoms/cm², compared to 10^{25} atoms/cm² for a solid target.

In order to enhance the target areal density, the gas is injected in a storage cell collinear to the lepton beam [85, 86]. A schematic view of this storage cell is

¹ In this chapter, lepton refers to either a positron or an electron.

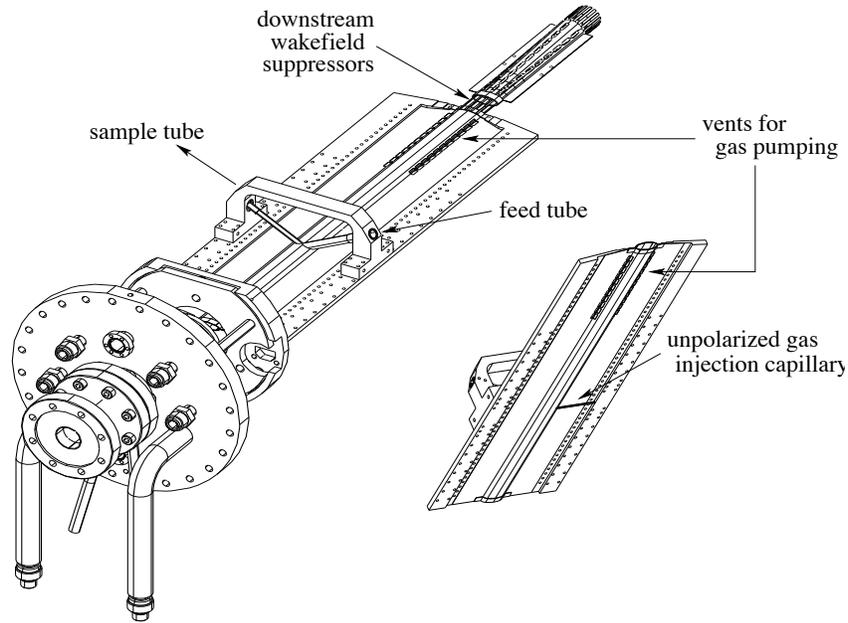


Figure 3.2: Schematic view of the HERMES storage cell.

presented in Fig. 3.2. It is constructed out of a very thin ($75\ \mu\text{m}$) aluminum sheets to minimize particle interactions. It is 400 mm long and has an elliptical cross section of $21 \times 8.9\ \text{mm}^2$. This shape matches the lepton beam profile. The use of such a storage cell constrains the gas to spread along the beam line, increases the areal density by two orders of magnitude. In addition, the storage cell was cooled to 100 K to reduce the thermal velocity of the target atoms. This reduce target depolarization due to recombination, while increasing the maximum possible gas density by a factor of approximately $\sqrt{3}$ compared to room temperature. For deuteron, target densities of $2 \times 10^{14}\ \text{atoms/cm}^2$ have been reached. To suppress depolarization of the target atoms due to wall collisions, the storage cell was coated with a layer of Drifilm [87]. The storage cell ends in a 147 mm long extension to ensure all scattered particles encounter the same amount of target wall material.

Wake field suppressors up- and downstream of the storage cell prevent heating of the target cell due to beam wake fields, while heating due to synchrotron photons is avoided by a system of two collimators.

Gas polarization is obtained using an atomic beam source (ABS) which is described in detail in Ref. [88]. Up to 2000, a longitudinal polarization was maintained by means of a superconducting magnet. This magnet was replaced for the 2002-2005 run by a transverse magnet. In 2006-2007, polarized running was not possible anymore due to the presence of a recoil detector [89–91]. For unpolarized

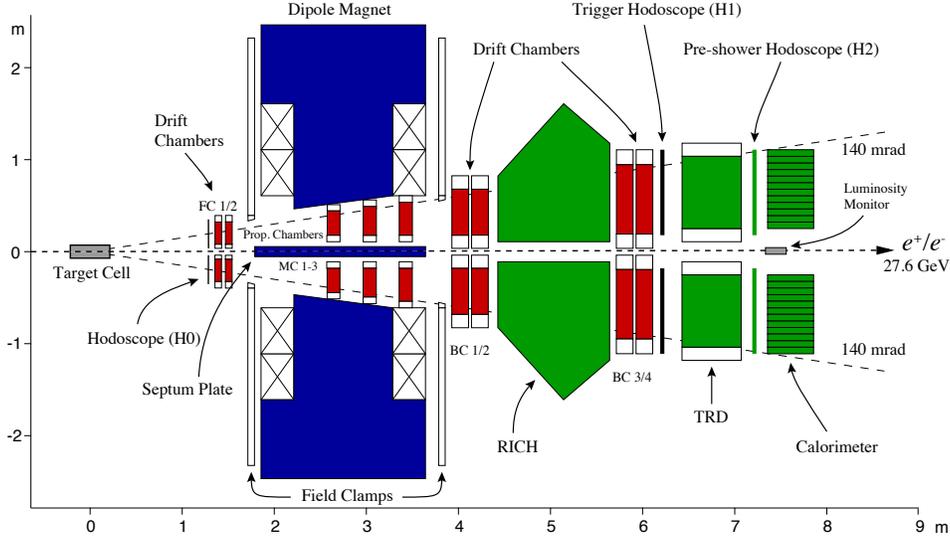


Figure 3.3: Side view of the HERMES spectrometer.

runs, as well as the high-density runs mentioned in Section 3.1, the gas was injected with an unpolarized gas feed system (UGFS). The target gas is continuously monitored by a target gas analyzer (TGA) and a Breit-Rabi polarimeter (BRP) [92].

3.3 The HERMES Spectrometer

The HERMES spectrometer [81] consists of a large dipole magnet for momentum separation, supported by tracking and particle identification detectors. It is a forward-spectrometer with identical top and bottom parts, and is designed for the reconstruction of inclusive and semi-inclusive DIS events. Figure 3.3 presents a side view of the spectrometer.

The beam pipe traverses the center of the spectrometer, providing a natural definition of the z -axis. The y -axis is pointing upwards, while the x -axis lies in the horizontal plane, pointing outwards, i.e. away from the HERA center. The origin is located in the center of the target cell. The spectrometer is therefore mirror-symmetric in the xz -plane.

The beam pipes are shielded from the 1.5 T m field of the spectrometer magnet by means of large steel septum plates. The angular acceptance of the spectrometer, constrained by the spectrometer magnet and the septum plates, is then given by $|\theta_x| < 170$ mrad in the horizontal plane, 40 mrad $< |\theta_y| < 140$ mrad in the vertical plane.

3.3.1 Track Reconstruction

A system of drift chambers is used to determine the origin, momentum and trajectory for charged particles. The vertex position and polar and azimuthal scattering angles are constrained by the front chambers (FC1/2). The spectrometer magnet provides an integrated deflection power of 1.5 T m, causing the charged particle trajectories to bend in the horizontal plane. The deflected trajectory is measured by two pairs of drift chambers (BC1/2 and BC3/4), which are installed respectively in front of, and behind, the ring imaging Čerenkov detector. The deflection angle determines the particle momentum and charge. This HERMES reconstruction program (HRC) accomplishes this by means of a pattern-matching algorithm [93], with a resolution of $\delta\theta < 1.8$ mrad and $\delta p/p < 2.6\%$.

From 2002 to 2005, a transversely polarized target was used. This affected the momentum reconstruction because the magnetic holding field is not accounted for in HRC. This is corrected for by the transverse magnet correction (TMC) software.

3.3.2 Particle Identification

A very clean lepton-hadron separation is essential for the reconstruction of inclusive DIS events. Furthermore, semi-inclusive measurements require that different hadron types can be identified.

The particle identification (PID) system at HERMES consists of a lead-glass calorimeter, a pre-shower hodoscope and a transition-radiation detector (TRD) and a ring imaging Čerenkov detector (RICH). The signals of all four detectors are combined for the lepton-hadron separation, while hadrons are subsequently identified with from their signal in the RICH. The PID algorithms are discussed in Section 4.3.

Lead-glass Calorimeter

The electromagnetic calorimeter, illustrated in Fig. 3.4, Each calorimeter half consists of a 42×10 lead-glass block array, Particles incident on the calorimeter deposit part of their energy in form of electromagnetic showers in the material. These showers are read out by means of photo-multiplier tubes (PMTs).

Electrons and positrons deposit almost their entire energy in the calorimeter blocks of about 18 radiation lengths. This yields a ratio of deposited energy over reconstructed momentum $E/p \approx 1$. Hadrons only deposit a fraction of their energy through ionisation energy loss. The resulting hadron-rejection factor is 100.

Besides lepton-hadron separation, the calorimeter also provides a fast first-level trigger for scattered leptons, and allows for the determination of the energy of

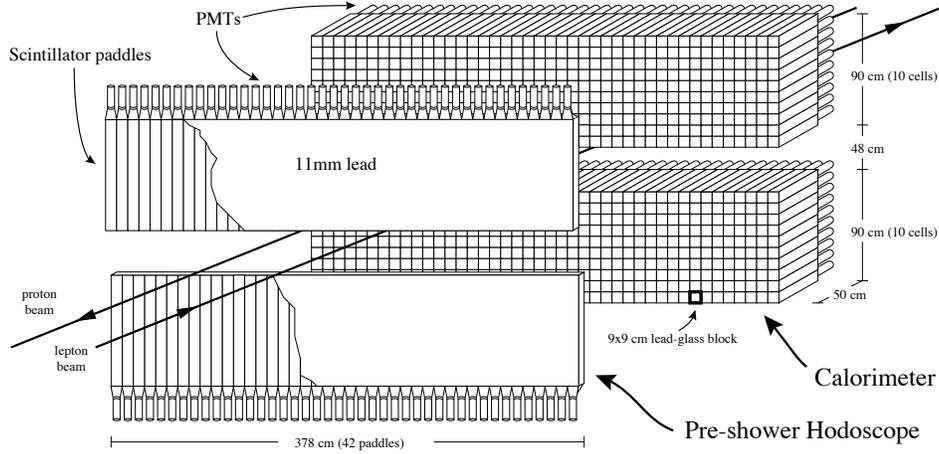


Figure 3.4: Illustration of the combined calorimeter (back) and the pre-shower (front) system. (front).

real photons originating, e.g., from π^0 decay. This feature of the calorimeter is leveraged in Chapter 9.

The calorimeter energy response for lepton showers is found to be linear in the range 1 GeV to 30 GeV, with an energy resolution that can be parameterized as [94],

$$\frac{\sigma(E)}{E} [\%] = \frac{5.1 \pm 1.1}{\sqrt{E} [\text{GeV}]} + (2.0 \pm 0.5) + \frac{(10.0 \pm 2.0)}{E [\text{GeV}]} \quad (3.1)$$

The lepton hit position on the calorimeter surface can be obtained with a resolution $\sigma_x = \sigma_y \approx 7$ mm by means of the energy-weighted average of the nine blocks containing the electromagnetic shower.

Pre-shower Hodoscope

The pre-shower hodoscope H2, illustrated in Fig. 3.4 consists of a scintillator hodoscope preceded by two radiation lengths of lead. Hadrons behave as minimum ionizing particles in this detector, only depositing a small fraction of their energy. Leptons start their electromagnetic shower in the lead curtain, depositing on average ten times more energy in the scintillator. It provides a hadron-rejection factor of 10 at an efficiency of about 95 %.

Transition Radiation Detector (TRD)

Transition radiation is emitted when charged particles with relativistic energies traverse the boundary between two media with different dielectric constants. The energy of the emitted radiation is proportional to its Lorentz factor $\gamma = E/mc$ [95].

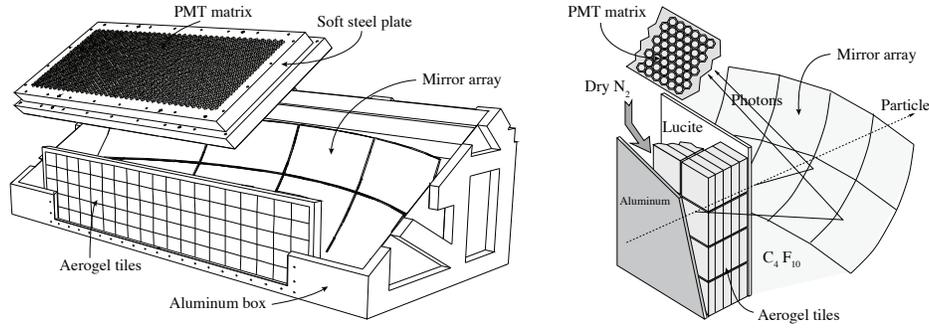


Figure 3.5: The left panel shows the full configuration of the upper RICH detector, while the right panel provides a more schematic overview.

The radiated energy for hadrons and leptons at the same energy therefore differs by several orders of magnitude. Due to the low emission probability, the TRD is composed of six modules, each consisting of a polyethylene fiber radiator and adjacent proportional wire chamber. Both the hadrons and leptons produce a signal in the wire chambers, the lepton energy deposition being approximately twice as large due to the transition radiation. Combining the information of all six modules with a truncated mean method results in a hadron-rejection factor of over 100 at an efficiency of about 90 %.

Ring Imaging Čerenkov Detector (RICH)

Charged particles that traverse a medium with a velocity higher than the speed of light in the medium emit electromagnetic Čerenkov radiation. The angle with which this radiation is emitted depends on the refractive index of the radiators and the **velocity** of the particle. Different particles therefore have a different **momentum** dependence of the Čerenkov angles θ_c due to their respective masses.

The HERMES RICH detector, shown in Fig. 3.5, leverages this principle to allow for the separation of pions, kaons and protons in the momentum range 2.0 GeV to 15.0 GeV. To cover this range, the detector combines a combination of two radiators: a clear silica aerogel, and a C_4F_{10} gas radiator. The resulting distribution of the Čerenkov angles is shown in Fig. 3.6. See Ref. [96] for more details.

3.3.3 Trigger and Data Acquisition

The data acquisition system reads out the detector signals when one of the triggers fire. The trigger to filter out DIS events from the background requires that,

1. There is a coincident signal in the hodoscopes H0, H1 and H2

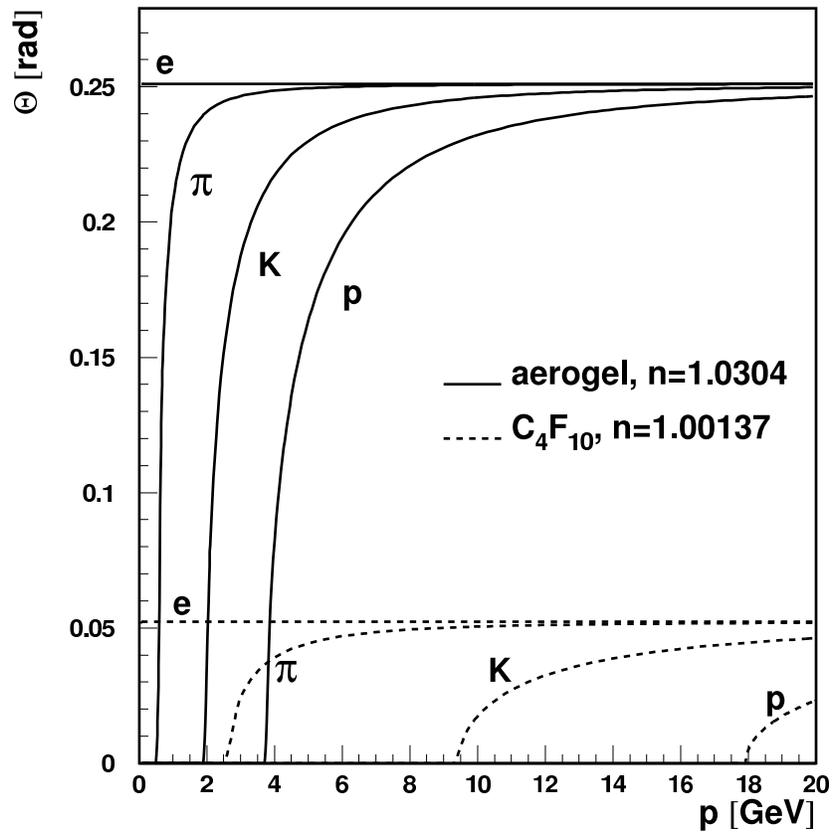


Figure 3.6: The hadron identification with the RICH detector leverages the momentum dependence of the Čerenkov angle θ for different particle masses.

2. The energy deposition in the calorimeter exceeds a certain threshold. The values of this threshold are 1.4 GeV for polarized and 3.5 GeV for unpolarized target operation. This signal is typically caused by the scattered beam particle.
3. A reasonable timing of the signals. This requirement filters out, e.g., particle showers caused by the proton beam, which travel backwards through the spectrometer.

Due to the large bunch spacing in HERA, and the relatively low luminosity, the occurrence of multiple simultaneous DIS events in the spectrometer is negligible [81].

4 Multiplicity Analysis

Both the proton and deuteron multiplicities are extracted from three different datasets, taken in 2000, 2004 and 2005. Some base characteristics are presented in Table 4.1. The high statistics of these datasets, together with the powerful particle identification capabilities of the RICH detector allow in principle for a highly precise measurement of charge and flavor separated SIDIS multiplicities in a multidimensional binning.

The small size of the statistical uncertainties on the raw measurement leads to a result that is inherently systematics dominated, complicating the extraction of the final multiplicities significantly. This requires a sophisticated analysis that suppresses the systematic uncertainty where possible, while providing a careful estimate of the remaining systematic uncertainty (Section 4.9).

The final results are corrected RICH misidentification (Section 4.3.2), contamination by charge-symmetric background events (Section 4.4.1), trigger inefficiencies (Section 4.4.2), QED radiative effects, limited acceptance and detector smearing (Section 4.6) and contamination by exclusive vector mesons (Section 4.7). During the entire analysis, all six datasets are treated separately to correctly account for slightly different experimental conditions. The data combination occurs only in the final stages of the multiplicity extraction.

This chapter describes the entire analysis chain from the candidate DIS events to the fully corrected final multiplicity results.

Target	Production	Polarization	Statistics
H	2000 (e1)	unpol.	5.0×10^6 DIS
	2004 (c1)	transv. pol.	1.2×10^6 DIS
	2005 (c2)	transv. pol.	4.0×10^6 DIS
D	2000 (e1)	long. pol.	5.8×10^6 DIS
	2004 (d2)	unpol.	0.8×10^6 DIS
	2005 (d2)	unpol.	4.6×10^6 DIS

Table 4.1: Data productions

4.1 Data Selection

The data productions in Table 4.1 were chosen because, for each of the productions, the HERMES spectrometer was running with a consistent and well-understood setup.

The data collected before 2000 either do not have enough statistics to allow for the sophisticated correction method described in Section 4.6 (1998-1999), or lack the precise hadron identification first introduced by the RICH detector (prior to 1998).

The data collected in 2006-2007, while similar in size to the 2000-2005 data, have systematic uncertainties that are less well-understood due to major changes in the spectrometer. A recoil detector was added, resulting in a changed acceptance, as well as small fringe magnetic fields that were detectable up to the electromagnetic calorimeter at the back of the spectrometer. It was decided that the small (factor of two) increase in statistical precision did not warrant the large systematic complications that would be introduced by these datasets.

Within the chosen data samples, the data has to be selected to ensure absolute data integrity. This means that, for the entire data sample, the HERMES spectrometer was consistently running within normal operational parameters. This was accomplished using the data collected by the **slow control** part of the data acquisition system, that periodically writes out the experimental conditions (target mode, beam conditions, etc.), combined by the result of several dedicated data quality studies. The time scale of the data quality, called a **burst** is defined by the period of the slow control system.

For this analysis in particular, the following values were checked for each burst:

1. the burst length has to be reasonable
2. the trigger dead time has to be reasonable
3. the beam current has to be reasonable
4. the burst cannot be marked as a bad burst by the data quality experts (this also ensures that the spectrometer was minimally functional)
5. the particle identification system has to be functioning correctly
6. there cannot be any dead blocks in the calorimeter, preshower hodoscope and luminosity monitor
7. the TRD has to be functioning correctly
8. the RICH has to be functioning correctly

1. Trigger method	21
2. Tracking method	NOVC
3. Event momentum	$P_{\text{tot}} < 28 \text{ GeV}$

(a) Event selection

4. Vertex	$ z_{\text{vert}} < 18 \text{ cm}$ $d_{\text{vert}} < 0.75 \text{ cm}$
5. Septum plate	$y_{\text{sept}} > 7 \text{ cm}$
6. Front field clamp	$ x_{\text{ffc}} < 31 \text{ cm}$
7. Rear field clamp	$ x_{\text{rfc}} < 100 \text{ cm}$ $ y_{\text{rfc}} < 54 \text{ cm}$
8. Calorimeter	$ x_{\text{cal}} < 175 \text{ cm}$ $ y_{\text{cal}} > 30 \text{ cm}$ $ y_{\text{cal}} < 108 \text{ cm}$

(b) Track selection

Table 4.2: Experimental cuts

4.2 Experimental Cuts

The cuts in Table 4.2, accept tracks that originate from the target cell (4) and passed through the fiducial volume of the HERMES spectrometer (5-8). In particular (8) rejects tracks that passed too close to the edges of the calorimeter, ensuring that the entire electromagnetic shower is contained in the lead-glass blocks. The event-level cuts (1-3) ensure data consistency by checking the trigger values and track reconstruction method.

4.3 Particle Identification

This analysis is made possible by the precise particle identification (PID) capabilities of the HERMES spectrometer. The PID process is split in two parts: lepton-hadron separation, and hadron type identification. The lepton-hadron separation is performed combining the signals of four detectors: a lead-glass calorimeter, a pre-shower hodoscope, a transition-radiation detector (TRD) and a dual-radiator RICH (ring imaging Čerenkov detector). The hadrons are then further identified as pions, kaons or protons using the opening angles of the Čerenkov photons in the RICH.

See Section 3.3.2 for a description of these detectors.

4.3.1 Lepton-Hadron Separation

Define the conditional probability that a track is a lepton (hadron), given the momentum p , polar angle θ and energy response $\vec{E} \equiv E_1, \dots, E_N$ in PID detectors D_1, \dots, D_N can be written as $P(l(h) | \vec{E}; p, \theta)$. The PID value is defined as the logarithmic ratio of these conditional probabilities,

$$\text{PID} \equiv \log_{10} \left(\frac{P(l | \vec{E}; p, \theta)}{P(h | \vec{E}; p, \theta)} \right). \quad (4.1)$$

If the observed particle is more consistent with the lepton (hadron) hypothesis, the PID value will be positive (negative). This is the basis of the lepton-hadron separation at HERMES.

Application of Bayes' theorem on numerator and denominator of Eq. (4.1) yields,

$$\text{PID} = \log_{10} \frac{P(l | p, \theta) \cdot P(\vec{E} | l, p)}{P(h | p, \theta) \cdot P(\vec{E} | h, p)} \quad (4.2)$$

$$= \log_{10} \frac{P(l | p, \theta)}{P(h | p, \theta)} + \log_{10} \left(\prod_i \frac{P_{D_i}(E_i | l, p)}{P_{D_i}(E_i | h, p)} \right). \quad (4.3)$$

The first term in Eq. (4.3) is the ratio the **prior** probabilities $P(l(h) | p, \theta)$ of having a lepton (hadron) with a given p and θ . In this case, the prior probabilities are equivalent to the hadron fluxes ϕ_l and ϕ_h . Defining the flux factor Φ as ϕ_h/ϕ_l , the first term of Eq. (4.3) can be replaced with,

$$\log_{10} \frac{P(l | p, \theta)}{P(h | p, \theta)} \equiv -\log_{10} \Phi(p, \theta). \quad (4.4)$$

More details regarding the flux factor and the iterative approach to its calculation in HERMES can be found in Ref. [97].

The second term of Eq. (4.3) depends on the probability $P_D(E | l(h), p)$ that a lepton (hadron) with momentum p deposits E in detector D is commonly referred to as the **parent distribution**. In principle, these parent distributions could be determined using various lepton and hadron test-beams. In order to account for varying running conditions, changes in the PID system and detector aging, the parent distributions at HERMES have been extracted from their respective data sample through hard PID cuts on the other detectors.

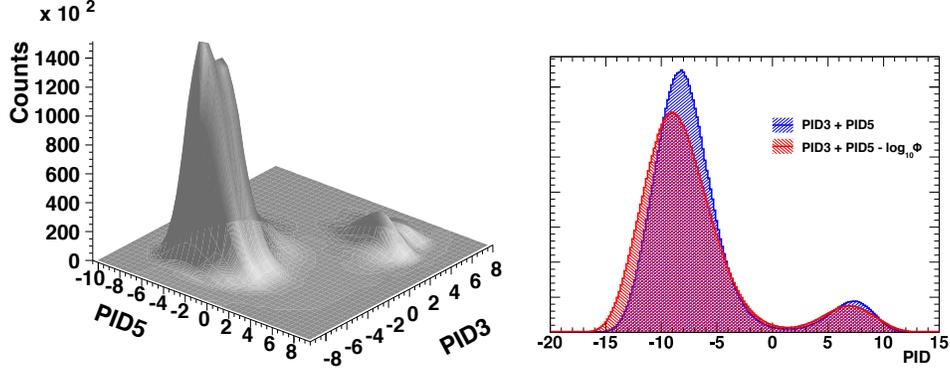


Figure 4.1: Result of the lepton-hadron PID algorithm. The large peak corresponds to identified hadrons, while the smaller peak at positive PID values corresponds to identified leptons. As can be seen from the right panel, the impact of the flux factor Φ is relatively small. Figures from [98] and [99].

Introducing,

$$\text{PID}_D \equiv \log_{10} \frac{P_D(E|l, p)}{P_D(E|h, p)}, \quad (4.5)$$

and using Φ from Eq. (4.4), Eq. (4.3) can be rewritten as,

$$\text{PID} = \sum_i \text{PID}_{D_i} - \log_{10} \Phi. \quad (4.6)$$

In HERMES, the PID values for the different detectors are combined in two groups,

$$\text{PID3} \equiv \text{PID}_{\text{cal}} + \text{PID}_{\text{pre}} + \text{PID}_{\text{RICH}}, \quad (4.7)$$

$$\text{PID5} \equiv \text{PID}_{\text{TRD}} = \sum_i \text{PID}_{\text{TRD}_i}, \quad (4.8)$$

where the sum in Eq. (4.8) runs over the six TRD modules.

Using these definitions in Eq. (4.6), the formula for lepton-hadron separation at HERMES finally becomes,

$$\text{PID} = \text{PID3} + \text{PID5} - \log_{10} \Phi. \quad (4.9)$$

Fig. 4.1 shows the lepton-hadron separation in action. Overall, the lepton identification efficiency at HERMES is 98 %, with a hadron contamination much smaller than 1 %. The identification of hadrons in coincidence with a scattered lepton has an efficiency of 99 % with a lepton contamination smaller than 1 %. See Refs. [97, 100] for a detailed description of the HERMES lepton-hadron PID framework.

9. Hadron Momentum	$P_h > 2 \text{ GeV}$ $P_h < 15 \text{ GeV}$
10. EVT Quality	$rQp > 0$

Table 4.3: Hadron PID cuts

4.3.2 Determining Hadron Flavor

The opening angle of the Čerenkov photons hitting the PMT matrix in the RICH detector provides information on the hadron type. An event-level direct ray tracing algorithm (EVT) simulates the detector response for all possible track hypotheses. Comparing the likelihoods L_H for the observed RICH response to originate from event hypothesis H determines whether a hadron track is identified as a pion, kaon, proton, or other (referred to as X). The EVT algorithm is discussed in detail in Refs. [101, 102].

The momentum dependence of the Čerenkov angles allows for hadron identification between 2 GeV (1 GeV for pions) and 15 GeV. This enforces an additional momentum cut on the hadron tracks. Additionally, checking the EVT quality parameter rQp , defined as the (logarithmic) ratio of the likelihoods of the two most likely hypothesis, rejects tracks where the hadron identification result is ambiguous. These cuts are summarized in Table 4.3.

RICH Unfolding

The efficiency of the hadron identification can be encoded in a set of 3×4 matrices as a function of momentum, particle charge and event topology (number of tracks in one detector half). These matrices relate the fluxes for the identified particle type I_h to the true hadron fluxes T_h ,

$$\begin{pmatrix} I_\pi \\ I_K \\ I_p \\ I_X \end{pmatrix} = \begin{pmatrix} P_\pi^\pi & P_K^\pi & P_p^\pi \\ P_K^K & P_K^K & P_p^K \\ P_p^p & P_K^p & P_p^p \\ P_X^X & P_K^X & P_p^X \end{pmatrix} \cdot \begin{pmatrix} T_\pi \\ T_K \\ T_p \end{pmatrix}, \quad (4.10)$$

with P_i^j the probability that a hadron of true type t is identified as a hadron of type i . These matrices are extracted from a Monte-Carlo simulation of the detector, by comparing the results of the PID algorithm with the true particle type. As example, the matrices for single-track events are shown in Fig. 4.2.

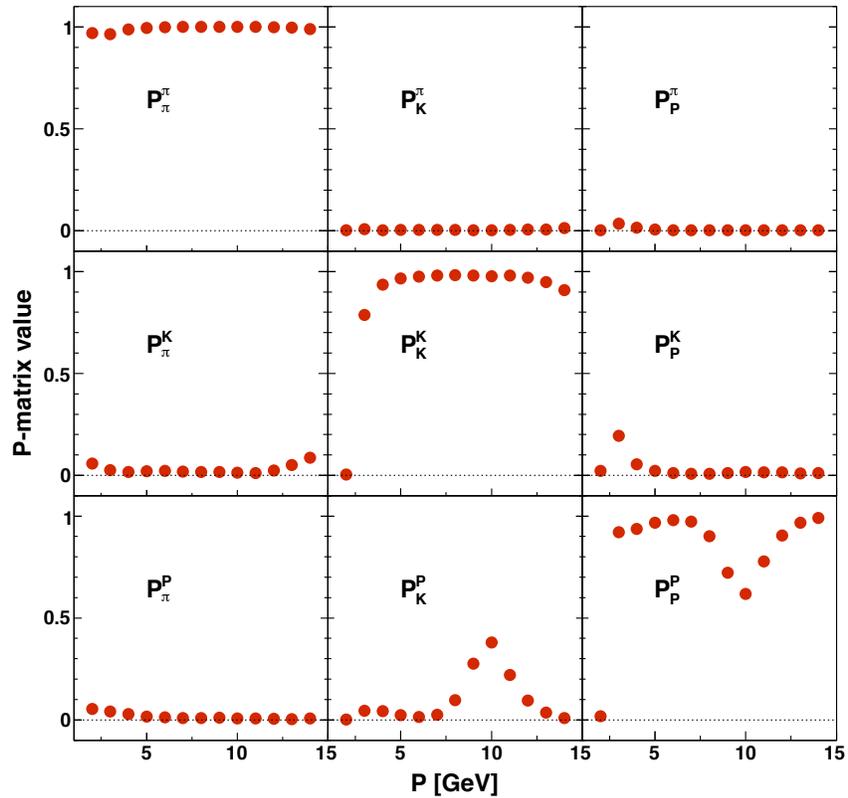


Figure 4.2: Graphical representation of the 3×3 truncated RICH efficiency matrices for single-track events in the 2000 geometry as a function of momentum. The diagonal panels show the identification efficiency, while the off-diagonal elements show contamination caused by misidentification. Figure from [101].

11. DIS region	$Q^2 > 1 \text{ GeV}^2$
12. Exclude resonances	$W^2 > 10 \text{ GeV}^2$
13. Good momentum resolution	$y > 0.1$
14. Suppress radiative effects	$y < 0.85$

(a) Event selection

15. Suppress target remnant	$z > 0.2$
16. Exclude exclusive region	$z < 0.8$

(b) Integrated multiplicities

Table 4.4: DIS event selection

Truncating the X row, and inverting the matrix yields the relation,

$$\vec{T} = P_{\text{trunc}}^{-1} \cdot \vec{I}. \quad (4.11)$$

Each element of P_{trunc}^{-1} corresponds to the weight with which each identified hadron is counted as pion, kaon, and proton.

RICH unfolding corrects for most of the hadron identification inefficiencies, leaving PID uncertainties of less than 0.5 % for pions and 1.5 % for kaons. These uncertainties are discussed in detail in Section 4.9.1.

4.4 DIS Event Selection

Events are required to fulfill the kinematic conditions $Q^2 > 1 \text{ GeV}^2$ to select the DIS region, and $W^2 > 10 \text{ GeV}^2$ to suppress the nucleon resonance region and enhance the geometric separation of the fragmenting quark from the target remnant (see Chapter 8, especially Sections 8.1.2 and 8.1.3). The upper limit on the fractional energy transfer to the target, $y < 0.85$ excludes the region where the trigger efficiencies have not yet reached a plateau as a function of momentum. This upper y limit also avoids the kinematic region where the radiative corrections to the inclusive cross section become very large. The (redundant) lower limit $y > 0.1$ excludes the region where the momentum resolution starts to degrade. Finally, hadrons that are not explicitly binned in the fractional hadron energy z are constrained to $z > 0.2$ to suppress contamination of the target remnant, as well as $z < 0.8$ to exclude the region where the fractional contribution from exclusive processes becomes sizable.

These cuts are summarized in Table 4.4 and Fig. 4.3.

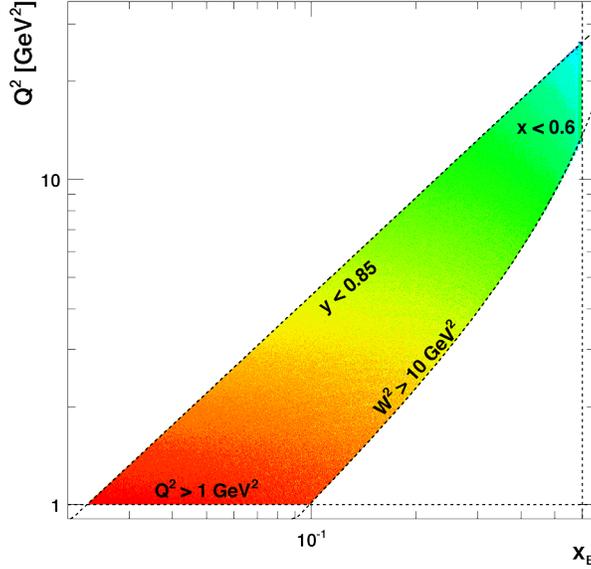


Figure 4.3: Born level inclusive cross section as a function of x and Q^2 . The DIS cuts at HERMES kinematics introduce an effective correlation between both quantities. The lower y cut is made redundant by the relatively high W^2 requirement. The cross section drops off exponentially as a function of Q^2 , from red (high) to blue (low).

4.4.1 Charge Symmetric Background

High-energy leptons from charge-symmetric QED processes, e.g., photon conversion into e^+e^- pairs or the π^0 Dalitz decays $\pi^0 \rightarrow e^+e^-\gamma$, can produce a signature indistinguishable from that of a DIS event in the spectrometer. This background is most significant at low Q^2 . Leveraging the charge symmetry of this background, it is taken into account by subtracting the number of events with a DIS signature, where the leading lepton has a charge opposite to that of the beam particles, from the measured DIS or SIDIS yields. The impact of this correction on the multiplicities is generally much smaller than 1 %, and rises up to 2 % at very low $z < 0.2$.

4.4.2 Trigger Inefficiencies

The required DIS trigger combines information from the three hodoscopes and the calorimeter. The efficiency of the individual detectors are extracted using special calibration triggers, yielding an overall efficiency that ranged from 95 % to 99 %, as a function of the track momenta and event topology [103]. The events are weighted with the inverted efficiency factor, impacting the DIS yield as well as the SIDIS yields. Because of this, the net impact on the multiplicities was found to be very small.

4.5 Kinematic Coverage and Binning

As will be explained in detail in Section 4.6, in order to correct for the limited detector acceptance in a model-independent way, the inclusive and semi-inclusive yields extraction should be fully differential in all relevant variables, i.e., Q^2 and x for the DIS yield and Q^2 , x , z , $P_{h\perp}$ and ϕ_h for the SIDIS yield.

Because of the shape of the kinematic coverage of the HERMES spectrometer, a strong correlation between x and Q^2 is present (cf. Fig. 4.3), making simultaneous binning in both variables unnecessary. A ϕ_h binning was omitted as well, because of limited statistical precision. The impact the implicit integration over ϕ_h is discussed in Section 4.9.2.

The total statistical precision of the datasets constrain the minimum granularity in the remaining three variables (x , z and $P_{h\perp}$ or Q^2 , z and $P_{h\perp}$). In order to accommodate the widest possible range of use-cases, the multiplicities are extracted in five different specialized binnings, defined in Table 4.5. An illustration of how the kinematic coverage constrains the choice of binning is shown in Fig. 4.4.

The results presented as a function of only one or two kinematic variable are all integrated versions of their corresponding three dimensional binning.

4.6 Multidimensional Smearing-Unfolding

QED radiative effects, the limited acceptance of the HERMES spectrometer and the finite detector resolution significantly complicate the interpretation of the extracted yields in Section 4.5. In order to access the underlying QCD processes from a measurement, it is crucial to properly account for these effects. For this analysis, a matrix-based smearing-unfolding technique was used to deconvolute the underlying, experiment independent cross sections (colloquially called 4π -Born cross sections) from the experimental effects.

4.6.1 Formalism

To construct the formalism behind the unfolding procedure it is instructive to start from the end result and construct a **measured** cross section $d\sigma_M$ from a Born-level cross section $d\sigma_B$. The bin-integrated cross sections $\sigma(i)$ are then given by:

$$\sigma_B(i) \equiv \int_{\text{bin } i} dX \frac{d\sigma_B}{dX}, \quad (4.12)$$

$$\sigma_M(i) \equiv \int_{\text{bin } i} dX \frac{d\sigma_M}{dX}, \quad (4.13)$$

Q^2 [GeV ²]	> 1
x	0.023–0.085–0.6
z	0.1–0.15–0.2–0.25–0.3–0.4–0.5–0.6–0.7–0.8–1.1
$P_{h\perp}$ [GeV]	0.0–0.1–0.3–0.45–0.6–1.2

(a) Three-dimensional binning for the multiplicities as a function of z

Q^2 [GeV ²]	> 1
x	0.023–0.085–0.6
z	0.1–0.2–0.3–0.4–0.6–0.8–1.1
$P_{h\perp}$ [GeV]	0.0–0.1–0.2–0.3–0.4–0.5–0.6–0.7–0.8–1.2

(b) Three-dimensional binning for the multiplicities as a function of $P_{h\perp}$ and z

Q^2 [GeV ²]	> 1
x	0.023–0.04–0.055–0.075–0.1–0.14–0.2–0.3–0.4–0.6
z	0.1–0.2–0.3–0.4–0.6–0.8–1.1
$P_{h\perp}$ [GeV]	0.0–0.3–0.5–0.7–1.2

(c) Three-dimensional binning for the multiplicities as a function of x and z

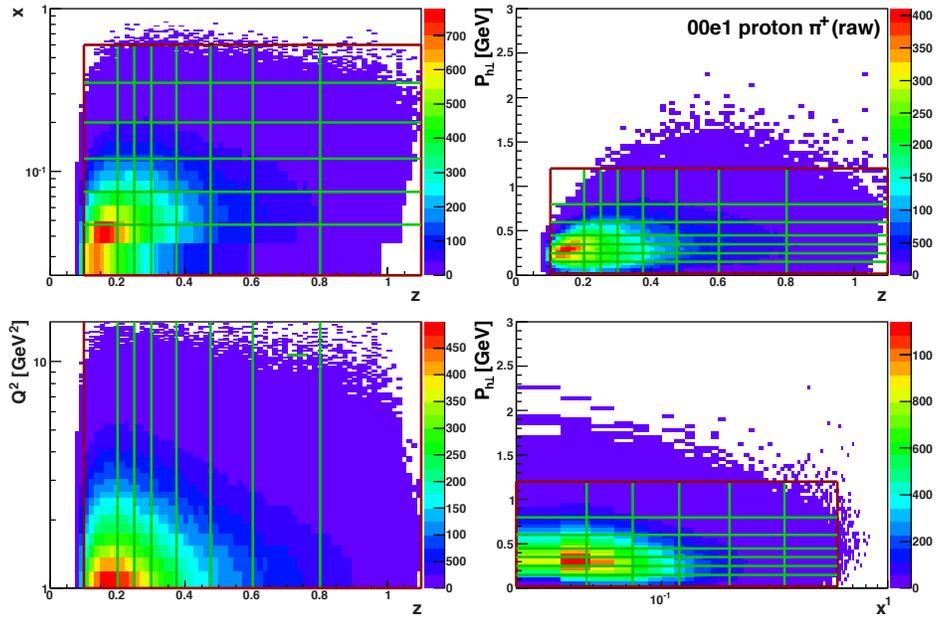
Q^2 [GeV ²]	1–1.25–1.5–1.75–2.0–2.25–2.5–3.0–5.0–15.0
x	0.023–0.6
z	0.1–0.2–0.3–0.4–0.6–0.8–1.1
$P_{h\perp}$ [GeV]	0.0–0.3–0.5–0.7–1.2

(d) Three-dimensional binning for the multiplicities as a function of Q^2 and z

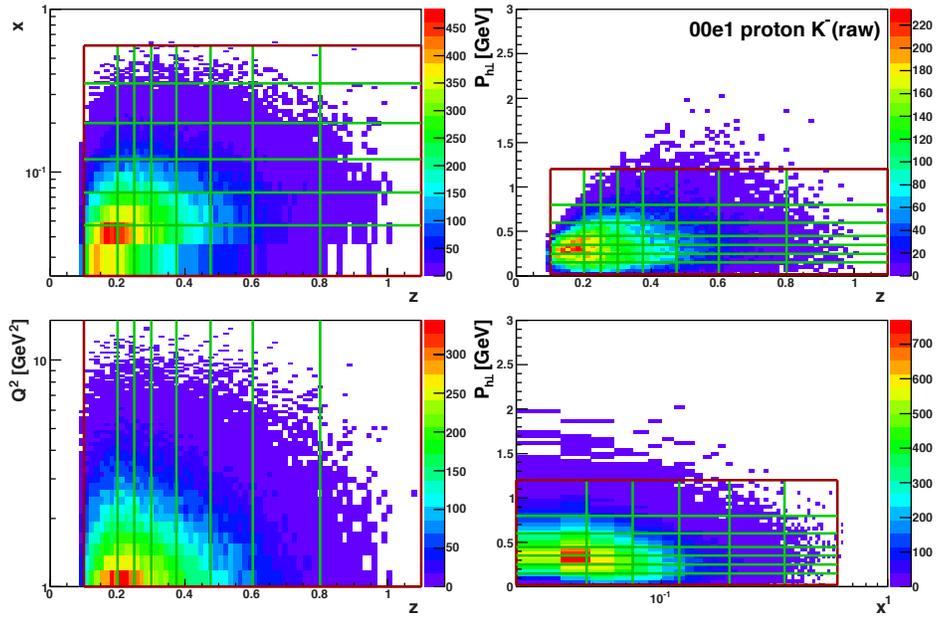
Q^2 [GeV ²]	> 1
x	0.023–0.047–0.075–0.12–0.2–0.35–0.6
z	0.1–0.2–0.25–0.3–0.375–0.475–0.6–0.8–1.1
$P_{h\perp}$ [GeV]	0.0–0.15–0.25–0.35–0.45–0.6–0.8–1.2

(e) Three-dimensional binning the multiplicities fully differential in x , z and $P_{h\perp}$

Table 4.5: Overview of the three-dimensional binnings



(a) 2000 proton raw π^+ yields



(b) 2000 porton raw K^- yields

Figure 4.4: To optimally construct the specialized three-dimensional binnings, the semi-inclusive yields were collected in various two-dimensional projections to study the occurrence of low statistics corners. The K^- has the lowest statistics, and is therefore the limiting factor. These figures show the result of these studies for the construction of the binning in Table 4.5e. The K^- (bottom panels) are clearly the limiting factor, especially when compared with the π^+ (top panels).

where X stands for all relevant kinematic variables (e.g., Q^2 and x for inclusive DIS, or Q^2 , x , z , $P_{h\perp}$ and ϕ_h in SIDIS). The yields are related to the bin-integrated cross section by the integrated luminosity L .

