UNIVERSITEIT GENT

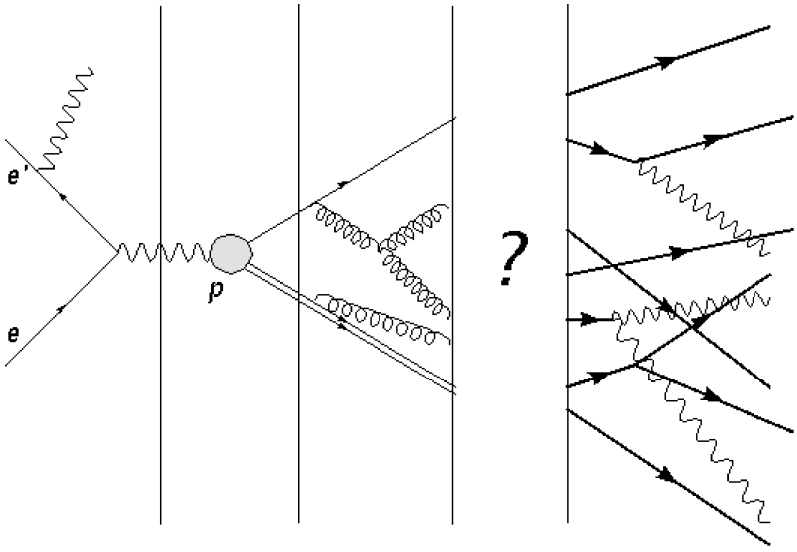
Vakgroep Subatomaire en Stralingsfysica
Academiejaar 2005-2006

Faculty of Sciences

Hadronization in electron - proton scattering at HERMES

Hadronizatie in elektron - proton verstrooiing in HERMES

Bino Maiheu



Promotor: Prof. Dr. Dirk Ryckbosch

Proefschrift ingediend tot het verkrijgen van de graad van
Doctor in de Wetenschappen : Natuurkunde

"There is a theory which states that if ever anybody discovers exactly what the Universe is for and why it is here, it will instantly disappear and be replaced by something even more bizarre and inexplicable. There is another theory which states that this has already happened."

Douglas Adams

Contents

Contents **i**

1 Introduction and motivation **1**
1.1 In the beginning, there was... 1
1.2 Probing the nucleon structure... 4

2 Semi-inclusive deep-inelastic scattering **7**
2.1 Deep-inelastic scattering 7
2.1.1 The DIS cross section 9
2.1.2 Interpretation in the Quark Parton Model (QPM) 11
2.2 Perturbative QCD extension of the Quark Parton Model 13
2.3 Semi-inclusive Deep Inelastic Scattering 15
2.3.1 Factorization 17
2.4 Aspects of hadronization 18
2.4.1 Introduction 19
2.4.2 The LUND symmetric string model 22
2.4.2.1 Symmetric fragmentation 23
2.4.2.2 Baryon production 25
2.4.3 Independent fragmentation 26
2.4.4 Cluster fragmentation 27
2.4.5 Parameterizations for $D_q^h(z, Q^2)$ fragmentation functions . . 28
2.4.6 About favored and unfavored fragmentation 30
2.5 Fragmentation in lepton-nucleon scattering 31

3 The HERMES experiment **35**
3.1 The DESY facility and the HERA storage ring 35
3.2 The target section 37
3.3 The luminosity monitor 38
3.4 The HERMES spectrometer 39
3.4.1 The HERMES dipole magnet 39
3.4.2 The tracking system 40

3.4.2.1 Detectors involved . . . 40
3.4.2.2 The tracking algorithm . . . 41
3.4.3 Particle identification and energy measurement . . . 42
3.4.3.1 The Transition Radiation Detector . . . 42
3.4.3.2 The Electromagnetic Calorimeter . . . 43
3.4.3.3 The Hodoscopes . . . 44
3.4.3.4 The PID algorithm . . . 45
3.4.3.5 The muon detection system . . . 47
3.4.3.6 The Ring Imaging Čerenkov Detector . . . 49
3.4.3.7 RICH particle identification . . . 52
3.5 Triggering . . . 52
3.6 Data handling at HERMES . . . 53
3.7 Calibration of the Hodoscopes . . . 54
3.7.1 Fitting procedure . . . 54
3.7.2 The calibration input . . . 58

4 Experimental charge and flavor separated hadron multiplicities 61
4.1 Data selection . . . 61
4.1.1 Data Quality Cuts . . . 61
4.1.2 Event and Track level Cuts . . . 64
4.1.3 Lepton Hadron Separation . . . 66
4.2 Hadron selection and PID correction . . . 67
4.2.1 Hadron kinematical cuts . . . 67
4.2.2 RICH unfolding . . . 67
4.2.2.1 About $\mathcal{P}$ and $\mathcal{Q}$ matrices . . . 67
4.2.2.2 The unfolding procedure . . . 70
4.2.2.3 Generation of the $\mathcal{P}$ matrix and systematic uncertainties involved. . . . 70
4.2.2.4 Propagation of the systematical uncertainties . . . 71
4.3 Experimental hadron multiplicity distributions . . . 73
4.4 Correction for pair creation $\gamma \rightarrow e^+ + e^-$ . . . 75

5 Monte Carlo Simulations 77
5.1 Introduction . . . 77
5.2 Monte Carlo in HERMES . . . 78
5.2.1 Overview . . . 78
5.2.2 The `gmc_disNG` generator . . . 81
5.2.2.1 QED radiative effects . . . 81
5.2.2.2 QCD effects . . . 83
5.2.2.3 The 'new' sea quark treatment and the nucleon remnant . . . 84
5.2.2.4 Intrinsic transverse momentum . . . 85
5.2.2.5 Tuning the fragmentation model . . . 85

5.2.2.6 Other parameterizations in the Monte Carlo 89
5.2.2.7 Lepton box cut 90
5.2.3 The gmc_pythia6 implementation 91
5.3 Discussion on target fragmentation relying on Monte Carlo 91

6 Corrected multiplicity distributions **97**
6.1 Contamination from diffractive processes 97
6.1.1 Extraction of the background fraction 98
6.1.2 Error propagation of the diffractive background correction 100
6.2 Unfolding the radiative and experimental effects 103
6.2.1 The formalism 104
6.2.2 Monte Carlo input to the correction procedure 107
6.2.2.1 Extraction of the migration matrix $n(i,j)$ 107
6.2.2.2 Smearing of DIS events 109
6.2.3 Unfolding in multiple dimensions 109
6.2.4 Treatment of the uncertainties involved 110
6.2.4.1 Propagation of uncertainties on the experimental input data 111
6.2.4.2 Intrinsic uncertainty due to limited Monte Carlo statistics 112
6.3 Q^2 scaling of the multiplicities 113
6.3.1 Uncertainty on the Q^2 correction 116
6.4 Multiplicity results and discussion 119
6.4.1 Multiplicities versus z 120
6.4.2 Comparison of z dependence to published data 123
6.4.3 The residual x dependence of the pion multiplicities 126

7 Influence of $\langle \cos\phi \rangle_{UU}$ moments on the acceptance correction **131**
7.1 Introduction, problem statement 131
7.1.1 Some theoretical considerations 131
7.1.2 Description of the azimuthal ϕ distribution in the Monte Carlo 134
7.2 Extraction of the $\cos\phi$ moments 135
7.2.1 The methods compared 136
7.2.2 Binning and extraction 141
7.3 Reweighting the Monte Carlo 142
7.4 Influence on the acceptance correction 144

8 Summary and discussion **149**

9 Nederlandstalige samenvatting **153**

Bibliography **161**

Acknowledgments **169**

A Calculation of statistical uncertainties on multiplicity distributions **i**

B Data tables for multiplicity analysis **iii**
B.1 Experimental Data Multiplicities iii
B.2 Diffractive background $\sigma^h_{diff}/\sigma^h_{tot}$ vi
B.3 Born level multiplicity distributions versus z without Q^2 correction vii
B.4 Q^2 scaling correction factors xii

C Data tables for $\langle\cos\phi\rangle_{UU}$ and $\langle\cos 2\phi\rangle_{UU}$ extraction **xix**

Chapter 1

Introduction and motivation

"Several billion trillion tons of super hot exploding hydrogen nuclei rose slowly above the horizon and managed to look small, cold and slightly damp."

Douglas Adams
The Hitchhiker's Guide to the Galaxy

1.1 In the beginning, there was...

Approximately 15 billion years ago something rather dramatic has happened. Suddenly, for no clear or apparent reason the perfect homogeneity of nothingness was brutally disrupted in a cataclysmic explosion. Enormous amounts of energy were unleashed from a tiny singularity. What caused or triggered this spectacular event will probably remain a mystery, but we do know what resulted from it. In fact, we see it all around us. Every living thing, every tree, every rock, every grain of sand on the beach once took part in this, what people somewhat derogatory call, "*Big Bang*". It is a fundamental part of human existence to stop and wonder "where it all comes from". What is the origin of matter, how is it constructed ? What, if any, are its most fundamental building blocks and what can we learn about their behavior? The events that took place all these billions of years ago and created the entire universe as we observe it today, are very much tied in with the search for the fundaments of matter. By far, one does not have to be a physicist to appreciate this fascinating endeavor. The answers and questions that arise from this quest for knowledge have a profound impact on how we see the world and maybe even each other. Let us take a small trip into the early universe and see in what

tiny corner this work can contribute to the gigantic puzzle laid out all around us.

The laws of physics as we know them cannot describe times smaller than the Planck scale ($\sim 10^{-43}$ s). We are therefore unable to say anything about the workings of the big bang itself. However what happened right after this moment is a most intriguing story. In the period from 10^{-43} to 10^{-34} s, the universe was extremely hot and dense, with a temperature over 1000 million million million million K. All forces, matter and energy that exist or are supposed to exist today, were indistinguishably united and squeezed together into a volume not larger than a minuscule subatomic particle.

At about 10^{-34} s later, a first phase transition occurred in the newborn universe. The gravitational force, before united with all other forces, separated at this time. The other ones, however, remained unified and therefore the **quarks** and **leptons** which occupied the early universe, could not be told apart by their interactions. This period is called the Grand Unified Epoch and is described by Grand Unified Theories (GUT). It is also during this time that presumably a small asymmetry of matter over antimatter caused the universe we observe today to consist of matter alone. As it got colder in the early universe, also the strong and the electroweak force separated. Now we are at 10^{-10} s after the big bang occurred. The strong force governs the interaction between the quarks and is mediated by gluons (g). Basically, it is the force which holds the nucleons together. The electroweak force is the unification of the weak and electromagnetic force. The latter one is exchanged by photons (γ), particles of light, and is known very well in our everyday lives. The weak interaction takes responsibility for radioactive decay and is able to convert different quark flavors into each other. This force is carried by particles called $W^{\pm}$ and Z^0 bosons. The collection of particles and force carriers described here, is known in present day physics as the **Standard Model** (see figure 1.1). There are 6 flavors of quarks, 3 with fractional electric charge of $+\frac{2}{3}$, namely **up** (u), **charm** (c) and **top** (t) and 3 with electromagnetic charge $-\frac{1}{3}$: **down** (d), **strange** (s) and **bottom** (b). Their masses range from a couple of MeV for the lightest quarks (u, d) to about 175 GeV for the heaviest (t). As shown in figure 1.1 they are organized into 3 families. Corresponding with these quark families, there are also 3 generations of leptons, each consisting of a negatively charged particle accompanied by a neutral, weak interacting neutrino. The most prominent member of the leptons of course is the electron. Its heavier counterparts are called the muon (μ^-) and the tau-particle (τ^-). To this day these 12 particles together with their anti-particles are assumed to be the fundamental building blocks of all matter known to mankind.

A little later, at around 10^{-7} s, the universe was cool enough not to be able to sustain a state where quarks and gluons move freely and unbound. This stage is known in astro-particle physics as the **baryogenesis**, quarks and gluons lock themselves together into baryons, the most important examples of which are of course the proton and the neutron. From this point onwards the universe will be dominated by matter whereas before it was radiation that flooded its confinements. At the timescale of 1 second almost all anti-matter

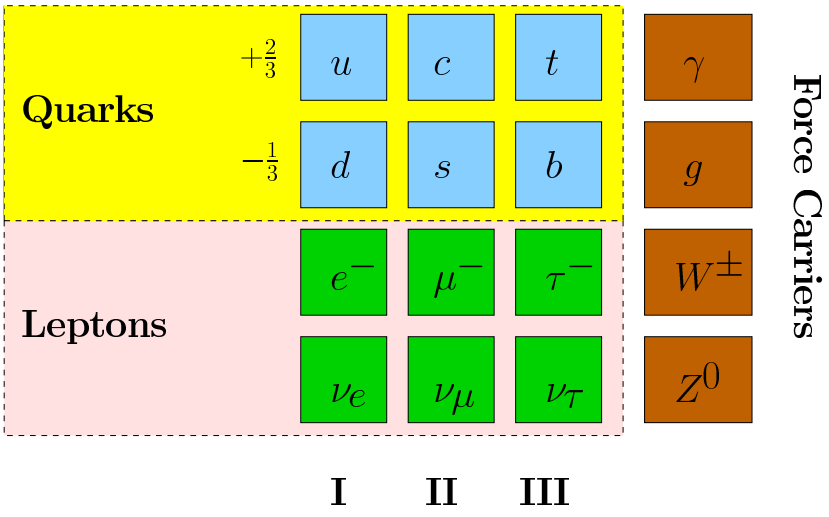


Figure 1.1: The Standard Model of particle physics

was supposed to have annihilated thus leaving an excess of electrons as well. At these early stages of the universe the freshly formed matter was subject to constant radiation bombardments; however, as it further expanded and cooled down, this radiation subsided. Today its traces are still detectable as the cosmic microwave background. About 300.000 years after the big bang, electrons, protons and neutrons were able to combine and form hydrogen and helium atoms and rhe universe became transparent. After this stage the evolution of the universe gets pretty dull[1]. Gravity pops up into the picture again and over the course of millions of years, atoms stick together to form dust clouds, nebulae and proto-galaxies which eventually lead to the formation of for instance our very own milky way, the solar system, the earth and even our very own bodies.

Why tell this fascinating story? Obviously, physicists cannot travel back in time to study first hand the events that took place at the beginning of our universe. So they have found a way to somehow reproduce on a much lesser scale here on earth the conditions that were in effect in the universe during its first fractions of a second. In giant accelerator machines, elementary particles are sped up to enormous energies and are smashed into one another. By increasing the energy of these collisions physicists can study different timescales of the early universe and at the same time penetrate deeper into the structure of matter. Giant particle accelerator facilities are built all over the world to undertake this endeavor. Some of the most well known include of course the HERA facility in DESY, Hamburg where this work is situated, CERN in Geneva, SLAC in Stanford, California, KEK in Japan and so on...

The phase of the early universe where quarks and gluons confine themselves into baryons, is in fact at the very core of this work. As mentioned above, quarks interact with one another through the strong interaction, mediated by gluons. Due to the very peculiar nature of this interaction, quarks –

[1]At least from a particle physicists point of view ;-)

unlike electrons and neutrinos – cannot exist as free particles, unless the energy scale is very high as in the early universe. Nowadays they always come in triplets (baryons) or doublets of a quark and an antiquark (mesons) bound together by gluons. All matter that consists out of quarks is labeled as **hadronic matter.** Baryons and mesons are the two subclasses of the hadrons.

In high energy collisions these 'early universe' conditions are recreated. At the HERA collider for instance, high energy electrons impact on oppositely moving high energy protons. At the moment of such a collision the energy is so high that the quarks inside the proton do behave themselves as if they are free particles. However as the impact of the initial collision fades away the quarks which are kicked out somehow have to '**rehadronize**' since they cannot freely continue on their way. This hadronization or **fragmentation** process is exactly what has happened on a much larger scale during the *baryogenesis* we talked about above.

Obviously this hadronization process has been the subject of many theoretical and experimental studies. Especially in e^+e^- colliders, where an electron and a positron are annihilated to form a quark and an antiquark moving at high energy in opposite directions. These experiments, however, typically took place at very high energies. This work will try to contribute something in the sense that it will study the hadronization process at a somewhat more moderate energy. The main question which arises then is whether or not the framework developed for this process is still valid. But before we come to that, let us maybe shed a little more light upon the experimental setup at the HERA facility where this study was undertaken.

1.2 Probing the nucleon structure...

In fact the quarks mentioned above were only discovered in the early 1970s. The theory describing their dynamics is called quantum chromodynamics or QCD. As quarks are locked up inside protons (uud) and neutrons (udd), QCD in fact tells the story about the inner structure of the nucleons. Experimentally we can investigate the internal workings of the proton by scattering off its constituents with a high energy probe. Much like with common microscopes, the higher the energy of the probe the higher the resolution, so the smaller the details that will be visible. However, because of the self-coupling of gluons[2] a perturbative approach, as is common in most field theories, is only possible at high energy where one can treat the quarks as quasi-free particles. If the energy is too low, the coupling constant α_S for the strong interaction rises and we observe the quarks no longer as free, but confined inside the baryon.

One of the most elusive properties of the proton is its intrinsic spin of $\frac{1}{2}\hbar$. Ever since the discovery of quarks and gluons, we know that the proton is not an elementary particle. Therefore the questions arises how its building blocks make up its spin. The first experiments were carried out at SLAC

[2]Gluons do carry strong (color) charge as opposed to photons, which are electrically neutral.

[1] with subsequent and more precise measurements at CERN with the EMC experiment [2], where the contributions of the 3 valence quarks spins to the nucleon spin were measured. Surprisingly it was found that by far these did not account for the total spin of $\frac{1}{2}\hbar$, initiating what was called the "**Spin Crisis**". Now we know that we have to take into account the gluons and the sea quarks as well, not to forget their orbital angular momentum. The total proton spin puzzle can be written as

$$\frac{1}{2} = \frac{1}{2}\Delta\Sigma + \Delta G + L_q + L_g, \tag{1.1}$$

where $\Delta\Sigma$ is the total contribution of the spin of the valence and sea quarks, ΔG the contribution of the spin of the gluons and L_q, L_g the orbital angular momenta of respectively quarks and gluons.

In 1995, a second generation experiment called HERMES[3] took its first data, with as primary goal further investigation of this puzzle. It uses a polarized target in combination with the polarized 27.5 GeV lepton beam at the HERA facility in DESY, Hamburg. From the asymmetries between the scattering cross sections for different relative spin orientations of beam and target, it was able to measure the total spin structure function of the proton to great precision [3]. The individual quark contributions were measured from the semi-inclusive data set [4]. However, next to the proton spin structure, a wide variety of physics topics are studied at the HERMES experiment. The fragmentation process is one of these. As HERMES uses a fixed target the center of mass energy is only about 7 GeV. It provides us therefore with a unique possibility to study hadronization right on the edge of the region where perturbative QCD is well established. Do all of the models and parameterizations still hold? This is a particularly important question as quite some of the analyses performed in HERMES rely on a good understanding of the fragmentation process at the energy scale considered.

At first we will outline some theoretical aspects of the hard scattering of leptons (e^+) off quarks inside the nucleon. This process is called **deep-inelastic scattering** (chapter 2). Obviously we will have to define a number of quantities which will help us to better understand and describe the nucleon as well as the production of hadrons through fragmentation. Next (chapter 3) the experimental setup will be explained briefly. Special attention will be given to detectors which enable us to distinguish between different hadron flavors and an algorithm which enhances the precision of this determination. This is really one feature which makes HERMES unique. As quite some attention during the course of this Ph.D. went to maintaining and calibrating of the trigger detectors, section 3.7 will spend some words on that. We will proceed in chapter 4 with the extraction of experimental multiplicities, which are quantities that can learn us a great deal about the fragmentation process. Obviously these experimental quantities are subject to any number of inefficiencies and instrumental effects and will therefore be in need of correction. These will be handled in chapter 6, but before that we will have to describe the used Monte

[3]HERMES stands for HERa MEasurement of Spin

Carlo simulations which will enable us to do these corrections (chapter 5). In chapter 7 we will extensively deal with one special non-trivial systematic effect which was conjectured to have an impact on the analysis. A brief summary will conclude this work.

Chapter 2

Semi-inclusive deep-inelastic scattering

Before we are able to present any results in this thesis, we must understand the underlying process of deep inelastic scattering. We will start by introducing a number of kinematical variables and physical quantities that enable us to probe and describe matter at its most microscopic level. Note that in what follows we will use natural units, so $\hbar = 1$ and $c = 1$. This implicitly means that masses, energies as well as momenta will be given in GeV. Although positrons were used to probe the target nucleons, we will use the terms 'electron', 'positron' or even 'lepton' without distinction unless explicitely mentioned.

2.1 Deep-inelastic scattering

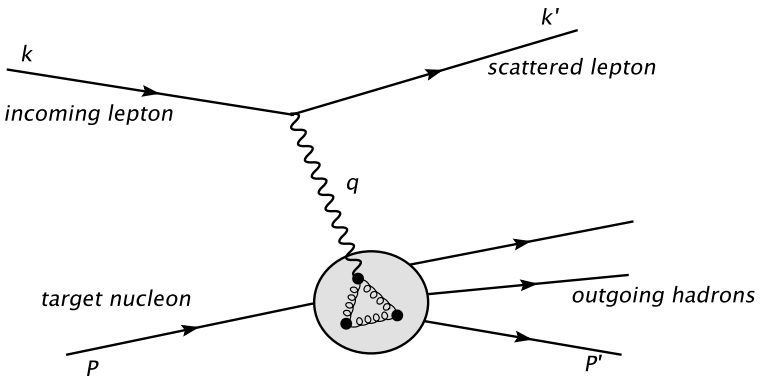


Figure 2.1: Schematic representation of the deep inelastic scattering process. The symbols in the figure denote the fourmomentum vectors of the interacting particles.

In lowest order lepton-nucleon scattering, a target nucleon is probed by an neutral electroweak (γ, Z^0) boson with a resolution of Q^2 where

$$Q^2 \equiv -q^2, \qquad \text{and} \quad q = k - k' \tag{2.1}$$

with q the fourmomentum of the probe. At high enough energy the resolution of the probe is so high that it is able to resolve substructures inside the nucleon. In order to understand this process of **deep-inelastic scattering** (DIS) we have to introduce a number of kinematical variables. A right-handed reference system was chosen having the z axis along the lepton beam direction and the y axis pointing upwards. Assume that the incoming lepton[1] is incident along the z axis with an energy E, so $k = (E, 0, 0, E)$. The lepton scatters under an angle θ with the z axis and has fourmomentum vector $k' = (E', \vec{k'})$. In the HERMES experiment, we work with a fixed target, so in our laboratory frame the expressions will by greatly simplified assuming the target nucleon has $P = (M, 0, 0, 0)$, with M its mass. As already stated, the momentum of the exchanged boson $q = k - k'$, so that in the lab frame

$$Q^2 \approx 4EE' \sin^2 \frac{\theta}{2} \tag{2.2}$$

Furthermore we define the energy transfer ν from the incoming lepton to the target nucleon

$$\nu \equiv \frac{Pq}{M} \underbrace{= E - E'}_{lab} \tag{2.3}$$

and the total invariant mass W of the hadronic final state P'

$$W^2 = P'^2 = (P + q)^2 = M^2 + 2M\nu - Q^2 \tag{2.4}$$

From this definition of W^2 we can construct a criterion for **elastic** scattering. In this case the nucleon remains in its ground state after the interaction with the electron and we have, $W = M$ and (see equation 2.4) :

$$2M\nu - Q^2 = 0 \tag{2.5}$$

For inelastic scattering we will have $W > M$ resulting in a second degree of freedom as well. Let us introduce a parameter which measures the **inelasticity** of the process namely

$$x \equiv \frac{Q^2}{2M\nu} \tag{2.6}$$

which will be 1 for elastic reactions and $0 < x < 1$ for inelastic. This scaling parameter is called the **Bjorken** x parameter. Together with this parameter x another scaling parameter is usually introduced, namely y, defined as

$$y \equiv \frac{Pq}{Pk} \underbrace{= \frac{\nu}{E}}_{lab} \tag{2.7}$$

We can also calculate the center of mass energy s in the scattering reaction as

$$s = (k + P)^2 \underbrace{= 2ME - M^2}_{lab} \tag{2.8}$$

[1] In our case this will be a positron coming from the HERA beam. However there are also DIS experiments that use muons, like e.g. the EMC experiment at CERN.

so that for a fixed proton target and a HERA beam energy of 27.57 GeV we obtain $\sqrt{s} \approx 7.1$ GeV which is well below the invariant mass for the Z^0 boson[2] so for HERMES its contribution will be negligible.