QED Radiative Effects

QED radiative effects include vertex corrections, and initial- and final-state radiation of real photons by the incoming and/or outgoing lepton. These effects alter the hard scattering amplitude and therefore mask the **true** kinematics of the $l\gamma^*$ vertex. It should be noted that these effects distort the SIDIS cross section only through this prime $l\gamma^*$ vertex — the impact of QED radiative effects on hadrons is negligible because of the larger mass and (comparatively) much bigger contribution of various QCD effects.

The influence of radiative effects on the cross section can be described by the probability $Q(\bar{X}|X)$ that an event has the apparent $l\gamma^*$ kinematics \bar{X} provided the true kinematics were X . This then modifies the cross section,

$$\frac{d\bar{\sigma}_B}{d\bar{X}} = \int dX Q(\bar{X}|X) \frac{d\sigma_B}{dX}. \quad (4.14)$$

It is interesting to note at this point that radiation of a real photon, both before or after the hard vertex, will carry away a fraction of the lepton energy. This leads to a higher apparent energy transfer $\bar{\nu}$. There are no effects that go in the opposite direction, and therefore,

$$\bar{\nu} \geq \nu. \quad (4.15)$$

An asymmetric smearing of ν to larger values leads to a smearing of x to lower values. Elastic events ($x = 1$) in particular are smeared downwards and therefore contaminate the inclusive DIS sample.

Limited Acceptance and Detector Smearing

An event with radiatively smeared kinematics \bar{X} has a certain probability of giving a signal in the spectrometer due to the limited geometric acceptance and generally a less than perfect detector efficiency. This information can be encoded in a probabilistic **acceptance** function $A(\bar{X})$. In addition to the acceptance, the detector also has a finite **resolution** $R(Y|\bar{X})$, which is the probability that an event with kinematics \bar{X} is observed with kinematics Y . The final measured cross section is then given by,

$$\frac{d\sigma_M}{dY} = B(Y) + \int d\bar{X} dXR(Y|\bar{X})A(\bar{X})Q(\bar{X}|X)\frac{d\sigma_B}{dX}, \quad (4.16)$$

where B represents background contributions that are indistinguishable from the signal.

The Unfolding Problem as a Linear Matrix Transformation

The bin-integrated cross section is given by,

$$\sigma_M(i) = B(i) + \int_{\text{bin } i} dY \int d\bar{X} dXR(Y|\bar{X})A(\bar{X})Q(\bar{X}|X)\frac{d\sigma_B}{dX}, \quad (4.17)$$

$$= B(i) + \int_{\text{bin } i} dY \sum_j \left(\int d\bar{X} \int_{\text{bin } j} dXR(Y|\bar{X})A(\bar{X})Q(\bar{X}|X)\frac{d\sigma_B}{dX} \right). \quad (4.18)$$

On the last line of Eq. (4.18) the integration over the true kinematics X has been split up into a sum of integration as the first step towards a matrix-formulation of the smearing-unfolding problem.

Multiplying the inner summation with $1 = \sigma_B(j)/\sigma_B(j)$ yields,

$$\sigma_M(i) = B(i) + \sum_j \frac{\left(\int_{\text{bin } i} dY \int d\bar{X} \int_{\text{bin } j} dXR(Y|\bar{X})A(\bar{X})Q(\bar{X}|X)\frac{d\sigma_B}{dX} \right) \sigma_B(j)}{\int_{\text{bin } j} dX \frac{d\sigma_B}{dX}}, \quad (4.19)$$

$$= \sum_j S(i, j) \sigma_B(j) + B(i), \quad (4.20)$$

with the smearing matrix S defined as,

$$S(i, j) \equiv \frac{\int_{\text{bin } i} dY \int d\bar{X} \int_{\text{bin } j} dXR(Y|\bar{X})A(\bar{X})Q(\bar{X}|X)\frac{d\sigma_B}{dX}}{\int_{\text{bin } j} dX \frac{d\sigma_B}{dX}}. \quad (4.21)$$

At first sight, this formalism appears to be of limited usefulness, due to the $d\sigma_B$ -dependence of S . However, if the bins can be chosen such that the Born-level cross section becomes constant within each bin, it can be moved out of the integrals,

$$\int_{\text{bin } j} dX f(X) \frac{d\sigma_B}{dX} \approx \sigma_B(j) \int_{\text{bin } j} dX f(X). \quad (4.22)$$

This implies that in the limit of small bins, the cross section terms in Eq. (4.21) drop out of the fraction, and the smearing matrix becomes (approximately) independent

of $d\sigma_B$. This crucial **model-independence** allows for S to be calculated from a Monte-Carlo simulation.

A similar reasoning can be used in the case of a flat acceptance, as long as the bins are **large** enough so that the amount of smearing is negligible.

Finally, σ_B can be obtained from σ_M by inverting the smearing matrix in Eq. (4.20),

$$\sigma_B(j) = \sum_i S^{-1}(j, i) (\sigma_M(i) - B(i)). \quad (4.23)$$

For more information regarding unfolding techniques, including a generalized likelihood-based formulation that can overcome potential complications that arise with the simple matrix-based approach above in the case of very large smearing at low statistics, see, e.g., Ref. [104].

Propagating the Covariance Matrix

The error propagation through a simple linear matrix transformation of the form $\vec{y} = A \cdot \vec{x}$ is given by [104],

$$\Sigma_y = A \cdot \Sigma_x \cdot A^T. \quad (4.24)$$

Assuming both B and S in Eq. (4.23) are known to an arbitrary precision, the error propagation of the unfolding reduces to this simple case with $A = S^{-1}$,

$$\Sigma_B = S^{-1} \cdot \Sigma_M \cdot (S^{-1})^T, \quad (4.25)$$

with Σ_B and Σ_M the covariance matrices for respectively $\vec{\sigma}_B$ and $\vec{\sigma}_M$.

Different events are statistically independent, resulting in a diagonal Σ_M . The unfolding procedure corrects for the bin-to-bin smearing by moving events to their correct bins, resulting in Born-level bins that are statistically correlated. This is immediately apparent Eq. (4.25): smearing implies a non-diagonal S , translating in off-diagonal elements in Σ_B . When analyzing an unfolded dataset, these correlations have to be accounted for, to correctly retrieve the original statistical power of the measurement.

4.6.2 Unfolding in Practice

The smearing matrix Eq. (4.21) is calculated from two separate Monte-Carlo simulations, both using the same LEPTO/JETSET [105, 106] event generator. The first simulation includes QED radiative effects through the RADGEN [107] generator, and the produced particles were tracked through a GEANT3 [108] model of the

HERMES spectrometer. Track reconstruction and data production was done with the same algorithms used for real data [81].

This first Monte-Carlo simulation contains both the generated and reconstructed properties needed to calculate the numerator of Eq. (4.21), defined as $n_{M,MC}^h(i, j)$ for the hadron yields, and $n_{M,MC}^l(i, j)$ for the inclusive yield. Additionally, it also includes the background events $B_{M,MC}^h(i)$ and $B_{M,MC}^l(i)$ generated outside of the acceptance that migrate into it. This also includes elastic and quasi-elastic events that smear into the DIS region.

The denominator in Eq. (4.21) is obtained from the second Monte-Carlo simulation, generated without radiative and instrumental effects. These values are referred to as $n_{B,MC}^h(j)$ and $n_{B,MC}^l(j)$. To summarize,

$$S^{h(l)}(i, j) = \frac{n_{M,MC}^{h(l)}(i, j)}{n_{B,MC}^{h(l)}(j)}. \quad (4.26)$$

The background subtraction in the unfolding requires the measured yields to be normalized to the Monte Carlo. This is accomplished by multiplying the ratio of the total inclusive cross section from the reconstructed Monte Carlo over the total measured inclusive yield,

$$R \equiv \frac{\sum_k \sigma_{M,MC}^l(k)}{\sum_k N_M^l}, \quad (4.27)$$

with the measured yields $N(i)$. In fact, $R = 1/L$ with L a Monte-Carlo based estimator of the total integrated luminosity of the data. The HERMES Monte Carlo is able to reproduce the inclusive cross section with a very high accuracy, and the normalized yields are therefore a very good estimator of the measured bin-integrated cross sections,

$$n_M^{h(l)}(i) = RN_M^{h(l)}(i) \approx \sigma_M^{h(l)}(i). \quad (4.28)$$

Using the unfolding formalism Eq. (4.23) then yields,

$$\vec{n}_B^{h(l)} = [S^{h(l)}]^{-1} \cdot (\vec{n}_M^{h(l)} - \vec{B}_{M,MC}^{h(l)}). \quad (4.29)$$

4.6.3 Importance of a Multi-Dimensional Approach

An earlier publication [109] used a very preliminary version of the x -dependent kaon multiplicities from the 2000 dataset to perform a leading-order extraction of the strange-quark parton distribution. When this extraction was repeated with the results of this work, it became immediately apparent that the input multiplicities

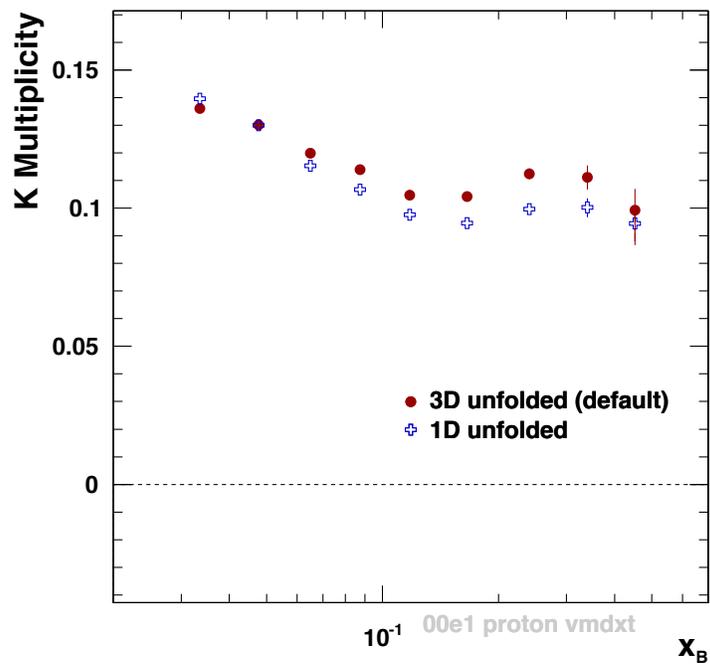


Figure 4.5: The x -dependence of the $(K^+ + K^-)$ -multiplicities differs drastically between a proper three-dimensional analysis (red), compared to a simple one-dimensional extraction (blue). This illustrates the large systematic uncertainty introduced by not considering the proper kinematic dependencies during the analysis, in particular the acceptance correction. The points in this figure were extracted from the 2000 proton sample.

differed far more than what would be expected from the refinements in the analysis.

This was traced back to the fact that the smearing-unfolding was performed one-dimensionally as a function of x , as shown in Fig. 4.5. The SIDIS cross section falls exponentially as a function of z , and the acceptance function for SIDIS hadrons is a strongly rising function of z . Because of this, reasoning behind Eq. (4.22) does not apply anymore when working with z -integrated yields, and the result becomes strongly dependent on the Monte-Carlo model.

In the limit where the Monte-Carlo cross section would be identical to the real cross section, this problem would be present, but not noticeable. While the highly tuned LEPTO/JETSET-based Monte Carlo used for the unfolding reproduces the measured SIDIS cross section very well, the strong z -dependence of both cross section and acceptance function amplify the small differences, leading to a noticeable discrepancy in Fig. 4.5.

To understand the exact shape of the discrepancy, it should be noted that, starting at medium values of x , the semi-inclusive kaon cross section in the Monte-Carlo model overshoots the measured cross section at medium-to-high z . The acceptance function is also a rising function of z , sampling the right side of the z -range more than the left side in the numerator of Eq. (4.21). The Born-level cross sections are not weighted by an acceptance function and therefore sample the left and right side of the z -range equally. This leads to inflated values in the smearing matrix, and therefore to an inverted smearing matrix with values that are too low. Application of Eq. (4.23) with values of S^{-1} systematically too low explains why the results of the one-dimensional extraction undershoot the results of a proper three-dimensional extraction, starting from values of $x \approx 0.1$.

The results of the re-evaluation of the strange-quark parton distributions are expected to be released for publication by the HERMES collaboration in the near future.

It should be noted that the result of a one-dimensional analysis, as used for Ref. [109], is in principle not incorrect, as long as the Monte-Carlo dependence is correctly accounted for in the systematic uncertainty. Of course, in this case the very large systematic uncertainty would make the results unusable, and the issue lies in the fact that this uncertainty was overlooked. This potential issue is, to a lesser extent, also present in a fully differential analysis, and a systematic uncertainty to account for finite bin width should in all cases be considered. This is discussed in detail in Sections 4.9.2 and 4.9.3.

4.6.4 The HERMES Effect

Unlike QED radiative effects associated with inelastic processes, which is predominantly radiated in the direction of either the beam lepton (ISR) or the scattered lepton (FSR), hard photons associated with elastic and quasi-elastic scattering off the target involve a negligible momentum transfer to the target. The reason for this Compton peak is discussed in detail in Ref. [110]. This implies that the transverse momentum of the scattered lepton has to be balanced by that of the radiated hard photon,

$$(1 - y) \sin \theta_{l'} = y \sin \theta_{\gamma}. \quad (4.30)$$

These effects predominantly occur at large values of **apparent** y , causing the angle of the high-energy photon with the beam line to be very small. Because of the mirror-symmetric open geometry of the HERMES spectrometer, these photons have a high probability of hitting the detector frames surrounding the beam line in front of the dipole magnet, producing extensive electromagnetic showers. These showers blast the tracking detectors with a flood of tracks, making track reconstruction impossible, dropping these type of events from the measured data sample.

As mentioned before, elastic and quasi-elastic scattering events that get smeared into the kinematic acceptance are indistinguishable from DIS events, providing a background to the DIS measurement. On its own, the loss of some of these events is therefore not a problem. However, due to historical reasons, the GEANT simulation of the spectrometer in the reconstructed Monte Carlo that was used for the unfolding does not simulate the showering of beam-line photons — these events are very computationally expensive, especially in the context of SIDIS. Therefore the tracked Monte-Carlo sample does in fact contain these events, incorrectly subtracting them from the DIS cross section during the unfolding procedure.

This unintentional overcorrection for radiative effects was only discovered after it artificially introduced a strong nuclear dependence in the ratio $R = \sigma_L/\sigma_T$, leading to the publication of Ref. [111], and its subsequent retraction in Ref. [110]. Before its retraction, the effect was commonly referred to as the “HERMES effect”.

The relatively high W^2 cut of 10 GeV^2 , chosen to select a clean SIDIS sample in this work, suppresses the impact of radiative effects, compared to a pure inclusive analysis, where a lower W^2 cut of 4 GeV^2 is used. Accordingly, the impact of this “HERMES effect” on the multiplicity analysis is relatively small.

In order to correct for this issue, the missing events are added to the measured inclusive yield prior to subtracting them again during the unfolding as described in Eq. (4.23).

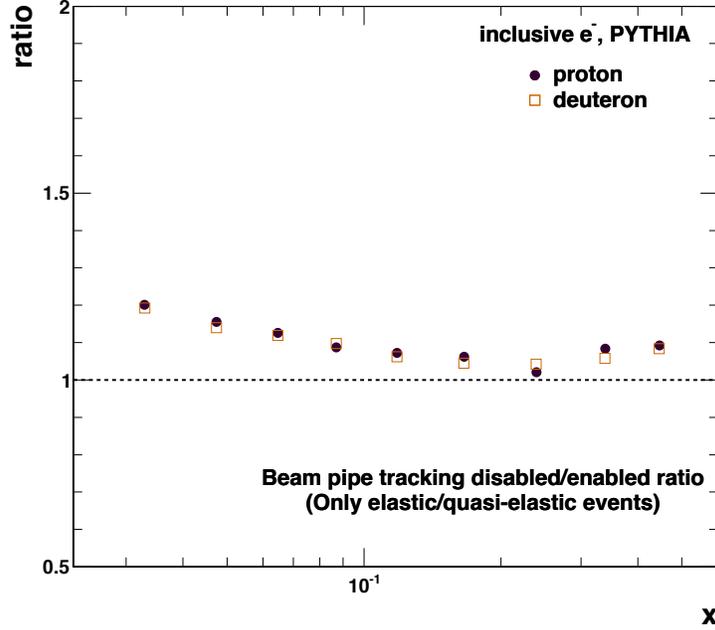


Figure 4.6: The HERMES effect inflation factor h for both proton and deuteron, as a function of x . The ratios were extracted from two inclusive Monte-Carlo simulations, as explained in the text. Due to the kinematics of the radiated hard photon, the necessary inflation is the largest at low x .

The inflation factor h describing the fraction of elastic events that should have been lost in the reconstructed Monte Carlo, is obtained from two Monte-Carlo simulations that of elastic and quasi-elastic scattering in HERMES, one without beam-pipe tracking, and one with beam-pipe tracking,

$$h(Q^2, x) \equiv \frac{\sigma_{MC, nobpt}^{ela}}{\sigma_{MC, bpt}^{ela}}, \quad (4.31)$$

with $\sigma_{MC, (no)bpt}^{ela}$ the observed inclusive cross section, from the Monte-Carlo simulation with beam-pipe tracking enabled (disabled). This factor is shown in Fig. 4.6.

The reconstructed Monte-Carlo sample used for the unfolding yields the fraction of elastic and quasi-elastic events in the measured inclusive sample f^{ela} ,

$$f^{ela}(Q^2, x) \equiv \frac{\sigma_{M, MC}^{ela}}{\sigma_{M, MC}^{ela} + h(Q^2, x) \cdot \sigma_{M, MC}^l}, \quad (4.32)$$

with $\sigma_{M, MC}^l$ ($\sigma_{M, MC}^{ela}$) the contribution of the DIS (elastic and quasi-elastic) cross sections to the total inclusive cross section. In Eq. (4.32), h is used to remove the unwanted ‘HERMES effect events’ that are not present in the elastic fraction in the measurement. The fraction is presented as a function of x in Fig. 4.7. It is of the

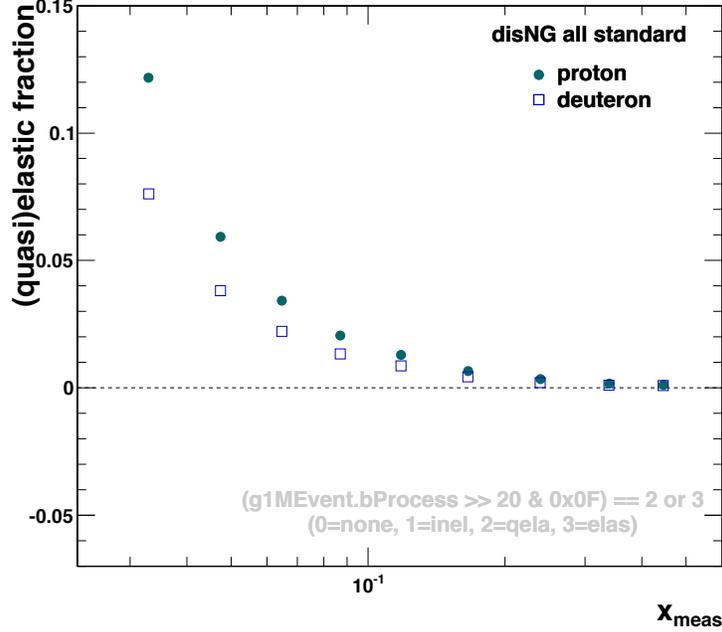


Figure 4.7: The fraction of elastic and quasi-elastic events in the inclusive sample, from reconstructed Monte Carlo. Compared to the proton, the ratio of elastic and quasi-elastic events to the total inclusive sample for the deuteron is visibly suppressed.

order of 1 % to 2 %, rising up to 12 % at low x .

Combining Eq. (4.31) and Eq. (4.32), the measured inclusive yield N_M^l can then be corrected,

$$N_{M, \text{corr}}^l(Q^2, x) = N_M^{\text{DIS}} + h \cdot N_M^{\text{ela}} \quad (4.33)$$

$$= \left((1 - f^{\text{ela}}) + h \cdot f^{\text{ela}} \right) \cdot N_M^l. \quad (4.34)$$

This correction only impacts the multiplicities at low x , as shown on Fig. 4.8. This is not unexpected, as the kinematics of elastic events with a high-energetic photon reconstructs as a low x event.

4.7 Contamination by Exclusive Channels

Several exclusive processes have a final state that cannot be distinguished from semi-inclusive pion and kaon production. These include direct exclusive production of pseudoscalar mesons (e.g. $\gamma^* p \rightarrow \pi^+ n$), as well as the exclusive production of vector mesons, that subsequently decay into lighter hadrons (e.g. $\gamma^* p \rightarrow p \rho^0 \rightarrow p \pi^+ \pi^-$). The cross section for these processes is suppressed by $1/Q^4$ compared to the leading behavior of the SIDIS cross section [112].

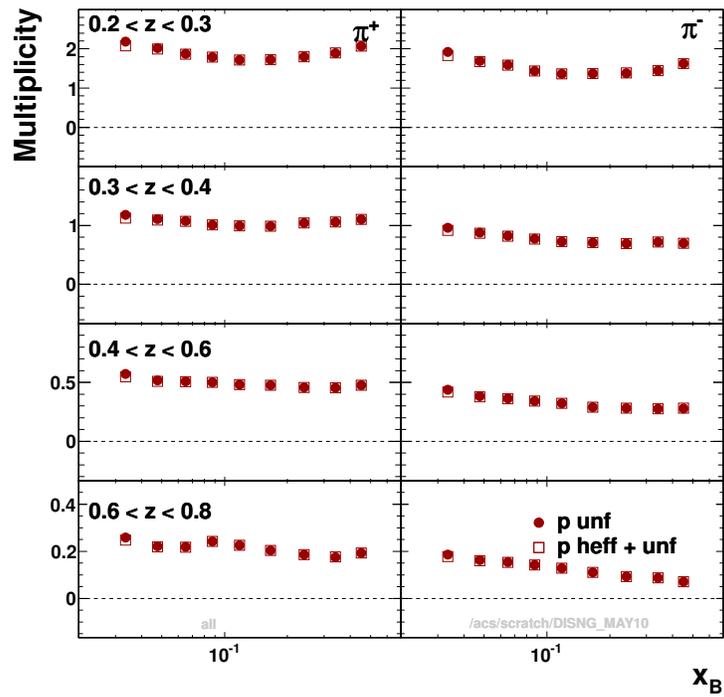


Figure 4.8: Unfolded multiplicities with (empty squares) and without (full circles) the HERMES effect correction, in a projection versus z and x . The correction flattens the multiplicities at low x somewhat. The overall impact is very small, due to the relatively large $W^2 > 10 \text{ GeV}^2$ cut.

This suppression is not very large at the HERMES, where $\langle Q^2 \rangle \sim 2.5 \text{ GeV}$. Additionally, the invariant mass of the inclusive final state,

$$M_X^2 = Q^2 \frac{1-z}{x} + M^2 - (q - P_h)^2, \quad (4.35)$$

is not very large, especially at high values of z . This severely restricts the amount of individual channels that contribute to the SIDIS cross section. Hence, in certain kinematic areas, the cross section for certain individual exclusive channels can become comparable in size to the total SIDIS cross section.

Direct exclusive production of scalar mesons is restricted to high values of z because of Eq. (4.35). For typical HERMES kinematics, this corresponds to $z > 0.94$, far outside the recommended range of $0.2 < z < 0.8$.

The contribution due to the decay of exclusive vector mesons has a very different z -distribution, as the energy of the vector meson is shared by multiple decay products. The dominant channels at HERMES energies,

$$lp \rightarrow lp\rho^0 \rightarrow lp\pi^+\pi^-, \quad (4.36)$$

$$lp \rightarrow lp\phi \rightarrow lpK^+K^-, \quad (4.37)$$

providing a sizable background to the measured pion and kaon yields [112, 113].

The formation of the decay products from exclusive vector mesons does not involve the fragmentation of quarks originating from the target nucleon. Therefore, if not properly accounted for, fragmentation functions extracted from multiplicities that include a significant contribution from these processes may not be process independent, violating universality. However, especially at larger values for Q^2 , these each individual exclusive channel could be treated as a power correction, and as such they could potentially be included in a leading-twist analysis. In fact, there is no way to a priori determine if the subtracted cross section is better-suited for an interpretation in a leading-twist SIDIS formalism (and vice-versa), especially because the derivation of the factorization theorem for SIDIS itself requires the sum over all final states to be complete [112]. At the other hand, at HERMES kinematics, the contributions of exclusive ρ^0 -production to the π^+ and π^- cross section becomes dangerously large for higher values of z [114], making the use of a leading-twist interpretation undesirable in this region.

In light of this discussion, it was decided to present the results in this work with and without a correction for the exclusive vector meson contribution. In fact, the comparison of the both versions therefore constitutes a valuable experimental result in itself.

The exclusive production of ρ^0 -, ω - and ϕ - mesons can be described in the

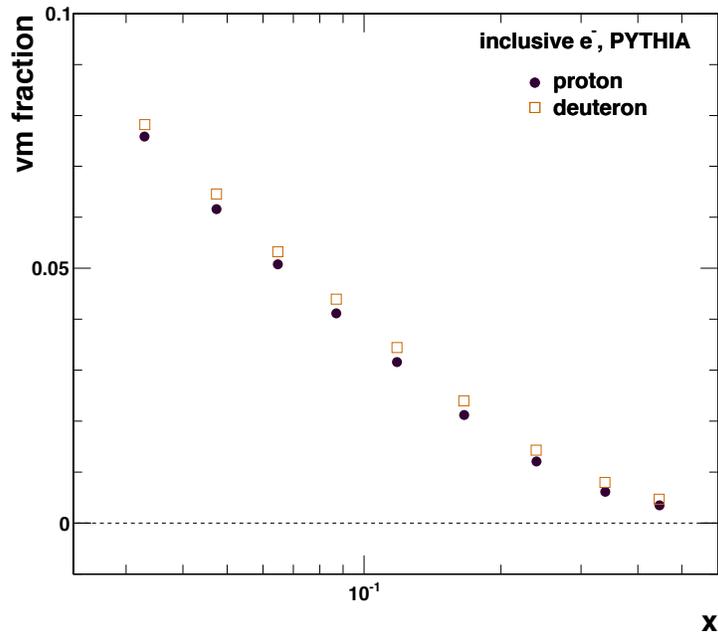


Figure 4.9: Fractional contribution of exclusive vector meson production to the inclusive cross section in the HERMES acceptance, from the VMD model in PYTHIA. The fractions are projected versus x .

vector meson dominance (VMD) model as the fluctuation of the virtual photon in a $q\bar{q}$ pair with compatible quantum numbers. This is discussed in detail in Ref. [113].

The fraction of final-state hadrons from exclusive production, as well as the fraction of exclusive events incorrectly counted as SIDIS events, was estimated as a function of all kinematic variables using a modified PYTHIA 6 Monte-Carlo generator [115] that incorporates a VMD model tuned to describe exclusive ρ^0 and ϕ production at HERMES [113, 116]. PYTHIA can only simulate proton or neutron targets, and the values for the deuteron were constructed by combining the results of both nucleons.

The Monte-Carlo generator for the smearing-unfolding does not include these contributions. Because the acceptance for exclusive events is not necessarily identical to that for SIDIS events, the correction is applied prior to the smearing-unfolding. It was however found that subtracting the vector meson fractions after unfolding to 4π -Born did not significantly alter the results.

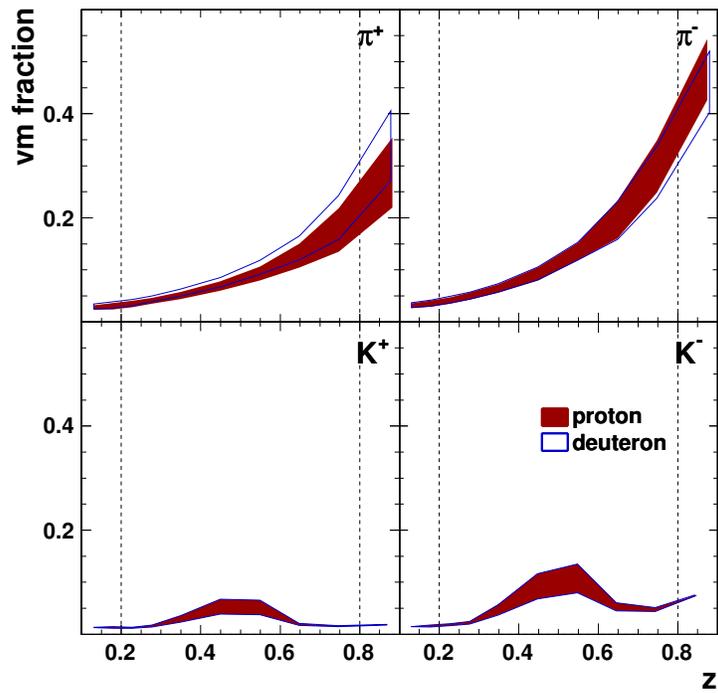


Figure 4.10: Fractional contribution of exclusive vector meson production to the the SIDIS cross section in the HERMES acceptance, from the VMD model in PYTHIA. The fractions are projected versus z . The bands correspond with the 1σ uncertainty of the VMD model tuning. The vertical dashed lines are the limits in z used in the multiplicity extraction.

The vector meson fractions are defined as,

$$f^{h,\text{vm}} \equiv \frac{\sigma_{\text{M, PYTHIA}}^{h,\text{VMD}}}{\sigma_{\text{M, PYTHIA}}^h} \quad (4.38)$$

$$f^{l,\text{vm}} \equiv \frac{\sigma_{\text{M, PYTHIA}}^{l,\text{VMD}}}{\sigma_{\text{M, PYTHIA}}^l}, \quad (4.39)$$

with $\sigma_{\text{M, PYTHIA}}$ the relevant cross sections from PYTHIA. PYTHIA does not simulate elastic (or quasi-elastic) processes, and therefore the fraction Eq. (4.39) only pertains to the other contributions to the measured inclusive cross section. The elastic fraction from Eq. (4.32) can be easily modified to include the vector meson contribution,

$$\tilde{f}^{\text{ela}}(Q^2, x) \equiv f^{\text{ela}}(Q^2, x) \cdot \frac{1}{1 + \frac{\sigma_{\text{M, PYTHIA}}^{l,\text{VMD}}}{\sigma_{\text{M, MC}}^l}}, \quad (4.40)$$

where $\sigma_{\text{M, MC}}^l$ is extracted from the Monte-Carlo sample used for the smearing-unfolding.

The vector meson subtracted yields are then given by,

$$N_{\text{M, corr}}^h(Q^2, x, z, P_{h\perp}) = (1 - f^{h,\text{vm}}) \cdot N_{\text{M}}^h \quad (4.41)$$

$$N_{\text{M, corr}}^l(Q^2, x) = \left((1 - \tilde{f}^{\text{ela}}) \cdot (1 - f^{l,\text{vm}}) + h \cdot \tilde{f}^{\text{ela}} \right) \cdot N_{\text{M}}^l, \quad (4.42)$$

where Eq. (4.42) combines the vector meson subtraction with the HERMES effect correction Eq. (4.34).

Figures 4.9 and 4.10 show the inclusive and SIDIS fractions as a function of respectively x and z . The bands correspond to a 1σ variation in the parametrization of the exclusive cross sections.

The impact of the vector meson subtraction is presented in Figs. 4.11 to 4.13. The inclusive and semi-inclusive corrections Eq. (4.41) and Eq. (4.42) mostly cancel each other in the multiplicities, and the impact of the correction is therefore relatively small, with the exception of pions at high z .

Pions originating from ρ^0 decay have relatively high velocity in the center-of-mass system of the decay. When boosted to the lab frame, this translates into a pion distribution mostly concentrated at low and high z . At the same time, the SIDIS cross section drops exponentially as a function of z , yielding a very large fractional contribution of pions from ρ^0 decay, up to 50% near $z = 1$. The mass difference between ϕ -mesons and kaons is much smaller, leading to a more isotropic decay distribution and a much less pronounced effect on the multiplicities.

The figures in this work, unless explicitly mentioned, use the vector-meson-

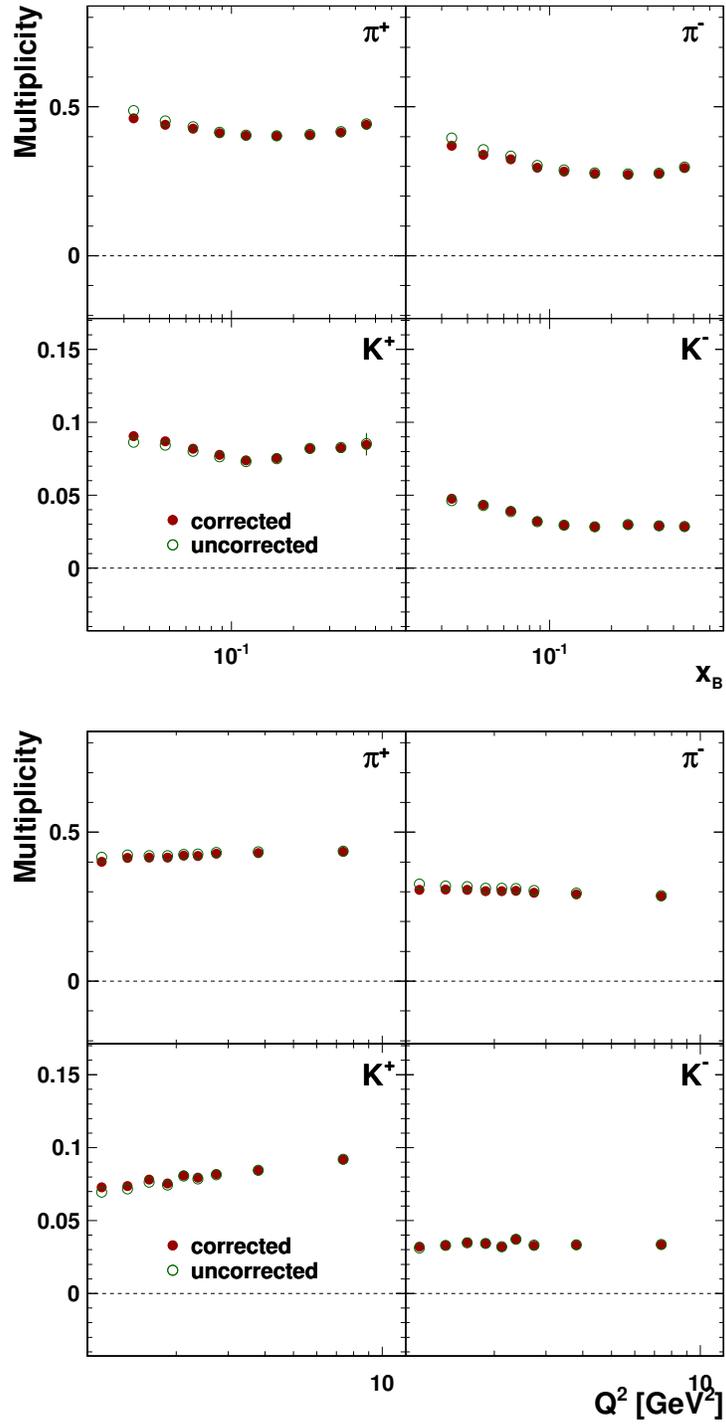


Figure 4.11: Comparison of the final unfolded multiplicities for a proton target with and without the correction for exclusive vector mesons as a function of x (top panels) and Q^2 (bottom panels). Due to cancellations between the inclusive and semi-inclusive corrections, the net contribution of the correction is relatively small.

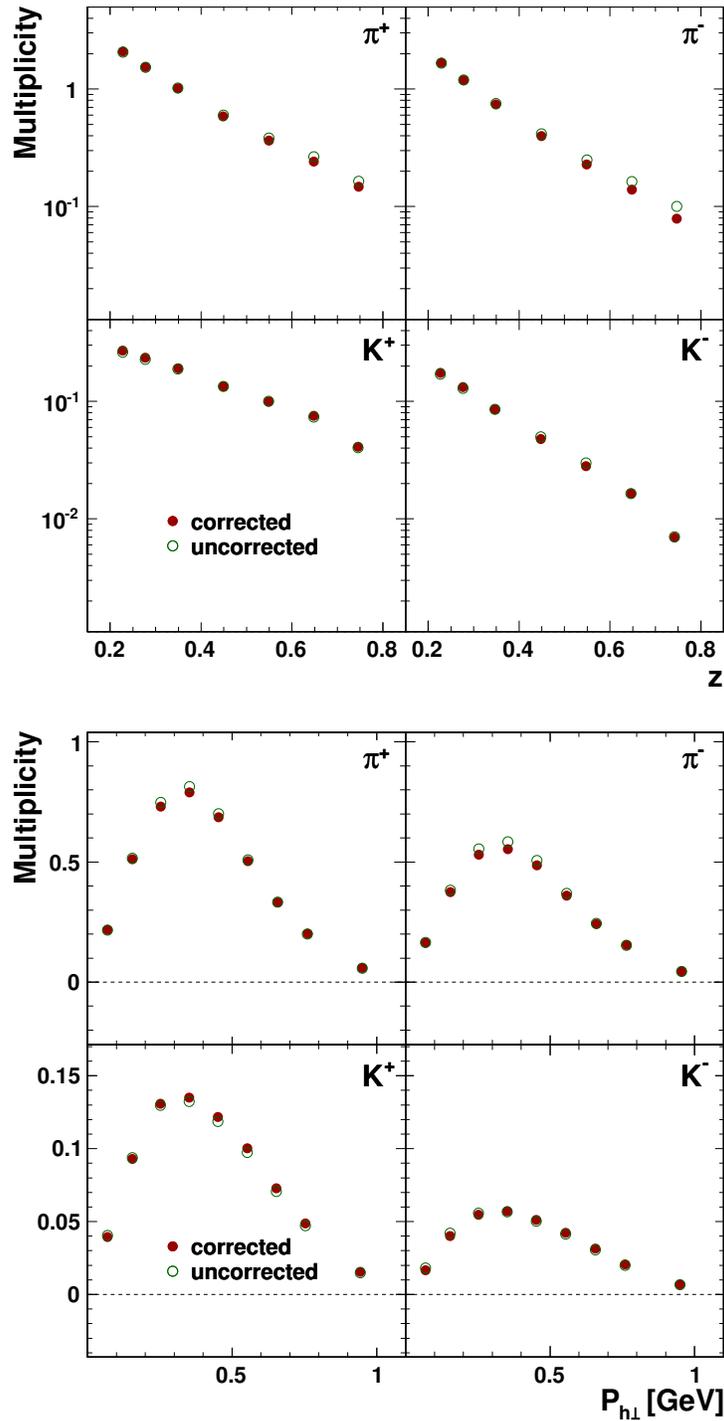


Figure 4.12: Comparison of the final unfolded multiplicities for a proton target with and without the correction for exclusive vector mesons as a function of z (top panels) and $P_{h\perp}$ (bottom panels). Due to cancellations between the inclusive and semi-inclusive corrections, the net contribution of the correction is relatively small.