2.1.1 The DIS cross section

If one calculates (see e.g. [5]) the cross section for the process depicted in figure 2.1 one traditionally introduces two tensor objects: $L_{\mu\nu}$ to describe the leptonic interaction part $l \rightarrow \gamma^* + l'$ and another one to describe the hadronic part, $W_{\mu\nu}$ for $P + \gamma^* \rightarrow P'$. The expression obtained for the differential cross section for finding the scattered lepton inside the solid angle $d\Omega$ and with energy between E' and $E' + dE'$ is :

$$\frac{d^2\sigma}{dE'd\Omega} = \frac{\alpha^2}{MQ^4}\frac{E'}{E}L^S_{\mu\nu}W^{S,\mu\nu} \tag{2.9}$$

where α is the electromagnetic coupling constant ($\approx 1/137$). As in classical Rutherford scattering, a typical Q^{-4} dependence is obtained here. In expression 2.9 we explicitely summed over the spin degrees of freedom and hence are left with symmetric quantities $L^S_{\mu\nu}$ and $W^{S,\mu\nu}$. Within QED the tensor $L^S_{\mu\nu}$ can be exactly calculated as

$$L^S_{\mu\nu} = 2[k_\mu k'_\nu + k_\nu k'_\mu - g_{\mu\nu}k \cdot k']. \tag{2.10}$$

The hadronic tensor, however, cannot be exactly calculated and one has to introduce two structure functions F_1 and F_2, such that

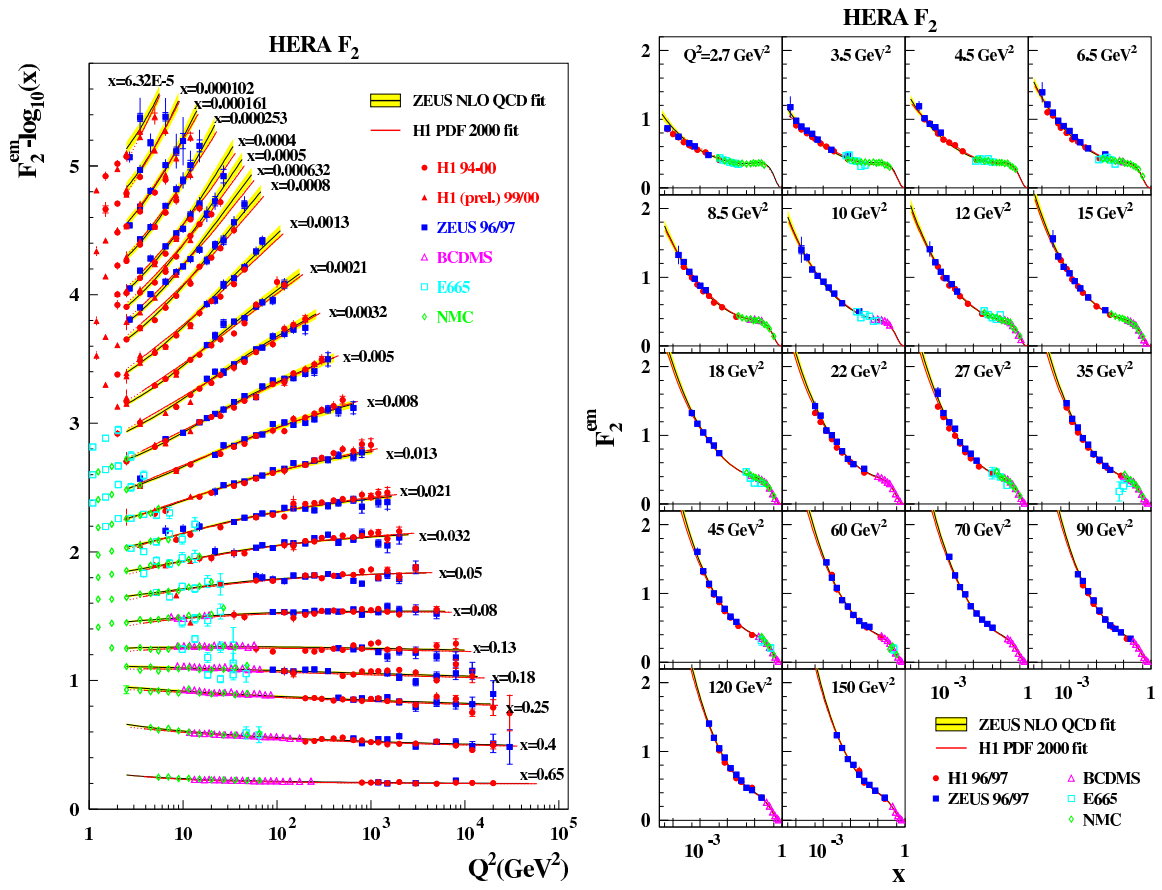
$$W^S_{\mu\nu} = F_1 \cdot \left(\frac{q_\mu q_\nu}{q^2} - g_{\mu\nu}\right) + F_2 \cdot \frac{1}{P \cdot q}\left(P_\mu - \frac{P \cdot q\, q_\mu}{q^2}\right)\left(P_\nu - \frac{P \cdot q\, q_\nu}{q^2}\right) \tag{2.11}$$

These two structure functions F_1 and F_2 contain all the information about the spin independent structure of the nucleon. Note that in the general case, where one does not sum over the spin states we obtain two extra structure functions g_1 and g_2 describing the spin structure of the nucleon, one of HERMES' main endeavors. Also, should we not have restricted ourselves to one-photon exchange and include contributions of Z^0 exchange, one obtains 4 extra structure functions : one unpolarized F_3 and three polarized g_3, g_4 and g_5 [6].

Experimentally, it was observed that these structure functions depend mainly on the Bjorken scaling variable x and only weakly or not at all on Q^2. This is the so-called **scaling behavior** of the structure functions. They are also subject to the *Callan-Gross* relation [7] which states that

$$2xF_1(x) = F_2(x). \tag{2.12}$$

Therefore one usually only talks about the $F_2(x, Q^2)$ structure function of the nucleon. Experimentally, the F_2 structure function is well known from the H1 and ZEUS data at HERA. Figure 2.2 shows the world data on $F_2(x, Q^2)$ for the



(a) versus Q^2 for different x (b) versus x for different Q^2

Figure 2.2: World data on $F_2(x, Q^2)$ from H1, ZEUS, NMC, E665 and the BCDMS collaborations. Figure is publicly available from the ZEUS web site (`http://www-zeus.desy.de`).

proton together with a NLO QCD fit provided by the ZEUS collaboration. Also a number of parameterizations are available like the ALLM97 parameterization [8], which manages to describe the F_2 structure function well over a kinematic range of $3 \cdot 10^{-6} < x < 0.85$ and $0 \leq Q^2 < 5000$ GeV^2 [9].

In summary, we can write the final expression for the DIS cross section (equation 2.9) as

$$\frac{d^2\sigma}{dxdQ^2} = \frac{4\pi\alpha^2}{Q^4}\left[\left(1 - y - \frac{Mxy}{2E}\right)\frac{F_2(x,Q^2)}{x} + y^2 F_1(x,Q^2)\right] \tag{2.13}$$

The earliest DIS experiments were carried out at SLAC. In scattering of 7 to 17 GeV electrons off protons the observed cross section was normalized to the well known Mott cross section. It was seen that for increasing W the Q^2 dependence seemed to flatten out. In particular above the resonance region ($W > 2$ GeV) the Q^2 dependence was much weaker than what would have been expected from elastic scattering or Δ-resonance excitation of the proton [10]. Together with the experimental confirmation of the Callan-Gross relation, this led in the beginning of the seventies to the development of the so-called Quark

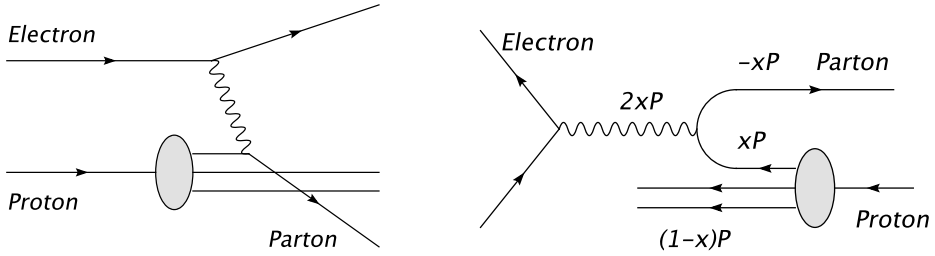
[2] at 91.2 GeV

Parton Model (QPM) by R. Feynman [11], J.D. Bjorken [12], [13]. This model was based on the quark hypothesis introduced in the mid-sixties by M. Gell-Mann [14], which states that a nucleon is made up out of tiny point like spin $\frac{1}{2}$ particles, quarks, glued together by the gauge bosons of the underlying quantum field theory of **quantum chromodynamics** (QCD). These gauge bosons were called **gluons**.

We will see in the next section that within this QPM model the $F_2(x, Q^2)$ structure function has a very straightforward interpretation.

2.1.2 Interpretation in the Quark Parton Model (QPM)

As the physics should be independent of the frame of reference, we will choose a specific frame which allows us to handle the equations more easily. More specifically, we choose a frame where the transverse momentum of the point-like constituents can be neglected and the nucleon has very high momentum. A very straightforward interpretation for the Bjorken x variable emerges. Namely, it represents the fraction x of the total nucleon momentum P carried by the struck **parton**. These so-called partons here are in fact the point-like constituents seen earlier and encompass the different quarks and gluons as introduced by Gell-Mann. A frame in which these conditions are met is the so-called *Breit-frame* as shown in figure 2.3. In the QPM interpretation we can



(a) Laboratory Frame (b) Breit Frame

Figure 2.3: Schematic representation of DIS in the lab frame and in the Breit frame where the transferred energy ($q_0 = 0$) by the virtual photon is 0 and the momentum of the struck quark just changes sign, but remains the same in magnitude after absorption of the virtual photon.

write the $F_2(x, Q^2)$ structure function as

$$F_2(x, Q^2) = x \cdot \sum_f e_f^2 \cdot q_f(x, Q^2) \tag{2.14}$$

where e_f is the fractional charge carried by the considered quark flavor and $q_f(x, Q^2)$ is the so-called **parton density function** (pdf) which is in fact an expectation value for the number of quarks of type f to be found in the nucleon with a momentum fraction between x and $x + dx$. The sum in equation 2.14 runs over the quark flavors which are relevant for the particular energy at which we look to the nucleon and over the respective antiquarks as well. For HERMES the relevant flavors are **up** (u), **down** (d) and **strange** (s). The u quark

has a fractional charge of $+\frac{2}{3}$, the d and s quarks $-\frac{1}{3}$. Therefore we can write the structure function in terms of the individual quark contributions. We will denote e.g. the u-quark distribution $q_u(x)$ simply with $u(x)$. For the proton and the neutron we have in the most extensive form :

$$\frac{1}{x}\cdot F_2^p = \left[\frac{4}{9}\cdot(u_v^p+u_s+\bar{u}_s)+\frac{1}{9}\cdot(d_v^p+d_s+\bar{d}_s)+\frac{1}{9}\cdot(s_s+\bar{s}_s)\right] \quad (2.15)$$

$$\frac{1}{x}\cdot F_2^n = \left[\frac{4}{9}\cdot(u_v^n+u_s+\bar{u}_s)+\frac{1}{9}\cdot(d_v^n+d_s+\bar{d}_s)+\frac{1}{9}\cdot(s_s+\bar{s}_s)\right] \quad (2.16)$$

Where the subindex v denotes the **valence quark** distributions. These quarks determine the '*macroscopic*' quantum numbers of the proton. The subindex s refers to the **sea quark** distributions. These quarks are spontaneously created and annihilated in the strong color field which binds the valence quarks. The proton and the neutron are partners in an isospin doublet ($I = 1/2$). Therefore their quark distributions are subject to a number of symmetry relations

$$u_v^p(x) = d_v^n(x) \quad (2.17)$$

$$d_v^p(x) = u_v^n(x) \quad (2.18)$$

$$u_s^p(x) = d_s^p(x) = d_s^n(x) = u_s^n(x) \quad (2.19)$$

Since sea quarks are always created in quark - antiquark pairs of the same flavor we can equate $u_s(x) = \bar{u}(x)$ and analogous for the other quark flavors. Therefore we can write a number of sum rules simply in terms of the total quark and anti-quark distributions. For the proton :

$$\int_0^1 dx\,[u(x)-\bar{u}(x)] = \int_0^1 dx\,u_v(x) = 2 \quad (2.20)$$

$$\int_0^1 dx\,\left[d(x)-\bar{d}(x)\right] = \int_0^1 dx\,d_v(x) = 1 \quad (2.21)$$

$$\int_0^1 dx\,[s(x)-\bar{s}(x)] = 0 \quad (2.22)$$

Recent data from di-muon production in ν_μ deep inelastic scattering provided by the NuTeV collaboration shows that the strange quark density $s(x)$ is not necessarily equal to the anti-strange $\bar{s}(x)$. There are signs for a very small negative asymmetry in the strange sea [15]. Although discussion is still ongoing [16].