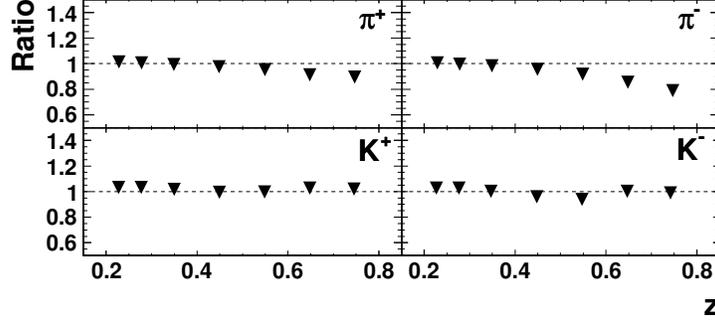


Figure 4.13: Fractional impact of the vector meson correction on the multiplicities for a proton target as a function of z . The ratio is calculated between the corrected multiplicities over the uncorrected multiplicities. The deviations from unity become more significant for higher values of z .

corrected version of the multiplicities.

4.8 Data Combination and Projections

As already mentioned at the start of this chapter, to account for slight changes in the experimental conditions, the three datasets on each target are treated as independent measurements. The last step in the extraction framework combines measured normalized yields by means of a weighted average,

$$n_{\text{B}}^{h(l)}(i) = \frac{\sum_y w_y \cdot n_{\text{B},y}^{l(h)}(i)}{\sum_y w_y}, \quad (4.43)$$

where the summation runs over the three data productions y . The total integrated raw inclusive yield provides a measure of the statistical power of each data sample, and therefore constitutes a good choice for the weight factor w_y ,

$$w_y \equiv \sum_i N_{\text{M},y}^i. \quad (4.44)$$

A similar equation as Eq. (4.43) can be written for the average kinematic variables in bin i .

The multiplicity M_n^h of hadrons of type h off a target n , defined in Eq. (2.33), can be expressed in terms of the normalized yields from Section 4.6.2 as,

$$M_n^h(Q^2, x, z, P_{h\perp}) = \frac{1}{\frac{d^2 n_{\text{B}}^l(Q^2, x)}{dQ^2 dx}} \int_0^{2\pi} \frac{d^5 n_{\text{B}}^h(x, Q^2, z, P_{h\perp}, \phi_h)}{dQ^2 dx dz dP_{h\perp} d\phi_h} d\phi_h, \quad (4.45)$$

or in terms of the binned results,

$$M_n^h(Q_i^2, x_i z_i, P_{h\perp,i}) = \frac{1}{\Delta z(i) \Delta P_{h\perp}(i)} \frac{n_B^h(i)}{n_B^l(i)}, \quad (4.46)$$

where indexed kinematic variables $(Q_i^2, x_i, z_i, P_{h\perp,i})$ refer to the average kinematics in bin i . The main fraction is normalized to the bin width of the semi-inclusive variables in bin i (respectively $\Delta z(i)$ and $\Delta P_{h\perp}(i)$). A similar normalization factor for the inclusive variables cancels between the semi-inclusive numerator and the inclusive denominator.

Using the results from the weighted average Eq. (4.43) in Eq. (4.46) then yields the three-dimensional multiplicities for both targets in all five binnings.

To obtain the one- and two-dimensional projections, the inclusive and semi-inclusive normalized yields are integrated separately over the relevant variables, e.g., both are integrated to project out x , while only the SIDIS yield is integrated to project out z . The resulting multiplicity is then only differential in the remaining semi-inclusive variables. Therefore, after removal of the relevant Δ -factors, Eq. (4.46) is still relevant for the integrated yields.

Define $X(i, j)$ as a bin average kinematic variable in a bin (i, j) , where j stands for the bins that are projected out during the integration, and i for the remainder. The average kinematic variable $\bar{X}(i)$ for the projected multiplicity is then given by weighing $X(i, j)$ with the normalized yields $n_B(i, j)$,

$$\bar{X}(i) = \frac{\sum_j n_B(i, j) \cdot X(i, j)}{\sum_j n_B(i, j)}, \quad (4.47)$$

where n_B is the inclusive yield for inclusive variables, and the semi-inclusive yield for semi-inclusive variables.

4.9 Systematic Uncertainties

The precise estimation of the systematic uncertainties on the multiplicities is crucial because of the small size of the statistical uncertainties. The extraction method was designed to optimally avoid a major loss of precision due to systematic limitations. The dominant sources of systematic uncertainties are mostly linked to limitations on the correction methods, in particular the RICH unfolding from Section 4.3.2 and the smearing unfolding from Section 4.6.

Figures 4.14 to 4.17, present a breakdown of the main contributions to the systematic uncertainty:

1. Azimuthal asymmetries in the unpolarized cross section (blue),

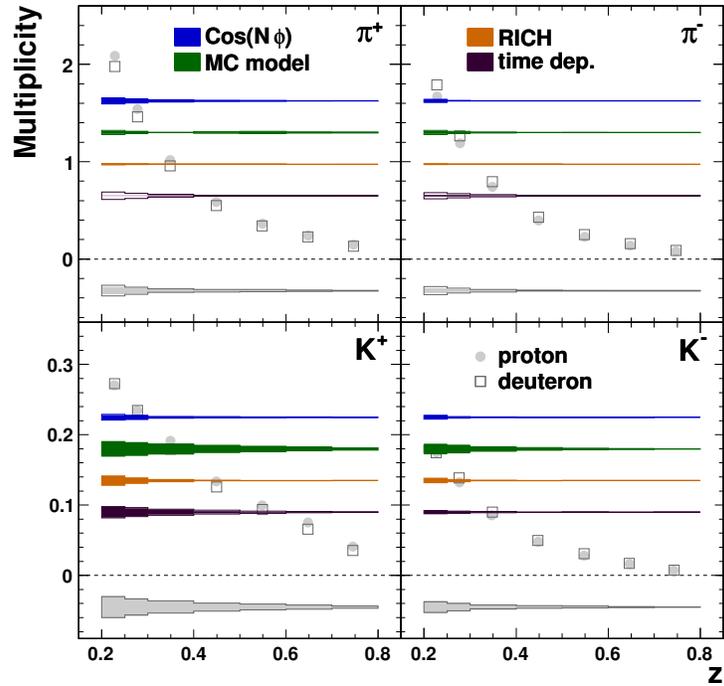


Figure 4.14: Breakdown of the systematic uncertainty into its major components as a function of z . The colored bands correspond to the contributions due to the azimuthal asymmetries in the unpolarized cross section (blue), the Monte-Carlo model-dependence due to the finite bin width (green), the RICH unfolding (orange) and the time stability (purple). The gray bands at the bottom are the total systematic uncertainty. Proton and deuteron uncertainties are shown as respectively full and empty bands. The proton (deuteron) multiplicity results are represented as full circles (empty squares).

2. Finite bin width and Monte-Carlo model-dependence (green),
3. RICH unfolding (orange),
4. Time stability (purple).

The main contributions are discussed in detail in Sections 4.9.1 to 4.9.4. Additional contributions that were considered, but ultimately found to be of secondary importance, are discussed in Section 4.9.5.

The systematic uncertainties have a high degree of bin-to-bin correlation. For the statistical uncertainties, the bin-to-bin correlation due to the unfolding procedure is exactly known. This is not the case for the systematic uncertainties, complicating the error propagation during the projection procedure multiplicities. Treating the systematics as a Gaussian uncertainty would most certainly overestimate (or underestimate) the actual systematic uncertainty.

To evaluate the systematic uncertainties, several reasonable variations of the multiplicities are compared. The difference between these results is then used to assign the systematic uncertainty for that particular contribution. This procedure is repeated for the projected multiplicities (as well as for the derived quantities in Chapters 5 and 8). Evaluating the systematic uncertainties only at the last possible stage, rather than using a Gaussian (or linear) error propagation, ensures that the bin-to-bin correlations of the systematic uncertainties are taken into account.

4.9.1 RICH Unfolding

The EVT algorithm uses a background function B to take into account PMT hits due to physical particles not tracked by the spectrometer (e.g. from $\pi^0 \rightarrow \gamma\gamma$ decays), as well as well as altered detector behavior due to dead and hot PMTs. Events with no reconstructed tracks in one of the spectrometer halves allow for this background to be extracted from the average number of hits in each PMT in this detector half.

The RICH efficiency matrices Eq. (4.10) were extracted from a Monte-Carlo simulation. When applying the EVT algorithm to the results of this simulation, the background function has to be chosen. The background function extracted from Monte-Carlo simulations using different generators (LEPTO/JETSET [105, 106] or PYTHIA 6 [115]) was found to differ from each other. These variations are used to extract four different efficiency matrices in order to determine the systematic uncertainty on the RICH unfolding:

1. extracted from LEPTO/JETSET with its own background (center/default)
2. extracted from LEPTO/JETSET with the data background

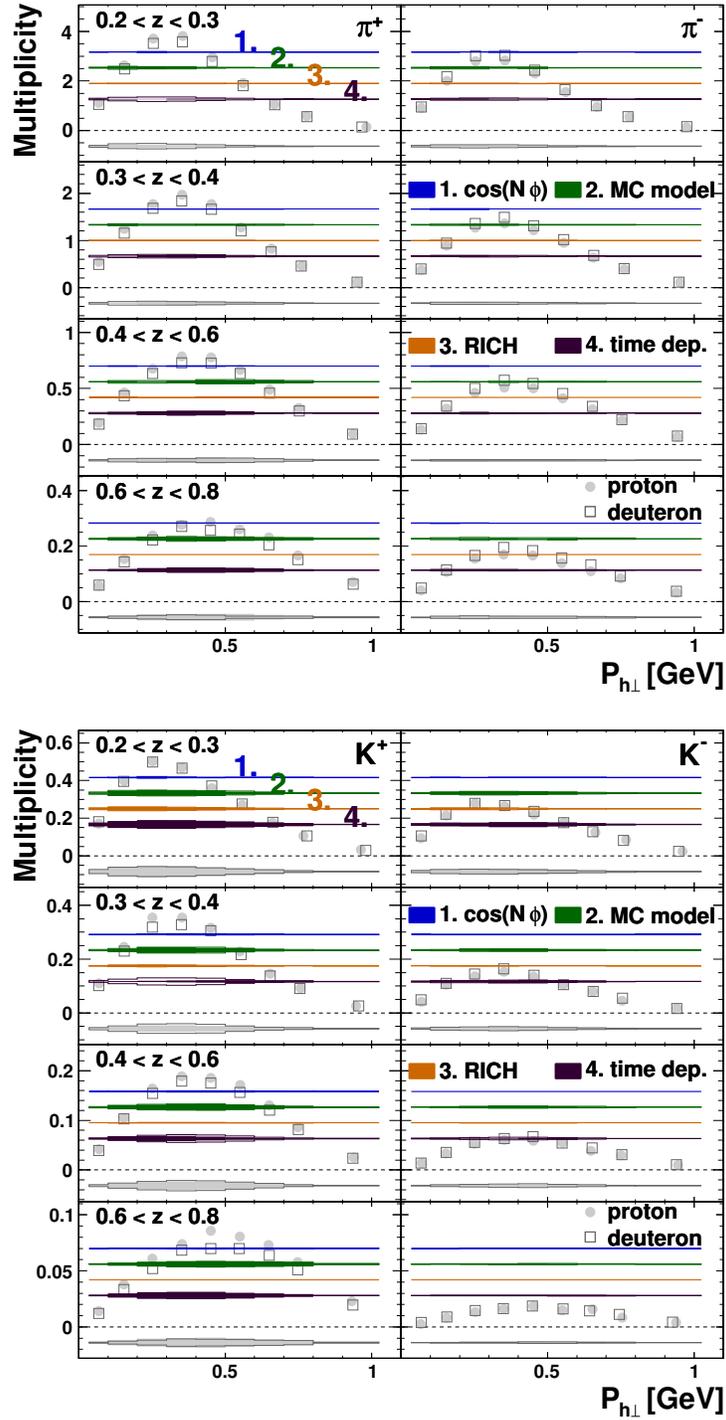


Figure 4.15: Breakdown of the systematic uncertainty into its major components as a function of $P_{h\perp}$ in z -slices for pions (top panels) and kaons (bottom panels). The bands, and markers are defined in the caption of Fig. 4.14. Generally, the systematic uncertainty is dominated by the Monte-Carlo model-dependence and the time stability uncertainty.

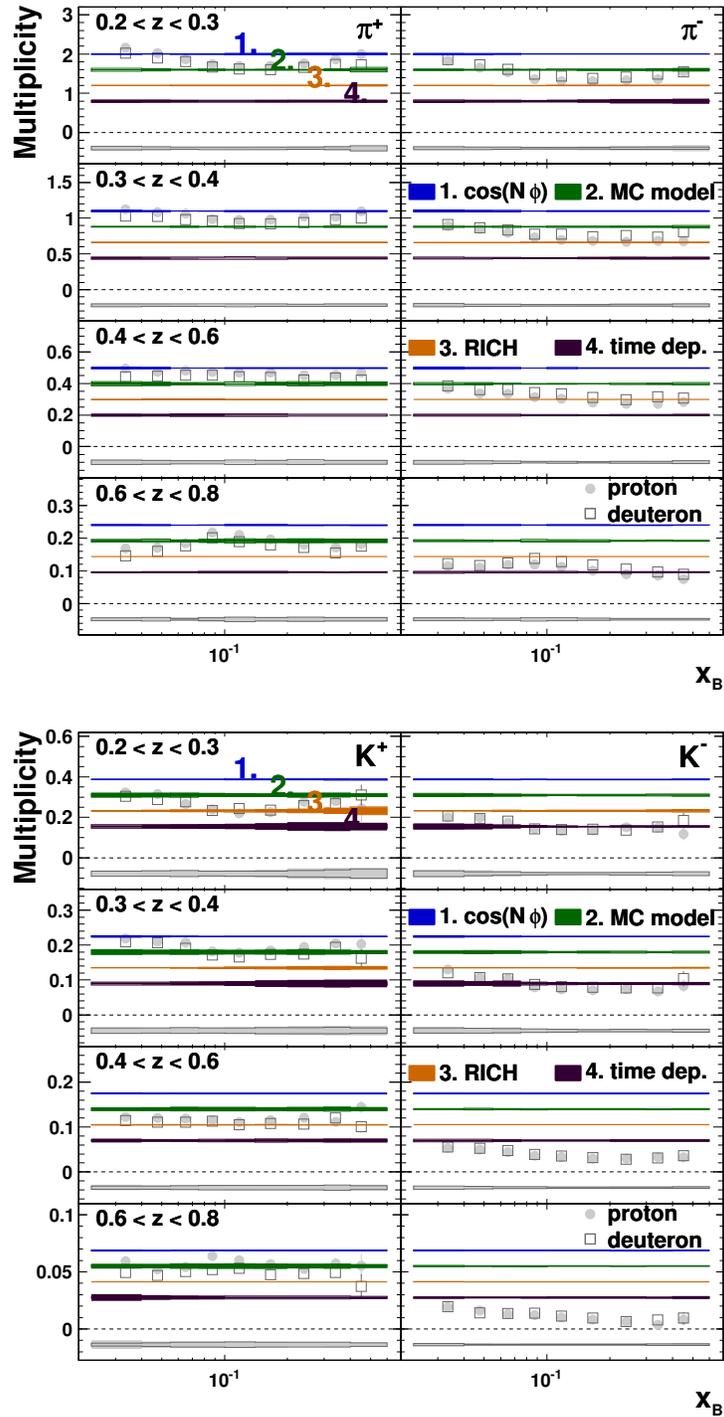


Figure 4.16: Breakdown of the systematic uncertainty into its major components as a function of x in z -slices for pions (top panels) and kaons (bottom panels). The bands, and markers are defined in the caption of Fig. 4.14. Generally, the systematic uncertainty is dominated by the Monte-Carlo model-dependence and the time stability uncertainty.

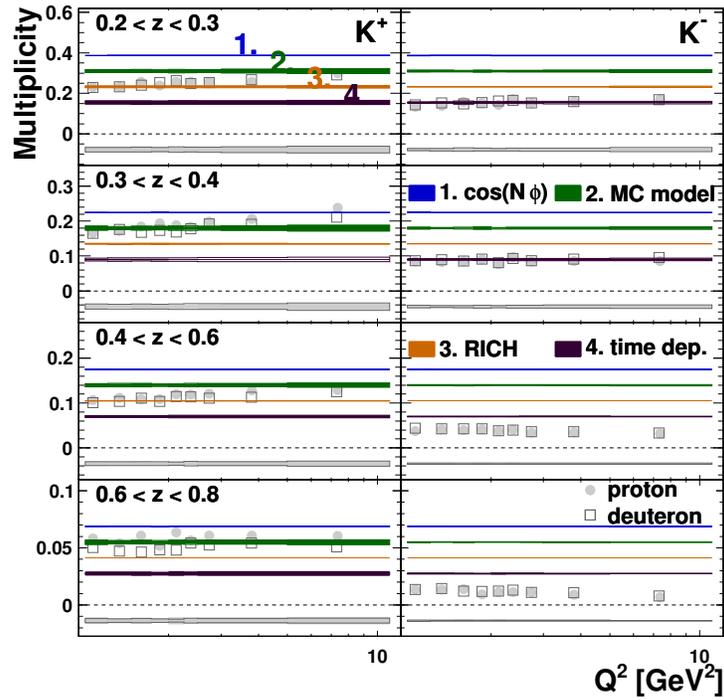
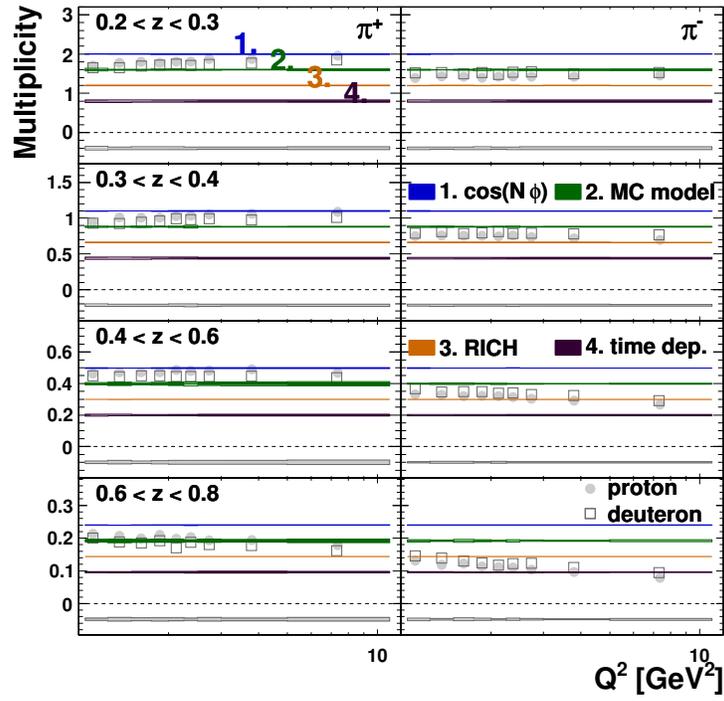


Figure 4.17: Breakdown of the systematic uncertainty into its major components as a function of Q^2 in z -slices for pions (top panels) and kaons (bottom panels). The bands, and markers are defined in the caption of Fig. 4.14. Generally, the systematic uncertainty is dominated by the Monte-Carlo model-dependence and the time stability uncertainty.

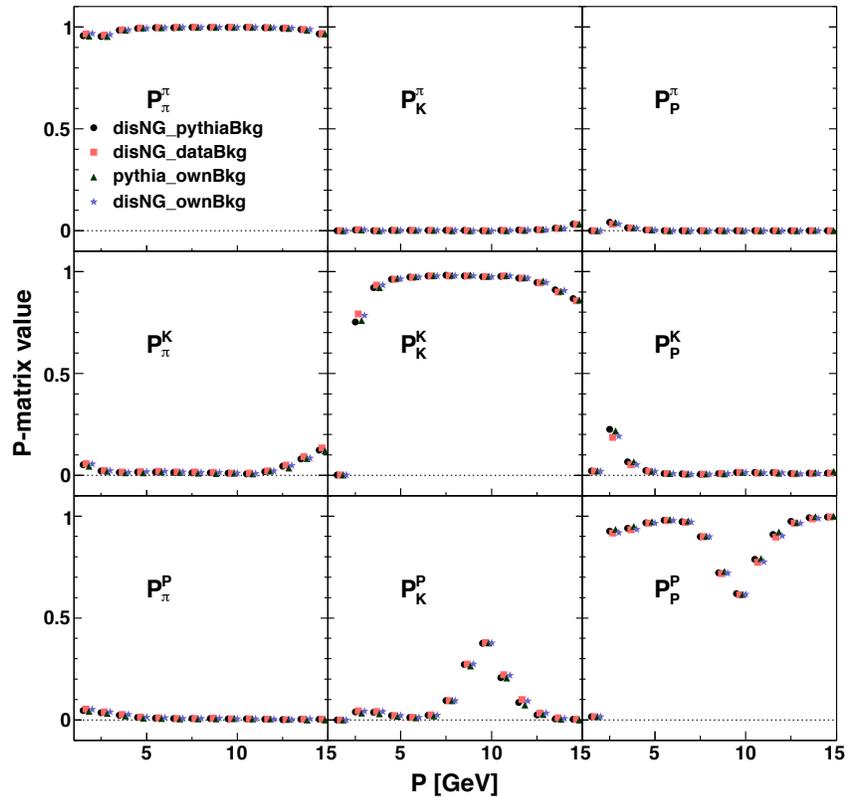


Figure 4.18: Comparison of the four versions of the RICH efficiency matrices for single-track events in the 2000 geometry that are used to calculate the RICH unfolding systematic uncertainty. The differences between these matrices are discussed in the text. Figure from [101].

3. extracted from LEPTO/JETSET with the PYTHIA background
4. extracted from PYTHIA with the LEPTO/JETSET background

The four versions of the RICH efficiency matrices are presented in Fig. 4.18. Other variations considered, including changing mirror roughness parameters in the simulation of the RICH detector, were found to have a negligible impact on the efficiency matrices.

The impact of this systematic uncertainty on the multiplicities, illustrated in Figs. 4.14 to 4.17, is typically less than 0.5 % for pions and 1.5 % for kaons. For more information regarding the extraction of the RICH efficiency matrices, and its systematic uncertainties, see Ref. [101].

4.9.2 Azimuthal Asymmetries in the Unpolarized Cross Section

Fully differential, the unpolarized semi-inclusive cross section is modulated in the azimuthal angle ϕ_h due to terms in $\cos \phi_h$ and $\cos 2\phi_h$ [117, 118]. These terms can be explained by the Cahn [119–121] and the Boer-Mulders [122, 123] effects. The azimuthal $\cos \phi_h$ and $\cos 2\phi_h$ modulations have been extracted at HERMES [102].

The multiplicity extraction is not differential in ϕ_h , because of statistical constraints (cf. Section 4.5). Due to the functional dependence of the acceptance on ϕ_h , this could introduce a systematic Monte-Carlo model dependence in the unfolded multiplicities, similar to the issues addressed in Section 4.6.3.

The ϕ_h modulations of the unpolarized cross section are not included in the Monte-Carlo simulation, potentially exacerbating the issue. In order to study the impact of these omissions, an additional extraction was performed, using a Monte-Carlo simulation where the events are reweighed, on the Born level, with a four-dimensional parametrization of the $\cos \phi_h$ and $\cos 2\phi_h$ extracted from HERMES data (see, e.g., Ref. [101]).

The difference between the results of this second extraction, and the original results provides a good estimate of the systematic uncertainty introduced by the azimuthal modulations.

The uncertainty is typically less than 2 %, except for a few bins at higher values of x (cf. Figs. 4.14 to 4.17). The parametrization used to reweigh the Monte-Carlo was extracted on a more restricted x -range than used for this work, and the rising uncertainty at high x is therefore due to extrapolation of the polynomial four-dimensional parametrization. In fact, these extrapolations are the reason that the introduction of an additional systematic uncertainty was preferred over a simple correction.

4.9.3 Finite Bin Width and Monte-Carlo Model-Dependence

The smearing matrix that is used for the unfolding correction in Section 4.6, can be extracted from a set of Monte-Carlo simulations if the bin size is **small enough** for Eq. (4.22) to be approximately valid.

The finite detector resolution imposes a natural lower limit on the bin size. Additionally, the dramatic increase in detector smearing near this intrinsic limit introduces instability issues with the straightforward maximum variance unfolding technique utilized for the extraction (cf. Ref. [104]). The available statistics in the dataset further restrict the minimum bin size.

Additionally, background contributions that smear into the observed cross section are directly subtracted from the measured yields in Eq. (4.23), providing an additional source of Monte-Carlo model dependence. This contribution is strongly suppressed by the presence of extra bins below the $z > 0.2$ edge (cf. Table 4.5).

The potential impact of any residual model dependence is estimated by scanning over the JETSET parameters which control the Monte-Carlo event generator. These parameters are determined with a χ^2 -fit to the measured multiplicities [113, 124, 125]. This fit was found to be particularly sensitive to a set of nine parameters which constrain various features of the fragmentation process.

An eigenvector-basis approach to the Hessian method [126] is used to calculate a complete set of orthogonal nine-parameter vectors in JETSET parameter space [125]. The intersections of these eigenvectors with the 1σ -contour of the χ^2 -fit provides eighteen sets of JETSET parameters corresponding to the uncertainties in the Monte-Carlo model.

The average deviation of the baseline multiplicities with those extracted with each of the eighteen alternate Monte-Carlo model assumptions is taken as a systematic uncertainty. This uncertainty is presented in Figs. 4.19 to 4.21. Figures 4.14 to 4.17 compare this uncertainty with the other major contributions to the systematic uncertainty.

The Monte-Carlo model-dependence results in a systematic uncertainty up to 3% to 4%, with a relative size that is fairly constant for all binnings and projections.

4.9.4 Time Stability

As mentioned in Section 4.1, both the proton and deuteron samples consist of data taken over three different years, between 2000 and 2005. The running conditions were not exactly identical, e.g., beam conditions and beam charge, different target polarization and densities (cf. Table 4.1). Some degree of detector degradation is also to be expected over the course of five years.

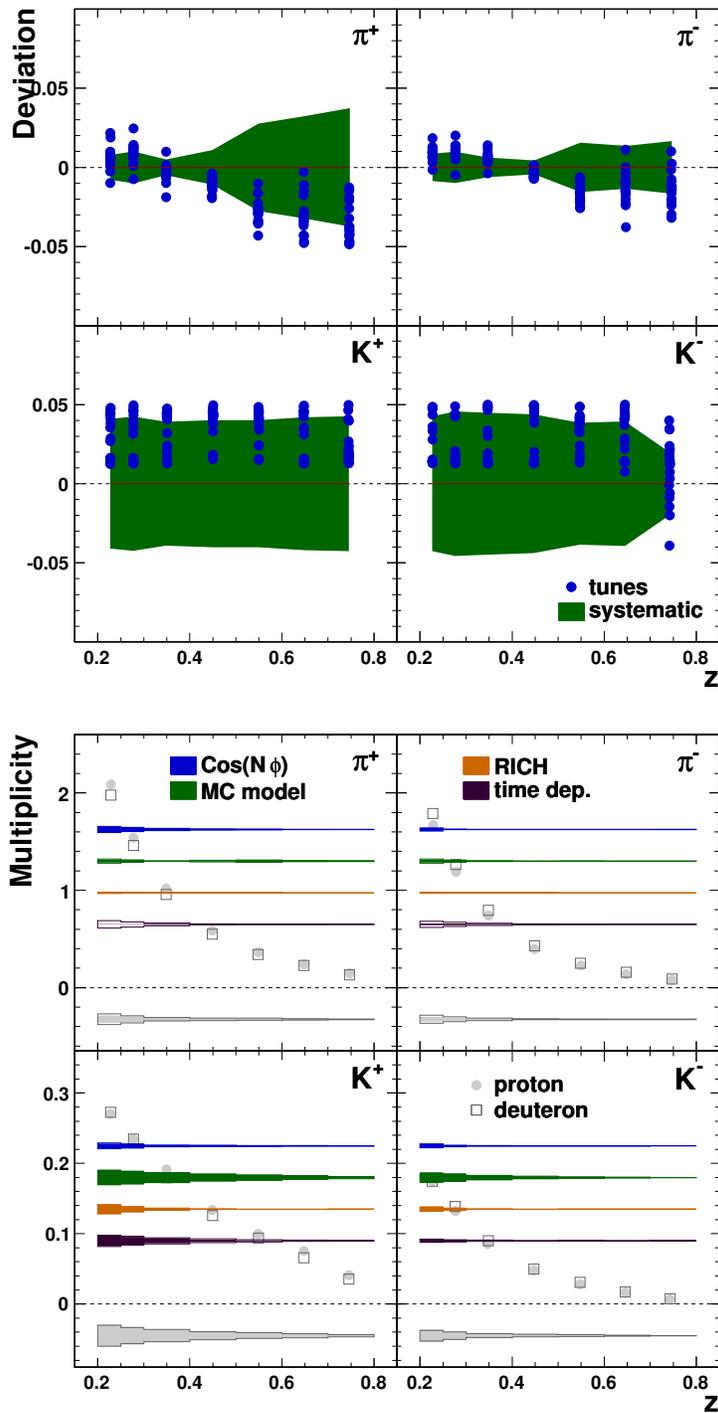


Figure 4.19: The top panels show the relative deviation (full circles) of the eighteen multiplicities obtained with the alternate Monte-Carlo model assumptions to the baseline multiplicities in various projections as a function of z . The final systematic uncertainty, obtained by averaging these deviations, is shown as a green band. The bottom panels compare this uncertainty with the other major contributions to the systematic uncertainties. The symbols are defined in Fig. 4.14.

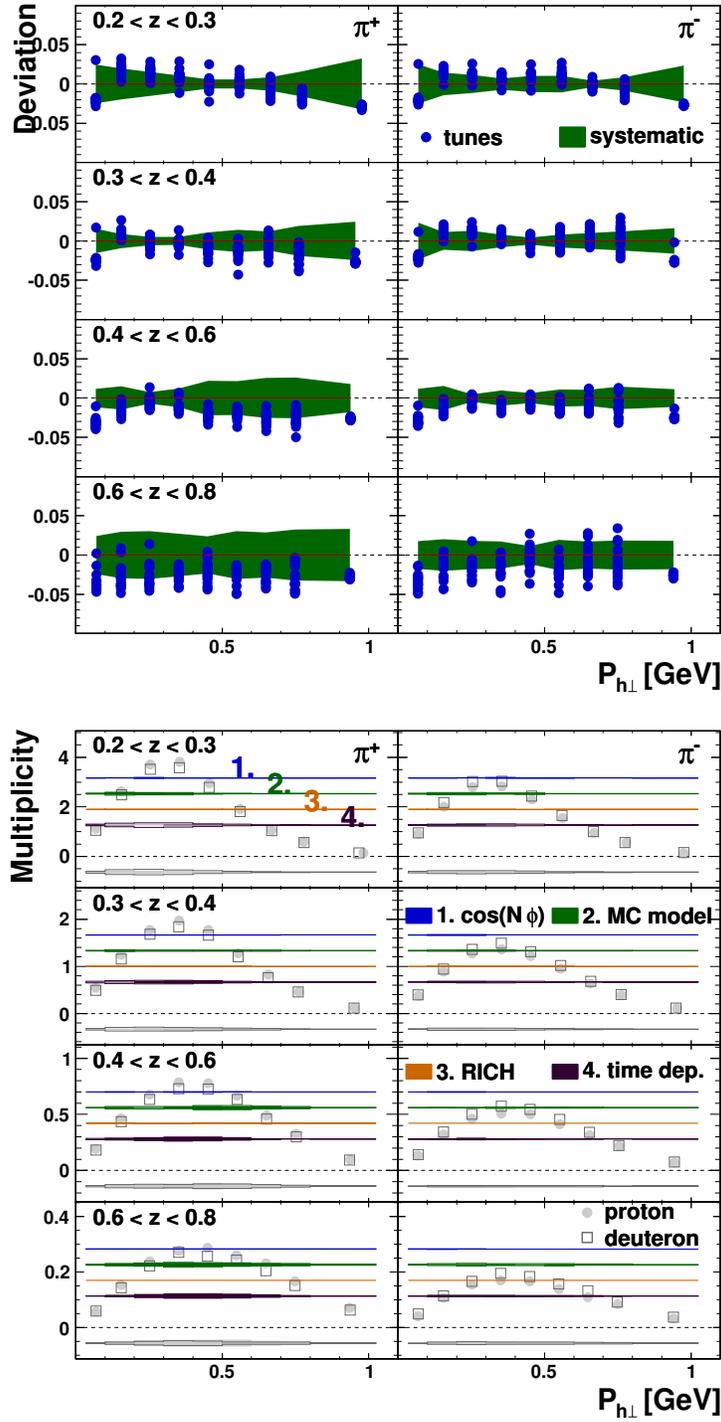


Figure 4.20: The top panels show the relative deviation (full circles) of the eighteen multiplicities obtained with the alternate Monte-Carlo model assumptions to the baseline multiplicities in various projections for pions as a function of $P_{h\perp}$ in four z -bins. The final systematic uncertainty, obtained by averaging these deviations, is shown as a green band. The bottom panels compare this uncertainty with the other major contributions to the systematic uncertainties. The symbols are defined in Fig. 4.14.

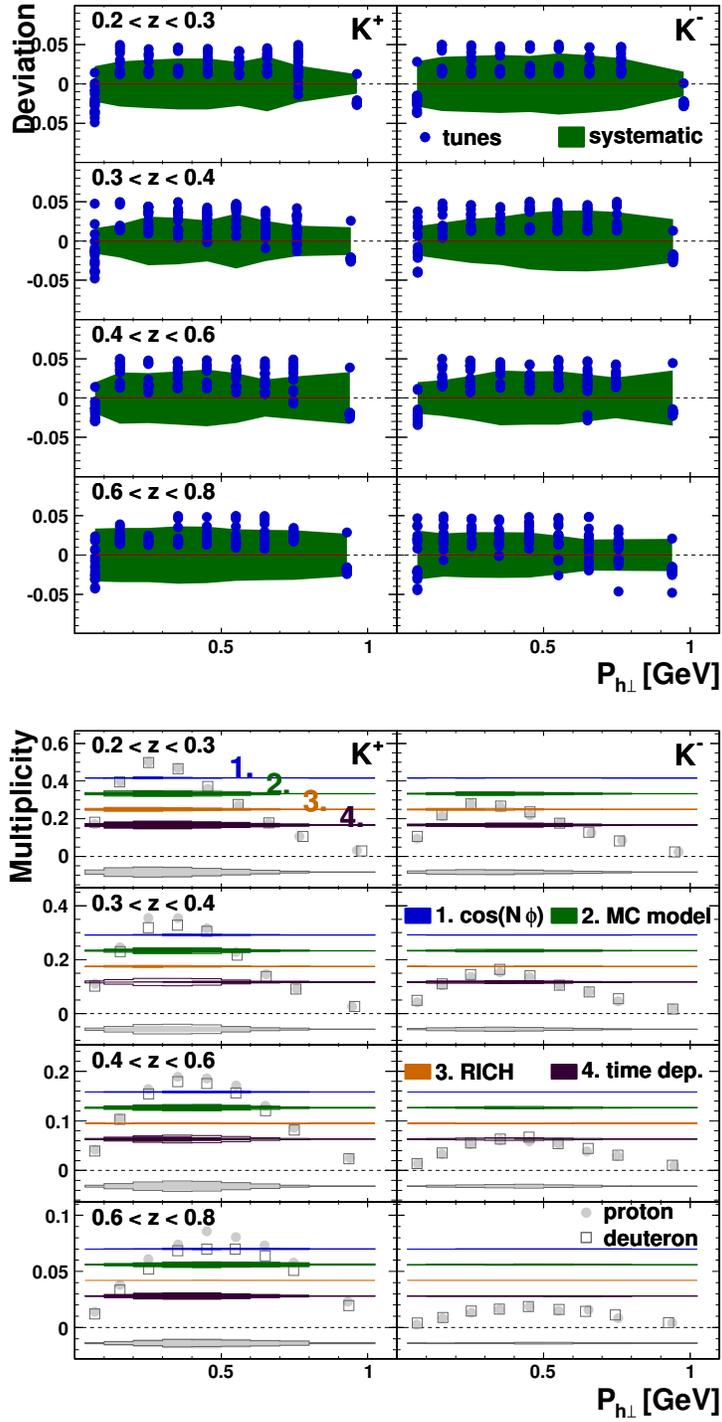


Figure 4.21: The top panels show the relative deviation (full circles) of the eighteen multiplicities obtained with the alternate Monte-Carlo model assumptions to the baseline multiplicities in various projections for kaons as a function of $P_{h\perp}$ in four z -bins. The final systematic uncertainty, obtained by averaging these deviations, is shown as a green band. The bottom panels compare this uncertainty with the other major contributions to the systematic uncertainties. The symbols are defined in Fig. 4.14.

The results from the individual years are compared to each other to investigate if their differences could be explained by statistical fluctuations by means of a two-tailed t -test¹[104]. In this case, the null hypothesis states that the weighted deviation δ_{y_1,y_2} between the productions y_1 and y_2 ,

$$\delta_{y_1,y_2}(i) \equiv \frac{M_{y_1}(i) - M_{y_2}(i)}{\sqrt{\sigma_{y_1}^2(i) + \sigma_{y_2}^2(i)}}, \quad (4.48)$$

be compatible with zero. When comparing n bins, the t -value for a two-tailed t -test with $(n-1)$ degrees of freedom is given by,

$$t = \sqrt{n} \frac{\bar{\delta}_{y_1,y_2}}{\hat{\sigma}_{y_1,y_2}}, \quad (4.49)$$

with $\bar{\delta}_{y_1,y_2}$ the mean of Eq. (4.48) for all bins, and $\hat{\sigma}_{y_1,y_2}$ the estimator for its uncertainty,

$$\hat{\sigma}_{y_1,y_2} = \sqrt{\frac{\sum_i (\delta_{y_1,y_2}(i) - \bar{\delta}_{y_1,y_2})^2}{n-1}}. \quad (4.50)$$

To identify potential problem regions, the t -test is performed for different slices in two-dimensional projections in addition to the full datasets. One of these tests, comparing the 2000 and 2005 deuteron data, is illustrated in Figs. 4.22 and 4.23. The different years were found not to be compatible with each other at the 95 % confidence level, necessitating the introduction of a systematic uncertainty to account for time instabilities. The discrepancy are mainly concentrated in the low-momentum region.

This uncertainty is calculated by taking the weighted average of the relative deviations Δ_y ,

$$\Delta_y(i) \equiv \frac{M_y(i) - M(i)}{M(i)}, \quad (4.51)$$

with M the average multiplicity from Section 4.8, and M_y the multiplicity from the year y , for the bin i . To avoid having the systematic uncertainties be influenced by statistical fluctuations, these relative deviations are smoothed by means of a three-dimensional second order polynomial. The systematic uncertainty is then given by,

$$\text{syst} = M(i) \frac{\sum_y w_y(i) \tilde{\Delta}_y(i)}{\sum_y w_y(i)}, \quad (4.52)$$

¹A simple χ^2 -test is sign-agnostic, and therefore unable to resolve small but systematic shifts

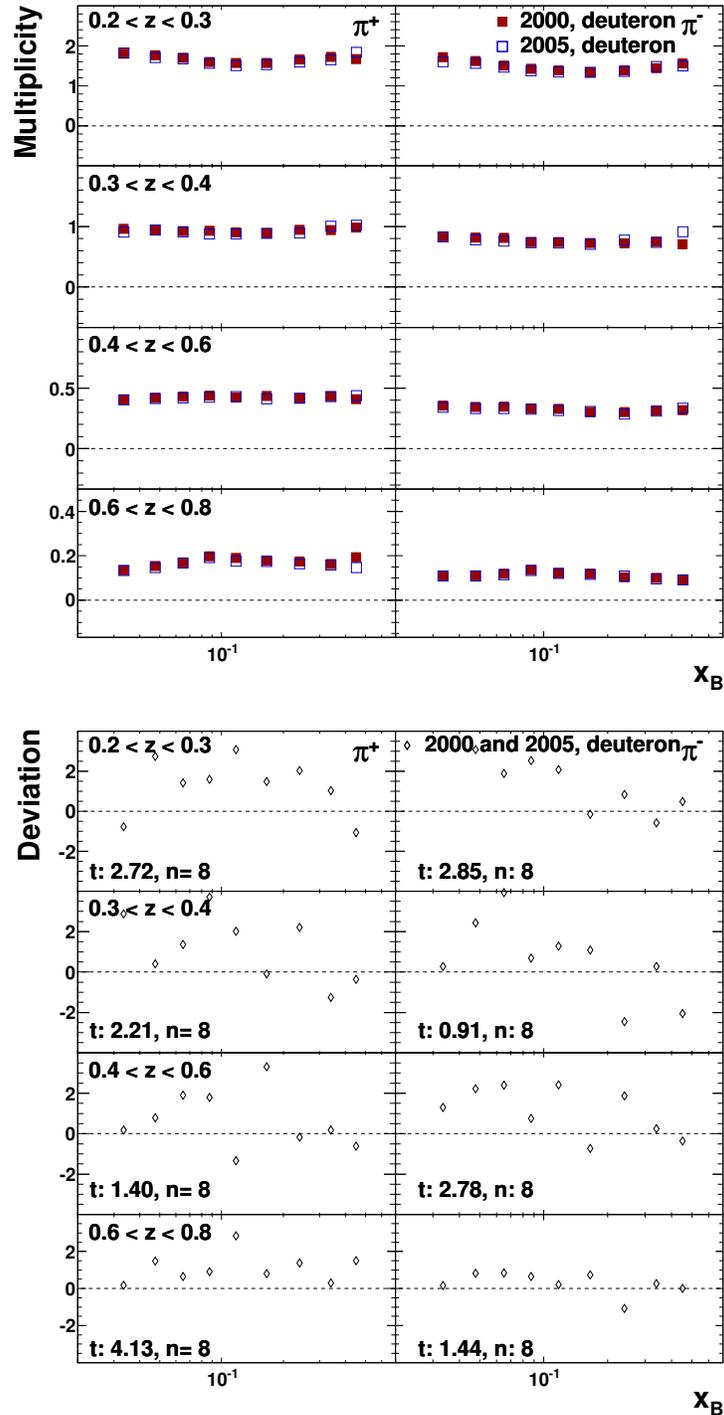


Figure 4.22: Comparison of the 2000 and 2005 deuteron results for pions as a function of x in z -slices. The bottom panels show the results of the t -test to test their statistical compatibility. For agreement within the 95 % confidence level, the t -value for eight degrees of freedom (nine x bins) has to be smaller than 2.306. This is clearly not the case for a significant amount of kinematic regions. The datasets are judged to be not compatible and the introduction of a systematic uncertainty to account for the time fluctuations has to be introduced.

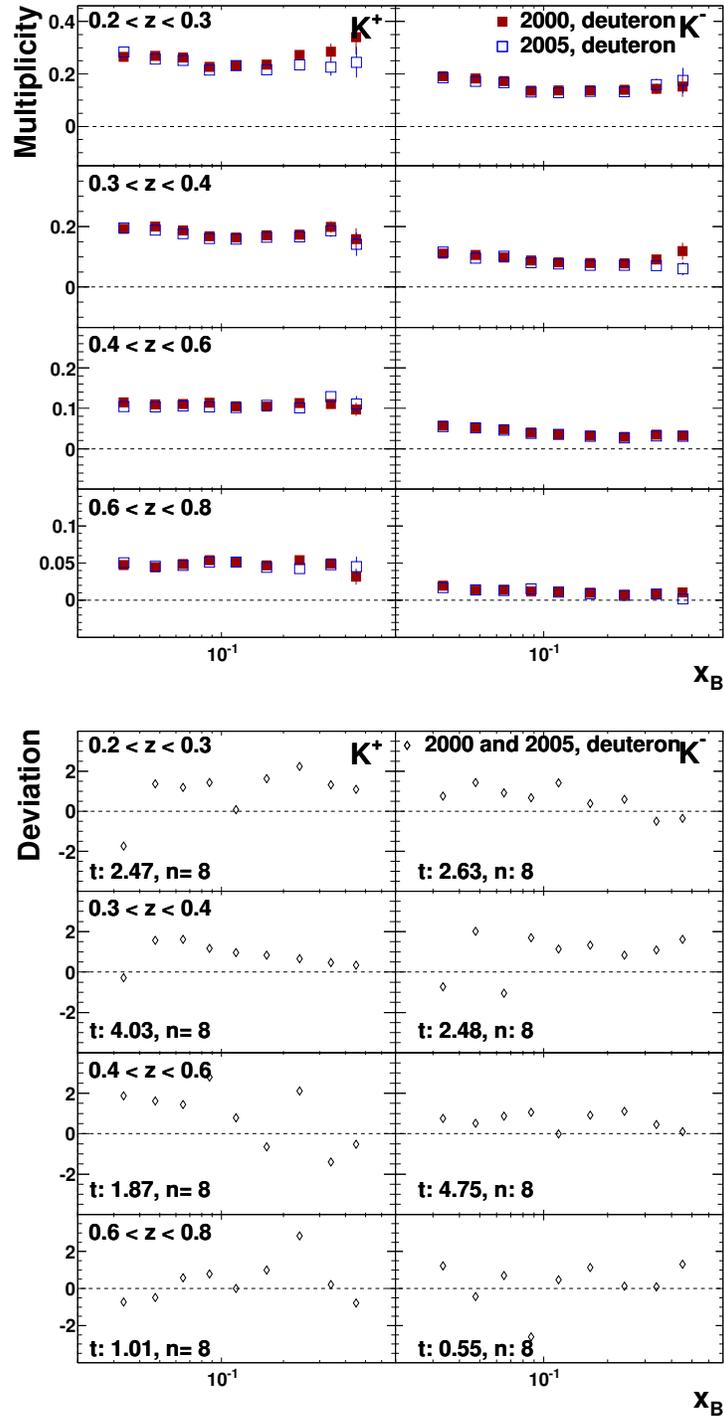


Figure 4.23: Comparison of the 2000 and 2005 deuteron results for kaons as a function of x in z -slices. The bottom panels show the results of the t -test to test their statistical compatibility. For agreement within the 95 % confidence level, the t -value for eight degrees of freedom (nine x bins) has to be smaller than 2.306. This is clearly not the case for a significant amount of kinematic regions. The datasets are judged to be not compatible and the introduction of a systematic uncertainty to account for the time fluctuations has to be introduced.

with $\tilde{\Delta}$ the smoothed relative deviations, and $w_y(i)$ as in Eq. (4.43). The size of this uncertainty is of the order of 2 % of the measured value. Figures 4.14 to 4.17 compare the impact of the time stability systematic to the other major contributions of the systematic uncertainty.

4.9.5 Minor Contributions

Finite Monte-Carlo Statistics

The Monte-Carlo simulations that are used to calculate the smearing matrices have a number of simulated events. This will lead $n_{M, MC}^{l(h)}(i, j)$ and $n_{B, MC}^{l(h)}(i)$ in Eq. (4.26), and therefore the smearing matrices themselves, to have a finite statistical uncertainty.

These effects were evaluated by modifying Eq. (4.25) to take into account these uncertainties. The resulting uncertainties were found to be smaller than 0.01 % for all bins of all binnings, as well as in all the projections. This is negligible compared to the other systematic uncertainties.

Exclusive Vector Meson Contribution

The 1σ uncertainty on the VMD-fit to the exclusive vector meson cross sections at HERMES is known. The influence of this uncertainty on the extracted fractions is shown in Fig. 4.10.

To evaluate the impact of this uncertainty on the vector meson subtracted multiplicities, the extraction was repeated using exclusive vector meson fractions from several PYTHIA Monte-Carlo simulations, varying the VMD tune on the 1σ band. In fact, this is how the bands in Fig. 4.10 were obtained.

The correction for contamination by exclusive vector meson production has a relatively small impact on the multiplicities because it affects both the numerator and denominator in a similar way. Furthermore, the uncertainty on the fraction is already relatively small, and therefore the systematic uncertainty connected to the correction was found to be significantly smaller than 1 %, except for a few bins at low x and high z , where the impact of the correction is at its largest. Even in these bins, the major contributions to the systematic uncertainties are comparatively much larger.

4.9.6 Comparison with the Statistical Uncertainties

Adding the different contributions to the systematic in quadrature yields the full systematic uncertainty. Figures 4.24 to 4.26 compare the systematic uncertainty to the size of the statistical uncertainty. The systematic uncertainty is typically 1

to 10 times the size of the statistical uncertainty. This is even the case without considering the (largely negative) bin-to-bin correlations in the covariance matrix that increase the effective statistical precision of the result. The size of the systematic uncertainty illustrates how the multiplicity extraction pushes the HERMES spectrometer to its experimental limits. It should be noted again that systematic uncertainties are highly correlated, and more complicated to interpret than a simple Gaussian statistical uncertainty. For this reason the systematic uncertainty is re-evaluated for all projections, as well as the derived quantities presented in the following chapters (cf. Section 4.8).

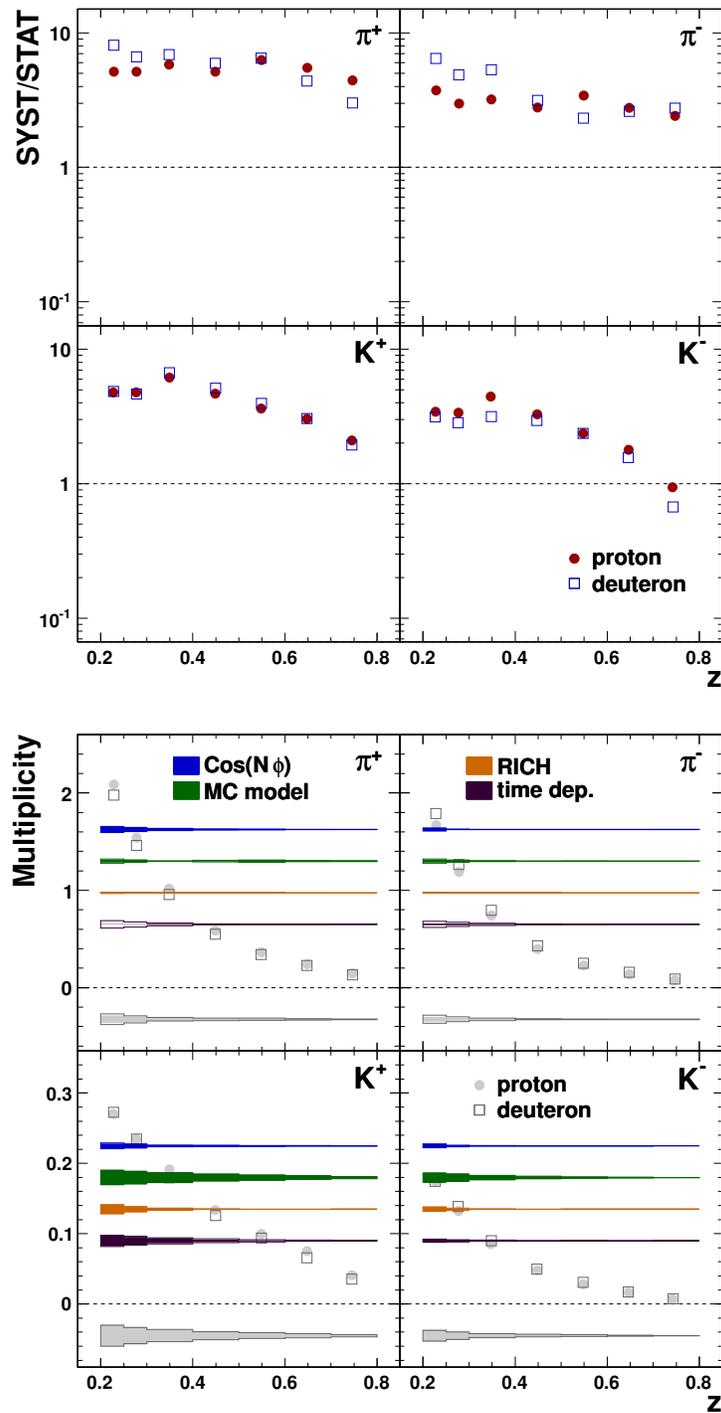


Figure 4.24: The top panels show the ratio of the final systematic uncertainty over the statistical uncertainty as a function of z . For most of the kinematic space, the value ranges between 1 to 10. Even with the sophisticated extraction procedure discussed in this chapter, the systematic uncertainty is dominant. Because of this, the precision of the results presented in this work is the highest that can be attained at the HERMES experiment.

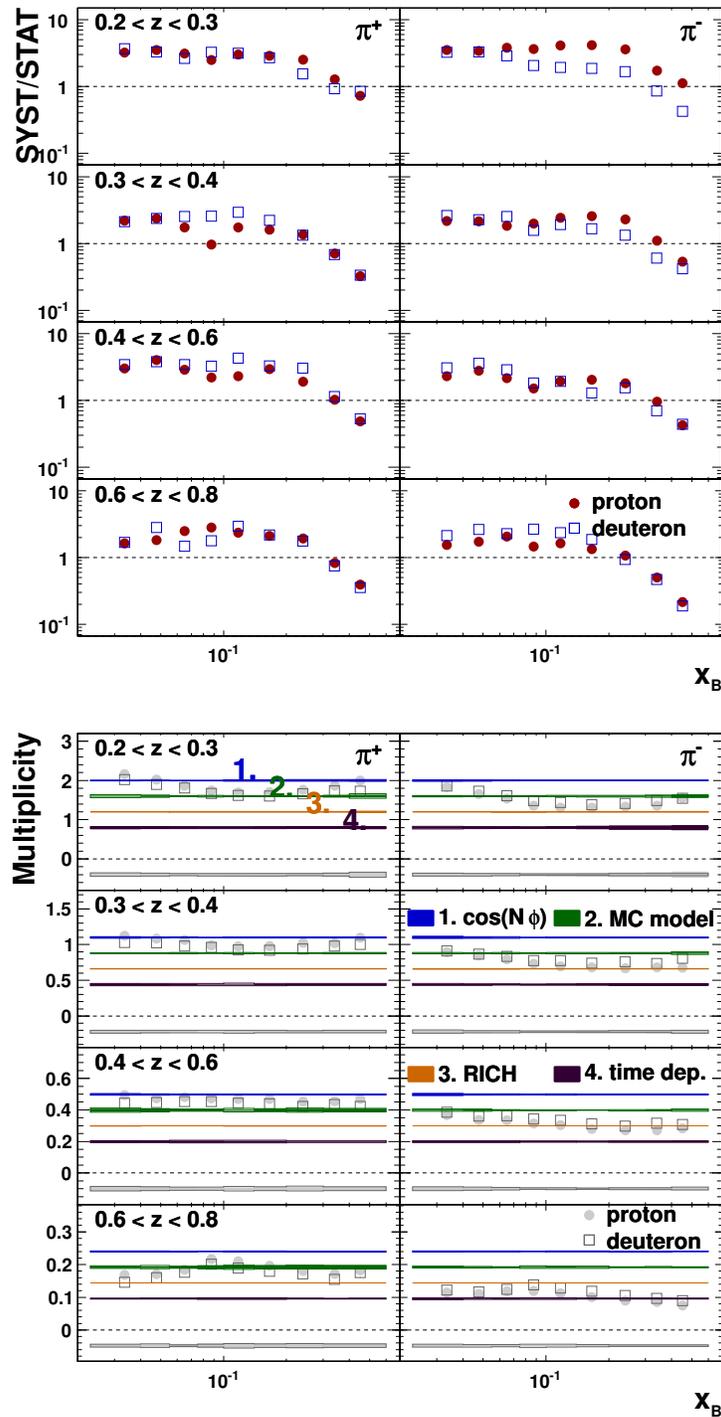


Figure 4.25: The top panels show the ratio of the final systematic uncertainty over the statistical uncertainty as a function of x in four z -bins for pions. For most of the kinematic space, the value ranges between 1 to 10. Even with the sophisticated extraction procedure discussed in this chapter, the systematic uncertainty is dominant. Because of this, the precision of the results presented in this work is the highest that can be attained at the HERMES experiment.

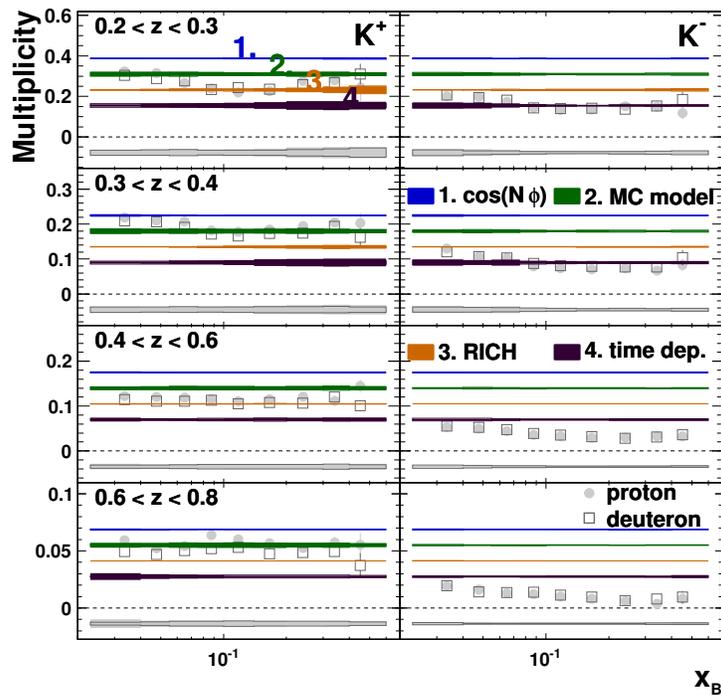
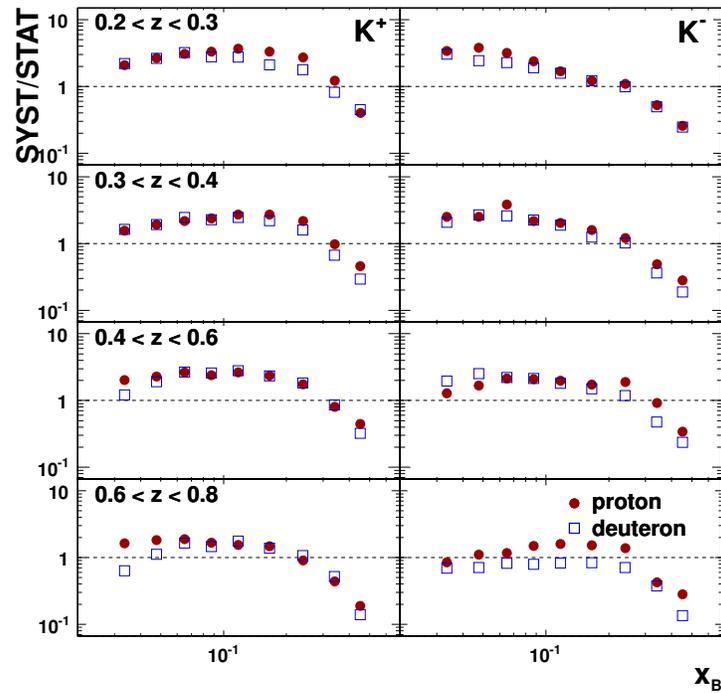


Figure 4.26: The top panels show the ratio of the final systematic uncertainty over the statistical uncertainty as a function of x in four z -bins for kaons. For most of the kinematic space, the value ranges between 1 to 10. Even with the sophisticated extraction procedure discussed in this chapter, the systematic uncertainty is dominant. Because of this, the precision of the results presented in this work is the highest that can be attained at the HERMES experiment.

5 Hadron Multiplicity Results

This chapter presents the Born-level multiplicities in various one- and two-dimensional projections. As mentioned in Section 4.7, all the results in this work are corrected for the contamination due to exclusive vector meson production. The impact of these results is discussed in Chapters 6 to 8, in the context of collinear DIS, transverse-momentum dependence and the limits of the factorization theorem. The results from this work are published in Ref. [127].

5.1 One-Dimensional Projection versus z

Figure 5.1 presents the Born-level multiplicities as a function of z . The four panels correspond to π^+ , π^- , K^+ and K^- . In each of the panels, the data taken with a hydrogen target (full circles) are compared with those taken on a deuterium target (empty squares). The statistical error bars are generally too small to be visible. The systematic uncertainties are given by the error bands. The z -bins for this projection are defined in Table 4.5a.

The pion results in a one-dimensional projection (top row in Fig. 5.1) show a good level of agreement with the results of a more older one-dimensional extraction of pion multiplicities from the (much smaller) pre-2000 dataset at HERMES published in Ref. [78].

The proton π^+ multiplicities are slightly larger than those of the deuteron, the reverse being true for the π^- multiplicities. Additionally, the π^+ multiplicities are significantly larger than the π^- multiplicities, with a π^+/π^- ratio for the proton (deuteron) ranging from 1.2 (1.1) at low z to 2.6 (1.8) at high z .

The fragmentation process $u \rightarrow \pi^-(\bar{u}d)$ requires both a \bar{u} - and d -quark from the vacuum, while $u \rightarrow \pi^+(u\bar{d})$ only requires a \bar{d} -quark from the vacuum. This makes the former process energetically **unfavored**, and the latter energetically **favored**. The fraction of u -quarks for the proton is larger than for the deuteron. As a result, the favored fragmentation of a u -quark into a π^+ leads to a larger multiplicity for the proton than for the deuteron. Likewise, the increased fraction of d -quarks in the neutron, coupled with the favored fragmentation of a d -quark into a π^- , explains the higher π^- multiplicities for the deuteron.

The values for the π^+/π^- ratio can be attributed to the dependence of the DIS

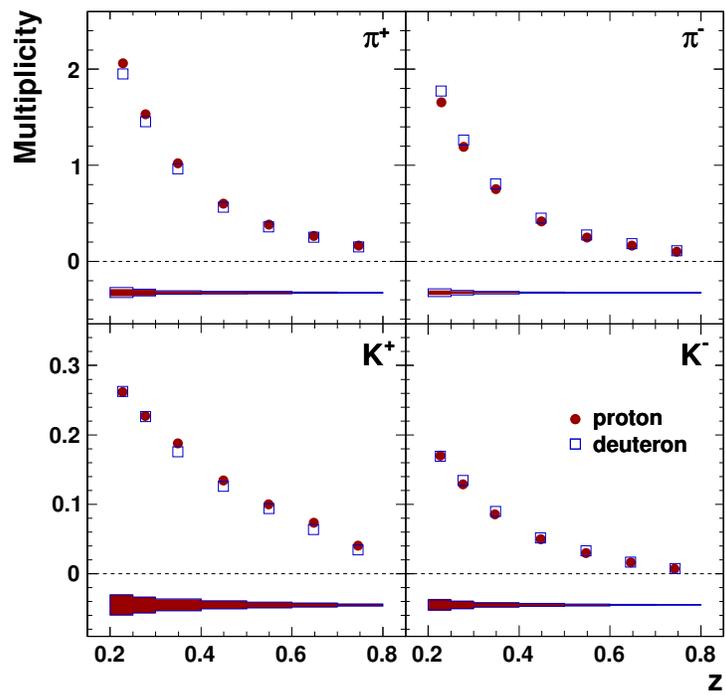


Figure 5.1: Pion and kaon multiplicities off the proton (full circles) and the deuteron (empty squares), presented as a function of z . Statistical error bars are too small to be visible. The systematic uncertainties are given by the error bands.

cross section on the square of the electrical charge of the quarks e_q (cf. Eq. (2.16)),

$$d\sigma^J \propto \sum_q e_q^2 f_q(x) \otimes \hat{\sigma}. \quad (5.1)$$

Therefore, the contribution to the cross section from scattering off a u -quark (charge $+2/3$) is enhanced by a factor of four compared to the contribution from scattering off a d -quark (charge $-1/3$). The combination of this u -quark dominance with energetically favored and unfavored fragmentation explains why the π^+ multiplicities are larger than the π^- multiplicities. Furthermore, at higher values of z , the produced hadron is automatically closer related to the scattered quark. Consequently, the ratio of favored-to-unfavored fragmentation rises as a function of z , accounting for the observed increase in the π^+/π^- ratio.

The proton K^+ multiplicities are slightly larger than those of the deuteron. The K^- multiplicities on both targets are consistent with each other. The K^+ multiplicity is much larger than the K^- multiplicity, with the ratio K^+/K^- for the proton (deuteron rising from 1.5 (1.3) at low z to 5.7 (4.6) at high z). The large difference in size between the K^+ and K^- multiplicities reveals the K^- as a pure sea object, as it cannot be produced through favored fragmentation of a valence quark.

5.2 The Multiplicity Target Asymmetry

At first sight, the differences between the multiplicities results appear to be not very significant due to the size of the systematic uncertainty. However, these systematic uncertainties are highly correlated with each other the systematic uncertainties on derived quantities, e.g., differences and ratios, should be re-evaluated rather than propagated (cf. Section 4.9). To illustrate this, it is instructive to consider the multiplicity target asymmetry,

$$A_{d-p}^h \equiv \frac{M_d^h - M_p^h}{M_d^h + M_p^h}. \quad (5.2)$$

The results for this asymmetry in a one-dimensional projection versus z is given in Fig. 5.2. The sign of the π^+ and π^- asymmetries reflects the different valence-quark structure of both targets, as explained in Section 5.1. The K^+ asymmetry is similar to the π^+ asymmetry due to u -quark dominance, while the K^- asymmetry is near zero, suggesting the lower sensitivity to the valence-quark content of the target discussed above.

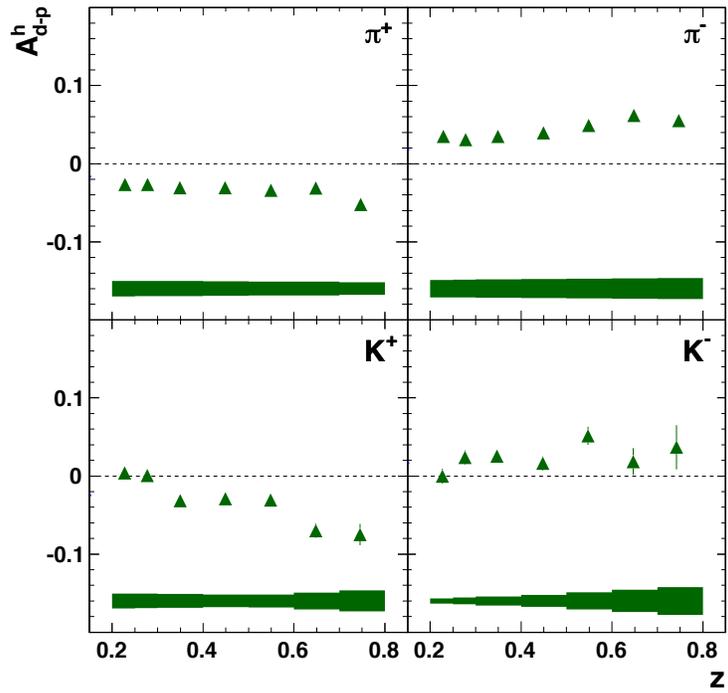


Figure 5.2: The multiplicity target asymmetry $A_{d-p}^h \equiv (M_d^h - M_p^h)/(M_d^h + M_p^h)$ for pions and kaons, presented as a function of z . The statistical uncertainty is shown by the error bars, while the systematic uncertainty is given by the error bands.

5.3 Two-Dimensional Projections

Figures 5.3 to 5.5 present the multiplicities as a function of $P_{h\perp}$, x and Q^2 in four z -bins. The relevant binnings are defined in Table 4.5b for $P_{h\perp}$, Table 4.5c for x and Table 4.5d for Q^2 . The two- and three-dimensional binnings are designed to optimally leverage the available statistics to provide insights in the correlations between kinematic variables (cf. Section 4.5). This allows for the separation of the effects of PDFs and FFs (cf. Chapters 6 and 8). Additionally, these multi-dimensional results also provide a new handle, useful for the quest to disentangle the transverse momentum structure of the nucleon from the transverse momentum that is generated by the fragmentation process. This is discussed in Chapter 7.

In context of the multiplicities extracted as a function of x , it should be noted that at HERMES kinematics, there exists a strong correlation between x and Q^2 (cf. Fig. 4.3.). Within the range of $0.023 < x < 0.6$, the value of Q^2 rises from 1 GeV^2 at low x to 10 GeV^2 at high x . The exact shape of the kinematic coverage should always be taken into account when analyzing these results.

The multiplicities display a strong kinematic dependence on z , while they are only weakly dependent on x and Q^2 . This qualitatively agrees the expectation from standard collinear factorization into universal PDFs and FFs, providing evidence that these assumptions apply for SIDIS at the low energy scales of HERMES, where $\langle Q^2 \rangle \approx 2.5 \text{ GeV}^2$ and $\langle W^2 \rangle \approx 10 \text{ GeV}^2$. This is the topic of Chapter 8.

5.4 The Multiplicity Website

The multiplicities are extracted in five different three-dimensional binnings (cf. Section 4.5) and various one- and two-dimensional projections, for four particles on two targets. Additionally, because of the correlations between systematic uncertainties for different hadrons and targets, the target asymmetries (discussed in Section 5.2), the average $P_{h\perp}$ distributions and the hadron charge asymmetries (discussed in Chapter 7), the kaon-to-pion ratios (discussed in Chapter 6) and the “ d_v/u_v ” π^\pm ratio (discussed in Chapter 8) are also provided. All of these results are extracted in two versions: with and without a correction for the contamination due to exclusive vector-meson production.

The author has created a dedicated website for the multiplicity results [128], allowing end-users to easily locate and download the files they are looking for from the hundreds of tabulated results. This includes a web-based viewer for the available figures. It also provides an overview and brief explanation of the available binnings and their ideal usage.

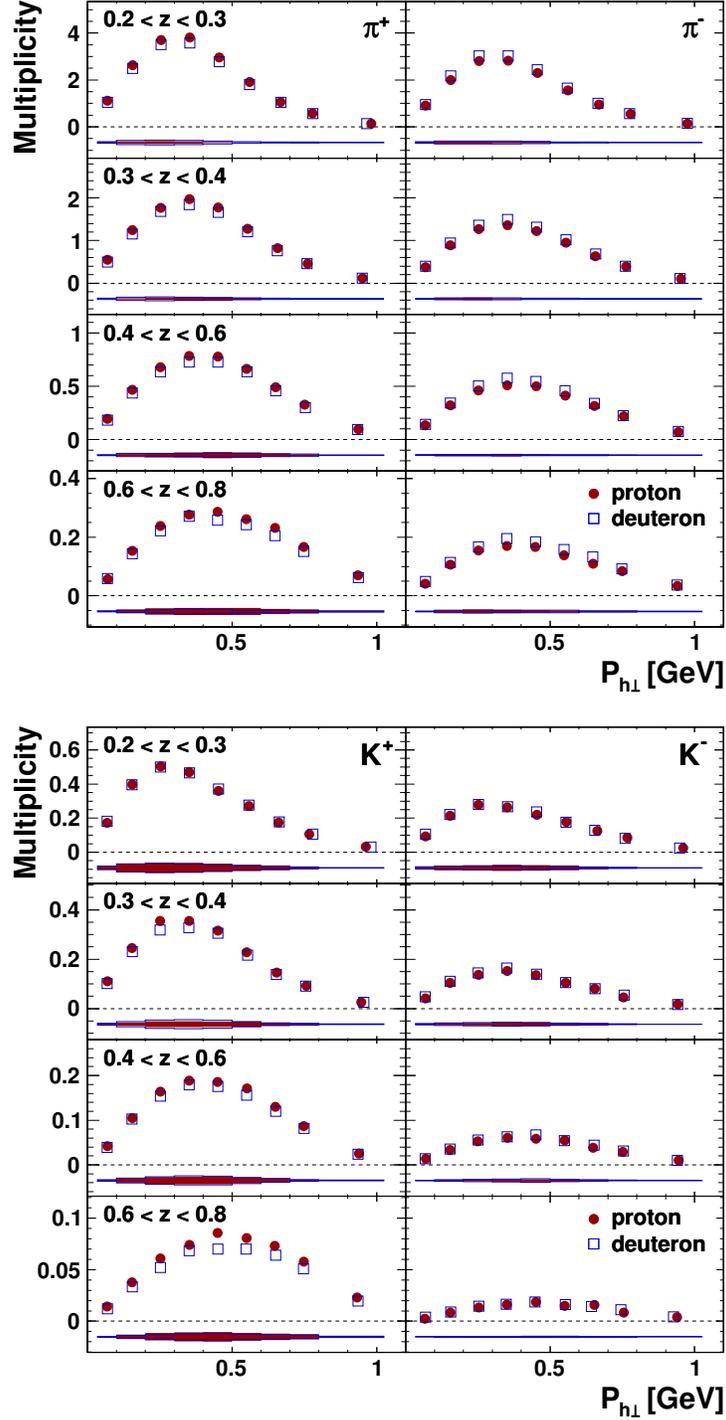


Figure 5.3: Pion multiplicities (top panels) and kaon multiplicities (bottom panels) off the proton (full circles) and the deuteron (empty squares), presented as a function of $P_{h\perp}$ in four z -bins. Statistical error bars are too small to be visible. The systematic uncertainties are given by the error bands.

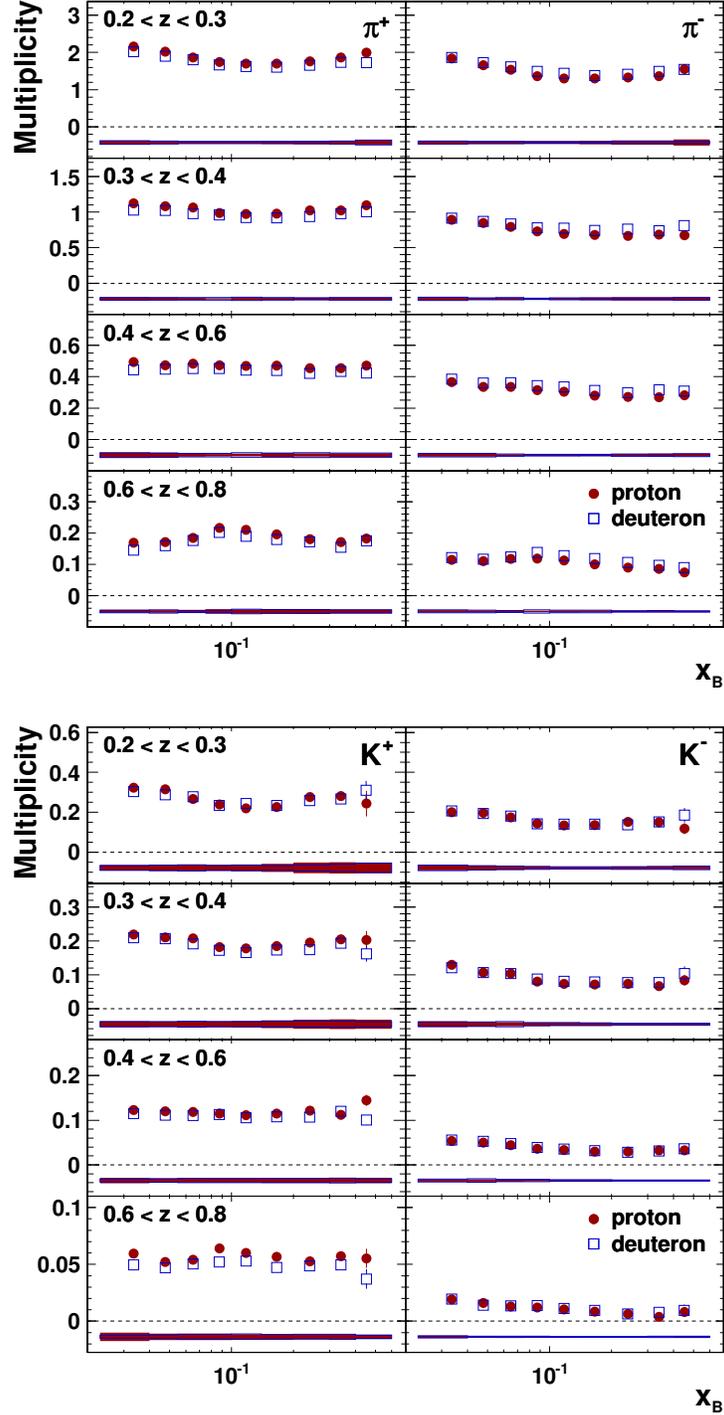


Figure 5.4: Pion multiplicities (top panels) and kaon multiplicities (bottom panels) off the proton (full circles) and the deuteron (empty squares), presented as a function of x in four z -bins. Statistical error bars are too small to be visible. The systematic uncertainties are given by the error bands.

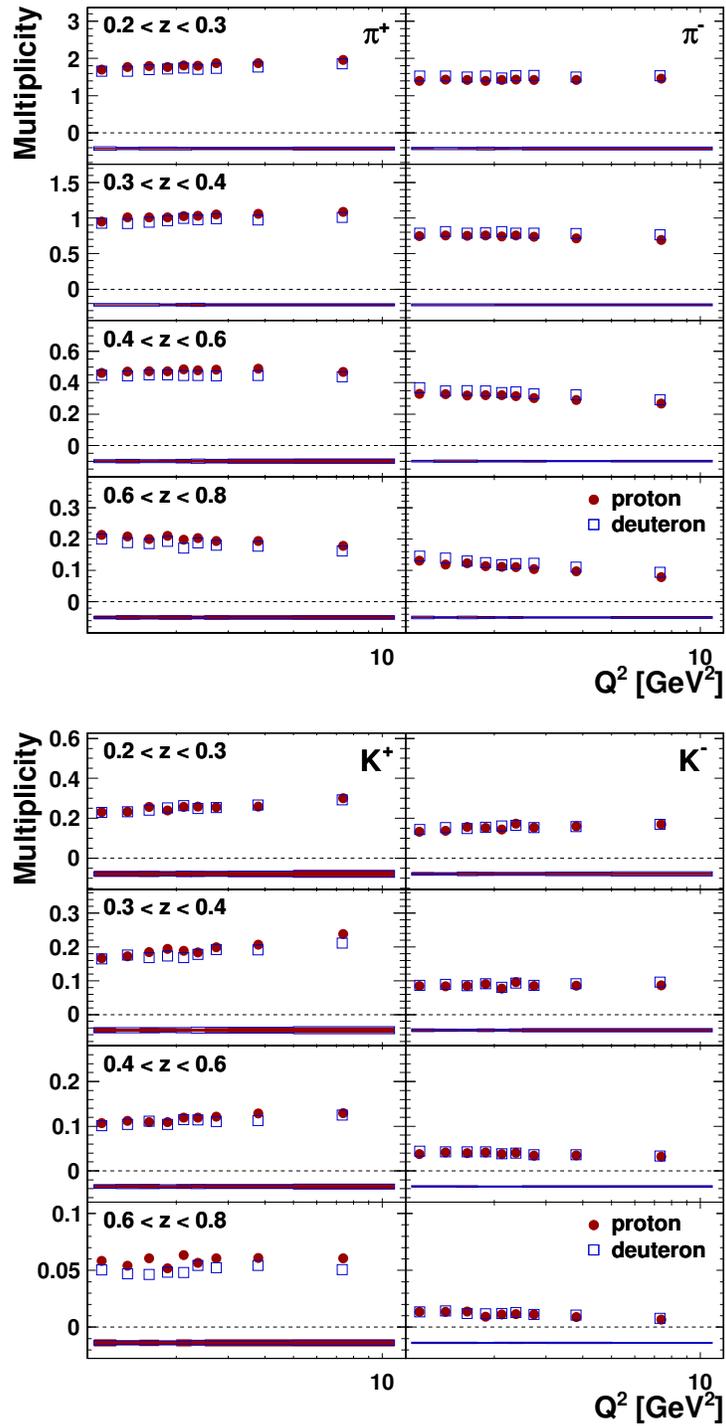


Figure 5.5: Pion multiplicities (top panels) and kaon multiplicities (bottom panels) off the proton (full circles) and the deuteron (empty squares), presented as a function of Q^2 in four z -bins. Statistical error bars are too small to be visible. The systematic uncertainties are given by the error bands.