As far as the light sea quarks are concerned we know that flavor symmetry is not conserved there and $\bar{u} < \bar{d}$. This behavior is known as the **Gottfried Sum Rule** (GSR) violation [17]. If one subtracts F_2^n from F_2^d, one obtains the **Gottfried integral**

$$S_G = \int_0^1 \frac{1}{x}\left[F_2^p(x)-F_2^n(x)\right]dx = \frac{1}{3}(2-1)+\frac{2}{3}\int_0^1\left[\bar{u}(x)-\bar{d}(x)\right]dx \quad (2.23)$$

$$= \frac{1}{3} \qquad \text{only if} \quad \bar{u}(x)=\bar{d}(x) \quad (2.24)$$

Experimental verification has been summarized very nicely in [18] and it is shown that the GSR is violated with 4σ significance. The experimental value obtained by the NMC Collaboration is $S_G = 0.235 \pm 0.026$ [18]. This GSR violation is still a hot topic to this date.

2.2 Perturbative QCD extension of the Quark Parton Model

If one sums the individual quark contributions one obtains only half of the total nucleon momentum, thus about 50 % of the nucleon's momentum must be carried by the gluons [19]. This leads to the following momentum sum rule for the nucleon

$$\int_0^1 dx\, x \left[g(x) + \sum_f e_f^2 q_f(x) \right] = 1 \tag{2.25}$$

The gluons are actually responsible for the scaling violations in the structure function. In section 2.1.1 we mentioned the fact that the F_2 structure function depends basically only on x. However, this scaling is not completely fulfilled and, especially at low x, a residual Q^2 dependence can clearly be seen from figure 2.2.

This scaling violation was explained in QCD by the existence of the gluons, as being the quantum field gauge bosons of the strong force. As opposed to the gauge bosons of the electromagnetic interaction (photons), gluons do carry (color-)charge and are therefore able to interact with the quarks **as well as** with each other. This results in the non-Abelian character of the theory. The four basic gluon interactions are depicted in figure 2.4. The strong coupling

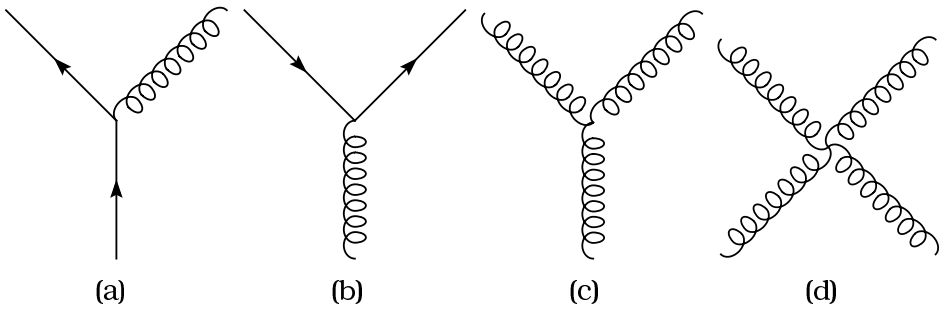


Figure 2.4: The four basic gluon interactions. a) gluon radiation by a quark, b) splitting of a gluon in a quark anti-quark pair. c) splitting of a gluon in two gluons and d) a gluon four-vertex.

constant α_S is given up to leading order by

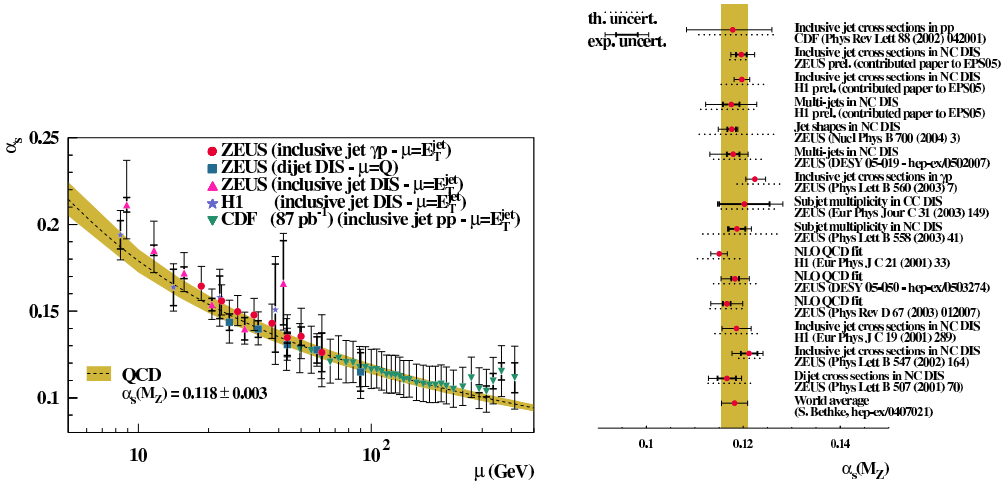
$$\alpha_S(Q^2) = \frac{12\pi}{(33 - 2N_f)\ln(Q^2/\Lambda^2)} \tag{2.26}$$

where N_f is the number of involved quark flavors[3] and Λ ($\approx$ 200 MeV) the

[3]depending on the energy 3 to 6

scale where perturbative QCD can be applied. It can be seen that α_S will rise as Q^2 decreases. This property shows the tendency towards **confinement** and is essential to understand the fragmentation process of quarks into hadrons which will be dealt with later on. At low energy we find the quarks bound very strongly together into a color singlet. If we increase Q^2, however, we see that the coupling constant decreases and approaches 0 as $Q^2 \to \infty$. Quarks behave as if they move freely and unbound. This property is called **asymptotic freedom**. (A discovery for which the Nobel prize was awarded to David J. Gross, H. David Politzer and Frank Wilczek in 2004.)

The variation itself of the strong coupling constant with energy scale has been confirmed experimentally e.g. by making ratios of multi-jet events in high energy collider experiments like ZEUS and H1. Figure 2.5 summarizes the ZEUS, CDF and H1 data about α_S.



(a) The running Coupling Constant

(b) α_S at the Z^0 mass

Figure 2.5: Experimental measurements from ZEUS, CDF and H1 for the strong coupling constant and it's scaling with energy scale ($\mu \equiv Q^2$ for DIS). On the right hand side a selection of world data for α_S at the Z^0 mass is displayed. Figures are taken from the ZEUS web site (`www-zeus.desy.de`).

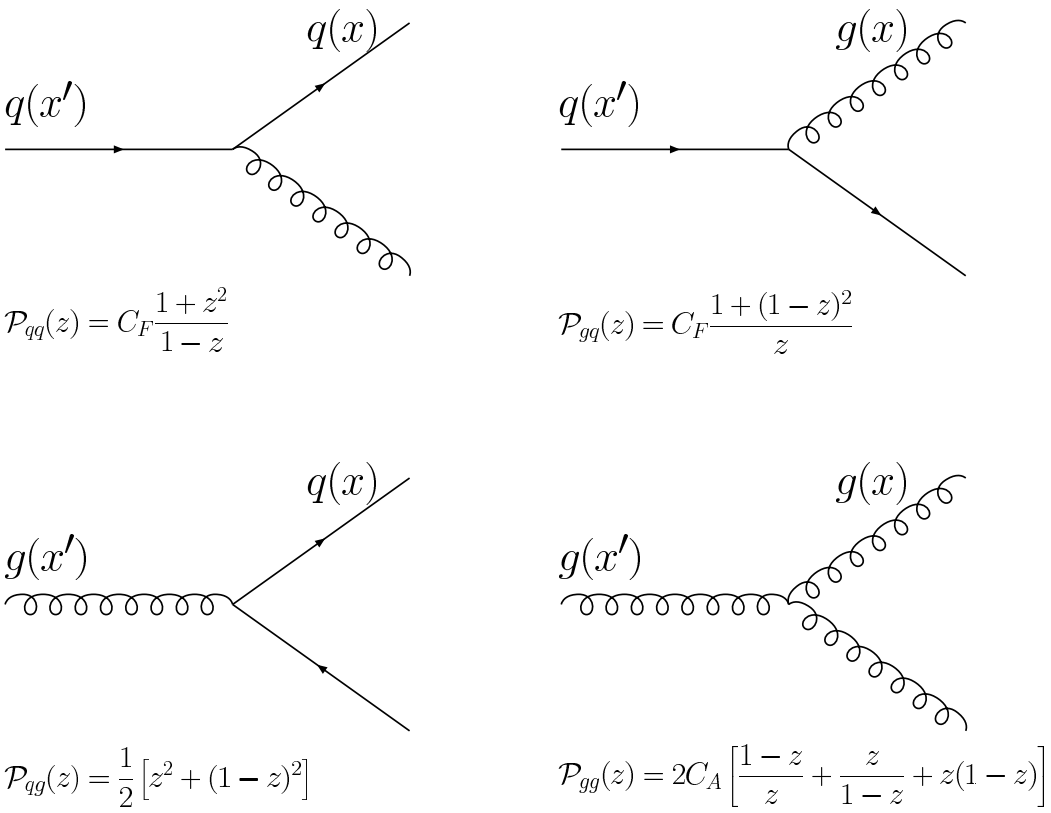
The variation or **evolution** of the parton density functions is understood in QCD by means of the Dokshitzer-Gribov-Lipatov-Altarelli-Parisi (DGLAP) equations [20],[21], [22] and [23].

$$\frac{\partial q_f(x,Q^2)}{\partial \ln Q^2} = \frac{\alpha_S(Q^2)}{2\pi}\int_x^1 \frac{dx'}{x'}\left[q_f(x',Q^2)\cdot \mathcal{P}_{qq}\left(\frac{x}{x'}\right) + g(x',Q^2)\cdot \mathcal{P}_{qg}\left(\frac{x}{x'}\right)\right] \quad (2.27)$$

$$\frac{\partial g(x,Q^2)}{\partial \ln Q^2} = \frac{\alpha_S(Q^2)}{2\pi}\int_x^1 \frac{dx'}{x'}\left[g(x',Q^2)\cdot \mathcal{P}_{gg}\left(\frac{x}{x'}\right) + \sum_f q_f(x',Q^2)\cdot \mathcal{P}_{qg}\left(\frac{x}{x'}\right)\right] \quad (2.28)$$

The **splitting functions** $\mathcal{P}_{ij}\left(\frac{x}{x'}\right)$ give the probability that a parton i with momentum fraction x was radiated from a parton j with momentum fraction x'

[24]. We will denote the fraction x/x' as z in the rest of this paragraph. It is important to be aware that these equations can describe how the quark and gluon distributions vary with Q^2, but they still need the shape of an initial parameterization at an input scale Q_0^2 to start from. Hence they do not predict the shape of the $q_f(x, Q^2)$ and $g(x, Q^2)$ ab initio, just their evolution with varying Q^2. Figure 2.6 shows the four splitting functions with their corresponding



$$\mathcal{P}_{qq}(z) = C_F \frac{1+z^2}{1-z} \qquad \mathcal{P}_{gq}(z) = C_F \frac{1+(1-z)^2}{z}$$

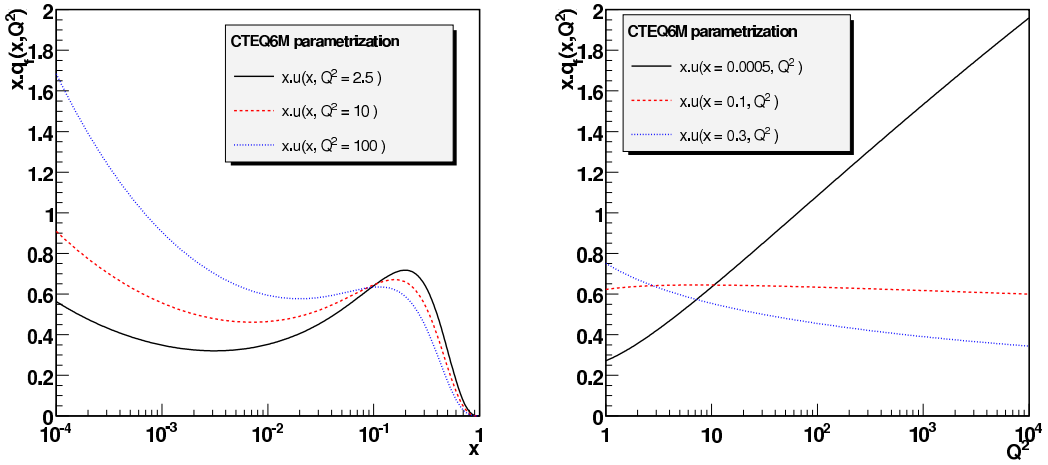
$$\mathcal{P}_{qg}(z) = \frac{1}{2}\left[z^2 + (1-z)^2\right] \qquad \mathcal{P}_{gg}(z) = 2C_A \left[\frac{1-z}{z} + \frac{z}{1-z} + z(1-z)\right]$$

Figure 2.6: The 4 splitting functions [24]. C_F and C_A are group-theoretical constants that follow from the underlying $SU(N)_{col}$ symmetry structure of QCD. As we know there are 3 colors, we have $C_A = N = 3$ and $C_F = \frac{N^2-1}{2N} = \frac{4}{3}$. Note that the equations are only valid for $z < 1$.

diagrams. Note the obvious symmetry relations $z \leftrightarrow (1-z)$ in the splitting functions $\mathcal{P}_{qg}$ and $\mathcal{P}_{gg}$ and the fact that $\mathcal{P}_{qq}(1-z) = \mathcal{P}_{gq}(z)$, which can be seen from the diagrams. In figure 2.7 an example of the Q^2 evolution was displayed using the CTEQ6 parameterization for the $u(x, Q^2)$ distribution. From 2.7(b) one can clearly see that the scaling violation is much stronger at low x as expected.