6 Interpretation in Leading-Order Collinear DIS

A comparison of the multiplicity results with various calculations in the framework of leading-twist, collinear factorization at leading order (LO) in the strong coupling α_s , illustrates the degree to which these results will further constrain future extractions of fragmentation functions (FFs), while highlighting specific weak points in the current generation of parameterizations [12, 29–32]. Additionally, an interesting observation in itself is the fact that this kind of LO comparison is at all applicable at HERMES energies ($Q^2 \sim 2.5 \text{ GeV}$). This phenomenon of precocious scaling is the topic of Chapter 8.

6.1 Calculating the Multiplicities in Leading-Order

In the framework of leading-twist collinear factorization, the $P_{h\perp}$ and ϕ_h degree of freedoms are integrated out of the cross sections. The cross section for semi-inclusive deep-inelastic production of a hadron $d\sigma^h$ is then proportional to certain combinations of the parton distribution functions (PDFs) in the nucleon with the FFs describing the formation of the hadron h [12, 29–32]. Similarly, the cross section for inclusive DIS $d\sigma^l$ is proportional to a certain combination of PDFs in the nucleon. At LO, the cross sections are given by (cf. Eqs. (2.23) to (2.25)),

$$\frac{d^3\sigma^h}{dx dQ^2 dz} \propto \sum_q e_q^2 f_q(x, Q^2) D_q^h(z, Q^2), \quad (6.1)$$

$$\frac{d\sigma^l}{dx dQ^2} \propto \sum_q e_q^2 f_q(x, Q^2), \quad (6.2)$$

where the common pre-factors have been omitted for clarity.

Traditionally, when comparing measurements to theoretical predictions, the data points themselves are “evolved” to a common value in Q^2 . However, this approach requires the use of an evolution factor extracted from a theoretical parameterization, introducing an unnecessary model-dependence. In principle, this complicates the comparison with several theoretical predictions at once. It is, at the other hand, straightforward to unambiguously evaluate a theoretical prediction over the kinematic space covered by the data, especially at LO.

In order to calculate the multiplicities at LO, Eqs. (6.2) and (8.14) are integrated

over the relevant portion of the kinematic space covered by the data, illustrated in Fig. 4.3. To calculate the multiplicities as a function of z , this gives

$$M^h(z) = \frac{\sum_q e_q^2 \int_{0.023}^{0.6} dx \int_{Q_{\min}^2(x)}^{Q_{\max}^2(x)} dQ^2 f_q(x, Q^2) D_q^h(z, Q^2)}{\sum_q e_q^2 \int_{0.023}^{0.6} dx \int_{Q_{\min}^2(x)}^{Q_{\max}^2(x)} dQ^2 f_q(x, Q^2)}, \quad (6.3)$$

with $Q_{\min}^2(x)$ constrained by the $Q^2 > 1 \text{ GeV}^2$ and $W^2 > 10 \text{ GeV}^2$ cuts, and $Q_{\max}^2(x)$ by the $y < 0.85$ cut,

$$Q_{\min}^2 = \begin{cases} \frac{x}{1-x} (10 \text{ GeV}^2 - M^2) & \text{if } > 1 \text{ GeV}^2 \\ 1 \text{ GeV}^2 & \text{otherwise} \end{cases} \quad (6.4)$$

$$Q_{\max}^2 = 2ME x \quad (6.5)$$

with M and E the target mass and beam energy. It should be noted that replacing the integral over Q^2 with a parametrization for $\langle Q^2(x) \rangle$ gives results that are nearly indistinguishable due to the similar Q^2 -dependence of the numerator and denominator in Eq. (6.3). The more integral-based approach, as explained in this section, is superior because it closely mimics the experimental measurement.

Similar to the one-dimensional case, at LO, the multiplicities as a function of x (or Q^2) in various z -bins is given by,

$$M^h(x, z_i) = \frac{\sum_q e_q^2 \int_{Q_{\min}^2(x)}^{Q_{\max}^2(x)} dQ^2 \int_{z_i}^{z_{i+1}} dz f_q(x, Q^2) D_q^h(z, Q^2)}{\sum_q e_q^2 \int_{Q_{\min}^2(x)}^{Q_{\max}^2(x)} dQ^2 f_q(x, Q^2)}, \quad (6.6)$$

with z_i and z_{i+1} the bin edges of the z -bin with index i .

Unfortunately, the statistical and theoretical uncertainties on the various LO parameterizations f_q^g and D_q^h are typically not available. The calculations in this chapter are therefore presented without uncertainties. This prevents a precise quantification of the level of agreement between the measurements and the calculations. The following discussion is therefore fully qualitative.

6.2 Comparison with Leading-Order Calculations

A comparison of the multiplicities as a function of x , for both proton and deuteron targets, is presented in Fig. 6.1. The LO calculations are performed using Eq. (6.3). The FFs are taken from three widely used analyses that provide a LO parameterization, labeled DSS [29], HKNS [36] and Kretzer [61], together with the PDFs from CTEQ6L [37]. Additional calculations were performed with the MRST PDFs [38],

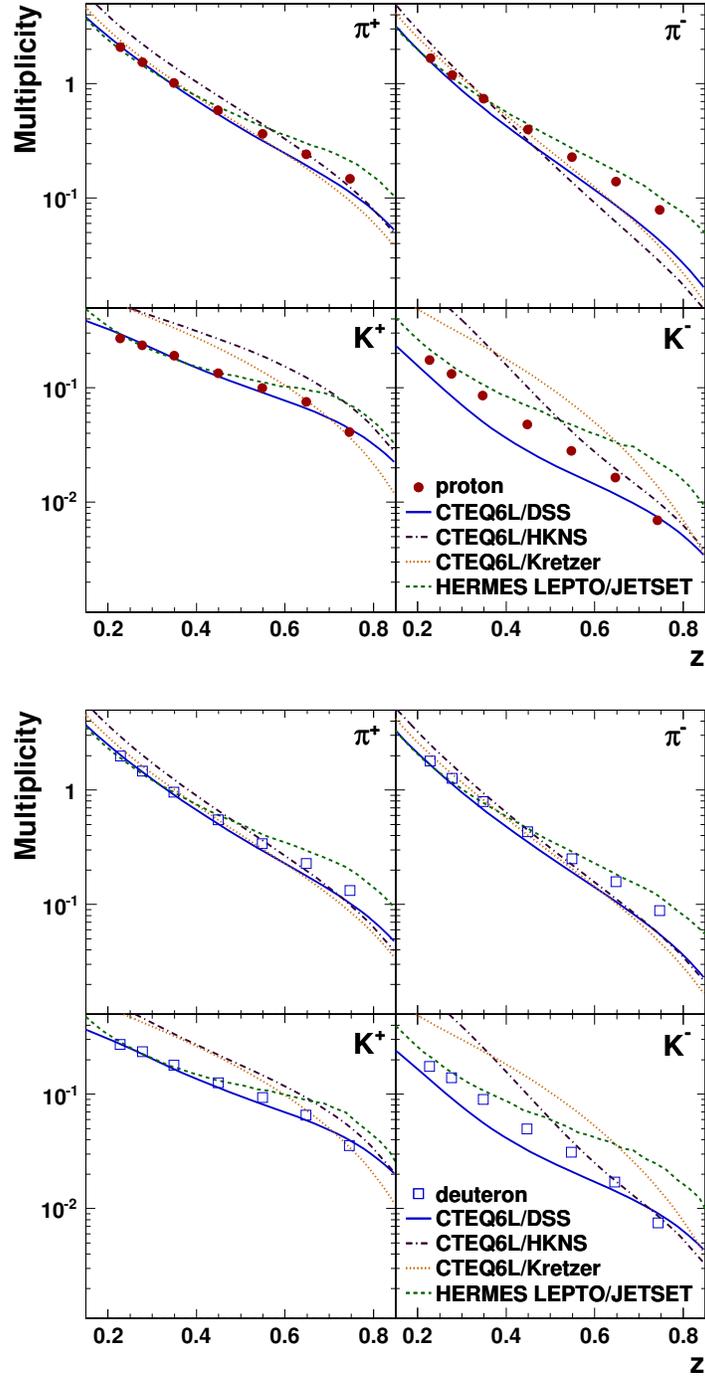


Figure 6.1: Comparison of the pion and kaon multiplicities measured on the proton (top panels) and deuteron (bottom panels), as a function of z , with LO calculations using CTEQ6L parton distributions [37] and three compilations of fragmentation functions [29, 36, 61]. A detailed description of the calculation is given in the text. Also shown are the values obtained from the HERMES Lund Monte Carlo [124]. Statistical error bars on the experimental points are too small to be visible. The corresponding systematic uncertainties are given in Fig. 5.1. Theoretical uncertainties on the LO calculations are not available.

yielding results indistinguishable from those with the CTEQ6L PDFs, testament to the high degree of precision of the current generation of PDFs.

It should be noted that the state-of-the-art DSS analysis included a very preliminary, unreleased version of the HERMES proton data in its database. Due to an incorrect momentum cut of 2 GeV at the Born level, as well as a less refined analysis, these unreleased multiplicities differ substantially from the official HERMES results. In fact, because these data introduced an unacceptable level of tension in the fit, only the fractions between the multiplicities were ultimately used [129].

For π^+ and K^+ on a proton target, the calculation using DSS results agrees reasonably well with the multiplicity results. This is not the case for the negatively charged particles. In particular the calculated K^- multiplicities are substantially lower than the measurement. The discrepancy for the π^- disappears at low values of z .

Section 5.1 explains how the DIS cross section is strongly dominated by scattering off u -quarks. In light of this, the disagreement in the negative sector could be explained by larger uncertainties on the largely disfavored production of d -, \bar{u} - and s -quarks. Alternatively, the case could be made that, especially for the Q^2 range of this measurement, the calculations and comparisons should be performed at NLO. The discrepancy could then be attributed to the possibility that NLO processes be proportionally more important for π^- and especially K^- production. Such calculations are beyond the scope of this work, and therefore left to the various theoretical groups studying the world database of fragmentation data. At the time of writing, the authors of Ref. [29] are re-evaluating their FF parameterizations using the data of this work, as well as the recently released BELLE data [130, 131].

The HKNS and Kretzer parameterizations perform relatively well for the pion data, but are unable to describe the kaon multiplicity measurements. In comparison, the DSS results represent the kaon data substantially better. The HKNS and Kretzer distributions are mostly constrained by results from high-energy e^+e^- annihilation. This clearly provides insufficient leverage to properly describe data taken at much lower energies. This is in contrast with the DSS analysis, that included a much wider variety of input data.

The predictions for the deuteron target in Fig. 6.1 are show a somewhat better level of agreement with the various predictions, except for the K^- results. Similar to the K^- off the proton, the available LO parameterizations are utterly unable to describe the K^- off the deuteron.

The top panels in Fig. 6.2 show the multiplicity target asymmetry from Section 5.2, compared with a LO calculation using the DSS FFs, as a function of z . What makes this comparison particularly interesting is the fact that the contribution from scattering off sea-quarks mostly cancels in the asymmetry. This leads to

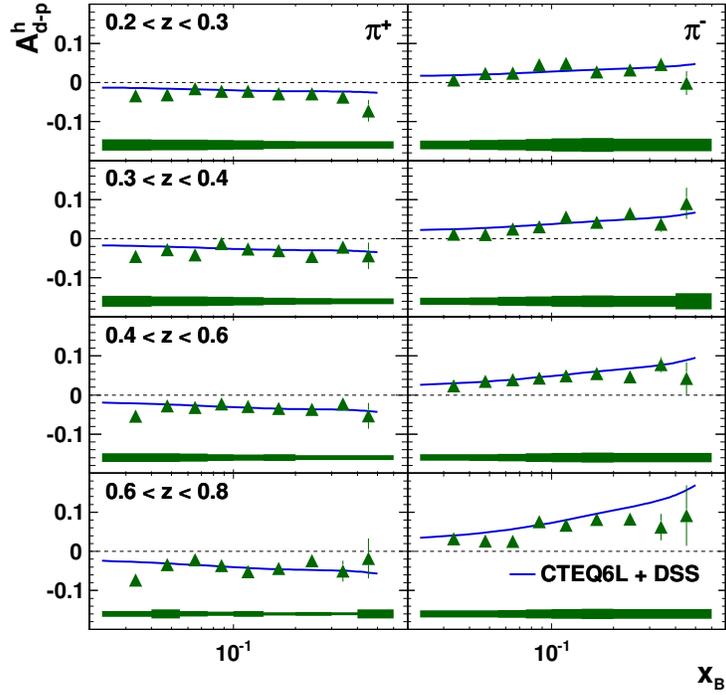
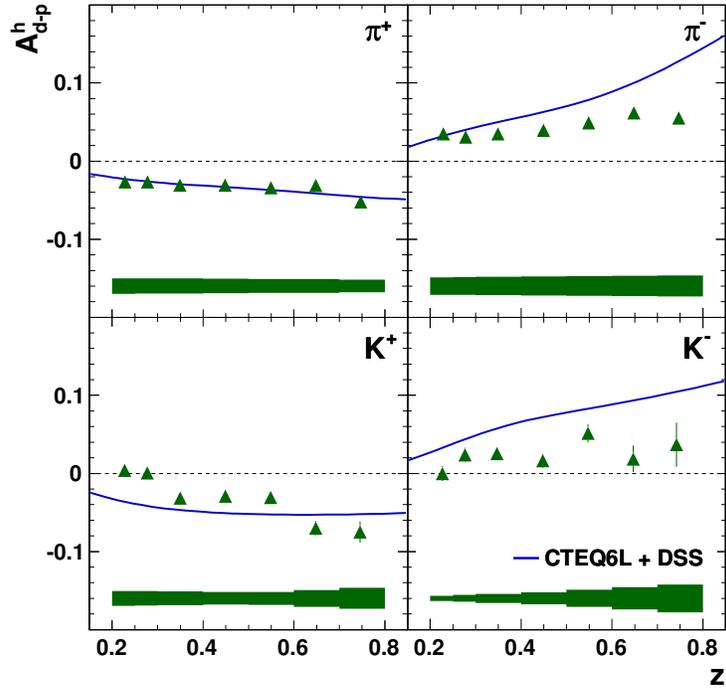


Figure 6.2: Comparison of the multiplicity target asymmetry from Section 5.2 with a LO calculation using CTEQ6L parton distributions [37] and fragmentation functions from DSS [29]. A detailed description of the calculation is given in the text. The statistical uncertainty is shown by the error bars, while the systematic uncertainty is given by the error bands. Theoretical uncertainties on the LO calculations are not available.

a comparison that is strongly sensitive of the fragmentation product from scattering off the valence quarks. Similar to Fig. 6.1, the π^+ and K^+ agree very well with the calculation. The curves increasingly overshoot the π^- results as a function of z . This is not surprising, in light of the better agreement of the parameterization with the deuteron results. The K^- asymmetry results are near-zero, while the calculation yields a curve not unlike the one for π^- . This strongly suggest the need to better constrain the fragmentation functions in the unfavored sector.

The bottom panels in Fig. 6.2 show the same comparison for pions as a function of x in four z -bins. The pion multiplicity target asymmetries A_{d-p}^π results show no significant x -dependence, and are fairly constant as a function of z . The discrepancy between the calculation and the π^- multiplicities is ostensibly only a function z .

Figures 6.3 and 6.4 present a comparison of a LO calculation using the DSS and Kretzer FFs with the multiplicities as a function of x and Q^2 in four z -bins. The LO calculations are performed using Eq. (6.6) for the x -dependence, and a analogous relation for the Q^2 -dependence. The DSS results describe the Q^2 -dependence of the HERMES results very well, with the discrepancies between measurement and calculation solely a function of z . The x -dependence at the other hand differs substantially at low x , especially so at lower values of z . This could indicate the need for a full NLO calculation, or given its occurrence at low values of z , some degree of factorization breaking. This will be discussed in detail in Chapter 8.

Superficially, the Kretzer results performed equally well as the DSS results to describe the z -dependence of the pion data in Fig. 6.1. The limits of this parameterization become more apparent in its inability to describe the Q^2 and x -dependence for both the pion and kaon multiplicity measurements. A approach beyond solely considering e^+e^- data at high energies is therefore a clear requirement for future parameterizations.

6.3 Strangeness Suppression

A final hadron with high z has a large probability of containing the struck quark. Because of u -quark dominance, this means that, for $z \rightarrow 1$, the K^+/π^+ multiplicity ratio essentially gives the extra cost of producing an $s\bar{s}$ -pair, compared to a $d\bar{d}$ -pair. This value is commonly referred to as the **strangeness suppression**. This K^+/π^+ multiplicity ratio is shown as a function of z in the top panel of Fig. 6.5. The K^-/π^- ratio is included for completeness.

The multiplicity results indicate a strangeness suppression at HERMES energies of about 0.3, significantly smaller than the commonly used value of about 0.5. This is apparent from the high- z data in Fig. 6.5, where the data drastically

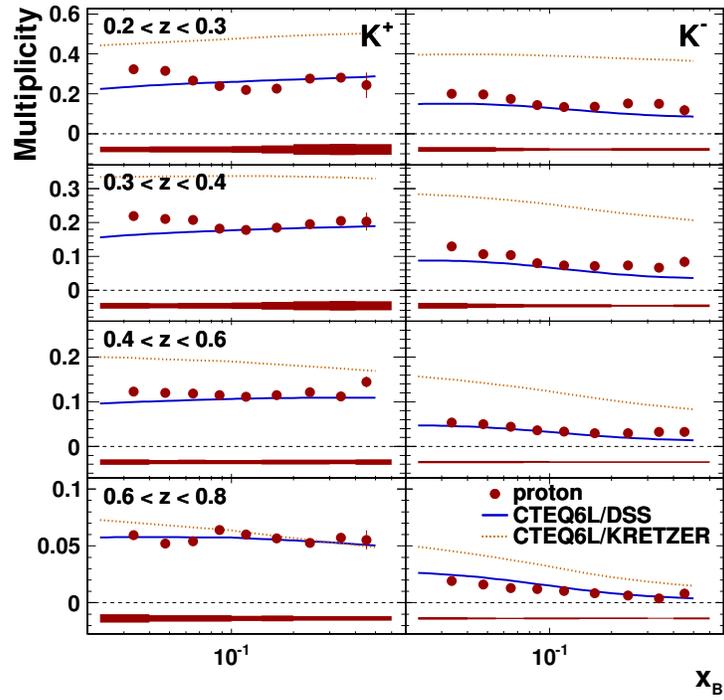
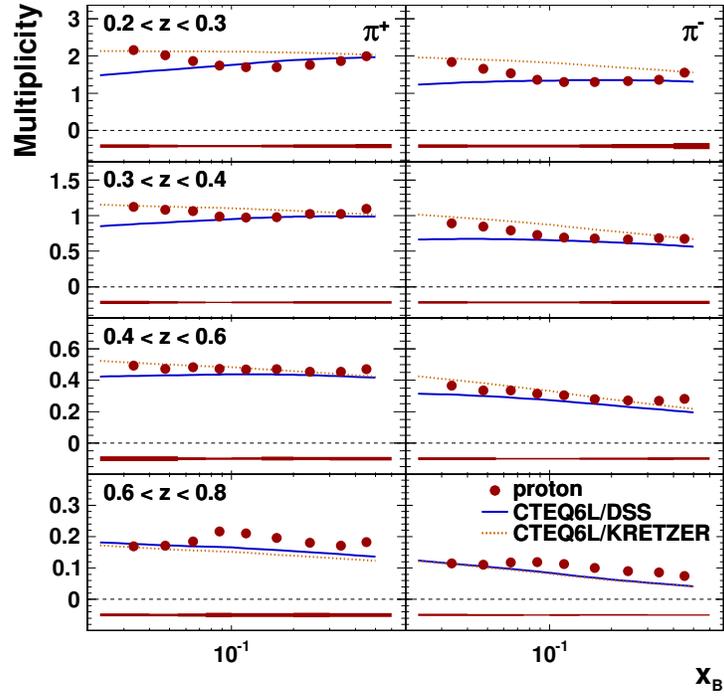


Figure 6.3: Comparison of the pion multiplicities (top panels) and kaon multiplicities (bottom panels) on the proton, as a function of x in four z -bins, with LO calculations using CTEQ6L parton distributions [37] and two compilations of fragmentation functions [29, 61]. A detailed description of the calculation is given in the text. Statistical error bars are too small to be visible. The systematic uncertainties are given by the error bands. Theoretical uncertainties on the LO calculations are not available.

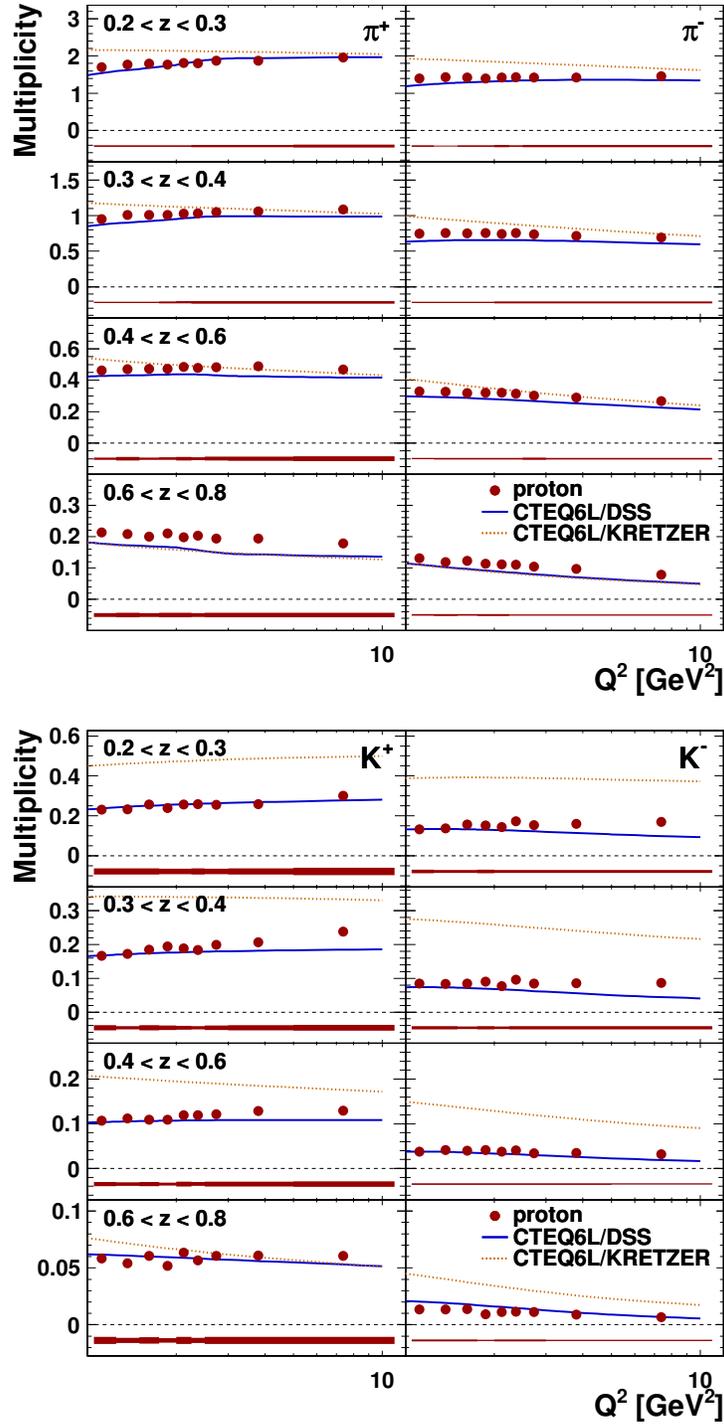


Figure 6.4: Comparison of the pion multiplicities (top panels) and kaon multiplicities (bottom panels) on the proton, as a function of Q^2 in four z -bins, with LO calculations using CTEQ6L parton distributions [37] and two compilations of fragmentation functions [29, 61]. A detailed description of the calculation is given in the text. Statistical error bars are too small to be visible. The systematic uncertainties are given by the error bands. Theoretical uncertainties on the LO calculations are not available.

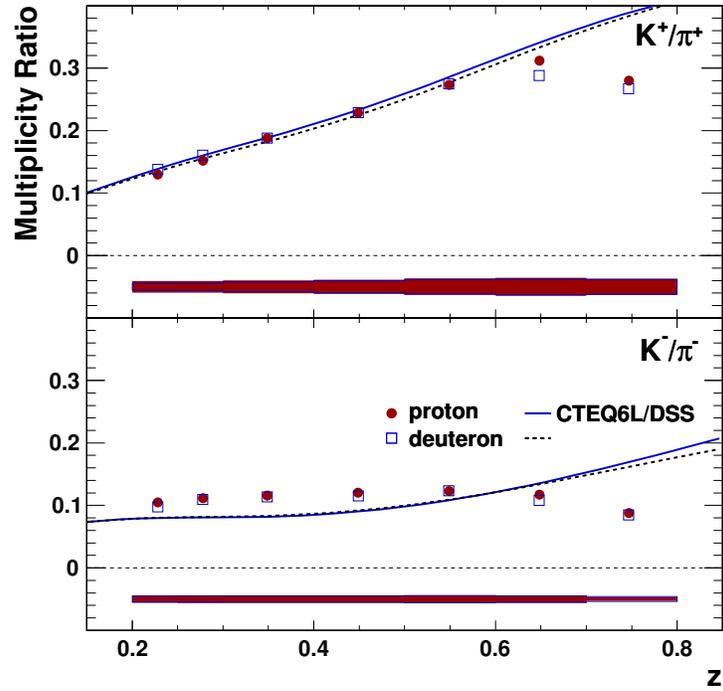


Figure 6.5: Comparison of the kaon-to-pion ratio on the proton and deuteron as a function of z with a LO calculation using CTEQ6L parton distributions [37] and fragmentation functions from DSS [29] for positive particles (top panel) and negative particles (bottom panel). The statistical uncertainty is shown by the error bars, while the systematic uncertainty is given by the error bands. Theoretical uncertainties on the LO calculations are not available. The K^+/π^+ results at high z indicate a larger degree of strangeness suppression than present in current parameterizations.

undershoots the results of a LO calculation using the DSS FFs.

The same K^+/π^+ multiplicity ratio is presented in Fig. 6.6 (left panels) as a function of x and Q^2 in four z -bins. The results are compatible with a flat x - and Q^2 -dependence, illustrating the strength of the u -quark dominance, even at low x where the sea quarks become more important. The fact that the proton and deuteron results also agree with each other strengthens this interpretation.

Finally, it is interesting to note that, except the high- z region, the LO calculation using the DSS parameterization predicts the K^+/π^+ and the K^-/π^- ratio very well, as function of x , Q^2 and z . This might be partially explained by the inclusion of the very preliminary, unreleased HERMES multiplicities in the DSS analysis (see Section 6.2).

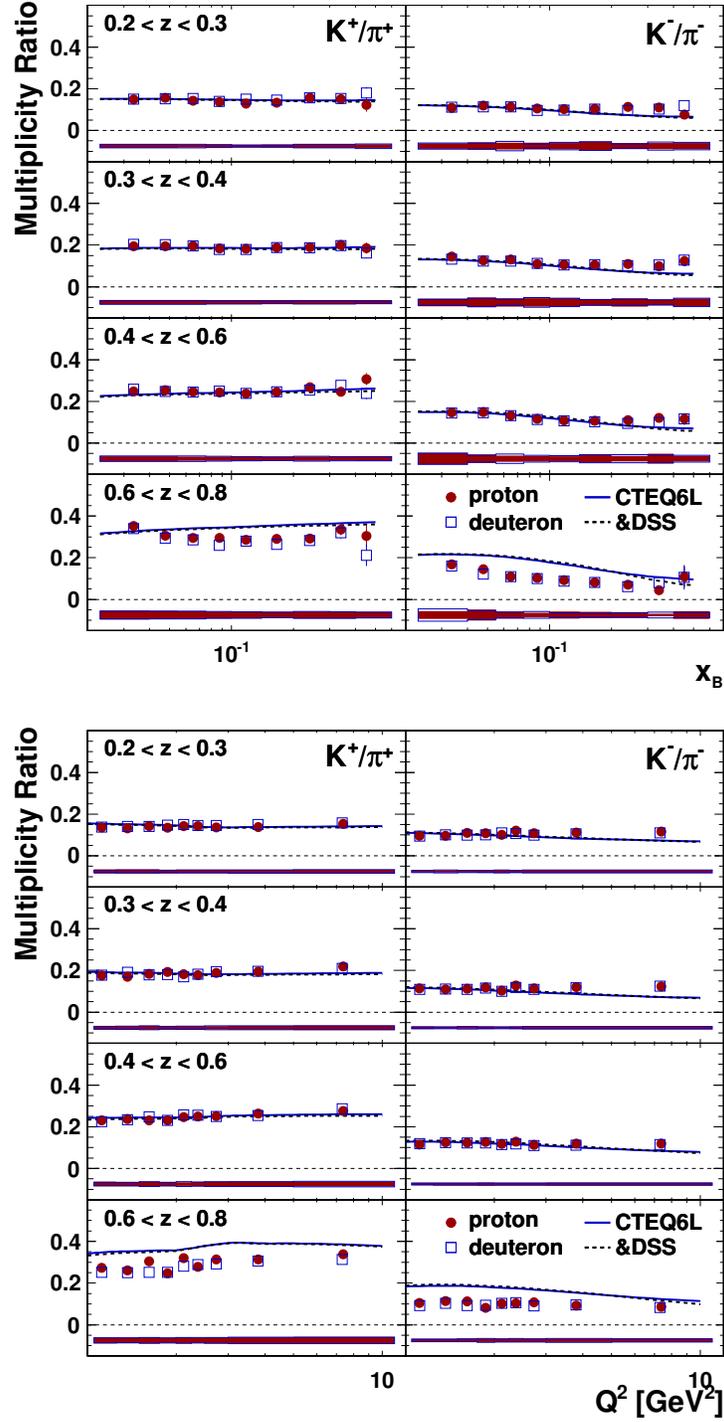


Figure 6.6: Comparison of the kaon-to-pion ratio on the proton and deuteron, as a function of x (top panels) and Q^2 (bottom panels) in four z -bins with a LO calculation using CTEQ6L parton distributions [37] and fragmentation functions from DSS [29]. The statistical uncertainty is shown by the error bars, while the systematic uncertainty is given by the error bands. Theoretical uncertainties on the LO calculations are not available. The LO parametrization predict the x and Q^2 -dependence very well up to medium z .

7 Transverse-Momentum Structure

With the inclusion of the kinematic dependence on $P_{h\perp}$, the multiplicity data reach beyond the standard collinear factorization, and hence access the transverse-momentum dependence (TMD) of the nucleon as well as the fragmentation process. The three-dimensional binning in $P_{h\perp}$, z , and x can be leveraged to constrain parameterizations in a TMD framework. Additionally, the availability of both pion and kaon results provides a handle on flavor separation. At the time of writing, these TMDs are only weakly constrained. The focus of this chapter lies therefore on a qualitative discussion of the $P_{h\perp}$ -dependent results.

7.1 Theoretical Background

The cross section for SIDIS hadron production with low transverse momentum, can be decomposed in terms of a set of structure functions [122, 132, 133]. At LO, and up to first subleading twist accuracy, these functions factorize into various combinations of transverse-momentum dependent parton distributions (TMDs) and fragmentation functions [134, 135], similar to the collinear case [11, 13]. In this framework, the four-times differential unpolarized SIDIS cross section is given by [117, 132],

$$\frac{d^4\sigma^h}{dQ^2 dx dz dP_{h\perp}^2} \propto \sum_q e_q^2 \int d^2\vec{p}_T d^2\vec{k}_T \delta^{(2)}(\vec{p}_T - \vec{k}_T - \vec{P}_{h\perp}/z) f_q(x, p_T^2) D_q^h(z, k_T^2), \quad (7.1)$$

with \vec{p}_T and \vec{k}_T defined as the transverse momentum of respectively the struck quark and the fragmenting quark, in a frame where the target nucleon and produced hadron are collinear. Additional pre-factors irrelevant to the current discussion have been omitted from Eq. (7.1) for clarity. The p_T dependence of f_q encodes the transverse momentum structure of the nucleon. The fragmentation function $D_q^h(z, k_T^2)$ is a probability density in the transverse momentum of the final-state hadron h relative to the fragmenting quark $k_T' = -zk_T$. Equation (7.1) convolutes p_T and k_T into the observable $P_{h\perp}$. This complicates the extraction of the unpolarized TMDs significantly.

At the time of writing, a common approach to unpolarized TMDs is to assume a Gaussian distribution multiplied with the standard collinear PDFs and FFs [136],

$$f_q(x, p_T^2) = \frac{f_q(x)}{\pi \langle p_T^2(x) \rangle^q} e^{-p_T^2 / \langle p_T^2(x) \rangle^q}, \quad (7.2)$$

$$D_q^h(z, k_T^2) = \frac{D_q^h(z)}{\pi \langle (k_T')^2(z) \rangle^q} e^{-z^2 k_T^2 / \langle (k_T')^2(z) \rangle^q}, \quad (7.3)$$

with $\langle p_T^2(x) \rangle^q$ and $\langle (k_T')^2(z) \rangle^q$ the average internal transverse momenta. After integration, Eq. (7.1) reduces to,

$$\frac{d^4 \sigma^h}{dQ^2 dx dz dP_{h\perp}^2} \propto \sum_q \frac{f_q(x) D_q^h(z)}{\pi (z^2 \langle p_T^2(x) \rangle^q + \langle (k_T')^2(z) \rangle^q)} e^{-P_{h\perp} / (z^2 \langle p_T^2(x) \rangle^q + \langle (k_T')^2(z) \rangle^q)}. \quad (7.4)$$

Additionally, if $\langle p_T^2(x) \rangle^q$ and $\langle (k_T')^2(z) \rangle^q$ are assumed to be independent of respectively x and z , they can be readily extracted from a $P_{h\perp}^2$ spectrum in various z -bins by means of a series of Gaussian fits.

7.2 Results for Transverse-Momentum Dependence

Figure 7.1 presents the pion and kaon multiplicities as a function of $P_{h\perp}$ in four z -bins. The data qualitatively agree with the Gaussian ansatz Eqs. (7.2) to (7.4). The distribution for negative kaons is markedly wider than for positive kaons. The K^- cannot be produced through favored fragmentation of a valence quark as discussed in Section 5.1. The extra steps necessary to fragment a valence quark into a K^- lead to additional smearing of the p_T contribution to $P_{h\perp}$, explaining the wider spectrum.

The functional behavior of $\langle P_{h\perp} \rangle$ versus x and z can be used to study possible kinematic dependences of $\langle p_T \rangle$ and $\langle k_T' \rangle$. As shown in Fig. 7.2, it rises as a function of z . This is to be expected in light of the δ -function in Eq. (7.1). The observed slope could be used to constrain $\langle p_T \rangle$, in case the z -dependence of $\langle k_T' \rangle$ can be safely neglected.

Additionally, the values for $\langle P_{h\perp} \rangle$ for kaons are significantly larger than those for pions, with a point-to-point significance of 2σ . The combination of u -quark dominance and strangeness suppression make the kaon multiplicities (slightly) more sensitive to scattering off sea-quarks. One could therefore attribute these higher values for kaons to a higher average p_T for sea quarks, consistent with the explanation for the difference in the observed Sivers amplitudes between pions and kaons in Ref. [137]. A similar increase in $\langle P_{h\perp} \rangle$ for kaons can be observed in Figs. 7.3 and 7.4 as a function of x and Q^2 in four z -bins. $\langle P_{h\perp} \rangle$ gently falls as a

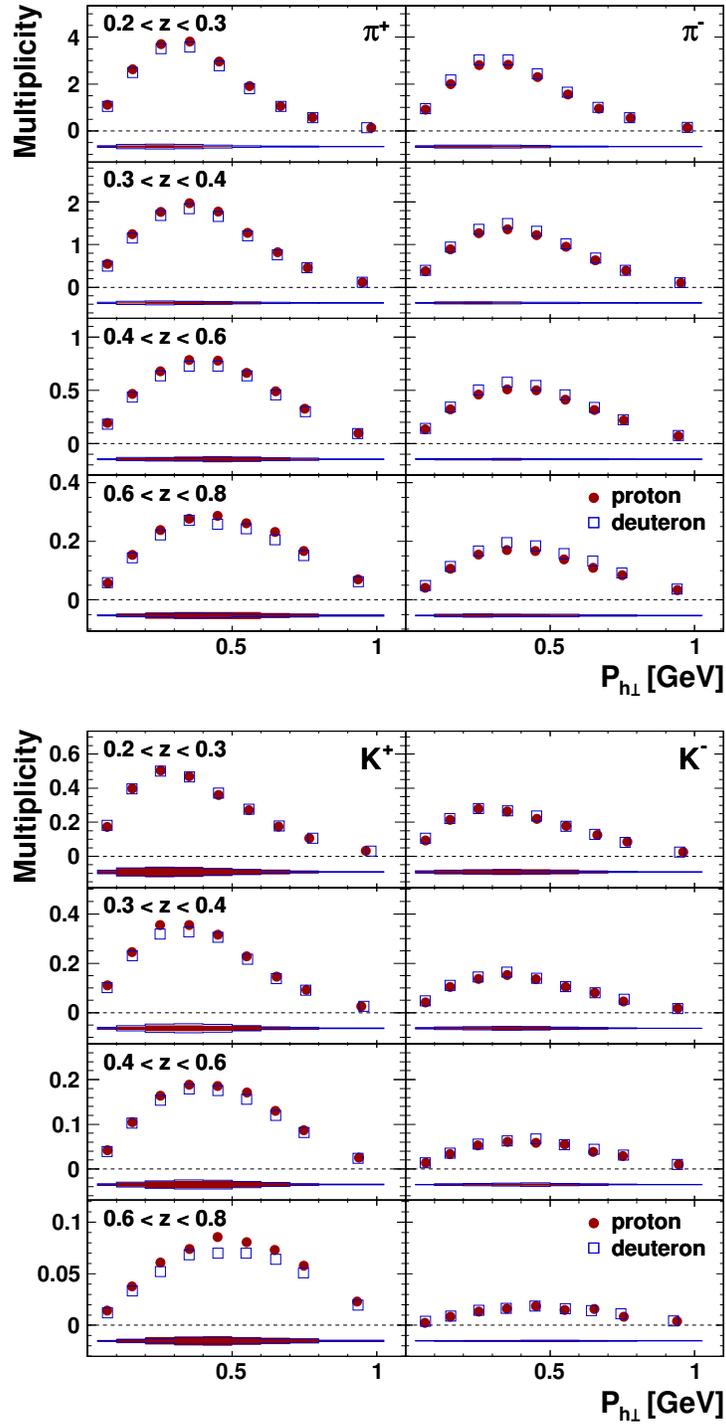


Figure 7.1: Pion multiplicities (top panels) and kaon multiplicities (bottom panels) off the proton (full circle) and the deuteron (empty square), presented as a function of $P_{h\perp}$ in four z -bins. Statistical error bars are too small to be visible. The systematic uncertainties are given by the error bands. The multiplicity results are superficially consistent with a Gaussian ansatz for the transverse momentum dependence of the PDFs and FFs.

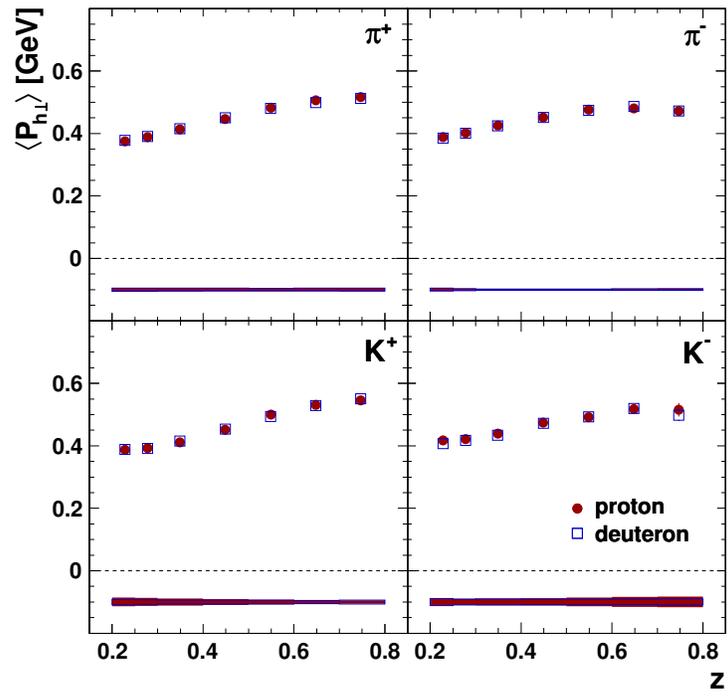


Figure 7.2: Average $P_{h\perp}$ for the pion and kaon multiplicities off the proton (full circles) and the deuteron (empty squares), presented as a function of z . The statistical uncertainty is shown by the error bars, while the systematic uncertainty is given by the error bands.

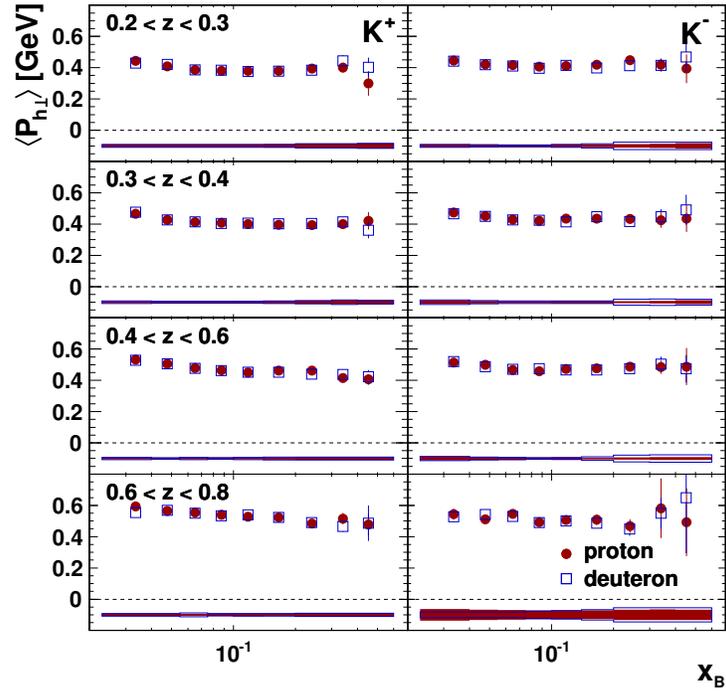
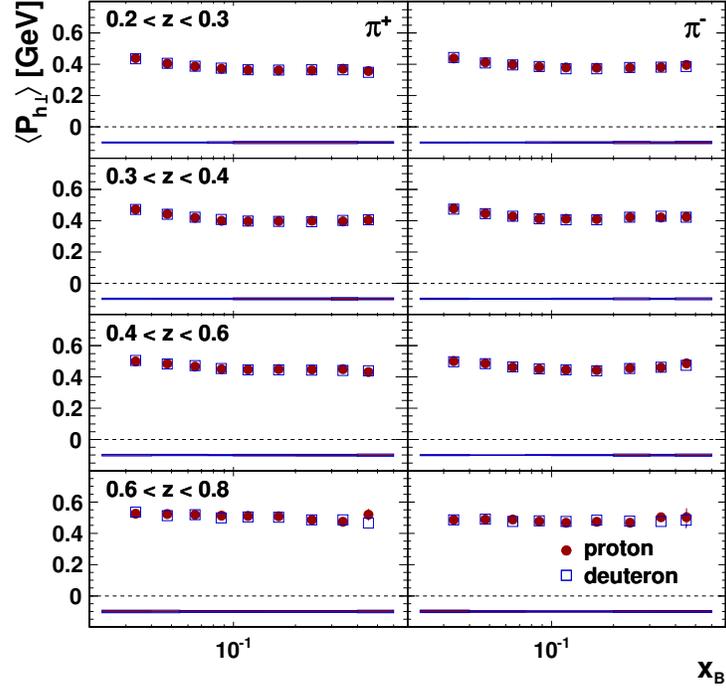


Figure 7.3: Average $P_{h\perp}$ for the pion multiplicities (top panels) and kaon multiplicities (bottom panels) off the proton (full circle) and the deuteron (empty square), presented as a function of x in four z -bins. The statistical uncertainty is shown by the error bars, while the systematic uncertainty is given by the error bands.

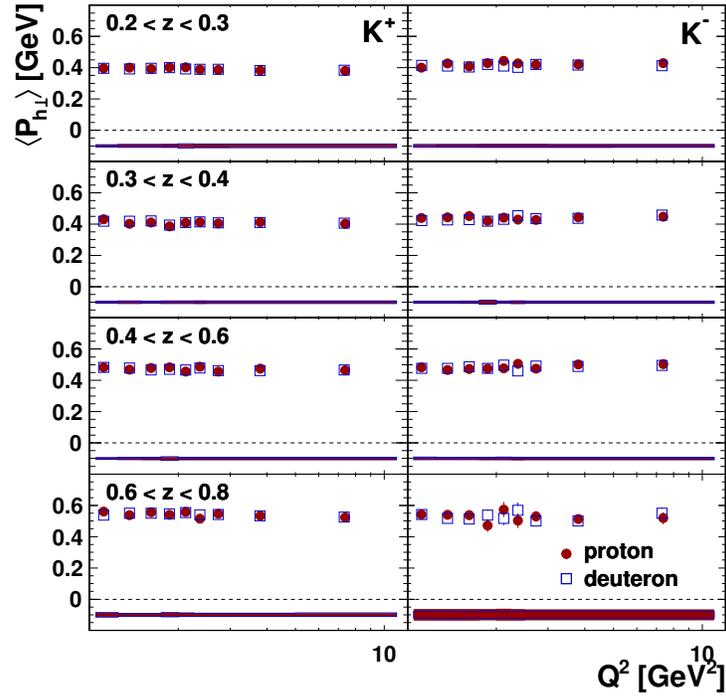
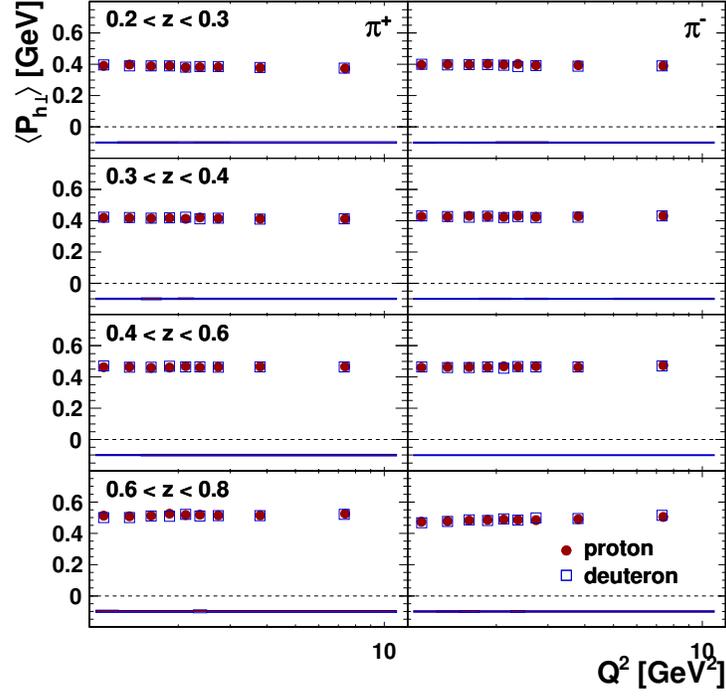


Figure 7.4: Average $P_{h\perp}$ for the pion multiplicities (top panels) and kaon multiplicities (bottom panels) off the proton (full circle) and the deuteron (empty square), presented as a function of Q^2 in four z -bins. The statistical uncertainty is shown by the error bars, while the systematic uncertainty is given by the error bands.

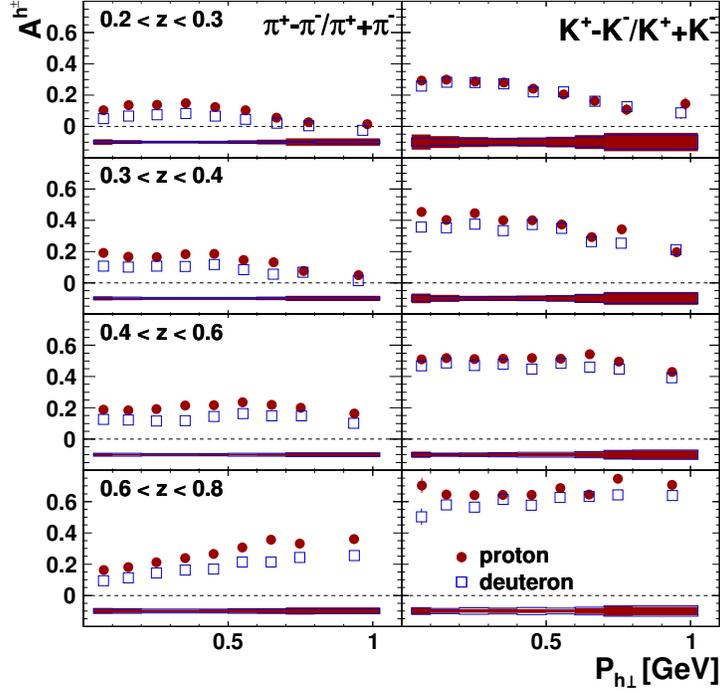


Figure 7.5: The multiplicity hadron charge asymmetry $A^{h^\pm} \equiv (M^{h^+} - M^{h^-}) / (M^{h^+} + M^{h^-})$ for pions (left panels) and kaons (right panels), presented as a function of $P_{h\perp}$. The statistical uncertainty is shown by the error bars, while the systematic uncertainty is given by the error bands. The numerator contains proportionally more contribution from valence quarks than the denominator, encoding information about the shape of the intrinsic quark p_T distribution in A^{h^\pm} .

function of x , corresponding to a slightly falling functional dependence of average p_T . The Q^2 dependence is consistent with a constant.

Finally, the multiplicity hadron charge asymmetry A^{h^\pm} ,

$$A^{h^\pm} \equiv (M^{h^+} - M^{h^-}) / (M^{h^+} + M^{h^-}), \quad (7.5)$$

is also available in the database. The numerator in this particular combination of multiplicities is more sensitive to valence contributions than the denominator. The results for A^{h^\pm} are shown in Fig. 7.5 as a function of $P_{h\perp}$ in four z -bins.

8 Precocious Scaling and the Limits of Factorization

Processes with a hadron in the final state can be described in the standard framework of collinear quantum chromodynamics (QCD), given a hard-scattering reaction. The cross sections can then be expressed in terms of a perturbative hard-scattering cross section, and certain nonperturbative but universal functions, the parton distribution functions (PDFs) and fragmentation functions (FFs) [11, 13]. The PDFs encode the inner partonic structure of the hadrons in the initial state, while the FFs describe the hadronization process after the hard-scattering takes place. Application of this framework to semi-inclusive deep-inelastic scattering (SIDIS) yields a cross section proportional to certain combinations of the PDFs and FFs [12, 29–32],

$$d\sigma^h \propto \sum_q e_q^2 f_q(x) \otimes \hat{\sigma} \otimes D_q^h(z). \quad (8.1)$$

Especially at leading order (LO) in the strong coupling α_s , this framework allows for a suggestive, quark-parton model (QPM)-like interpretation of the different components. Furthermore, because of universality, the results from e^+e^- annihilation, Drell-Yan and deep-inelastic scattering can be combined to optimally constrain each of these parts.

While very successful in its description of high energy reactions, the question arises how well it performs at intermediate energies ($Q \sim M$). This is especially relevant in light of ultra-high-precision measurements that will take place at Jefferson Lab after the 12 GeV CEBAF upgrade. Because of their unprecedented precision and multidimensional nature, the HERMES multiplicity results can be used to illustrate and estimate the sort of complications that may arise at intermediate energies.

8.1 Limits of Factorization

8.1.1 Theoretical Interpretation

Once the hard scale Q moves towards $Q \sim M$, Eq. (8.1) is not strictly valid anymore. Higher order terms in α_s become larger at these energies, making a LO

approach harder to defend. Additionally, the entire framework is only valid as long as Q is large enough. When this is not the case, higher twist effects, initial and final-state interactions and target mass effects start to play a larger role. When properly accounted for, these considerations from the theoretical side do not break down factorization, but rather demand a more sophisticated treatment. This would of course invalidate a simple QPM-like interpretation.

8.1.2 Target Remnant Influence in SIDIS Measurements

Experimentally, it is impossible to detect if a given particle is a **current particle** (semi-inclusive hadron produced off the struck quark), or part of the **target remnant**. The process independence necessary for universality requires the FFs to only correlate to the struck quark. This implies that the fragmentation function can only depend on the fractional hadron momentum z (given a fixed factorization scale Q). In principle, the effects of target remnant contamination could be accounted for by switching to a formalism in terms of “fracture functions” [32, 138], but this would make the powerful factorized approach Eq. (8.1) impossible. Lacking universality, the non-perturbative terms of the SIDIS cross section lose their predictive power across different lepton-hadron processes. Therefore, a clear experimental separation of the current and target “jets” appears to be necessary.

An ideal kinematic quantity to study and constrain the contamination by the target remnant is the center-of-mass frame rapidity (CM-rapidity) η_{CM} ,

$$\eta_{\text{CM}} \equiv \frac{1}{2} \ln \frac{E_h + P_{h\parallel}}{E_h - P_{h\parallel}}, \quad (8.2)$$

with E_h the hadron energy and $P_{h\parallel}$ the longitudinal momentum component of the hadron in the direction of the virtual photon, in the photon-nucleon center-of-mass frame. A positive value for η_{CM} indicates that a hadron is moving in the direction of the virtual photon, while a negative value indicates that the hadron is moving into the backward hemisphere. Therefore, the sign of η_{CM} is closely related to the current and target jets. A good separation of both jets in rapidity space would consequently suffice as a requirement for uncontaminated current fragmentation.

The range in the invariant mass of the photon-nucleon system, W , covered by the experiment, determines the maximum hadron momentum P_{max} , and accordingly the range in rapidity. For light hadrons, P_{max} is given by,

$$P_{\text{max}} \approx E_{\text{max}} = \frac{W}{2}. \quad (8.3)$$

The maximum CM-rapidity is reached when a hadron with $P = P_{\text{max}}$ moves along

the virtual photon direction,

$$\eta_{\text{CM,max}} = \frac{1}{2} \ln \frac{E_{\text{max}} + P_{\text{max}}}{E_{\text{max}} - P_{\text{max}}} \quad (8.4)$$

$$= \frac{1}{2} \ln \frac{(E_{\text{max}} + P_{\text{max}})(E_{\text{max}} + P_{\text{max}})}{(E_{\text{max}} - P_{\text{max}})(E_{\text{max}} + P_{\text{max}})} \quad (8.5)$$

$$\approx \frac{1}{2} \ln \frac{4P_{\text{max}}^2}{M_h^2} \quad (8.6)$$

$$= \ln \frac{W}{M_h}, \quad (8.7)$$

with M_h the hadron mass. The W -dependence of the rapidity distributions of SIDIS hadrons therefore provides access to the phenomenology of target fragmentation. This can be used to obtain the experimental requirements necessary for a good jet-separation in rapidity space.

This was studied in detail in the in the μ -scattering experiments E665 [139] and EMC [140]. It should be noted that, in the E665 results, the CM-rapidity is commonly referred to as y^* instead of η_{CM} .

Figure 8.1 presents the rapidity spectrum for positive hadrons from the E665 experiment, in three bins of W . All three panels show a strong peak in the backward hemisphere with a FWHM of approximately two units of rapidity. The mid-rapidity region is completely filled in, and only at higher values of W , a small bump in the forward direction becomes apparent.

It is instructive to use a chain-fragmentation model [23] to discuss these results. Such a model forms the basis for the Lund string model [20–22] used by the LEP-TO/JETSET [106] and PYTHIA [115] Monte-Carlo generators. In these models, hadron formation is explained by the creation of quark-antiquark pairs following the initial hard scattering. These final state mesons partition the energy and momentum of the color force field connecting the struck quark to the target remnant. In the center-of-mass frame, the target remnant moves into the backward hemisphere while the struck quark moves approximately parallel to the virtual photon, giving rises to peaks at respectively negative and positive rapidity. The additional hadrons produced during the fragmentation process lie between these extremes, and therefore explain the absence of a mid-rapidity gap.

Positive hadrons are more prevalent in the current jet than negative hadrons due to u -quark dominance. Similarly, the abundance of positive charge in the target causes the target remnant to be predominantly positively charged. The mid-rapidity region is filled by additional quark-antiquark pairs, which cannot influence the total hadronic charge. The hadronic net charge $1/N^\mu(dN^{h^+}/d\eta_{\text{CM}} - dN^{h^-}/d\eta_{\text{CM}})$ therefore provides a very clean way to access to the rapidity distribution of the primary jets. The hadronic net charge measured at E665 from μD - and μXe -scattering is

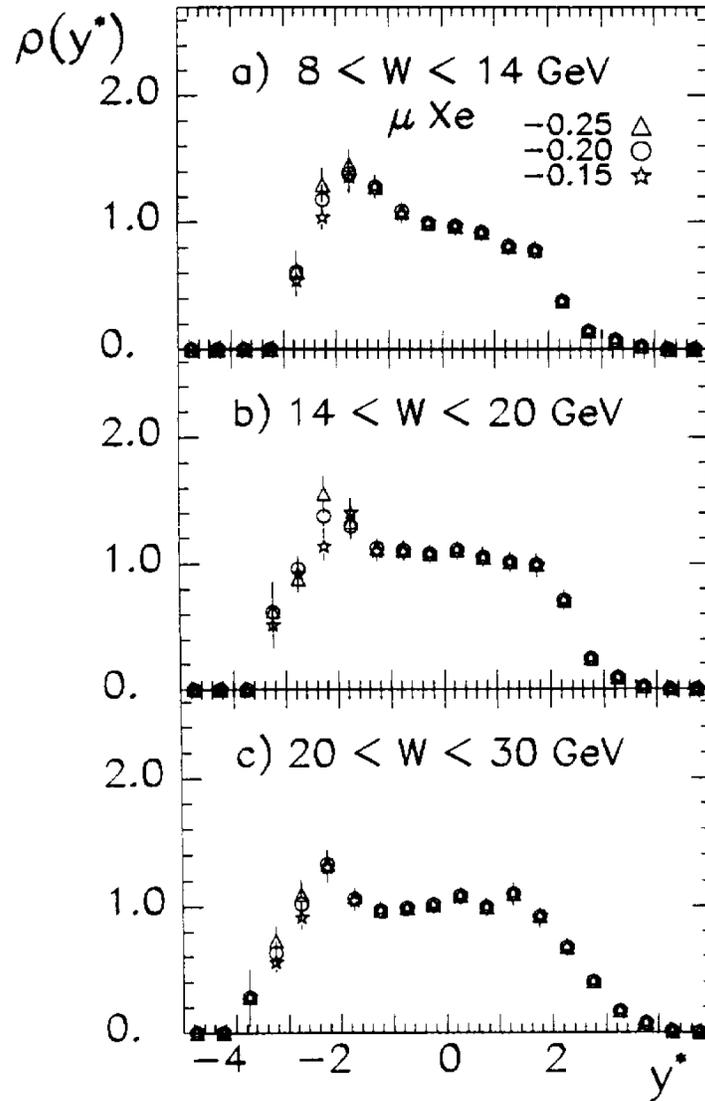


Figure 8.1: Normalized CM-rapidity distribution of positive hadrons in three bins of W from μXe -scattering at E665. The different markers refer to variants of the PID procedure not relevant to the current discussion. The target jet (negative rapidity) and current jet (positive rapidity) are hard to distinguish from each other due large amount of additional hadrons filling the gap between both jets. The situation becomes slightly better at higher values of W . See also Fig. 8.2. Figure from [139].

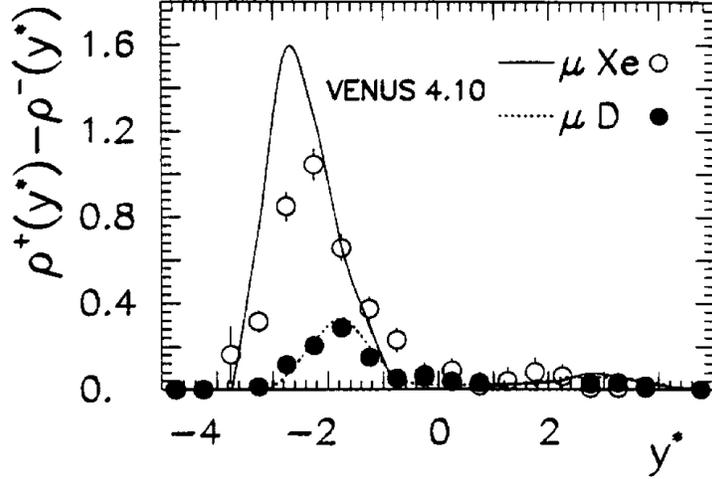


Figure 8.2: Normalized CM-rapidity distribution of the hadronic net charge $1/N^\mu(dN^{h^+}/d\eta_{\text{CM}} - dN^{h^-}/d\eta_{\text{CM}})$ from μD - and μXe -scattering at E665. The VENUS model predictions not relevant to the current discussion. The sharp peak at negative rapidity corresponds to the target remnant, while the shallow bump in the forward hemisphere corresponds to the current hadron. All peaks have a FWHM of approximately two units of rapidity. Figure from [139].

presented in Fig. 8.2. It features a sharp peak at negative rapidity, and a smaller peak at positive rapidity. Again, the FWHM of both peaks is two units of rapidity.

Figure 8.3 shows the hadronic net charge spectrum results from EMC, in three x -bins. This indirectly probes the W -dependence due to the x -dependence of $\langle W \rangle$. The EMC results show a strong target remnant peak in the backward hemisphere. The current fragmentation peak becomes more prominent in the valence region at mid-to-high x . This is to be expected, as the sea quarks, and the \bar{u} quark in particular, become more important at low x , and the hadronic net charge cannot be used to resolve the current fragmentation peak in the absence of u -quark dominance. Similar to Figs. 8.1 and 8.2, the FWHM of both peaks is about two units of rapidity. In the top panel, where $\langle W \rangle \sim 17 \text{ GeV}$, a $\eta_{\text{CM}} > 0$ cut is sufficient to avoid target remnant contamination. In terms of the commonly used Feynman scaling variable x_F ,

$$x_F \equiv \frac{2P_{\text{CM}}^{\parallel}}{W}, \quad (8.8)$$

this requirement becomes $x_F > 0$. For the other panels, a more stringent set of cuts is necessary to avoid the target remnant.

To summarize, the target remnant and current fragmentation jets are observed to have a constant FWHM of two units of rapidity. The lowest possible rapidity is reached when the target remnant moves opposite to the virtual photon direction.

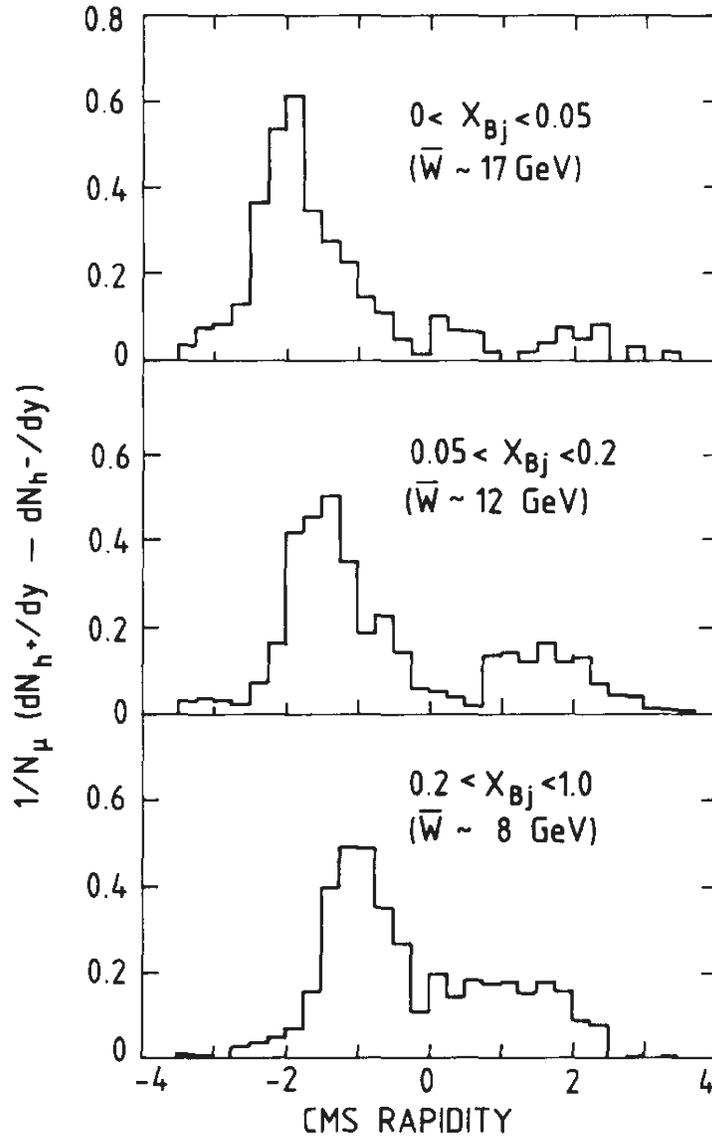


Figure 8.3: Distribution of the hadronic net charge $1/N^\mu(dN^{h^+}/d\eta_{CM} - dN^{h^-}/d\eta_{CM})$ as a function of the CM-rapidity in three different x regions from μp -scattering at EMC. The sharp peak at negative rapidity is clearly visible in all three x -bins. The peak at positive rapidity becomes more prominent in the valence region at mid-to-high x . All peaks are consistent with a FWHM of two units of rapidity. They move closer towards each other for smaller values of average W . The tails from both distributions strongly overlap each other in the bottom panel. Figure from [140].

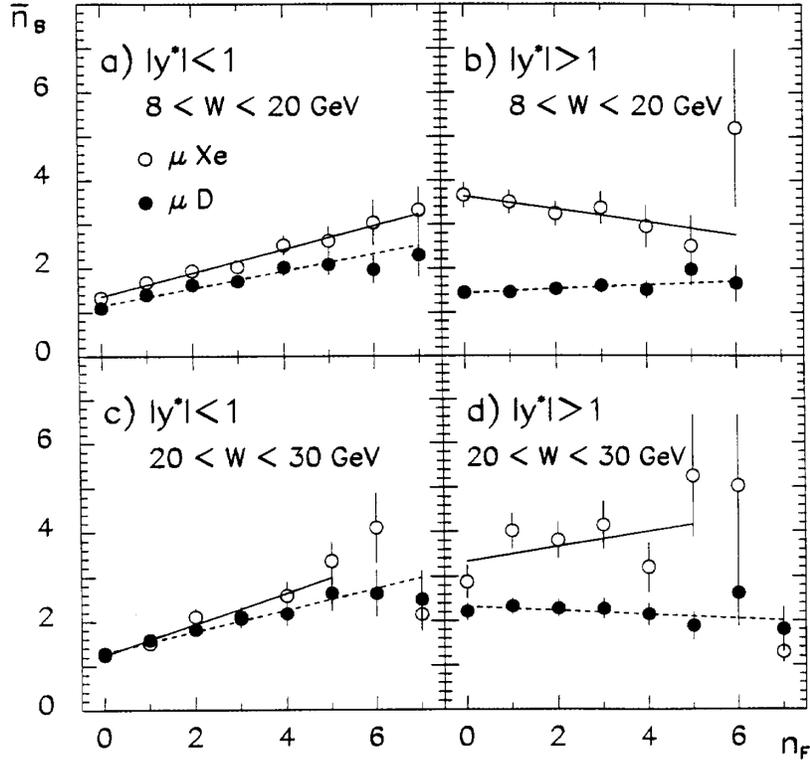


Figure 8.4: Average backward multiplicity as a function of the forward multiplicity for a central (left panels) and outer CM-rapidity range (right panels) from the E665 experiment, in two bins of W (top and bottom panels). The data are presented for μD (full circles) and μXe (empty circles), The lines are the results of straight-line fits. There is a clear correlation in the central rapidity region. The data in the outer rapidity region is consistent with no correlation for both bins in W . Figure from [139].

This value is determined by the hadron mass and W , similar to Eq. (8.7). The sufficient requirement that the target and current peaks are separated by four units of rapidity, or equivalently $\eta_{CM,max} = 4$, determines this lower bound on W as a function of hadron mass. This phenomenological requirement is commonly referred to as **the Berger criterion** [140–142].

The lower bound on W for pions and kaons that follows from this formulation of the Berger criterion is given by,

$$\pi \rightarrow W > 7.6 \text{ GeV}, \quad (8.9)$$

$$K \rightarrow W > 27 \text{ GeV}. \quad (8.10)$$

Figure 8.4 presents the average backward multiplicity as a function of the forward multiplicity from E665 for the central rapidity region $|\eta_{CM}| < 1$ (left panels) and the outer rapidity region $|\eta_{CM}| > 1$ right panels. The results show a strong

correlation between the forward and backward multiplicities in the central rapidity region, while the results are consistent with a flat dependence in the outer rapidity region. This absence of a correlation for $|\eta_{\text{CM}}| > 1$ provides further evidence for the notion of independent fragmentation, therefore supporting the phenomenological approach taken in the derivation of the Berger criterion.

8.1.3 Factorization at Intermediate Energies and Precocious Scaling

The upper limit in W is given by \sqrt{s} of the experiment. Scattering the 27.6 GeV electron/positron beam from HERA off the HERMES target yields,

$$W_{\text{max,HERMES}} = \sqrt{s} = 7.26 \text{ GeV}. \quad (8.11)$$

According to the Berger criterion (cf. Eqs. (8.9) and (8.10)), this upper limit is slightly below the **minimum** requirement $W > 7.6 \text{ GeV}$ needed for an uncontaminated pion measurement, and the kaon requirement $W > 27 \text{ GeV}$ is even more problematic. It should be noted that, for pions, a restriction of the forward hemisphere to $\eta_{\text{CM}} > 0.9$ lowers the W -limit to $W > 3.16 \text{ GeV}$ ($W^2 > 10 \text{ GeV}^2$), which is the lower limit used in this work. A similar cut for kaons does not exist at HERMES energies.

While at first sight problematic, the Berger criterion is a phenomenological guideline, rather than a hard requirement derived from first principle. When dealing with data at intermediate energies, the possibility of factorization breaking due to incorrect application of a simplified theoretical framework (Section 8.1.1), or contamination by the target remnant (Section 8.1.2) should be carefully studied. The identification of potential problem areas in multi-dimensional kinematic space provides information about the origin of the issues. This information can then be used to establish a “safe zone”, where there is no evidence of factorization breaking.

Such a test was performed on data taken during the first two years of HERMES [143]. To examine the accuracy of a factorized LO formalism, including the assumption that the fragmentation functions be perfectly symmetric under an isospin rotation, the light-sea flavor asymmetry,

$$\frac{\bar{d}(x) - \bar{u}(x)}{u(x) - d(x)}, \quad (8.12)$$

was extracted as a function of z in various slices of x .

The extraction procedure is similar to the d_v/u_v extraction described in Section 8.2.2. It requires π^+ and π^- cross sections (or multiplicities) for a proton and deuteron target, as well as the unfavored-to-favored fragmentation function ratio.

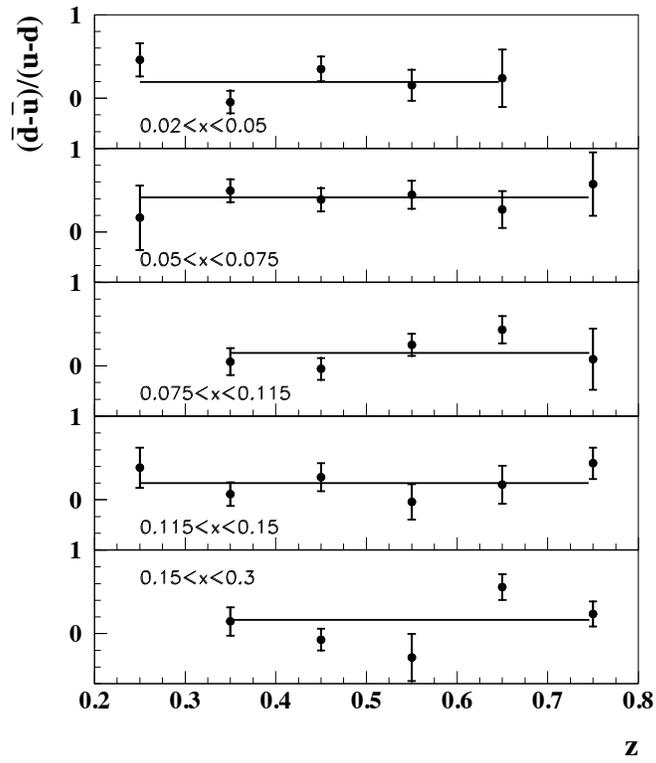


Figure 8.5: The distribution $(\bar{d} - \bar{u})/(u - d)$ as a function of z in five bins of x from HERMES. The points are fit to a constant for each x -bin. The error bars represent statistical and systematic uncertainties added in quadrature. Figure from [143].

The latter was obtained from a fit to a statistically separate data sample. The kinematic cuts are similar to those used in Section 4.4, with the exception of the less restrictive requirement $W > 2 \text{ GeV}$ and additional cut of $x_F > 0.25$. The lower cut in W does not significantly change the SIDIS sample because of the small momentum of the final state hadrons in these additional events. The spectrometer typically bends low-momentum hadrons out of the acceptance. The stringent cut of $x_F > 0.25$ is approximately equivalent with the requirement that $\eta_{\text{CM}} > 0.9$. According to the Berger criterion, the pion sample should therefore be clear of target remnant contamination.

The results in Fig. 8.5 show no z -dependence within the uncertainties, and consequently no evidence of factorization breaking at intermediate energies. It was later found that, even with a less restrictive cut of $z > 0.2$, approximately equivalent to $x_F > 0$, there is no evidence of a significant target remnant contamination at HERMES. Hence, fifteen years later, there is still no sign of any factorization breaking at HERMES. The apparent validity of a factorized approach when Q^2 is not a very large number, is sometimes referred to as **precocious scaling**. This precocious scaling seems to extend down to energy regimes very close to $Q \sim M$, meaning that $Q^2 > 1 \text{ GeV}^2$ is a sufficient requirement for the standard framework of leading-twist collinear QCD to be applicable.

After the 12 GeV upgrade, the CLAS12 and SOLID experiments at Jefferson Lab will be able to collect data with an unprecedented level of statistical precision. However, the W -range covered by these experiments will be lower than that in HERMES (cf. Eq. (8.11)),

$$W_{\text{max,JLAB}} = \sqrt{s} = 4.8 \text{ GeV}. \quad (8.13)$$

This combination of ultra-high-precision at a lower energy range forces a thorough evaluation of the applicability of a factorized approach. The new HERMES multiplicity results [127], presented in this work form the most precise SIDIS dataset currently available. A study of the limits of precocious scaling at HERMES with these results will hence provide valuable input to help better understand these future challenges at Jefferson Lab.

8.2 The Applicability of QPM-like Factorization and

$$d_v/u_v$$

In the standard framework of collinear QCD, the semi-inclusive cross section $d\sigma^h$ is given at LO by (cf. Eqs. (2.23) to (2.25)),

$$\frac{d^3\sigma^h}{dx dQ^2 dz} \propto \sum_q e_q^2 f_q(x, Q^2) D_q^h(z, Q^2). \quad (8.14)$$

where the pre-factors have been omitted for clarity. The factorized dependence of $d\sigma^h$ on the parton distribution functions $f_q(x, Q^2)$ and fragmentation functions $D_q^h(z, Q^2)$ offers the possibility to extract information about one if the other is known. For example, the DSS extraction of the FFs [29] uses SIDIS multiplicities in their global fit. The PDFs, which are extracted from inclusive data, are taken as input using the values obtained from independent global fits of the MRST [38, 144] and CTEQ [37] groups.

An alternative approach is to extract the PDFs by taking the FFs as input. Given the current uncertainties in our knowledge of the FFs, however, this approach is only profitable if symmetries between the FFs can be exploited. These symmetries are relations whose accuracy is presumed to be better than our knowledge of the individual FFs. As discussed in Section 8.1, this approach is only applicable in a regime where factorization holds. It is therefore perfectly suited to search for any evidence of factorization breaking, and to evaluate the accuracy of commonly assumed FF symmetries at HERMES kinematics, similar to the approach taken in Ref. [143].

In this section, various ratios of measured charged-pion yields are used to access the PDF combination d_v/u_v as a function of x and z . This extraction does not need any additional input by means of explicit FF ratios, unlike the extraction of the light-sea flavor-asymmetry, making it the more powerful tool in the hunt for factorization breaking at intermediate energies.

It should be noted that the d_v/u_v extraction is more sensitive to systematic issues with the underlying data, compared to the $(\bar{d} - \bar{u})/(u - d)$ extraction from Ref. [143]. These were not sufficiently known in the early years of HERMES. This finally changed with the release of the hadron multiplicities presented in this work, in particular due to the improved pion PID, the multi-dimensional smearing-unfolding, and the detailed knowledge of the systematic uncertainties and their correlations (cf. Chapter 4).

8.2.1 Fragmentation Function Symmetries

The current knowledge of the FFs [29, 35, 36, 61, 62] depends on global fits to data from various hard-scattering reactions, chiefly e^+e^- -annihilation from LEP and SLAC [63–67]. This inescapably limits the number of independent observables, making it impossible to determine $D_q^h(z, Q^2)$ independently for each flavor q of primary quark and each hadron of type h . To reduce the number of independent functions to a manageable set, symmetry relations between the FFs for similar hadron types are routinely invoked. Traditionally, the assumptions of charge-conjugation and isospin-rotation invariance have been made for charged pion production, reducing the number of independent FFs to three. However, due to the intrinsic isospin-symmetry violation (ISV) caused by the different charge and rest mass of the light quarks, a small ISV in the FFs is to be expected.

The charge-conjugation symmetry of the current-fragmentation process reduces the number of independent FFs for charged pion production to six,

$$D_u^{\pi^+} = D_{\bar{u}}^{\pi^-} \equiv D_1 \qquad D_d^{\pi^-} = D_{\bar{d}}^{\pi^+} \equiv D_2, \qquad (8.15)$$

$$D_u^{\pi^-} = D_{\bar{u}}^{\pi^+} \equiv D_3 \qquad D_d^{\pi^+} = D_{\bar{d}}^{\pi^-} \equiv D_4, \qquad (8.16)$$

$$D_s^{\pi^+} = D_{\bar{s}}^{\pi^-} \equiv D_{s,1} \qquad D_s^{\pi^-} = D_{\bar{s}}^{\pi^+} \equiv D_{s,2}. \qquad (8.17)$$

These three relations represent the traditional grouping of the fragmentation functions into favored Eq. (8.15), unfavored Eq. (8.16) and strange Eq. (8.17). If isospin symmetry is also imposed, only three independent FFs remain: $D_{\text{fav}} = D_1 = D_2$, $D_{\text{unf}} = D_3 = D_4$ and $D_s = D_{s,1} = D_{s,2}$. Here, and in the rest of this section the dependence of the FFs on Q^2 and z , as well as the dependence of the PDFs on Q^2 and x is implicitly assumed, unless noted otherwise.

The topic of ISV in the PDFs, often referred to as ‘‘charge-symmetry violation’’¹, has been explored for decades. It is estimated to be of order 1 % to 2 %, although experimental limits allow for values that are several times larger [145, 146]. The possibility of ISV in the FFs was explicitly addressed in the DSS fragmentation function fit [29]. As discussed in Section 6.2, this fit included a very preliminary version of the HERMES multiplicities. It was found to favor a 10 % ISV in the following FF combination at the input scale of 1 GeV,

$$\frac{D_{d+\bar{d}}^{\pi^+}}{D_{u+\bar{u}}^{\pi^+}} = \frac{D_2 + D_4}{D_1 + D_3} = 1.10 \text{ at } \mu_0 = 1 \text{ GeV}, \qquad (8.18)$$

¹Charge symmetry is a sub-group of isospin symmetry, which exchanges members of an isospin doublet, e.g. u and d , or p and n . It is easily confused with charge-conjugation symmetry, and therefore directly referred to as isospin symmetry in this work.

with the additional constraint that $D_u^{\pi^+} = D_d^{\pi^+}$ ($D_3 = D_4$). This 10 % effect was observed in both the LO and NLO fits. As noted by the authors, there is no particular reason for this ratio to be a constant; the available data were simply insufficient to fit a more complex function. Other symmetry violating parameters were also explored, e.g., between the disfavored fragmentation functions $D_u^{\pi^+}$ and $D_u^{\pi^-}$, but the fit was unable to either reveal or exclude other symmetry violations of similar size. If this result truly reflects a 10 % isospin violation in the fragmentation process, it is the first direct observation of a partonic ISV, and is considerably larger than the 1 % to 2 % effects seen in hadronic masses [147]. As described in Refs. [145, 146], 10 % partonic ISV effects are not excluded by current data or theoretical estimates, but are at the upper limit of the calculations. In addition, these calculations are focused entirely on quantifying ISV effects in the parton distribution functions; there is little theoretical guidance in the area of fragmentation functions.

In fact, the large isospin violation favored by the DSS fit is driven primarily by the (very preliminary) HERMES multiplicities, as they were the most accurate SIDIS results available for identified hadrons. The possibility cannot be neglected that the effect is, at least in part, caused by an inaccuracy in the SIDIS factorization assumption that is revealed at this level of precision. The cut $0.2 < z < 0.8$ should suppress the contamination by the target remnant, while avoiding the regions where semi-exclusive production becomes dominant.