2.3 Semi-inclusive Deep Inelastic Scattering

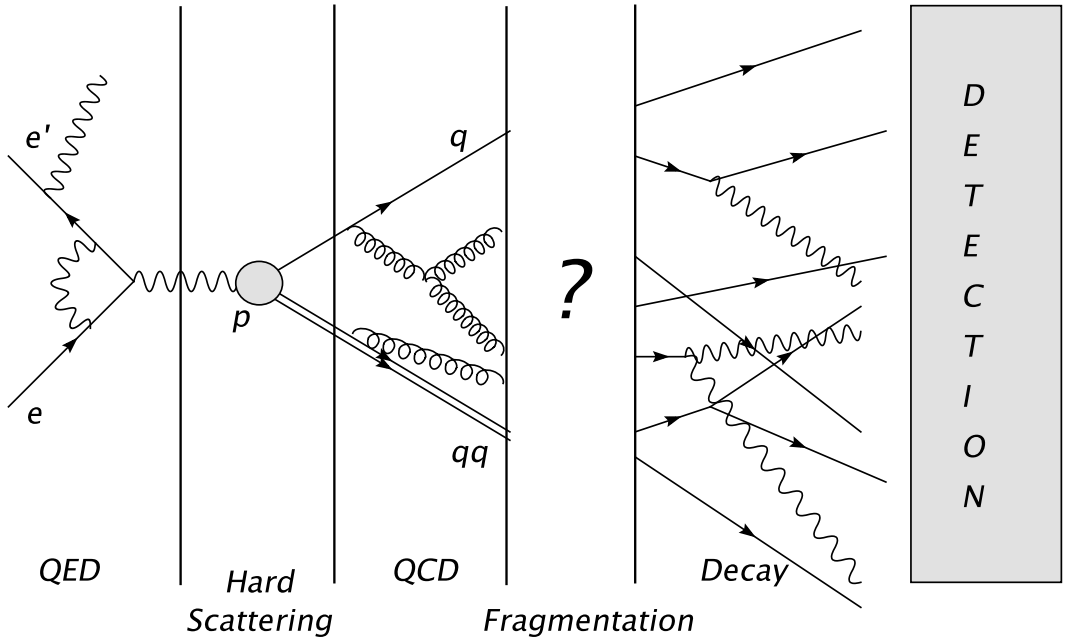
So far only the cross section integrated over all possible hadronic final states was considered, so we only regarded **inclusive scattering**. Let us now look into some more detail to the hadrons which are produced in the deep inelastic scattering reaction. The energy at which the scattering interaction occurs



(a) versus x (b) versus Q^2

Figure 2.7: The u-quark pdf versus x for three different Q^2-values on the left hand side. On the right hand side 3 x-values were chosen and the Q^2 dependence was plotted. The parameterization used was CTEQ6M.

is much higher than the force that holds the partons together, so one would expect them to be 'kicked out' of the nucleon. However, the principle of *confinement* dictates that quarks cannot exist as free particles in nature **(long distance behavior)**. They are always bound together into color neutral objects: baryons consisting of three differently colored quarks ($q_{c_1}q_{c_2}q_{c_3}$) or mesons, being a bound state of a quark and an antiquark ($q_c\bar{q}_{\bar{c}}$). So an obvious problem poses itself, which is in fact at the very hearth of this work. **How do the final hadrons that we detect emerge from the deep inelastic scattering interaction that involves just quarks and gluons** ? This process is called **fragmentation** or **hadronization** and cannot be tackled using perturbative QCD (pQCD) as the strong coupling constant α_S becomes too large at low energy (W), which is exactly where hadronization occurs. As a result the series expansion typical for any perturbative treatment converges extremely slow or not at all, which makes practical calculations impossible. The figure below gives a somewhat more graphical overview.



A lepton scatters off the proton by exchange of virtual photon. The kinematics of the scattered lepton, including radiative effects, can be completely derived from Quantum ElectroDynamics (QED). Inside the proton a quark is struck which will get kicked out of the proton and a color field is formed between this struck quark and the proton remnant. When the energy of the propagators is still high enough, we can still describe things in perturbative QCD. The partons are moving as quasi-free particles then. However, at the moment where the quarks 'freeze out' and form hadrons, the energy is too low and perturbative QCD does not work anymore. So we are unable to describe or in fact calculate the transition to hadrons from first principles. This is the fragmentation or hadronization region. Once the primary hadrons are formed, they will propagate and decay subsequently if they are unstable according to their own life expectancy. These decay products are what we observe in detectors.

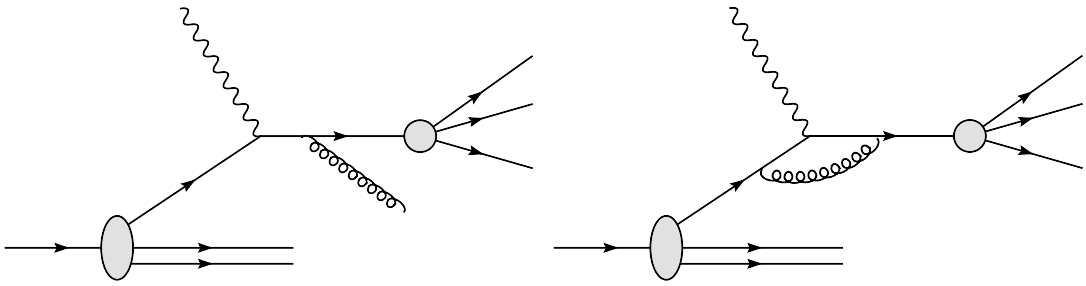
2.3.1 Factorization

We know that at short distances quarks behave as they were free particles (*asymptotic freedom*). A general cross section[4] combines the long (confinement) and short distance behavior. Factorization theorems [25], [26], [27] will try to separate (factorize) the two parts from each other in a systematic way. The factorization theorem for hadron production in semi-inclusive DIS is contained in the following expression for the hadron (h) production cross section σ^h:

$$\frac{d^3\sigma^h}{dxdQ^2dz} = \frac{\sum_f e_f^2 q_f(x,Q^2)\cdot D_f^h(z,Q^2)}{\sum_f e_f^2 q_f(x,Q^2)} \cdot \frac{d^2\sigma^{DIS}}{dxdQ^2} \tag{2.29}$$

where we can see that hadron production is given by the $D_f^h(z,Q^2)$ functions

[4] where also hadrons explicitely are part of the final state considered



(a) Scaling violation (b) Factorization breaking

Figure 2.8: Gluon radiation of the final quark line inducing scaling violations and a Q^2 dependence of the fragmentation functions on the left hand side and gluon exchange between the initial and final quark lines as an example of a QCD effect that could cause factorization breaking on the right hand side.

called **fragmentation functions**. The z scaling variable gives the longitudinal momentum fraction carried away by the produced hadron h and is defined as

$$z = \frac{E_h}{\nu}. \tag{2.30}$$

The fragmentation functions $D_f^h(z, Q^2)$ factorize from the parton distribution functions $q_f(x, Q^2)$ and depend mostly on the z variable. Note that for now we have integrated over all hadron transverse degrees of freedom such as transverse hadron momentum P_T and hadron azimuthal scattering angle ϕ. QCD effects, however, do induce scaling violations, e.g. by gluon emission as can be seen from figure 2.8(a) and thus a dependence on Q^2. It must be said as well that factorization is only approximate. Gluon exchange e.g. between the initial state and final quark lines can introduce a dependence on the Bjorken x-variable in the fragmentation functions (figure 2.8(b)), as observed experimentally [28], [29]. So QCD in fact predicts some amount of factorization breaking.

2.4 Aspects of hadronization

In this section we will discuss various fragmentation models and more in particular the **LUND string model** used in this analysis for a description of the fragmentation process. Also aspects like current and target fragmentation, and how to separate between them will be dealt with. We will start first with a little overview of what models exist.

As the fragmentation framework was originally developed for e^+e^- annihilation where

$$e^+ + e^- \rightarrow \begin{matrix} Z^0 \\ \gamma \end{matrix} \rightarrow q\bar{q} \rightarrow \text{hadrons} \tag{2.31}$$

we will at first talk about fragmentation in the e^+e^- case as the basic concepts for lepton-nucleon scattering are the same. It allows, however, for a cleaner treatment. In a separate section, we will go into more detail into what makes

hadronization in lepton-nucleon scattering different, like the fact that one has to deal with a target remnant. For this we will introduce two more variables here. The first one is the center of mass rapidity η (of η_{CM}):

$$\eta = \frac{1}{2} \ln \left[\frac{E_h + p_{\parallel}}{E_h - p_{\parallel}} \right] \tag{2.32}$$

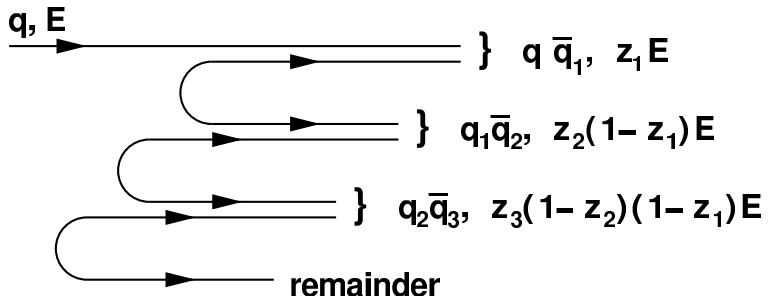
which is defined in the virtual photon - nucleon center of mass frame. $p_{\parallel}$ represents the hadron momentum component parallel to the direction of the virtual photon. The other variable which is often used is the so-called Feynman x_F variable, defined as:

$$x_F = \frac{2p_{\parallel}}{W}. \tag{2.33}$$

This enables us to define a forward region with both η and x_F positive and a backward region where the mentioned variables are negative. Around 0 we have the so-called central region.

2.4.1 Introduction

During the early seventies the first attempts to describe the fragmentation process came about. They were based on an iterative principle, an idea which can best be explained by the figure below.



Suppose we have a quark q produced with a specific momentum and energy E. At a certain point this quark q is 'split' into a hadron ($q\bar{q}_1$) and a remainder q_1. The hadron carries a fraction z_1 of the energy of the original quark, and the remainder of course $(1 - z_1)$ of this energy. This sharing of energy is given by some probabilistic function $f(z)$, where it is assumed that this function is independent of each step, thus independent of the remaining energy. Note that this z is not the same variable as one from equation 2.30. One defines [30]

$$f(z)dz = \textit{The probability that a primary meson leaves a fraction of the momentum between } z \textit{ and } z + dz \textit{ to the remaining cascade.} \tag{2.34}$$

This basic idea, together with some extensions for resonance production, strangeness production, decay, and baryon production formed the basis of the so-called

Field-Feynman model in 1978 [30]. This model was the starting point for a number of Monte Carlo programs which were used in e^+e^- collider experiments at PETRA and PEP in the beginning of the eighties [31], [32], [33]. This approach is called **independent fragmentation** (IF). The main disadvantage, however, was that a single jet by itself cannot conserve flavor or energy and momentum. So one has to do a balancing afterwards by hand between different jets to uphold the conservation laws. The most well know Monte Carlo program which is still in use that came forth from IF is called ISAJET [34], developed by F.E. Paige and S.D. Protopescu.

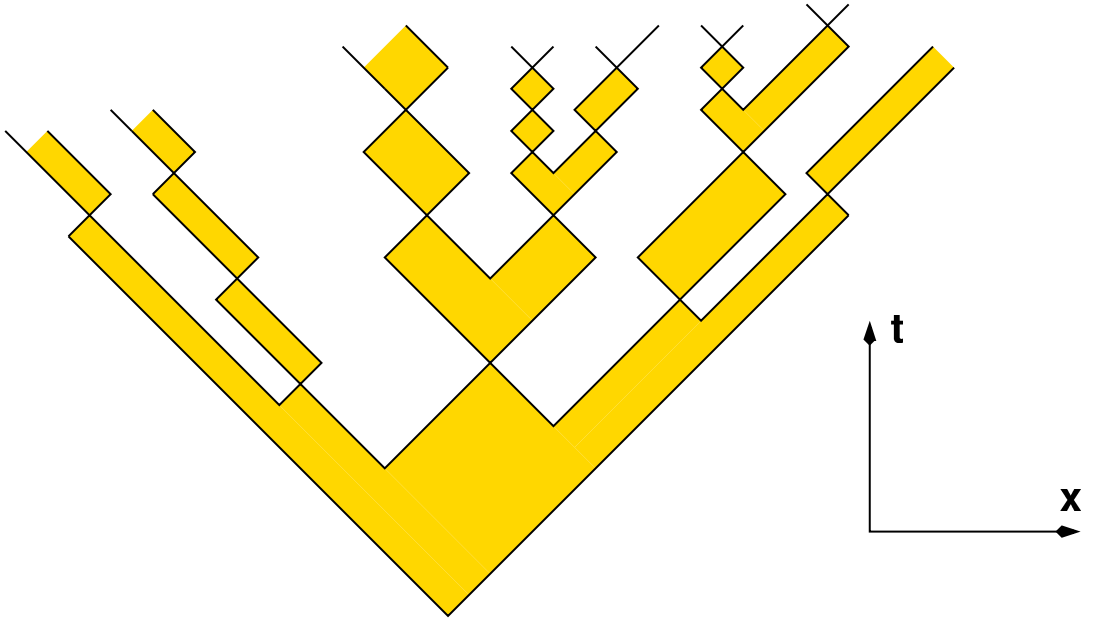


Figure 2.9: Breaking of a 1-D string in an xt diagram according to the Artru-Mennessier model [35]. In this model clusters may break up further on their own.

In this work, however, we will make use of a Monte Carlo model based upon a different principle. It started in 1974 when X. Artru and G. Mennesier published a first version of the so-called **string model** [35]. In this model one assumes a string is stretched between two emerging partons. This string has a uniform probability $P\,dA$ to break within a given space-time area dA, where each breakup corresponds to the production of a new $q\bar{q}$ pair from the QCD vacuum. In analogy with radioactive decay, a factor e^{-PA} is obtained for the probability that the string had not already broken. 'A' represents the space-time area of the light cone between the two ends (see figure 2.9). In this way, the original color singlet system with high mass is broken down into different smaller color singlet subsystems, **clusters**. Of course a cut-off mechanism is needed to stop this breakdown at a certain point.

Around the same time, an alternative approach [36] was devised by the LUND group. As with the Artru-Mennessier model, the LUND model is based upon **string fragmentation** (SF), but in this description the string fragments into hadrons of well-defined masses. Figure 2.10 shows the breaking of a one

dimensional string between a $q\bar{q}$ 'à la LUND'.

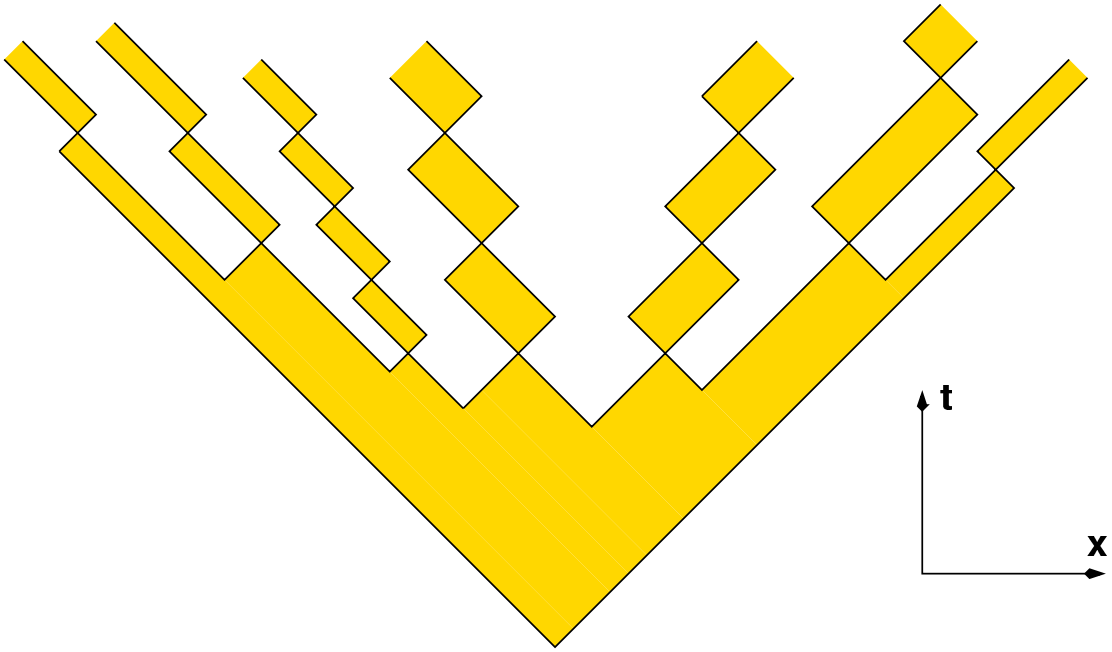


Figure 2.10: Breaking of a 1-D string in an xt diagram according to the LUND model. The formed hadrons have a well defined mass. The slow particles in the central region are produced first and the fragmentation spreads out.

Gluons in LUND string fragmentation are introduced as a 'kink' in the string between the q and $\bar{q}$. When the strings between the g and the q or $\bar{q}$ on either ends fragment, the transverse motion of these strings tends to boost the particles away from the central region between the q and $\bar{q}$. Not long after this effect, which is absent in independent fragmentation was observed experimentally by the JADE collaboration [37], [38] and [38]. It took a couple of years before this phenomenon and these results were generally accepted, but by halfway through the eighties one might say that SF replaced IF as being the standard fragmentation model.

To conclude this introductory paragraph, we would like to note that every fragmentation model has to start with a set of partons. The fragmentation models are most commonly implemented as Monte Carlo models. In these models one has two possibilities to proceed after the hard scattering occurred. One can calculated the final set of parton which need to fragment by means of perturbative QCD matrix elements, or one can use what is called **parton showering**. In e^+e^- scattering, one simply has in lowest order QCD $e^+e^- \rightarrow q\bar{q}$, which can be modified in first order by gluon radiation $e^+e^- \rightarrow q\bar{q}g$ and in second order by $e^+e^- \rightarrow q\bar{q}gg$ or $e^+e^- \rightarrow q\bar{q}q'\bar{q}'$. These cross sections can be perturbatively calculated. The parton shower approach is iteratively based upon the Altarelli-Parisi evolution equations (see 2.27) with their splitting kernels as displayed in figure 2.6. The probability that no branching occurs between t

and some t_{min} is then given by the so-called Sudakov form factor.

$$S_a(t, t_{min}) = \exp\left[-\int_{t_{min}}^{t} dt \int_{z_{min}}^{z_{max}} dz \frac{\alpha_S(Q^2)}{2\pi} \mathcal{P}_{a\to bc}(z)\right] \tag{2.35}$$

where $\mathcal{P}_{a\to bc}$ is the splitting kernel describing the branching considered. Figure 2.11 illustrates this process. The branch product may of course in turn

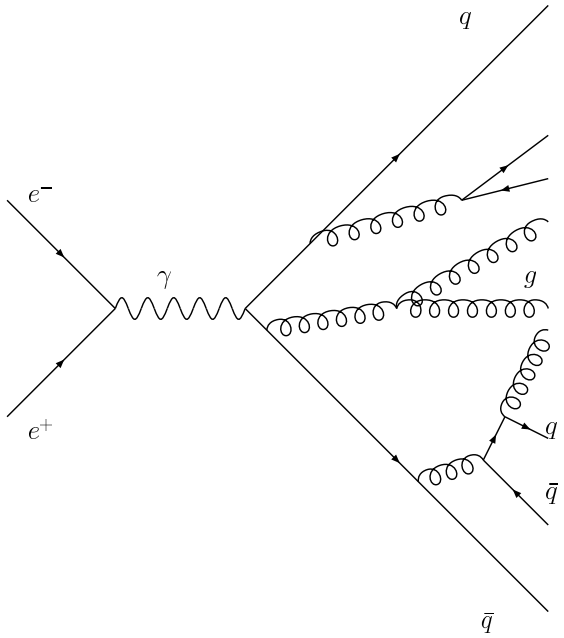


Figure 2.11: Schematic picture of parton shower evolution in e^+e^- events.

branch further until the parton mass ends up below a certain threshold and a final set of partons is given to a fragmentation model.

2.4.2 The LUND symmetric string model

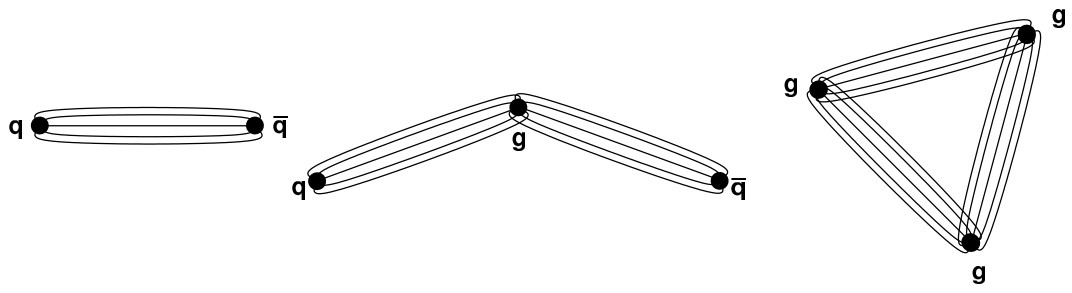


Figure 2.12: A color flux tube between a $q\bar{q}$ pair or with a gluon in between. A gluon is represented by a kink in the string with two flux tubes connected. Also closed gluon strings are possible as e.g. in the $\Upsilon \to ggg \to$ hadrons decay.

The physical picture of the string model (figure 2.12) is that of a color flux tube between a $q\bar{q}$ pair. However, in higher order other configurations are also possible like ggg in hadronic Υ decay. The tube is considered to be uniform

along its length with a typical transverse size of about 1 fm. This physical picture leads naturally to a linearly rising potential, and thus confinement:

$$V(r) = +\kappa \cdot r \tag{2.36}$$

where r is the separation between the partons and κ the string constant, or in other words, the **mass density** along the string. It is estimated to be in the order of ≈ 1 GeV/fm. A new $q\bar{q}$ pair will be formed when the two ends of a color singlet system are typically 1-5 fm apart in the $q\bar{q}$ rest frame. This means that the emerging hadrons are organized in a flavor chain in which each hadron can be given a rank. This is illustrated in figure 2.13.

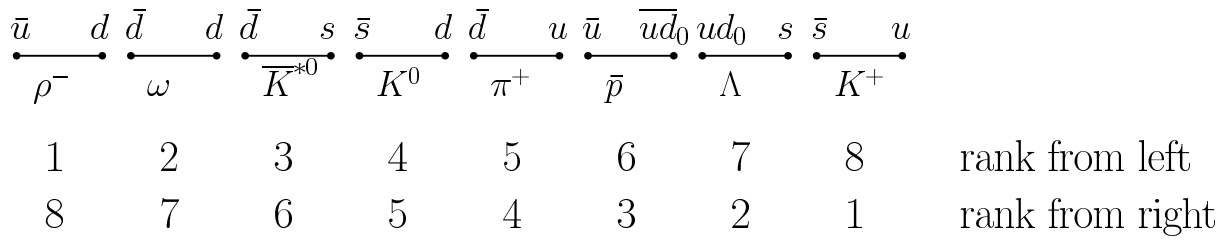


Figure 2.13: Flavor chain for primary hadrons in a fragmenting string.

A string is supposed to have no transverse excitations, which means that the buildup of transverse momentum in the fragmentation process has to come from the creation of $q\bar{q}$ pairs. Quantummechanically the q and $\bar{q}$ are created in the very same spot in the color field, but can each tunnel out to an allowed region in transverse momentum space. The probability for this is proportional to

$$e^{-\pi m_T^2/\kappa} \tag{2.37}$$

where the total transverse mass is given by the quark mass m and the intrinsic transverse momentum p_T, $m_T^2 = m^2 + p_T^2$ the latter one is compensated between the q and the $\bar{q}$.