Practical access to the PDFs via SIDIS multiplicities is thus contingent on the applicability of LO factorization Eq. (8.14), and consequently that the charge-conjugation symmetry of the FFs within this framework, Eqs. (8.15) to (8.17), holds to a sufficient degree of accuracy. Furthermore, the relatively unknown ISV in the FFs might introduce a z -dependence in the results, mimicking the signature of factorization breaking.

8.2.2 The Pion Yield Difference Ratio and d_v/u_v

Leveraging the charge-conjugation symmetry relations Eqs. (8.15) to (8.17), one can define a combination of charged pion cross sections on the proton and deuteron that results in a near-cancellation of the FFs,

$$R^\pi(z) \equiv 2 \frac{\int_{x_{\min}}^{x_{\max}} dx (\sigma_d^{\pi^+} - \sigma_d^{\pi^-})}{\int_{x_{\min}}^{x_{\max}} dx (\sigma_p^{\pi^+} - \sigma_p^{\pi^-})} - 1 \quad (8.19)$$

$$= \frac{\int_{x_{\min}}^{x_{\max}} dx (\sigma_n^{\pi^+} - \sigma_n^{\pi^-})}{\int_{x_{\min}}^{x_{\max}} dx (\sigma_p^{\pi^+} - \sigma_p^{\pi^-})}, \quad (8.20)$$

with the cross sections implicitly differential in x and z . Invoking the aforementioned charge-conjugation symmetry relations Eqs. (8.15) to (8.17), and assuming

isospin symmetry between the proton and neutron PDFs, R^π reduces to a simple form at LO,

$$R^\pi(z) \xrightarrow{\text{LO}} \frac{\int dx dQ^2 (u_v - 4kd_v + (s - \bar{s})k_s)}{\int dx dQ^2 (d_v - 4ku_v + (s - \bar{s})k_s)}, \quad (8.21)$$

where the integral runs over the full available x - and Q^2 -range (cf. Fig. 4.3). The PDFs in Eq. (8.21) depend on x and Q^2 , and

$$k(z, Q^2) \equiv \frac{D_1 - D_3}{D_2 - D_4}, \quad (8.22)$$

$$k_s(z, Q^2) \equiv \frac{D_{s,1} - D_{s,2}}{D_4 - D_2}. \quad (8.23)$$

The Q^2 boundaries in Eq. (8.21) are discussed in Section 6.1. Because of the near-equality of $D_{s,1}$ and $D_{s,2}$ in Eq. (8.17), and the very small size of $(s - \bar{s})$ [148], the impact of the $(s - \bar{s})k_s$ terms in Eq. (8.21) are negligible.

The relation Eq. (8.21) offer access to the valence quark PDFs,

$$\frac{4kR^\pi + 1}{4k + R^\pi} \xrightarrow{\text{LO}} \frac{\int dx dQ^2 d_v}{\int dx dQ^2 u_v}. \quad (8.24)$$

Assuming perfect isospin symmetry between the FFs, yields $k = 1$, removing the need to explicitly input FFs in the extraction. The uncertainty introduced by this assumption can be estimated by performing the extraction on multiplicities calculated with the DSS FFs (cf. Section 6.1), as the DSS extraction favored the ISV in Eq. (8.18).

If the model assumptions involved in Eq. (8.24) are valid, not only should the left-hand-side of Eq. (8.24) be independent of z , but the value for the integral over the full x -range should be close to 1/2. The HERMES coverage in x (0.023 to 0.6) spans almost the entire region of valence strength, and so the true ratio of valence-down to valence-up should be recovered. Integrating current LO PDF sets over the full range in x and Q^2 gives a value of, (cf. Fig. 4.3 and Section 6.1)

$$\frac{\int_{0.023}^{0.6} dx \int_{Q_{\min}^2(x)}^{Q_{\max}^2(x)} dQ^2 d_v(x, Q^2)}{\int_{0.023}^{0.6} dx \int_{Q_{\min}^2(x)}^{Q_{\max}^2(x)} dQ^2 u_v(x, Q^2)} \approx 0.42. \quad (8.25)$$

8.2.3 Results: d_v/u_v and Factorization Breaking

The measured result for the LO extraction of d_v/u_v (Eq. (8.24)) is presented as a function of z in Fig. 8.6. The superimposed dashed curve shows the CTEQ6L [37] value for Eq. (8.25). The solid blue curve is the result from the extraction

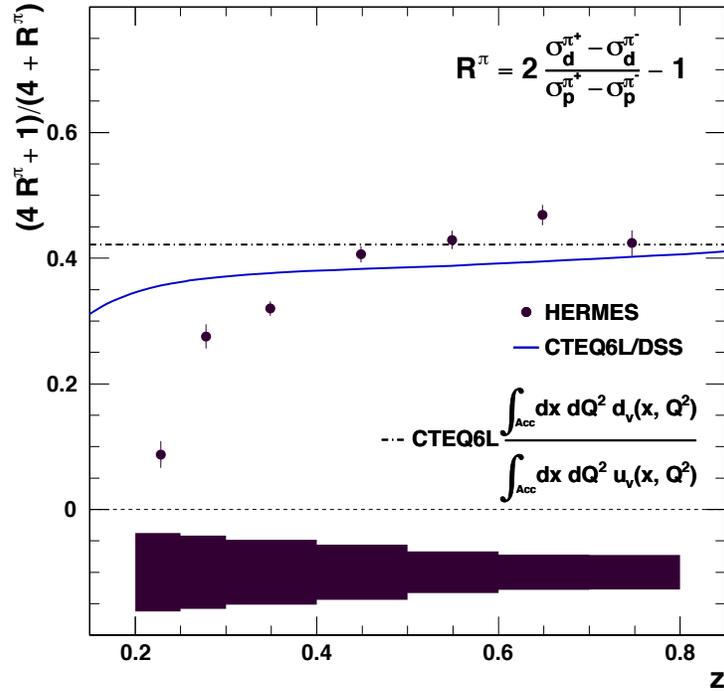


Figure 8.6: Comparison of $(4R^\pi + 1)/(4 + R^\pi)$ as a function of z (full circles) with a LO calculation using CTEQ6L parton distributions [37] with fragmentation functions from DSS [29] (solid curve). Also shown is $\int dx d_v / \int dx u_v$ from CTEQ6L (dash-dot line). The curves are integrated over the full range $0.023 < x < 0.6$, as well as the corresponding range in Q^2 (cf. Section 6.1). The statistical uncertainty is shown by the error bars, while the systematic uncertainty is given by the error bands. Theoretical uncertainties on the LO calculations are not available.

Eq. (8.24) from multiplicities calculated in LO from the DSS FFs in combination together with PDFs from CTEQ6L (cf. Chapter 6). The z -dependence of this curve is introduced by incorrectly assuming perfect isospin symmetry in the extraction, instead of an ISV of 10 %, consistent with Eq. (8.18). This provides a measure for the ISV-model dependence of the extraction.

Taking into account the systematic uncertainty, the measured result is in agreement with both theory curves, except for the lowest values of z . In particular, the discrepancy between the result for the lowest z -bin, and the dashed CTEQ6L curve, is larger than 3σ . The more realistic FF symmetry assumption in DSS, shown in the blue solid curve, lessens this discrepancy somewhat, but even then, this point strongly hints at problems with the extraction framework at low z .

Figure 8.7 presents the same results as a function of z in six x -bins. The extracted results deviate significantly at $z < 0.3$, while the results at mid-to-high z are consistent with the LO predictions. With the available precision clearly pushed to the limit in this comparison, it is hard to determine if the discrepancy shows any x -dependence.

To better resolve the x -dependence, the extraction was repeated as a function of x in four z -bins. The results from this study are presented in Fig. 8.8. The data are consistent with a discrepancy at $z < 0.3$, independent of x .

It is instructive to view these results in light of the discussion regarding the Berger criterion at HERMES in Section 8.1.3. This discrepancy **has** to occur at some point, as the target remnant grows in importance at very low z . The lower z -cut at 0.2 corresponds to a cut of $\eta_{CM} > 0$. At the W -range covered in HERMES, this is not sufficient to fully avoid contamination by the target remnant. This implies that, when performing an extraction that is sensitive enough to the target remnant, at a high enough level of precision, factorization breaking will occur. Even in this very sensitive quantity, the only evidence towards factorization breaking at HERMES occurs at $z < 0.3$. In fact, the Berger criterion suggests that, for pions at HERMES energies, $\eta_{CM} > 0.9$ completely avoids the target remnant. This cut corresponds to approximately $z > 0.3$, consistent with the observations in this section. Precocious scaling holds perfectly well for mid-to-high z .

Further evidence for the role of the target remnant in the observed factorization breaking below $z < 0.3$ is given by the HERMES tuned Lund (LEPTO/JETSET [105, 106]) Monte-Carlo generator [124]. The result for the LO extraction of d_v/u_v (Eq. (8.24)) from this Monte Carlo as a function of z , is compared to the measured results in Fig. 8.9. At higher values of z , the Lund model runs out of available phase space, and therefore automatically breaks factorization. Hence, a z -dependence at higher values of z is not unexpected. This is not unlike the expected issues for a measurement in the semi-exclusive region at high values of z , the reason behind the

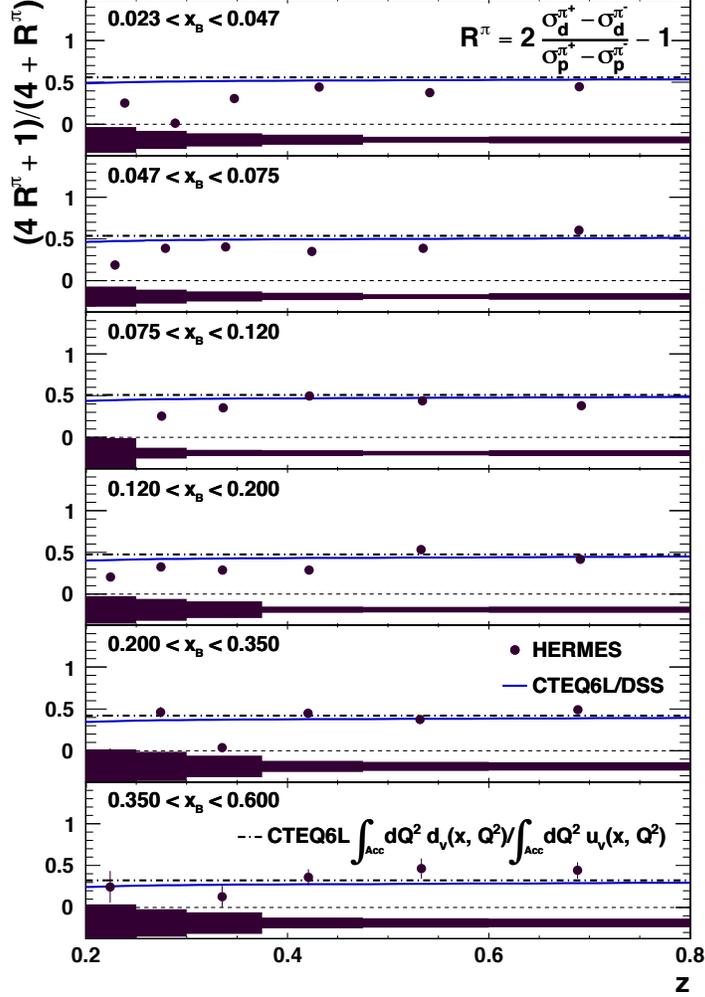


Figure 8.7: Comparison of $(4R^\pi + 1)/(4 + R^\pi)$ as a function of z in six x -bins (full circles) with a LO calculation using CTEQ6L parton distributions [37] with fragmentation functions from DSS [29] (solid curve). Also shown is $\int dx d_v / \int dx u_v$ from CTEQ6L (dash-dot line). The curves are integrated over the respective x -range, as well as the corresponding range in Q^2 (cf. Section 6.1). The statistical uncertainty is shown by the error bars, while the systematic uncertainty is given by the error bands. Theoretical uncertainties on the LO calculations are not available.

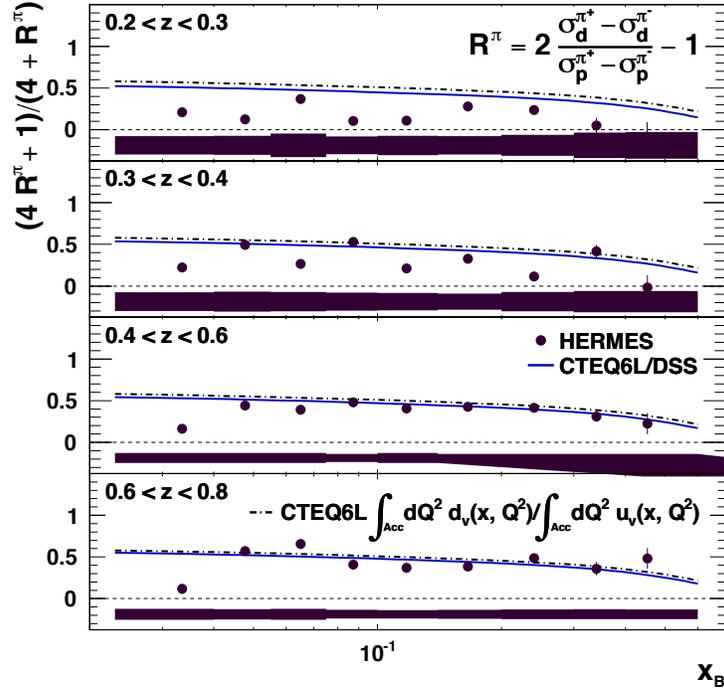


Figure 8.8: Comparison of $(4R^\pi + 1)/(4 + R^\pi)$ as a function of x in four z -bins (full circles) with a LO calculation using CTEQ6L parton distributions [37] with fragmentation functions from DSS [29] (solid curve). Also shown is d_v/u_v from CTEQ6L (dash-dot line). The curves are integrated over the respective z -range, as well as the range in Q^2 corresponding to the value of x (cf. Section 6.1). The statistical uncertainty is shown by the error bars, while the systematic uncertainty is given by the error bands. Theoretical uncertainties on the LO calculations are not available.

$z < 0.8$ requirement for SIDIS at HERMES. At the other hand, because both the simulation of the DIS process by LEPTO, as well as the simulation of the fragmentation process by JETSET, perfectly obey the charge-conjugation symmetry and isospin invariance used to derive Eq. (8.24), the z -dependence at lower values of z has to be caused by the target remnant, which is simulated by JETSET. Therefore, the Lund model predicts a signature for target remnant contamination in the form of a strong drop in the results for the LO extraction of d_v/u_v (Eq. (8.24)) at low z . While the Monte-Carlo simulation reproduces the measured results rather poorly in magnitude, both simulation and measurement exhibit this strong slope at low z , suggesting contamination by the target remnant at $z < 0.3$.

This measurement pushes the HERMES experiment to its limits, as exemplified by the size of the systematic uncertainties, especially when compared to the statistical uncertainties. The lesson to be drawn from this is that, to optimally leverage the ultra-high statistics in future experiments, a sophisticated multi-dimensional approach has to be taken. Such a fully differential approach with small bin-sizes, similar to what was done for this work, minimizes the dominant contributions to the systematic uncertainty, and therefore maximizes the attainable experimental precision.

Ultimately, rather than rigorously proved, the applicability of factorization at intermediate energies can only be supported by the absence of experimental evidence for factorization breaking.

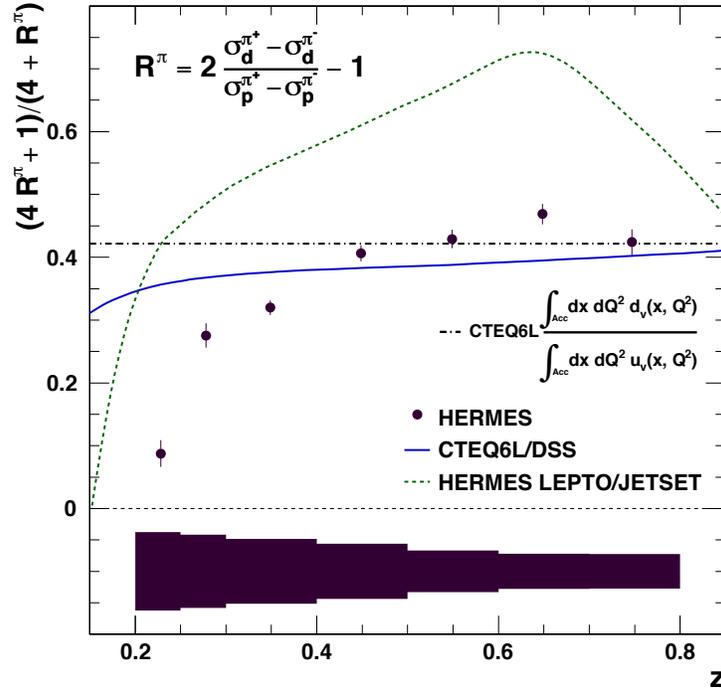


Figure 8.9: Comparison of $(4R^\pi + 1)/(4 + R^\pi)$ as a function of z (full circles) with a LO calculation using CTEQ6L parton distributions [37] with fragmentation functions from DSS [29] (solid curve), as well as the values obtained from the HERMES Lund Monte Carlo [124] (dashed curve). Also shown is $\int dx dQ^2 d_v(x, Q^2) / \int dx dQ^2 u_v(x, Q^2)$ from CTEQ6L (dash-dot line). The curves are integrated over the full range $0.023 < x < 0.6$, as well as the corresponding range in Q^2 (cf. Section 6.1). The statistical uncertainty is shown by the error bars, while the systematic uncertainty is given by the error bands. Theoretical uncertainties on the LO calculations are not available.

9 Neutral Pion Multiplicities

To constrain the fragmentation functions, data from single-inclusive e^+e^- annihilation, semi-inclusive deep-inelastic scattering (SIDIS) and single-inclusive hadron production in hadron-hadron collisions can be used (see e.g., Ref. [29]). The amount of available observables is insufficient to determine the FFs independently for each quark flavor q and hadron h . Symmetry relations between FFs for similar hadron types can be invoked to reduce this number to a more manageable set. For example, charge conjugation symmetry is manifest in quantum chromodynamics (QCD), and therefore automatically applies to the FFs. Typically, the FFs are also assumed to be invariant under isospin rotations.

Isospin is only an approximate symmetry in QCD, as it is automatically broken due to the difference in rest-mass of the light quarks. For example, the isospin symmetry violation in the parton distribution functions (PDFs) is estimated to be of order 1% to 2%, although values several times larger are not excluded [145, 146]. Section 8.2.1 explained how the DSS compilation of FFs [29] favored a 10% ISV between the FFs that describe the production of charged pions (see Eq. (8.18)).

The three pions form an isospin triplet state [9],

$$\pi^+ = |u\bar{d}\rangle \quad I_3 = 1 \quad (9.1)$$

$$\pi^0 = \frac{|u\bar{u}\rangle - |d\bar{d}\rangle}{2} \quad I_3 = 0 \quad (9.2)$$

$$\pi^- = |d\bar{u}\rangle \quad I_3 = -1, \quad (9.3)$$

allowing for a more straightforward study of the isospin symmetry in the FFs. In the case of perfect isospin invariance, the fragmentation functions describing pion production obey the relation [149–152],

$$D_q^{\pi^+}(z, Q^2) + D_q^{\pi^-}(z, Q^2) = 2D_q^{\pi^0}(z, Q^2), \quad (9.4)$$

for a each quark flavor q . Hence, the SIDIS multiplicity for neutral pions is predicted to be equal to the average of those for positive and negative pions. Results from ν Ne-scattering at BEBC [153] and μp -scattering at EMC [77] are consistent with this symmetry assumption. It should be noted that the relation Eq. (9.4) is

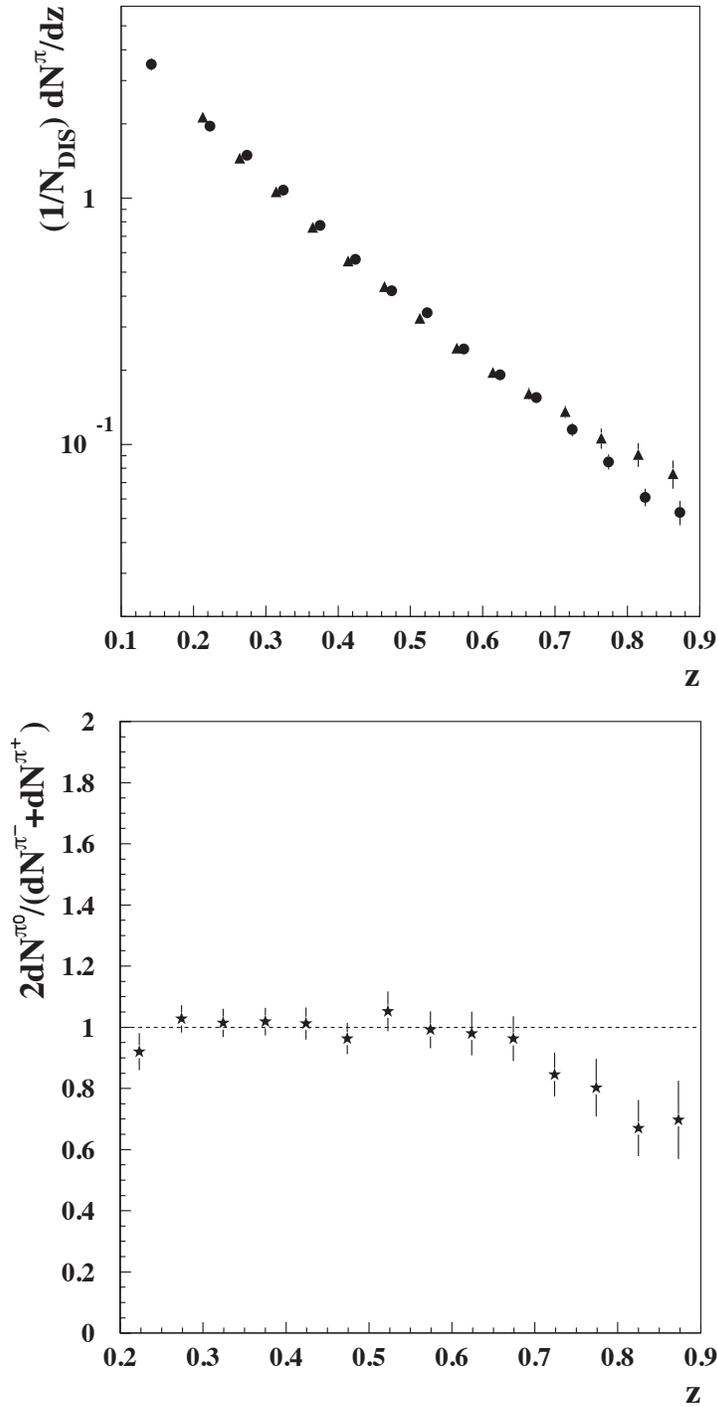


Figure 9.1: Neutral (circles) and average charged (triangles) pion multiplicities from the pre-2000 HERMES dataset are shown in the top panel. The error bars show the statistical uncertainty only. The systematic uncertainty for the average charged (neutral) pions is 7% (9%). The charged pion data have been shifted slightly in z to make them visible. The bottom panel present the ratio of neutral to average charged pion multiplicities. The systematic uncertainty on the ratio is 6%. Figure from [78].

expected to be correct to approximately the 3 % level, due to the mass difference between π^0 and π^\pm ,

$$\frac{M_{\pi^0}}{M_{\pi^\pm}} \approx \frac{135 \text{ MeV}}{139.6 \text{ MeV}} \approx 0.967 \rightarrow 3.3 \%. \quad (9.5)$$

Additionally, pions originating from exclusive ρ^0 production, a process that competes with SIDIS (see Section 4.7), strongly violate this symmetry in the kinematic regions where they become a dominant contribution to the cross section [112, 114].

Figure 9.1 presents HERMES results from an earlier publication on neutral pion production [78], which are extracted from the pre-2000 data sample. The top panel compares the neutral (circles) to the average charged (triangles) multiplicities, and the bottom panel shows the ratio between both. The multiplicities are consistent with isospin invariance below $z \sim 0.7$. It was suggested that this discrepancy at high z might be explained by radiative smearing, the limited detector resolution and exclusive processes. The extraction used a simple set of correction factors to account for the limited acceptance and radiative effects. Contrary to the smearing-unfolding technique approach, a set of multiplicative corrections does not account for the convolution between both effects (cf. Section 4.6). This chapter revisits the neutral pion analysis, leveraging the much larger 2000-2005 data sample, as well as the improved extraction techniques that were developed for the charged pion and kaon measurements presented in this work.

9.1 Measurement of the π^0 Yield

The π^0 has a mean life time of $\tau = (8.52 \pm 0.18) \times 10^{-17}$ s, It decays with a 98.8 % branching ratio into a photon pair $\pi^0 \rightarrow \gamma\gamma$ [9]. While Chapter 4 forms a good starting point for the π^0 extraction, the event reconstruction procedures, detection efficiencies and background corrections differ significantly from that for the charged pions and kaons.

The neutral pion yields for each kinematic bin are measured by means of the invariant mass $M_{\gamma\gamma}$ spectrum of the two-photon system,

$$M_{\gamma\gamma} = \sqrt{4E_{\gamma_1}E_{\gamma_2} \sin^2\left(\frac{\theta_{\gamma_1\gamma_2}}{2}\right)}. \quad (9.6)$$

The requirement that each of these spectra be sufficiently resolved, necessitates a relatively high amount of events per bin. Furthermore, the detection efficiency for measuring the π^0 decay photons in coincidence with the scattered lepton is an order of magnitude smaller than that for the charged pions. The extraction is therefore performed with a one-dimensional binning in z only, given in Table 9.1. Systematic

Q^2 [GeV ²]	> 1
x	0.023–0.6
z	0.1–0.15–0.2–0.25–0.3–0.4–0.5–0.6–0.7–0.8–1.1
$P_{h\perp}$ [GeV]	0.0–1.2

Table 9.1: One-dimensional binning versus z for the π^0 analysis.

1. Calorimeter position	$ x_{\text{cal}} < 125$ cm $ y_{\text{cal}} > 33$ cm $ y_{\text{cal}} < 105$ cm
2. Energy	$E_\gamma > 1$ GeV
3. Detector half	$y_{\gamma_1} \cdot y_{\gamma_2} > 0$

Table 9.2: Photon selection

effects due to this one-dimensional binning are expected to be acceptable because of the weaker dependence of the multiplicities on x , Q^2 and $P_{h\perp}$, compared to z .

Neutral particle reconstruction at HERMES is accomplished through the measurement of electromagnetic showers in the lead-glass calorimeters, that have no associated charged track. This allows for the reconstruction of the energy E_γ of the photon candidate, as well as the coordinates $(x_{\text{cal}}, y_{\text{cal}})$ of the shower center at the calorimeter surface ($z_{\text{cal}} = 747.5$ cm) [94]. The photon selection criteria are given in Table 9.2. For an accurate energy measurement, the entire photon shower has to be contained inside the calorimeter, requiring more strict cuts than those applied to charged particles (cf. Table 4.2b). The π^0 decay vertex is equal to the primary scattering vertex within the experimental resolution, due to the extremely short decay length $c\tau = 25.5$ nm.

This fully constrains the four-momentum of the detected photon. Applying Eq. (9.6) to all possible photon pairs yields the invariant mass of the π^0 candidates. In order to avoid discontinuities in the detector acceptance caused by the gap between both spectrometer halves, only pairs within the same half are considered. This requirement reduces the background by 50 % at low z , without reducing the reconstructed π^0 yield by more than 0.5 %.

Figure 9.2 shows the two-photon invariant mass spectra for three different z -bins. The π^0 decay width is much smaller than the observed width of the signal peak, which is therefore fully caused by the calorimeter resolution. Consequently, a Gaussian distribution provides a good description of the signal. Taking into account the combinatorial background, the combination of a Gaussian and Weibull

distribution,

$$\frac{N^{\pi^0} \Delta_M}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(M_{\gamma\gamma} - M)^2}{2\sigma^2}\right) + P_3(M_{\gamma\gamma} - P_4)^{P_5} + \exp(P_6 M_{\gamma\gamma}), \quad (9.7)$$

with Δ_M the bin width of the invariant mass histogram, describes the invariant mass spectrum near the π^0 peak even more accurately. The Gaussian parameters N^{π^0} , M and σ correspond to, respectively, the π^0 yield, the reconstructed π^0 mass and the calorimeter resolution. The start position of the Weibull distribution, given by P_4 , is fixed to 0.0004 GeV. The solid red curve in Fig. 9.2 shows the result of a χ^2 -fit of Eq. (9.7) to the invariant mass spectrum in three sample z -bins, restricted to the range $0.06 \text{ GeV} < M_{\gamma\gamma} < 0.22 \text{ GeV}$. The background estimate is given by the dashed green curve, while the shaded area indicates the reconstructed π^0 yield.

9.2 Corrections

The extraction of Born-level π^0 multiplicities mirrors Chapter 4 exactly, except for the yield reconstruction described in Section 9.1.

The contamination of the charged pion yields due to the exclusive production of ρ^0 mesons, and their subsequent decay to $\pi^+\pi^-$, strongly violates the isospin relation $2\sigma^{\pi^0} = \sigma^{\pi^+} + \sigma^{\pi^-}$, as there is no similar contribution to the π^0 yield [112]. It is therefore necessary to subtract the contribution by exclusive vector mesons when testing the relation Eq. (9.4).

The dominant exclusive contribution to the π^0 yield comes from the $\omega \rightarrow \pi^+\pi^-\pi^0$ and $\omega \rightarrow \pi^0\gamma$ channels. The subtraction procedure from Section 4.7 accounts for exclusive ω production. The resulting correction on the π^0 cross section is approximately 1 % to 2 %, never exceeding 5 %.

9.3 Systematic Uncertainties

The extraction of high-precision neutral pion multiplicities was not the primary goal of this work. A conservative estimate of the systematic uncertainty is therefore sufficient, instead of the precisely controlled systematic uncertainties on the charged pion and kaon multiplicities (cf. Section 4.9). The separate contributions are described below; addition in quadrature yields a total systematic uncertainty of about 7.8 %. This is significantly larger than the the systematic uncertainty on the charged pions, which ranges between 1 % to 4 % as a function of z (see Fig. 5.1).

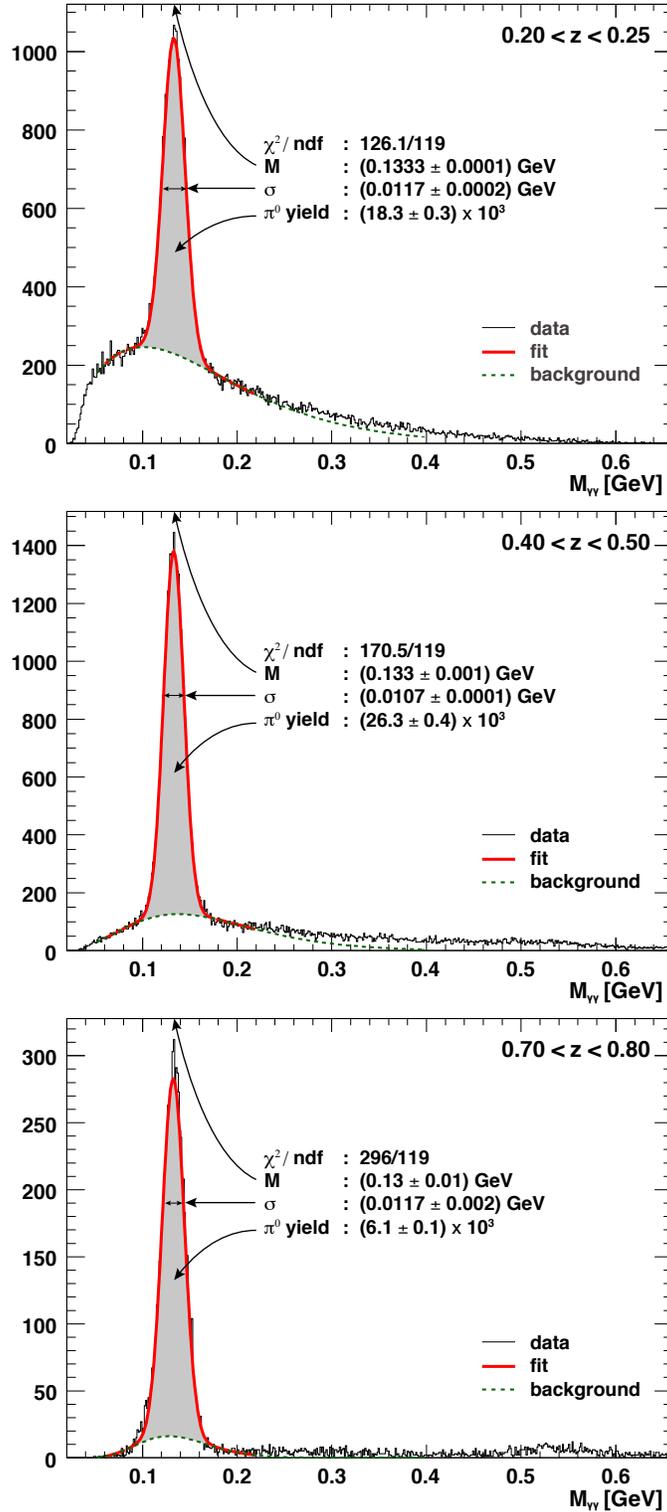


Figure 9.2: Invariant mass $M_{\gamma\gamma}$ spectrum of two-photon events in the calorimeter, for three different z -bins. The solid red curve is obtained through a χ^2 -fit to the spectrum, as described in the text. The dashed green curve shows the estimated background contributions to the spectrum. The shaded area corresponds to the measured π^0 yield.

9.3.1 Background Subtraction

The measured π^0 yields should not strongly depend on the choice of parameterization to describe the combinatorial background, nor should they depend on the specific range in $M_{\gamma\gamma}$ that is fit. To verify this, the invariant mass fits were repeated with two- and three-dimensional polynomials instead of the baseline Weibull distribution. Additionally, the fit range was expanded to $0.05 \text{ GeV} < M_{\gamma\gamma} < 0.40 \text{ GeV}$. The resulting variations do not exceed 2 %.

9.3.2 Calorimeter Calibration and Cluster Position

The two-photon invariant mass spectrum is typically found to be centered at $M_{\pi^0} = (133 \pm 1) \text{ MeV}$, as is illustrated in Fig. 9.2. This constitutes a 1.5 % difference from the value $M_{\pi^0} = (134.9766 \pm 0.0006) \text{ MeV}$ provided by the Particle Data Group [9].

The lead-glass calorimeter blocks are calibrated by means of an iterative procedure, which leverages the relation $E/p = 1$ between the electron energy deposition E in the calorimeter, and the electron momentum p from the standard tracking algorithm [94]. However, the decay photons from SIDIS π^0 production carry substantially less energy than the calibration particles. The position of the π^0 invariant mass peak in SIDIS has been found to change over the calorimeter surface, strongly correlating to the distribution of the calibration leptons, an artifact of this extrapolation. It should be noted that this problem is absent for π^0 from exclusive processes, which have much higher energies, similar to that of the calibration leptons. Work to improve the calorimeter precision at lower energies is currently ongoing, including a re-evaluation of the photon shower profile and the effects of the lead curtain in front of preshower hodoscope (cf. Fig. 3.3).

A shift in the calorimeter energy leads to a slight shift in $z = E_{\pi^0}/\nu$, which changes the measured π^0 kinematic distribution. Unfortunately, these issues are not exactly reproduced by the calorimeter simulation that is used for the HERMES Monte-Carlo. Hence, the issue survives the smearing-unfolding correction, influencing the Born-level multiplicities. To estimate the impact, the extraction was repeated with a calorimeter energy scaled with factors ranging between 0.95 to 1.05. Additionally, the geometrical z -position of the cluster was also changed over a range of -20 cm to 20 cm surrounding the standard position. This resulted in variations of up to 6% in the measured yield. This is the dominant contribution to the systematic uncertainty.

9.3.3 Azimuthal Asymmetries in the Unpolarized Cross Section

The approach from Section 4.9.2 cannot be taken here, as, at the time of writing, there exists no parametrization of the $\cos \phi$ and $\cos 2\phi$ moments for π^0 production. In light of the recent HERMES measurement for the azimuthal modulations of the unpolarized SIDIS cross section for charged pions and kaons [102], the effects on the π^0 extraction can be assumed to lie between those for the π^+ and π^- . For π^+ in a one-dimensional projection versus z , the systematic uncertainty was about 1 %, while the value for π^- is much smaller (cf. Fig. 4.14). The contribution to the systematic uncertainty was therefore estimated to be 1 %.

9.3.4 Monte-Carlo Model Dependence

The JETSET parameter scan from Section 4.9.3 is repeated for the π^0 extraction, resulting in a systematic uncertainty of 4 %.

9.3.5 Time Stability

This systematic uncertainty is estimated using the approach from Section 4.9.4. Its value does not exceed 2 %. It should be noted that a t -test reveals the 2000, 2004 and 2005 samples for both proton and deuteron targets to be (barely) consistent within a 95 % confidence level.

9.4 Neutral Pion Results

Figures 9.3 and 9.4 present the results of the π^0 extraction on respectively a proton and deuteron target. The neutral pions are shown as full circles. The average charged multiplicities have been extracted with the one-dimensional π^0 binning described in Table 9.1 (full squares), as well as with the standard three-dimensional binning described in Table 4.5a (empty squares). The difference between both results was found to be negligible.

The ratio $2M^{\pi^0}/(M^{\pi^+} + M^{\pi^-})$ in the bottom panels is consistent with isospin invariance in the low-to-mid z -range ($z < 0.7$). The values range between $(1.098 \pm 0.018 \pm 0.088)$ at $z = 0.22$ and $(0.841 \pm 0.021 \pm 0.068)$ at $z = 0.75$. The results below $z < 0.7$ are consistent with both unity as well as a modest ISV of the order of 3 % (cf. Eq. (9.5)), without favoring either. At higher z , the multiplicity for charged pions is larger than for neutral pions. Figure 9.5 compares the new values for the ratio (full triangles) to those previously published in Ref. [78] (empty stars). Both measurements agree with each other within their statistical uncertainties. The additional point at $z = 0.88$ in Fig. 9.5 shows a large discrepancy from unity, with a value of $(0.507 \pm 0.009 \pm 0.041)$. This high- z behavior, observed in both the old and

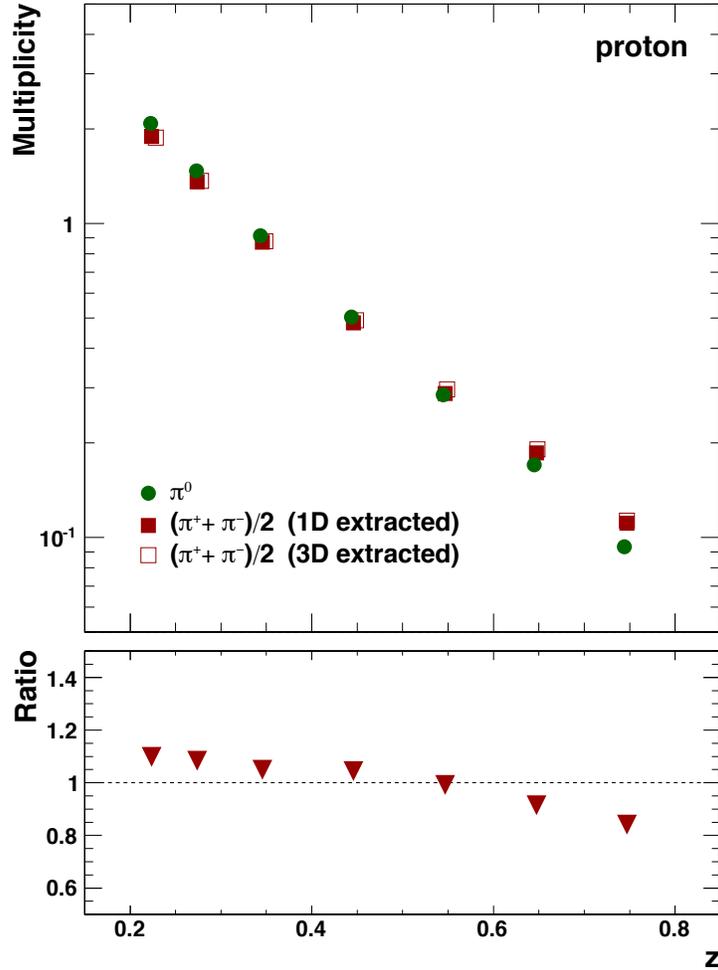


Figure 9.3: Neutral (circles) and average charged (squares) pion multiplicities off the proton as a function of z . The average charged multiplicities have been extracted with the one-dimensional binning described in Table 9.1 (full squares), as well as with the standard three-dimensional binning described in Table 4.5a (empty squares). Statistical error bars are too small to be visible. The systematic uncertainties for the neutral (average charged) pions is 7.8 % (2 % to 4 %). The bottom panels show the ratio of neutral to average charged multiplicities $2M^{\pi^0} / (M^{\pi^+} + M^{\pi^-})$. The systematic uncertainty on the ratio is 8 %.