2.4.2.1 Symmetric fragmentation

Suppose we have a quark - antiquark pair that fragments. The q and $\bar{q}$ will move in opposite directions. Let's call the quark q_0 with energy-momentum $(W^+, 0)$ and the anti quark $\bar{q}_0$ with energy-momentum $(0, W^-)$ in the light cone frame[5] [39]. At the production vertex V_i (see figure 2.14(a)) with light cone coordinates (x_i^+, x_i^-), the $q_i\bar{q}_i$ pair gets produced. This situation is now analogous as if the two states $(\bar{q}_0, q_i)$ and $(\bar{q}_i, q_0)$ had started in space-time points O_L and O_R, with energy-momenta given by

$$\begin{aligned} q_0 &: (W^+ - \kappa x_i^+, 0) \\ \bar{q}_i &: (0, \kappa x_i^-) \\ q_i &: (\kappa x_i^+, 0) \\ \bar{q}_0 &: (0, W^- - \kappa x_i^-) \end{aligned} \tag{2.38}$$

[5]Note that here $W^\pm = E \pm P$

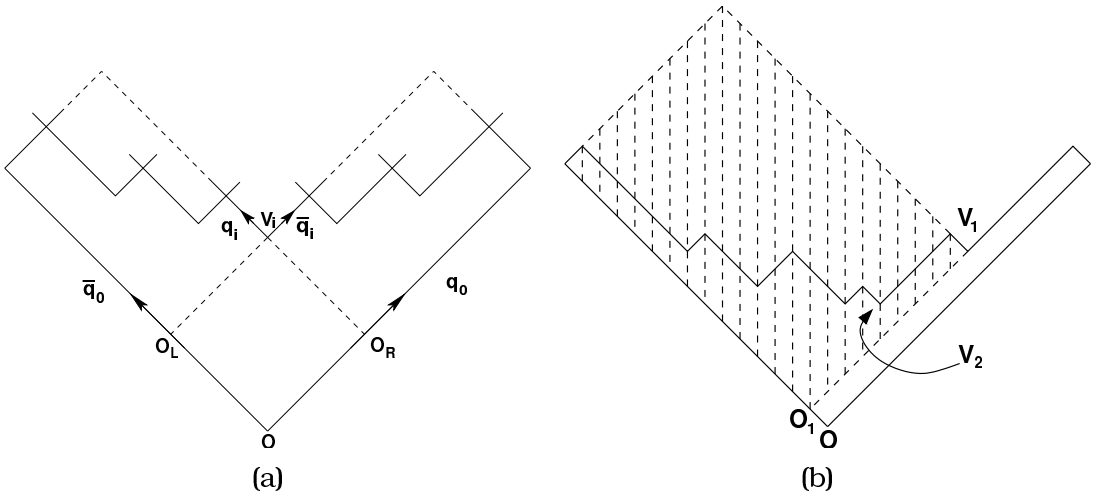


Figure 2.14: Production of a $q_i\bar{q}_i$ pair, which splits the system in two parts that fragment independently. Figures after [39].

where κ is of course again the string constant. This analogy is in fact at the very core of the discussion. The two subsystems are considered to fragment independently.

We define now two independent Lorentz invariant quantities upon which the rest of the treatment will be based. First of all we have the center of mass energy of the system

$$s = W^+W^- \tag{2.39}$$

and secondly the energy fraction

$$z = \frac{\kappa x_i^+}{W^+} \tag{2.40}$$

If we now consider vertex V_1 as displayed in figure 2.14(b), which occurs at space time point (x_1^+, x_1^-). At this point the system breaks into a meson of mass m and a remnant system, given by the shaded area. This point in fact is determined by the fraction z_1 of W^+ taken by the meson [39] :

$$x_1^+ = \frac{1}{\kappa}(1 - z_1)W^+ \tag{2.41}$$

$$x_1^- = \frac{m^2}{\kappa z_1 W^+} \tag{2.42}$$

If we call $s_0 = W^+W^-$, then we can say that the probability distribution for the vertex V_1 is generally given by

$$dP = f(z_1, s_0)dz_1 \tag{2.43}$$

As assumed we can now state that the remnant system will fragment independently staring at the space-time origin O_1 (see figure 2.14(b)) with a squared mass s_1 given by

$$s_1 = (1 - z_1)\left(s_0 - \frac{m^2}{z_1}\right) \tag{2.44}$$

the subsequent probability for a second vertex V_2 to occur is now given by

$$dP = f(z_2, s_1)dz_2 \tag{2.45}$$

where f is the exact same function as in 2.43. Thus the fragmentation process can iteratively be described by subsequent steps determined by one function $f(z, s)$. Two requirements will now give rise [39] to a very specific shape of this function.

1. For large energies there should be a central plateau in the rapidity distribution of the final state hadrons. This leads to the fact that for large s, the function $f(z, s)$ should become independent of s, or

$$\lim_{s\to\infty} f(z, s) = f(z) \tag{2.46}$$

2. The second requirement is that the fragmentation can proceed either from the quark end or from the anti quark end. There is no reason to expect otherwise. One can show [39] that this requirement leads to a general form for the $f(z)$ function being

$$f(z) = Nz^{-1}(1-z)^a e^{\frac{-bm_T^2}{z}} \tag{2.47}$$

where N is a normalization constant, m_T^2 the transverse mass and a and b are two parameters which need to be fixed by further dynamical input. This function is called the **LUND symmetric fragmentation function**.

In a Monte Carlo program there obviously should not be any artificial effect left in the treatment coming from the direction in which the fragmentation occurred. For the exact values of these parameters we refer to chapter 5.

2.4.2.2 Baryon production

As already illustrated in figure 2.13 a simple way of introducing baryons in string fragmentation is allowing for diquark - anti diquark pairs to be formed. When doing this, however, a number of parameters have to be introduced like the $qq\overline{qq}$ pair to $q\bar{q}$ pair production and strange di-quark to light di-quark production probability ratios and the suppression of spin 1 di-quarks to spin 0 diquarks.

Another mechanism however, which does not involve diquark creation from the vacuum and thus provides a more general approach, is the so-called **popcorn model** [40]. In this way of seeing, diquarks as such are never produced, but baryons are formed by successive $q\bar{q}$ pair creation. Figure 2.15 illustrates this tunneling mechanism. Suppose a red (r) quark q and anti-red ($\bar{r}$) anti-quark $\bar{q}$ fragment. Inside this color field it is quantummechanically possible to have color fluctuations of different color as illustrated by the blue anti-blue ($b\bar{b}$) and green anti-green ($g\bar{g}$) fluctuations of respectively $q'\bar{q}'$ and $q''\bar{q}''$. Suppose now that inside the field of one of these fluctuations $q_1(g)\bar{q}_1(\bar{g})$ yet another

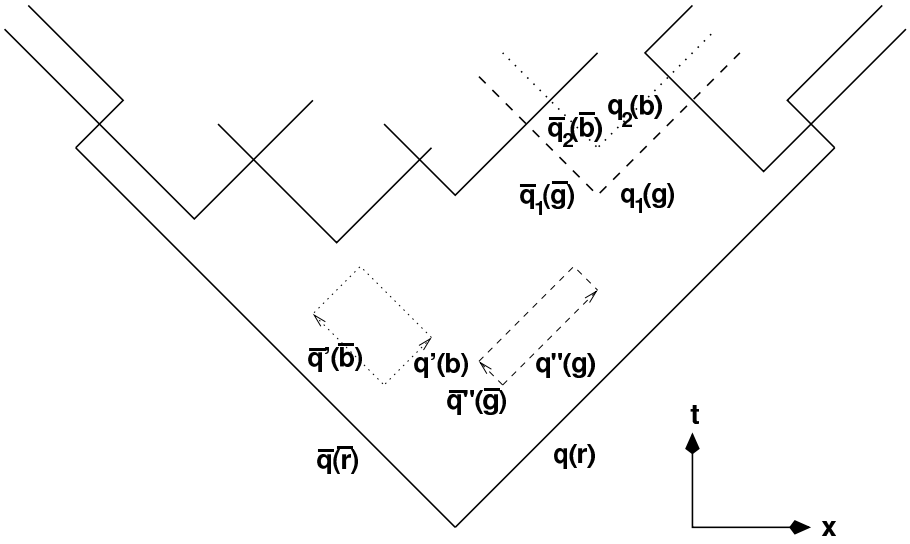


Figure 2.15: The popcorn mechanism, where a color field is stretched between a $q\bar{q}$ pair being red (r) and anti red ($\bar{r}$). Inside this field it is also possible to have $b\bar{b}$ and $g\bar{g}$ fluctuations (as with q' and q''). If a second pair is produced inside a region with the 'wrong' color, the field can break and a $B\bar{B}$ pair can be produced.

$q_2(b)\bar{q}_2(\bar{b})$ pair is produced of different color (b), then the q_2 will be dragged towards the rg-field and the $\bar{q}_2$ towards $\bar{r}\bar{g}$-field resulting in a baryon antibaryon ($B\bar{B}$) pair. Suppose now that more pairs get produced in between the $q_1\bar{q}_1$, then we obtain a meson in between the two baryons, resulting in $BM\bar{B}$ or even $BMM\bar{B}$ configurations.

2.4.3 Independent fragmentation

As already stated in the introductory section, the basis of any independent fragmentation model is the assumption that the partons all fragment independently from each other, where again each step is determined by a function $f(z)$. A number of $f(z)$ functions have been used in IF models. First of all we mention the so-called Field-Feynman parameterization [30] which is, in IF, often used for ordinary hadrons :

$$f(z) = 1 - a + 3a(1 - z^2) \tag{2.48}$$

where $a \approx 0.77$. The TASSO collaboration at the PETRA accelerator at DESY, however, found that a shape like

$$f(z) = (1 + c)(1 - z)^c \tag{2.49}$$

better described their data [41]. Also, for heavy quark fragmentation another form is often used [42], [43], proposed by Peterson et. al [44] :

$$f(z) = \frac{N}{z\left[1 - \frac{1}{z} - \frac{\varepsilon}{1-z}\right]^2} \tag{2.50}$$

where N is a normalization constant to fulfill $\int_0^1 f(z)dz = 1$ and $\varepsilon = (m_0/m_Q)^2$ with m_Q the heavy quark mass and m_0 a reference scale.

Independent fragmentation models do have some disadvantages, however. The original models have no baryon production mechanism, so additions had to be made later on analogous to the productions mechanisms in the LUND string model. Also there is no clear way how to treat the gluon inside these models, unlike with string models where a gluon is just a momentum carrying kink in the string. However, the most apparent problem is that these models do not fully conserve total flavor, energy and momentum [45] and tricks have to be used afterwards to restore the conservation. The use of independent fragmentation models in analysis typically also leads to different values for the strong coupling constant α_S needed to describe the data as one would find using string or cluster fragmentation models in the analyses [45].

2.4.4 Cluster fragmentation

Cluster fragmentation (CF) models leave behind them the fragmentation chains typically found in independent fragmentation and string fragmentation. Instead the produced partons are grouped in clusters of mass of a couple of GeV which subsequently decay to form hadrons. These decays are dominated by phase space considerations so that in fact the parton showers here are very important to determine the final energy spectrum of the hadrons. Inside the cluster rest frame the decay is also assumed to be isotropic, providing a unified description of transverse and longitudinal fragmentation properties, which is absent in SF and IF. The formed clusters usually do have a rather broad energy spectrum with a long tail towards high masses. Therefore it is needed that individual clusters can break up into smaller ones.

The HERWIG [46] Monte Carlo program is based upon a CF model by Webber and Maschesini [47], [48]. An example of a cluster mass spectrum is given in reference [47]. The so-called Caltech model holds the middle between CF and SF [49]. After parton showering evolution strings are defined between the partons and behave as in the LUND string model, however with the one difference that the produced hadrons do not have to end up on mass shell. Therefore one can use a simple areal breakup law, where the probability for a string break $dP_{break} \propto dA$ with dA a string area element. As in the Artru-Mennessier model this results in an exponential decay law. The endpoints of these breakups are then identified with clusters which are subject to decay laws as in the Webber CF model.

In the literature, one finds tons of comparisons between different fragmentation descriptions. Each model has its advantages and disadvantages, e.g. the large number of parameters in the LUND string model and independent fragmentation models makes them rather difficult to handle in a Monte Carlo

framework. On the other hand, cluster fragmentation models do a far worse job in describing baryon production as compared to the LUND model. In general, it is true that models using parton showering instead of matrix elements do a better job in describing the event shapes, especially in multi-jets events [45]. Also, the function $f(z)$ determines for IF and SF for the most part the momentum and rapidity distribution of the final state hadrons, whereas for CF it is in fact more the parton showering itself.

2.4.5 Parameterizations for $D_q^h(z, Q^2)$ fragmentation functions

Of course what we up till here have called *fragmentation function* ($f(z)$) is not the quantity found in the semi-inclusive deep inelastic cross section (eq. 2.29). As defined in (eq. 2.34) $f(z)$ does not contain any information on the parton type that initiated the fragmentation process or the final states hadrons which emerge. In Monte Carlo programs quark flavors and hadron types are added by using a number of parameters like strangeness suppression factors, baryon suppression factors and so on. The $f(z)$ functions are somewhat more of intrinsic functions in the models, describing the process at a microscopic level, whereas the fragmentation functions $D_q^h(z)$ describe the fragmentation process by their end products, and therefore somewhat more 'macroscopic'. For the latter ones there exist a number of parameterizations [50], [51], [52], [53].