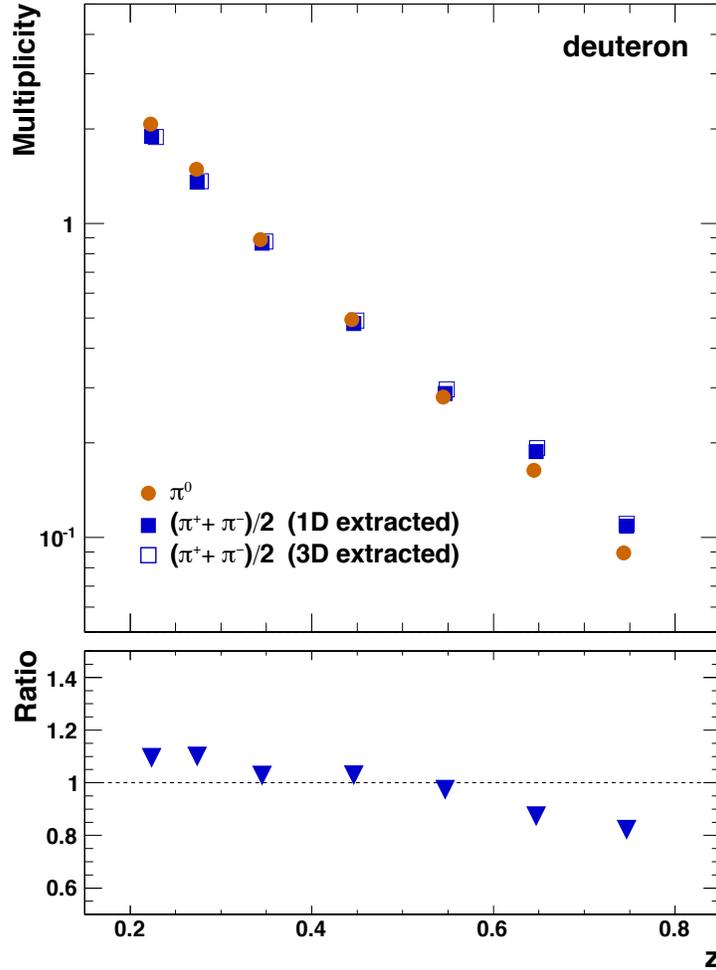


Figure 9.4: Neutral (circles) and average charged (squares) pion multiplicities off the deuteron as a function of z . The average charged multiplicities have been extracted with the one-dimensional binning described in Table 9.1 (full squares), as well as with the standard three-dimensional binning described in Table 4.5a (empty squares). Statistical error bars are too small to be visible. The systematic uncertainties for the neutral (average charged) pions is 7.8 % (2 % to 4 %). The bottom panels show the ratio of neutral to average charged multiplicities $2M^{\pi^0} / (M^{\pi^+} + M^{\pi^-})$. The systematic uncertainty on the ratio is 8 %.

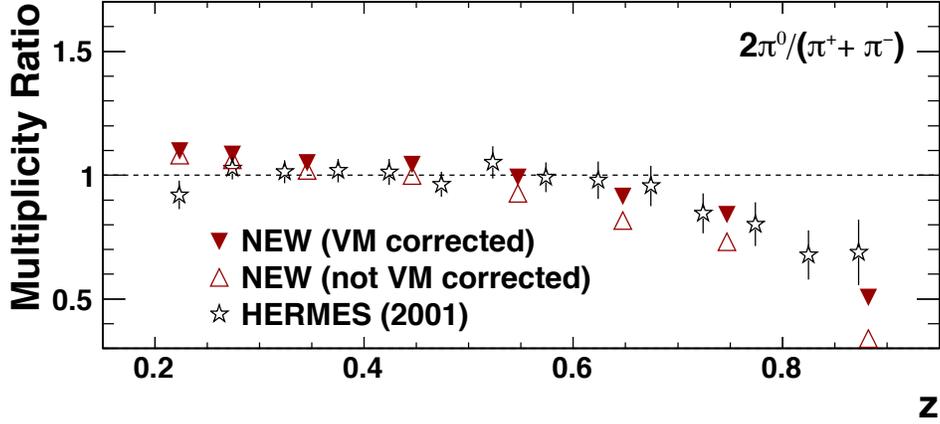


Figure 9.5: Comparison of the ratio of neutral to average charged multiplicities $2M^{\pi^0}/(M^{\pi^+} + M^{\pi^-})$ for a proton target, presented in this chapter (full triangles), to those previously published by HERMES in Ref. [78] (empty stars). Also shown are the new results without the correction for exclusive vector mesons (empty triangles). Statistical error bars on the new points are too small to be visible. The systematic uncertainty is 8 % for the new points and 6 % for the old points.

new results, can be attributed to a growing contribution from exclusive processes (cf. Section 4.7). In fact, the highest z -bin ($0.8 < z < 1.1$) contains the $z \sim 1$ region. The ISV-breaking contribution due to pions from the decay of exclusive ρ^0 , the dominant exclusive channel for this analysis [112, 114], is explicitly subtracted. The empty triangles on Fig. 9.5 show that, without this correction, the ratio at high z drops further, yielding $(0.731 \pm 0.018 \pm 0.06)$ at $z = 0.74$ and $(0.339 \pm 0.006 \pm 0.27)$ at $z = 0.88$.

It should be noted that, while the statistical precision of the new extraction has been greatly improved compared to Ref. [78] due to the much larger data sample used for the extraction, the systematic uncertainty is comparable in size. An improved calorimeter calibration, for example by using the π^0 invariant peak at lower energies ($E \sim 2$ GeV) as an additional constraint, would strongly enhance the precision of this measurement.

To conclude, the ratio of neutral to average charged pion multiplicities is consistent with isospin invariance in the typical SIDIS region ($0.2 < z < 0.7$). Additionally, the results from the very different extractions for π^0 and π^\pm at HERMES are shown to be internally consistent with each other.

Bibliography

- [1] J Bjorken et al. “Inelastic Electron-Proton and γ -Proton Scattering and the Structure of the Nucleon”. In: *Phys.Rev.* 185.5 (1969), pp. 1975–1982 (cit. on p. 6).
- [2] R.P. Feynman. “Photon-hadron interactions”. In: (1973) (cit. on p. 6).
- [3] C Callan et al. “High-Energy Electroproduction and the Constitution of the Electric Current”. In: *Physical Review Letters* 22.4 (1969), pp. 156–159 (cit. on p. 8).
- [4] H1 and ZEUS collaborations et al. “Combined measurement and QCD analysis of the inclusive $e^\pm p$ scattering cross sections at HERA”. In: *Journal of High Energy Physics* 2010.1 (2010), pp. 109–63 (cit. on p. 9).
- [5] L W Whitlow et al. “Precise measurements of the proton and deuteron structure functions from a global analysis of the SLAC deep inelastic electron scattering cross sections”. In: *Physics Letters B* 282.3-4 (1992), pp. 475–482 (cit. on p. 9).
- [6] A C Benvenuti et al. “A high statistics measurement of the proton structure functions $F_2(x, Q^2)$ and R from deep inelastic muon scattering at high Q^2 ”. In: *Physics Letters B* 223.3-4 (1989), pp. 485–489 (cit. on p. 9).
- [7] M R Adams et al. “Proton and deuteron structure functions in muon scattering at 470 GeV”. In: *Physical Review D* 54.5 (1996), pp. 3006–3056 (cit. on p. 9).
- [8] M Arneodo et al. “Measurement of the proton and deuteron structure functions, F_2^p and F_2^d , and of the ratio σ_L/σ_T ”. In: *Nuclear Physics B* 483.1-2 (1997), pp. 3–43 (cit. on p. 9).
- [9] J Beringer et al. “Review of Particle Physics (RPP)”. In: *Phys.Rev.* D86 (2012), p. 010001 (cit. on pp. 9, 11, 14–17, 122, 124, 128).
- [10] Davison E Soper. “Basics of QCD Perturbation Theory”. In: (1997) (cit. on p. 8).
- [11] John C. Collins et al. “Factorization of Hard Processes in QCD”. In: *Adv.Ser.Direct.High Energy Phys.* 5 (1988), pp. 1–91. arXiv: [hep-ph/0409313](https://arxiv.org/abs/hep-ph/0409313) [[hep-ph](https://arxiv.org/abs/hep-ph/0409313)] (cit. on pp. 8, 10, 16, 95, 102).

- [12] W Furmanski et al. “Lepton-hadron processes beyond leading order in quantum chromodynamics”. In: *Zeitschrift für Physik C Particles and Fields* 11.4 (1982), pp. 293–314 (cit. on pp. [10](#), [13](#), [84](#), [102](#)).
- [13] John C Collins et al. “Back-to-back jets in QCD”. In: *Nuclear Physics B* 193.2 (1981), pp. 381–443 (cit. on pp. [10](#), [16](#), [95](#), [102](#)).
- [14] G Curci et al. “Evolution of parton densities beyond leading order: The non-singlet case”. In: *Nuclear Physics B* 175.1 (1980), pp. 27–92 (cit. on pp. [10](#), [12](#)).
- [15] A D Martin et al. “Parton distributions for the LHC”. In: *The European Physical Journal C* 63.2 (2009), pp. 189–285 (cit. on pp. [11](#), [13](#)).
- [16] V.N. Gribov et al. “Deep inelastic e p scattering in perturbation theory”. In: *Sov.J.Nucl.Phys.* 15 (1972), pp. 438–450 (cit. on p. [11](#)).
- [17] L.N. Lipatov. “The parton model and perturbation theory”. In: *Sov.J.Nucl.Phys.* 20 (1975), pp. 94–102 (cit. on p. [11](#)).
- [18] G Altarelli et al. “Asymptotic freedom in parton language”. In: *Nuclear Physics B* 126.2 (1977), pp. 298–318 (cit. on p. [11](#)).
- [19] Yuri L. Dokshitzer. “Calculation of the Structure Functions for Deep Inelastic Scattering and e+ e- Annihilation by Perturbation Theory in Quantum Chromodynamics.” In: *Sov.Phys.JETP* 46 (1977), pp. 641–653 (cit. on p. [11](#)).
- [20] Bo Andersson et al. “On High-Energy Leptoproduction”. In: *Z.Phys.* C9 (1981), p. 233 (cit. on pp. [12](#), [104](#)).
- [21] Gunnar Ingelman et al. “Transverse Momentum Effects and Angular Energy Flow in Leptoproduction”. In: *Nucl.Phys.* B206 (1982), p. 239 (cit. on pp. [12](#), [104](#)).
- [22] Torbjorn Sjostrand. “The Lund Monte Carlo for Jet Fragmentation”. In: *Comput.Phys.Commun.* 27 (1982), p. 243 (cit. on pp. [12](#), [104](#)).
- [23] R D Field et al. “A Parametrization of the Properties of Quark Jets”. In: *Nucl.Phys.* B136 (1978), p. 1 (cit. on pp. [12](#), [104](#)).
- [24] P J Rijken et al. “Higher order QCD corrections to the transverse and longitudinal fragmentation functions in electron-positron annihilation”. In: *Nuclear Physics B* 487.1–2 (1997), pp. 233–282 (cit. on p. [12](#)).
- [25] P J Rijken et al. “ $O(\alpha_s)$ contributions to the longitudinal fragmentation function in e+e- annihilation”. In: *Physics Letters B* 386.1–4 (1996), pp. 422–428 (cit. on p. [12](#)).

- [26] W Furmanski et al. “Singlet parton densities beyond leading order”. In: *Physics Letters B* 97.3–4 (1980), pp. 437–442 (cit. on p. 12).
- [27] M Stratmann et al. “Next-to-leading order evolution of polarized and unpolarized fragmentation functions”. In: *Nuclear Physics B* 496.1–2 (1997), pp. 41–65 (cit. on p. 12).
- [28] R L Jaffe et al. “Spin, twist and hadron structure in deep inelastic processes”. In: Springer Berlin Heidelberg, 1997, pp. 178–249 (cit. on p. 12).
- [29] Daniel de Florian et al. “Global analysis of fragmentation functions for pions and kaons and their uncertainties”. In: *Physical Review D* 75.11 (2007), p. 114010 (cit. on pp. 12–14, 17, 84–88, 90–92, 94, 102, 112, 113, 116, 118, 119, 121, 122).
- [30] G Altarelli et al. “Processes involving fragmentation functions beyond the leading order in QCD”. In: *Nuclear Physics B* 160.2 (1979), pp. 301–329 (cit. on pp. 13, 84, 102).
- [31] P Nason et al. “Scaling violation in e^+e^- fragmentation functions: QCD evolution, hadronization and heavy quark mass effects”. In: *Nuclear Physics B* 421.3 (1994), pp. 473–517 (cit. on pp. 13, 84, 102).
- [32] Dirk Graudenz. “One-particle inclusive processes in deeply inelastic lepton-nucleon scattering”. In: *Nuclear Physics B* 432.1-2 (1994), pp. 351–373 (cit. on pp. 13, 84, 102, 103).
- [33] Daniel de Florian et al. “Global analysis of fragmentation functions for protons and charged hadrons”. In: *Physical Review D* 76.7 (2007), p. 074033 (cit. on p. 14).
- [34] S Albino et al. “AKK update: Improvements from new theoretical input and experimental data”. In: *Nuclear Physics B* 803.1-2 (2008), pp. 42–104 (cit. on p. 14).
- [35] S Albino et al. “Fragmentation functions for light charged hadrons with complete quark flavour separation”. In: *Nuclear Physics B* 725.1-2 (2005), pp. 181–206 (cit. on pp. 14, 113).
- [36] M Hirai et al. “Determination of fragmentation functions and their uncertainties”. In: *Physical Review D* 75.9 (2007), p. 094009 (cit. on pp. 14, 85, 86, 113).
- [37] J Pumplin et al. “Parton distributions and the strong coupling: CTEQ6AB PDFs”. In: *Journal of High Energy Physics* 0602 (2006), p. 032 (cit. on pp. 13, 85, 86, 88, 90–92, 94, 112, 115, 116, 118, 119, 121).

- [38] R.S. Thorne et al. “MRST parton distributions: Status 2006”. In: (2006), pp. 81–84. arXiv: [hep-ph/0606244](https://arxiv.org/abs/hep-ph/0606244) [[hep-ph](#)] (cit. on pp. 13, 85, 112).
- [39] DELPHI Collaboration et al. “Measurement of the quark and gluon fragmentation functions in Z^0 hadronic decays”. In: *The European Physical Journal C* 6.1 (1999), pp. 19–33 (cit. on p. 15).
- [40] R Barate et al. “Studies of Quantum Chromodynamics with the ALEPH detector”. In: *Physics Reports* 294.1-3 (1998), pp. 1–165 (cit. on p. 15).
- [41] B Adeva et al. “Measurement of the inclusive production of neutral pions and charged particles on the Z^0 resonance”. In: *Physics Letters B* 259.1-2 (1991), pp. 199–208 (cit. on p. 15).
- [42] Y Li et al. “Multihadron-event properties in e^+e^- annihilation at $\sqrt{s} = 52 - 57$ GeV”. In: *Physical Review D* 41.9 (1990), pp. 2675–2688 (cit. on p. 15).
- [43] D Bender et al. “Study of quark fragmentation in e^+e^- annihilation at 29 GeV: Global jet parameters and single-particle distributions”. In: *Physical Review D* 31.1 (1985), pp. 1–16 (cit. on p. 15).
- [44] G Abrams et al. “Measurements of charged-particle inclusive distributions in hadronic decays of the Z boson”. In: *Physical Review Letters* 64.12 (1990), pp. 1334–1337 (cit. on p. 15).
- [45] A Petersen et al. “Multihadronic events at $E_{c.m.} = 29$ GeV and predictions of QCD models from $E_{c.m.} = 29$ GeV to $E_{c.m.} = 93$ GeV”. In: *Physical Review D* 37.1 (1988), pp. 1–27 (cit. on p. 15).
- [46] OPAL Collaboration et al. “QCD studies with e^+e^- annihilation data at 130 and 136 GeV”. In: *Zeitschrift für Physik C Particles and Fields* 72.2 (1996), pp. 191–206 (cit. on p. 15).
- [47] OPAL Collaboration et al. “Measurements of flavour-dependent fragmentation functions in $Z^0 \rightarrow q\bar{q}$ events”. In: *The European Physical Journal C* 7.3 (1999), pp. 369–381 (cit. on p. 15).
- [48] OPAL Collaboration et al. “QCD studies with e^+e^- annihilation data at 172-189 GeV”. In: *The European Physical Journal C - Particles and Fields* 16.2 (2000), pp. 185–210 (cit. on p. 15).
- [49] OPAL Collaboration et al. “Charged particle momentum spectra in ee annihilation at $\sqrt{s} = 192 - 209$ GeV”. In: *The European Physical Journal C - Particles and Fields* 27.4 (2003), pp. 467–481 (cit. on p. 15).

- [50] OPAL Collaboration et al. “Scaling violations of quark and gluon jet fragmentation functions in e^+e^- annihilations at $\sqrt{s} = 91.2$ and 183-209 GeV”. In: *The European Physical Journal C - Particles and Fields* 37.1 (2004), pp. 25–47 (cit. on p. 15).
- [51] Koya Abe et al. “Production of π^+ , π^- , K^+ , K^- , p , and \bar{p} in light (uds), c , and b jets from Z^0 decays”. In: *Physical Review D* 69.7 (2004), p. 072003 (cit. on p. 15).
- [52] P Abreu et al. “A measurement of α_s from the scaling violation in e^+e^- annihilation”. In: *Physics Letters B* 398.1-2 (1997), pp. 194–206 (cit. on p. 15).
- [53] TASSO Collaboration et al. “Scale breaking in inclusive charged particle production by e^+e^- annihilation”. In: *Physics Letters B* 114.1 (1982), pp. 65–70 (cit. on p. 15).
- [54] TASSO Collaboration et al. “Global jet properties at 14-44 GeV center of mass energy in e^+e^- annihilation”. In: *The European Physical Journal C - Particles and Fields* 47.2 (1990), pp. 187–198 (cit. on p. 15).
- [55] H Aihara et al. “Charged-Hadron Inclusive Cross Sections and Fractions in e^+e^- Annihilation at $\sqrt{s} = 29$ GeV”. In: *Physical Review Letters* 61.11 (1988), pp. 1263–1266 (cit. on p. 15).
- [56] DELPHI Collaboration et al. “ π^\pm , K^\pm , p and \bar{p} production in $Z^0 \rightarrow q\bar{q}$, $Z^0 \rightarrow b\bar{b}$, $Z^0 \rightarrow u\bar{u}$, $d\bar{d}$, $s\bar{s}$ ”. In: *The European Physical Journal C* 5.4 (1998), pp. 585–620 (cit. on p. 15).
- [57] ALEPH Collaboration et al. “Inclusive π^pm , K^\pm and (p, \bar{p}) differential cross-sections at the Z resonance”. In: *The European Physical Journal C - Particles and Fields* 66.3 (1995), pp. 355–365 (cit. on p. 15).
- [58] ARGUS Collaboration et al. “Inclusive production of charged pions, charged and neutral kaons and antiprotons in e^+e^- annihilation at 10 GeV and in direct Upsilon decays”. In: *The European Physical Journal C - Particles and Fields* 44.4 (1989), pp. 547–555 (cit. on p. 15).
- [59] OPAL Collaboration et al. “Measurement of the production rates of charged hadrons in e^+e^- -annihilation at the Z^0 ”. In: *The European Physical Journal C - Particles and Fields* 63.2 (1994), pp. 181–195 (cit. on p. 15).
- [60] K Abe et al. “Production of π^+ , K^+ , K^0 , K^{*0} , ϕ , p and Λ^- in hadronic Z^0 decays”. In: *Physical Review D* 59.5 (1999), p. 052001 (cit. on p. 15).

- [61] S Kretzer. “Fragmentation functions from flavor-inclusive and flavor-tagged e^+e^- annihilations”. In: *Physical Review D* 62 (2000), p. 054001 (cit. on pp. 14, 85, 86, 90, 91, 113).
- [62] B A Kniehl et al. “Fragmentation functions for pions, kaons, and protons at next-to-leading order”. In: *Nuclear Physics B* 582.1–3 (2000), pp. 514–536 (cit. on pp. 14, 113).
- [63] D. Buskulic et al. “Inclusive π^\pm , K^\pm and (p, \bar{p}) differential cross-sections at the Z resonance”. In: *Z.Phys.* C66 (1995), pp. 355–366 (cit. on pp. 14, 113).
- [64] DELPHI Collaboration et al. “ π^\pm , K^\pm , p and \bar{p} production in $Z^0 \rightarrow q\bar{q}$, $Z^0 \rightarrow b\bar{b}$, $Z^0 \rightarrow u\bar{u}$, $d\bar{d}$, $s\bar{s}$ ”. In: *The European Physical Journal C* 5.4 (1998), pp. 585–620 (cit. on pp. 14, 113).
- [65] R. Akers et al. “Measurement of the production rates of charged hadrons in e^+e^- annihilation at the Z^0 ”. In: *Z.Phys.* C63 (1994), pp. 181–196 (cit. on pp. 14, 113).
- [66] The OPAL Collaboration et al. “Leading particle production in light flavour jets”. In: *The European Physical Journal C* 16.3 (2000), pp. 407–421 (cit. on pp. 14, 113).
- [67] K. Abe et al. “Production of π^+ , K^+ , K^0 , K^{*0} , ϕ , p and Λ^0 in hadronic Z^0 decays”. In: *Phys. Rev. D* 59 (5 1999), p. 052001 (cit. on pp. 14, 113).
- [68] S S Adler et al. “Midrapidity Neutral-Pion Production in Proton-Proton Collisions at $\sqrt{s} = 200\text{GeV}$ ”. In: *Physical Review Letters* 91.24 (2003), p. 241803 (cit. on p. 16).
- [69] WA 70 Collaboration et al. “Production of high transverse momentum prompt photons and neutral pions in proton-proton interactions at $280\text{ GeV}/c$ ”. In: *The European Physical Journal C - Particles and Fields* 38.3 (1988), pp. 371–382 (cit. on p. 16).
- [70] C Kourkoumelis et al. “A study of the production of high- p_T π^0 at the CERN intersecting storage rings”. In: *The European Physical Journal C - Particles and Fields* 5.2 (1980), pp. 95–104 (cit. on p. 16).
- [71] L Apanasevich et al. “Production of π^0 and η mesons at large transverse momenta in pp and $p\text{Be}$ interactions at 530 and 800 GeV/c ”. In: *Physical Review D* 68.5 (2003), p. 052001 (cit. on p. 16).
- [72] F Abe et al. “Transverse-momentum distributions of charged particles produced in $p\bar{p}$ interactions at $\sqrt{s} = 630$ and 1800 GeV ”. In: *Physical Review Letters* 61.16 (1988), pp. 1819–1822 (cit. on p. 16).

- [73] C Albajar et al. “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 (1990), pp. 261–287 (cit. on p. 16).
- [74] M Banner et al. “Inclusive π^0 production at the CERN $p\bar{p}$ collider”. In: *Physics Letters B* 115.1 (1982), pp. 59–64 (cit. on p. 16).
- [75] J Adams et al. “Identified hadron spectra at large transverse momentum in $p + p$ and $d + \text{Au}$ collisions at $\sqrt{s_{NN}} = 200\text{GeV}$ ”. In: *Physics Letters B* 637.3 (2006), pp. 161–169 (cit. on p. 16).
- [76] I Arsene et al. “Production of Mesons and Baryons at High Rapidity and High p_T in Proton-Proton Collisions at $\sqrt{s} = 200\text{GeV}$ ”. In: *Physical Review Letters* 98.25 (2007), p. 252001 (cit. on p. 16).
- [77] The European Muon Collaboration et al. “Production of π^0 mesons in muon-hydrogen interactions at 200 GeV ”. In: *The European Physical Journal C* 18.3 (1983), pp. 189–197 (cit. on pp. 17, 122).
- [78] HERMES Collaboration et al. “Multiplicity of charged and neutral pions in deep-inelastic scattering of 27.5 GeV positrons on hydrogen”. In: *The European Physical Journal C* 21.4 (2001), pp. 599–606 (cit. on pp. 17, 76, 123, 124, 129, 132).
- [79] J Ashman et al. “An investigation of the spin structure of the proton in deep inelastic scattering of polarised muons on polarised protons”. In: *Nuclear Physics B* 328.1 (1989), pp. 1–35 (cit. on p. 18).
- [80] J Ashman et al. “A measurement of the spin asymmetry and determination of the structure function g_1 in deep inelastic muon-proton scattering”. In: *Physics Letters B* 206.2 (1988), pp. 364–370 (cit. on p. 18).
- [81] K Ackerstaff et al. “The HERMES Spectrometer”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 417.2–3 (1998), pp. 230–265 (cit. on pp. 18, 21, 26, 42).
- [82] A.A. Sokolov et al. “On Polarization and spin effects in the theory of synchrotron radiation”. In: *Sov.Phys.Dokl.* 8 (1964), pp. 1203–1205 (cit. on p. 19).
- [83] D P Barber et al. “The first achievement of longitudinal spin polarization in a high energy electron storage ring”. In: *Physics Letters B* 343.1-4 (1995), pp. 436–443 (cit. on p. 19).

- [84] Jean Buon et al. “Hera variable-energy “mini” spin rotator and head-on ep collision scheme with choice of electron helicity”. In: *Nuclear Inst. and Methods in Physics Research, A* 245.2-3 (1986), pp. 248–261 (cit. on p. 19).
- [85] Avetik Airapetian et al. “The HERMES polarized hydrogen and deuterium gas target in the HERA electron storage ring”. In: *Nuclear Inst. and Methods in Physics Research, A* 540.1 (2005), pp. 68–101 (cit. on p. 19).
- [86] P Lenisa. “The HERMES transversely polarized hydrogen target”. In: *Nuclear Inst. and Methods in Physics Research, A* 536.3 (2005), pp. 244–247 (cit. on p. 19).
- [87] HERMES Collaboration. “Nuclear polarization of molecular hydrogen recombined on a non-metallic surface”. In: *The European Physical Journal D - Atomic, Molecular, Optical and Plasma Physics* 29.1 (2004), pp. 21–26 (cit. on p. 20).
- [88] A Nass et al. “The HERMES polarized atomic beam source”. In: *Nuclear Inst. and Methods in Physics Research, A* 505.3 (2003), pp. 633–644 (cit. on p. 20).
- [89] Björn Seitz. “The HERMES recoil detector: a combined silicon strip and scintillating fibre detector for tracking and particle identification”. In: *Nuclear Inst. and Methods in Physics Research, A* 535.1-2 (2004), pp. 538–541 (cit. on p. 20).
- [90] Wolf-Dieter Nowak. “HERMES run II”. In: *The European Physical Journal A* 19 (2004), pp. 261–265 (cit. on p. 20).
- [91] Charlotte Van Hulse et al. “The HERMES recoil detector”. In: *Nuclear Inst. and Methods in Physics Research, A* 623.1 (2010), pp. 359–361 (cit. on p. 20).
- [92] C Baumgarten et al. “An atomic beam polarimeter to measure the nuclear polarization in the HERMES gaseous polarized hydrogen and deuterium target”. In: *Nuclear Inst. and Methods in Physics Research, A* 482.3 (2002), pp. 606–618 (cit. on p. 21).
- [93] Wolfgang Wander. “Reconstruction of High Energy Scattering Events in the HERMES Experiment”. PhD thesis. Friedrich-Alexander-Universität Erlangen-Nürnberg, 1996 (cit. on p. 22).
- [94] H Avakian et al. “Performance of the electromagnetic calorimeter of the HERMES experiment”. In: *Nuclear Inst. and Methods in Physics Research, A* 417.1 (1998), pp. 69–78 (cit. on pp. 23, 125, 128).

- [95] Boris Dolgoshein. “Transition radiation detectors”. In: *Nuclear Inst. and Methods in Physics Research, A* 326.3 (1993), pp. 434–469 (cit. on p. 23).
- [96] N Akopov et al. “The HERMES dual-radiator ring imaging Cherenkov detector”. In: *Nuclear Inst. and Methods in Physics Research, A* 479.2–3 (2002), pp. 511–530 (cit. on p. 24).
- [97] J Wendland. *Particle identification for HERMES Run I*. HERMES Internal Report. 2001 (cit. on pp. 30, 31).
- [98] Markus Diefenthaler. “Signals for transversity and transverse-momentum-dependent quark distribution functions studied at the HERMES experiment”. PhD thesis. Friedrich-Alexander-Universität Erlangen-Nürnberg, 2008 (cit. on p. 31).
- [99] Sylvester Joosten. “An Improved Calibration of the HERMES Hodoscopes”. MA thesis. Ghent University, 2007 (cit. on p. 31).
- [100] J Wendland. “Improved particle identification at HERMES and polarized valence quark distributions in the proton”. MA thesis. Simon Fraser University, 1999 (cit. on p. 31).
- [101] Rebecca Lamb. “The Boer-Mulders and Cahn effects: azimuthal modulations in the spin-independent SIDIS cross section at HERMES”. PhD thesis. University of Illinois at Urbana-Champaign, 2010 (cit. on pp. 32, 33, 62, 63).
- [102] Avetik Airapetian et al. “Azimuthal distributions of charged hadrons, pions, and kaons produced in deep-inelastic scattering off unpolarized protons and deuterons”. In: *Physical Review D* 87.1 (2013), p. 012010 (cit. on pp. 32, 63, 129).
- [103] HERMES Collaboration et al. “Inclusive measurements of inelastic electron and positron scattering from unpolarized hydrogen and deuterium targets”. In: *Journal of High Energy Physics* 1105.5 (2011), p. 126 (cit. on p. 35).
- [104] Glen Cowan. *Statistical Data Analysis*. Oxford: Clarendon (Oxford), 1998 (ISBN: 0-19-850156-0 or 0-19-850155-2 in paperback) (cit. on pp. 41, 64, 68).
- [105] G Ingelman et al. “LEPTO 6.5 - A Monte Carlo generator for deep inelastic lepton-nucleon scattering”. In: *Computer Physics Communications* 101.1–2 (1997), pp. 108–134 (cit. on pp. 41, 58, 117).

- [106] Torbjorn Sjostrand. “The Lund Monte Carlo for jet fragmentation and e+e- physics - jetset version 6.2 ”. In: *Computer Physics Communications* 39 (1986), pp. 347–407 (cit. on pp. [41](#), [58](#), [104](#), [117](#)).
- [107] I. Akushevich et al. “RADGEN 1.0: Monte Carlo generator for radiative events in DIS on polarized and unpolarized targets”. In: *Proceedings of the Workshop on Monte Carlo Generators for HERA Physics, Hamburg, Germany* (1998). arXiv: [hep-ph/9906408 \[hep-ph\]](#) (cit. on p. [41](#)).
- [108] R. Brun et al. “GEANT: simulation program for particle physics experiments. user guide and reference manual”. In: *Report No. CERN-DD-78-2-REV* (1978) (cit. on p. [41](#)).
- [109] Avetik Airapetian et al. “Measurement of parton distributions of strange quarks in the nucleon from charged-kaon production in deep-inelastic scattering on the deuteron”. In: *Physics Letters B* 666.5 (2008), pp. 446–450 (cit. on pp. [42](#), [44](#)).
- [110] Avetik Airapetian et al. “Erratum to: “Nuclear effects on $R=\sigma_L/\sigma_T$ in deep-inelastic scattering”: [Phys. Lett. B 475 (2000) 386–394]”. In: *Physics Letters B* 567.3–4 (2003), pp. 339–346 (cit. on p. [45](#)).
- [111] K Ackerstaff et al. “Nuclear effects on $R=\sigma_L/\sigma_T$ in deep-inelastic scattering”. In: *Physics Letters B* 475.3-4 (2000), pp. 386–394 (cit. on p. [45](#)).
- [112] M Diehl et al. “Exclusive channels in semi-inclusive production of pions and kaons”. In: *Physical Review D* 72.3 (2005), p. 034034 (cit. on pp. [47](#), [49](#), [124](#), [126](#), [132](#)).
- [113] Patricia Liebing. “Can the Gluon Polarization in the Nucleon be Extracted from HERMES Data on Single High-p”. PhD thesis. Hamburg: University of Hamburg, 2004 (cit. on pp. [49](#), [50](#), [64](#)).
- [114] A Szczurek et al. “ $\bar{d} - \bar{u}$ asymmetry and semi-inclusive production of pions in deep inelastic scattering”. In: *Physical Review D* 63.11 (2001), p. 114005 (cit. on pp. [49](#), [124](#), [132](#)).
- [115] Torbjorn Sjostrand et al. “High-energy-physics event generation with Pythia 6.1”. In: *Computer Physics Communications* 135.2 (2001), pp. 238–259 (cit. on pp. [50](#), [58](#), [104](#)).
- [116] E C Aschenauer. *Improved PYTHIA Vector Meson Tune*. Private communication. 2012 (cit. on p. [50](#)).
- [117] Alessandro Bacchetta et al. “Semi-inclusive deep inelastic scattering at small transverse momentum”. In: *Journal of High Energy Physics* 2007.02 (2007), p. 093 (cit. on pp. [63](#), [95](#)).

- [118] T. P. Cheng et al. “Coincidence Electroproduction and Scaling in the Regge Region”. In: *Phys. Rev. D* 6 (3 1972), pp. 885–893 (cit. on p. 63).
- [119] R L Kingsley. “Azimuthal dependence of semi-inclusive deep-inelastic reactions in the parton model”. In: *Physical Review D* 10 (1974), pp. 1580–1586 (cit. on p. 63).
- [120] Robert N Cahn. “Azimuthal dependence in leptonproduction: A simple parton model calculation”. In: *Physics Letters B* 78.2–3 (1978), pp. 269–273 (cit. on p. 63).
- [121] Robert N Cahn et al. “Realistic equivalent-photon yields in heavy-ion collisions”. In: *Physical Review D* 42 (1990), pp. 3690–3695 (cit. on p. 63).
- [122] D Boer et al. “Time-reversal odd distribution functions in leptonproduction”. In: *Physical Review D* 57.9 (1998), pp. 5780–5786 (cit. on pp. 63, 95).
- [123] Daniël Boer et al. “Color gauge invariance in the Drell–Yan process”. In: *Nuclear Physics B* 569.1–3 (2000), pp. 505–526 (cit. on p. 63).
- [124] Achim Hillenbrand. “Messung und Simulation des Fragmentationsprozesses bei H”. PhD thesis. Friedrich-Alexander-Universität Erlangen-Nürnberg, 2005 (cit. on pp. 64, 86, 117, 121).
- [125] J Rubin. “Polarization, Motion and Fragmentation: Exploring the role of quarks in the nucleon through semi-inclusive longitudinal spin asymmetries at HERMES”. PhD thesis. University of Illinois at Urbana-Champaign, 2009 (cit. on p. 64).
- [126] J Pumplin et al. “Multivariate fitting and the error matrix in global analysis of data”. In: *Physical Review D* 65.1 (2001), p. 014011 (cit. on p. 64).
- [127] Avetik Airapetian et al. “Multiplicities of charged pions and kaons from semi-inclusive deep-inelastic scattering by the proton and the deuteron”. In: *Physical Review D* 87.7 (2013), p. 074029 (cit. on pp. 76, 111).
- [128] Sylvester Joosten. *Multiplicity Database*. 2013. URL: <http://www-hermes.desy.de/multiplicities> (cit. on p. 80).
- [129] Marco Stratmann. *Impact of the new HERMES multiplicities on a future global analysis of FFs*. Private communication. 2013 (cit. on p. 87).
- [130] Martin Leitgab. “Fragmentation Functions at Belle”. In: (2012), pp. 955–958. arXiv: [1210.2137 \[nucl-ex\]](https://arxiv.org/abs/1210.2137) (cit. on p. 87).
- [131] Martin Leitgab. “Precision measurement of charged pion and kaon multiplicities in e^+e^- annihilation at $Q = 10.52\text{GeV}$ ”. PhD thesis. University of Illinois at Urbana-Champaign, 2013 (cit. on p. 87).

- [132] P J Mulders et al. “The Complete tree level result up to order $1/Q$ for polarized deep inelastic lepton production”. In: *Nuclear Physics B* B461 (1996), pp. 197–237 (cit. on p. 95).
- [133] Daniël Boer et al. “Angular dependences in electroweak semi-inclusive lepton production”. In: *Nuclear Physics B* 564.3 (2000), pp. 471–485 (cit. on p. 95).
- [134] Xiangdong Ji et al. “QCD factorization for semi-inclusive deep-inelastic scattering at low transverse momentum”. In: *Physical Review D* 71.3 (2005), p. 034005 (cit. on p. 95).
- [135] Xiangdong Ji et al. “QCD factorization for spin-dependent cross sections in DIS and Drell–Yan processes at low transverse momentum”. In: *Physics Letters B* 597.3-4 (2004), pp. 299–308 (cit. on p. 95).
- [136] P Schweitzer et al. “Intrinsic transverse parton momenta in deeply inelastic reactions”. In: *Physical Review D* 81.9 (2010), p. 094019 (cit. on p. 96).
- [137] Avetik Airapetian et al. “Observation of the Naive-T-Odd Sivers Effect in Deep-Inelastic Scattering”. In: *Physical Review Letters* 103.15 (2009), p. 152002 (cit. on p. 96).
- [138] L Trentadue et al. “Fracture functions. An improved description of inclusive hard processes in QCD”. In: *Physics Letters B* 323.2 (1994), pp. 201–211 (cit. on p. 103).
- [139] M R Adams et al. “Production of charged hadrons by positive muons on deuterium and xenon at 490-GeV”. In: *Z.Phys.* C61 (1994), pp. 179–198 (cit. on pp. 104–106, 108).
- [140] T Sloan et al. “The Quark Structure of the Nucleon from the CERN Muon Experiments”. In: *Phys.Rept.* 162 (1988), pp. 45–167 (cit. on pp. 104, 107, 108).
- [141] Edmond L Berger. “Rapidity Correlations at Fixed Multiplicity in Cluster Emission Models”. In: *Nucl.Phys.* B85 (1975), p. 61 (cit. on p. 108).
- [142] P.J. Mulders. “Current fragmentation in semiinclusive lepton production”. In: *AIP Conf.Proc.* 588 (2001), pp. 75–88. arXiv: [hep-ph/0010199](https://arxiv.org/abs/hep-ph/0010199) [[hep-ph](https://arxiv.org/abs/hep-ph)] (cit. on p. 108).
- [143] K Ackerstaff et al. “Flavor Asymmetry of the Light Quark Sea from Semi-inclusive Deep-Inelastic Scattering”. In: *Physical Review Letters* 81 (1998), pp. 5519–5523 (cit. on pp. 109, 110, 112).

- [144] A D Martin et al. “Uncertainties of predictions from parton distributions I: Experimental errors”. In: *The European Physical Journal C* 28.4 (2003), pp. 455–473 (cit. on p. 112).
- [145] J T Londergan et al. “Charge symmetry at the partonic level”. In: *Reviews of Modern Physics* 82.3 (2010), pp. 2009–2052 (cit. on pp. 113, 114, 122).
- [146] J T Londergan et al. “Probing charge symmetry violating quark distributions in semiinclusive lepton production of hadrons”. In: *Journal of High Energy Physics* D54 (1996), pp. 3154–3161 (cit. on pp. 113, 114, 122).
- [147] G A Miller et al. “Charge symmetry, quarks and mesons”. In: *Phys.Rept.* 194 (1990), pp. 1–116 (cit. on p. 114).
- [148] H L Lai et al. “The Strange parton distribution of the nucleon: Global analysis and applications”. In: *Journal of High Energy Physics* 0704 (2007), p. 089 (cit. on p. 115).
- [149] Robert Cahn et al. “Consequences of Internal Symmetries for Inclusive Processes”. In: *Physical Review D* 4.11 (1971), pp. 3337–3344 (cit. on p. 122).
- [150] H Lipkin et al. “Internal Symmetries and Model-Independent Relations for Inclusive Processes”. In: *Physical Review Letters* 28.13 (1972), pp. 862–865 (cit. on p. 122).
- [151] J Honerkamp et al. “Bounds on pion multiplicities and fluxes from isospin invariance”. In: *Nuclear Physics B* 38.2 (1972), pp. 565–572 (cit. on p. 122).
- [152] C Llewellyn-Smith et al. “Isospin Constraints on Semi-inclusive Neutrino Reactions and Their Hadronic Analogs”. In: *Physical Review Letters* 28.13 (1972), pp. 865–869 (cit. on p. 122).
- [153] BEBC WA59 Collaboration et al. “Production of π^0 mesons and charged hadrons in $\bar{\nu}$ neon and ν neon charged current interactions”. In: *The European Physical Journal C* 40.2 (1988), pp. 231–251 (cit. on p. 122).