The Kretzer parameterization [52] is a.o. based on precision e^+e^- to $\pi^\pm$ and $K^\pm$ measurements at the Z^0 pole from the SLD collaboration [54] and inclusive data on charged hadrons from the ALEPH experiment [55], [56] and data at lower energy (29 GeV) from the TPC Collaboration [57]. A recent parameterization by Albino, Kniehl and Kramer [53] includes data obtained by the OPAL Collaboration [58] on light flavor separated measurements on light charged hadron production at the Z^0 pole in addition. It is essentially an updated NLO version of the KKP parameterization [51].

The Kretzer parameterization yields charge separated fragmentation functions using the following ansätze and symmetry relations [52]:

$$D_q^{h^+,h^-} = D_{\bar{q}}^{h^-,h^+}; h = \pi, K \tag{2.51}$$

$$D_d^{\pi^+} = D_{s,\bar{s}}^{\pi^+} < D_u^{\pi^+} = D_{\bar{d}}^{\pi^+} \tag{2.52}$$

$$D_{\bar{u}}^{K^+} = D_{d,\bar{d}}^{K^+} < D_u^{K^+} < D_{\bar{s}}^{K^+}. \tag{2.53}$$

The AKK and KKP parameterizations do not separate the hadron charges and thus make the following assumptions [53]

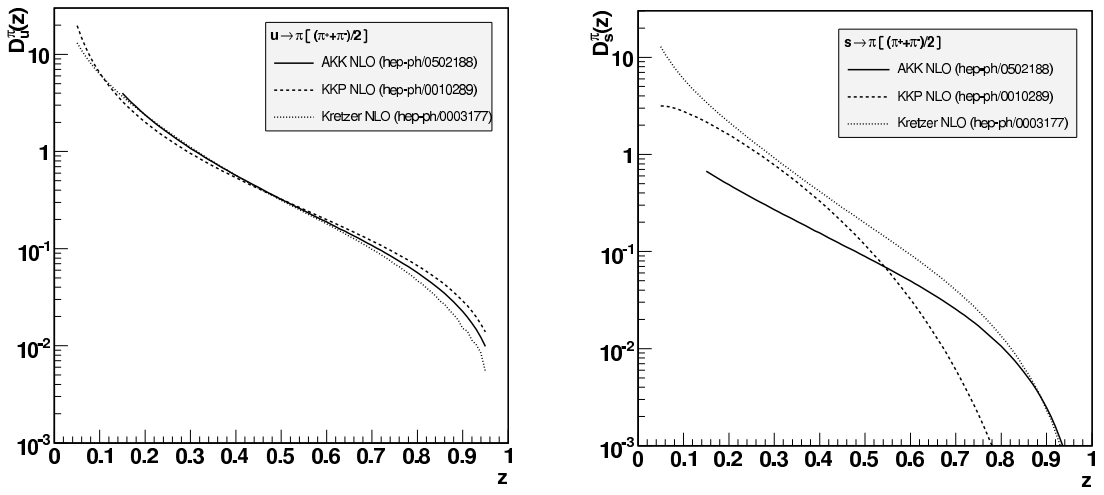
$$D_u^{\pi^\pm} = D_{\bar{d}}^{\pi^\pm}; D_u^{K^\pm} = D_s^{K^\pm} \tag{2.54}$$

and analogous for the anti-quark flavors.

All parameterizations take the form

$$D_i^h(z, \mu_0^2) = N_i z^{\alpha_i^h} (1-z)^{\beta_i^h} \tag{2.55}$$

at some input scale μ_0^2 as ansatz for their fitting procedure, where of course N_i, α_i^h and β_i^h are the parameters that need to be determined. The input scale for the Kretzer parameterization is $\mu_0^2 = 0.4$ GeV^2, whereas for AKK μ_0^2 = 2 GeV^2. The NLO behavior of the fragmentation functions comes from the fact that these quantities obey Altarelli-Parisi renormalization group equations identical to the ones given in 2.27 [52], [59].



(a) u quark fragmentation

(b) s quark fragmentation

Figure 2.16: Comparison of different NLO fragmentation function parameterizations for D_u^π where π stands for $\frac{1}{2}(\pi^+ + \pi^-)$ and D_s^π. The different models displayed are the KKP model by Kniehl, Kramer and Pötter [51], the Kretzer model [52] and the AKK (Albino, Kniehl and Kramer) model [53].

A comparison of a number of different parameterizations can be seen in figure 2.16, where we have plotted the NLO parameterizations from [51], [52] and [53] for the charge averaged π fragmentation functions D_u^π and D_s^π at HERMES average Q^2 of 2.5 GeV^2. One clearly sees that the different parameterizations agree fairly well as far as u quark fragmentation is concerned, but disagree rather dramatically for s quark fragmentation. This is basically due to the lack of charge and flavor separated experimental data. A possible input for this would be the extraction of charge separated kaon multiplicity distributions, which is one of the goals of this work. Kaon distributions are of course a lot more sensitive to strange quark fragmentation as the s and $\bar{s}$ are valence quarks for the K^+ and K^- respectively.

A comparison between LO and NLO parameterizations can be seen in figure 2.17 for the KKP and the Kretzer parameterization. From all the figures comparing different parameterizations we can see that it is at high z that the various approaches differ, mainly due to lack of precise data in this region. At high z statistics very low and also exclusive hadron production start to play an important role.

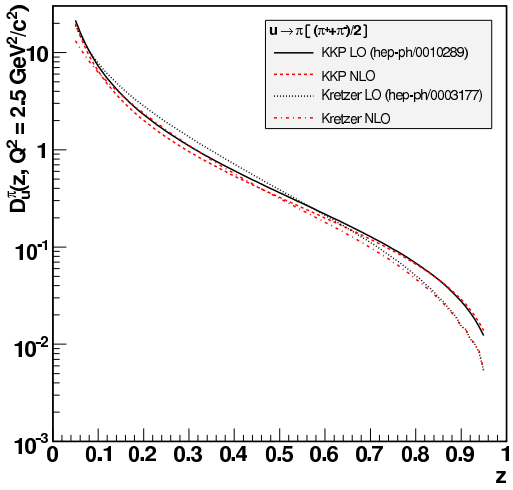


Figure 2.17: Comparison between the KKP and the Kretzer parameterizations for leading order and next to leading order. The scale of these curves are displayed for $Q^2 = 2.5$ GeV^2.

2.4.6 About favored and unfavored fragmentation

In literature and further on in this work one will encounter the terms **favored** and **unfavored fragmentation functions**. One refers to favored fragmentation where fragmentation functions D_f^h are considered where the quark (f) is one of the final state hadron's (h) valence quarks, e.g. $u \to \pi^+$ or $d \to \pi^-$. Obviously when this is not the case, one speaks about unfavored fragmentation. Often the favored pion fragmentation function is referred to as $D^1(z)$ or $D^+(z)$ and the unfavored $D^2(z)$ or $D^-(z)$. Figure 2.18 shows the ratio of the unfavored to favored pion fragmentation functions from both the Kretzer parameterization [52] and the EMC u-quark fragmentation functions [60]. As

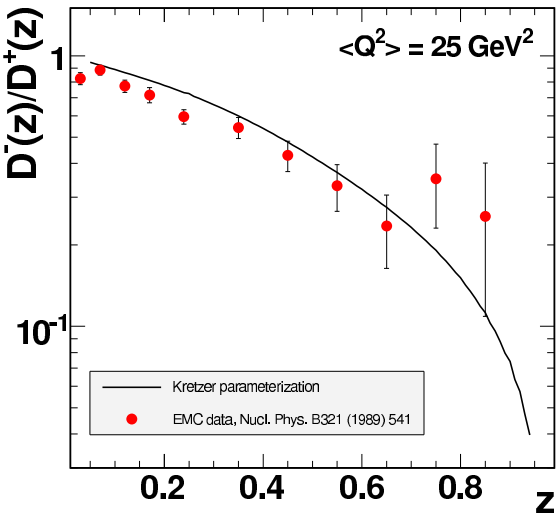
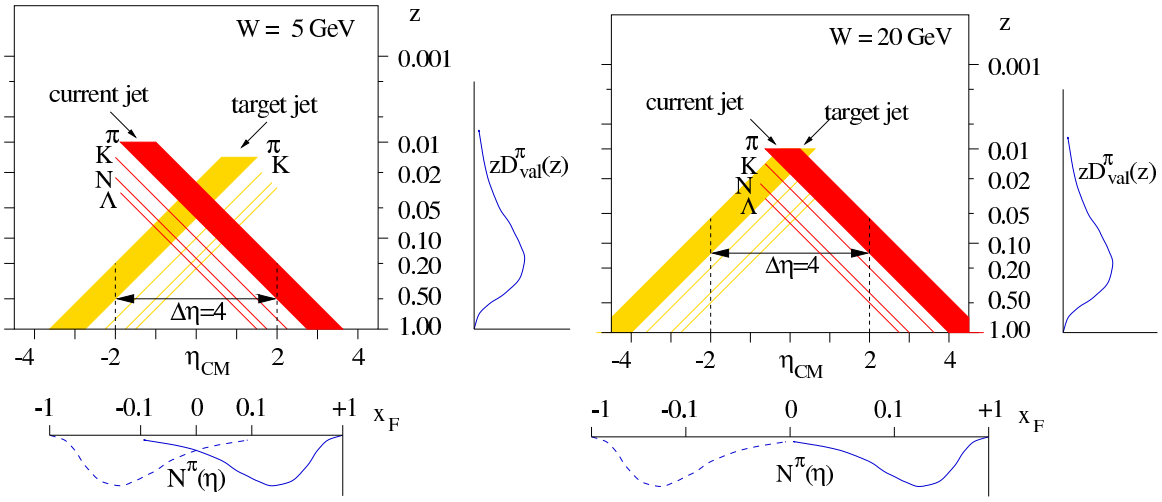


Figure 2.18: Ratio of unfavored to favored pion fragmentation functions comparing the Kretzer parameterization and the EMC u-quark fragmentation data points.

one might expect, we clearly see that unfavored fragmentation is suppressed at high z. However, this suppression decreases at low z as this corresponds to the central rapidity region.

2.5 Fragmentation in lepton-nucleon scattering

So far we have in fact dealt with the description of hadronization in e^+e^- scattering. Compared to this fragmentation in lepton-nucleon scattering (e.g. at HERA) brings along a couple of problems and ambiguities. The most important one is the treatment of the remnant of the nucleon after the scattering reaction. Hadrons coming from the fragmentation of the struck quark (**current jet**) in the nucleon are expected to be found in the forward region (see the beginning of section 2.4) whereas hadrons coming from the target remnant (**target jet**) are expected to be found in the backward region. Mixing of hadrons from the backward or target region is believed to lead to factorization breaking effects in the fragmentation of the struck quark. The so called **Berger criterion** provides us with a means of discussing this. As illustrated in figure 2.19 one can see the center of mass rapidity η plotted against the z variable for



(a) Low W case (b) High W case

Figure 2.19: Relation between the z variable in fragmentation and the rapidity η for both high and low W. This illustrates the Berger criterion [61]. See also [62].

both target and current jets. At high W, these two jets are clearly separated over almost the complete z range. The full rapidity range $|\eta_{max}| - |\eta_{min}|$ kinematically allowed is roughly given by $\ln W^2$. The typical hadronic correlation length expressed in rapidity is taken to be $\delta \simeq 2$ [63]. The Berger criterion [61] now states that in order to separate current and target regions it is necessary that

$$\Delta\eta \geq 2\delta \tag{2.56}$$

equivalently $W \geq 7.4$ GeV depending on z or x_F. However, this derivation neglects transverse mass effects, which are important at HERMES. As shown

on figure 2.19(a) one can impose x_F or z cuts to suppress target contamination. We see e.g. that for HERMES energies we have a clean separation between target and current jets for $z > 0.2$.

The target region itself is then described by so-called *fracture functions* $M^h_{f,N}(x, z, Q^2)$ [64]. These functions give the probability of finding within a target N a quark f and a hadron h. They depend on the Bjorken x as well as on the semi-inclusive variable z[6] and obey similar Altarelli-Parisi type evolution equations as the fragmentation functions do. They are in fact some sort of hybrid between structure functions and fragmentation functions and also obey the following sum rule :

$$\sum_h \int dz\, z M^h_{f,N} = (1 - x) \cdot q_f \tag{2.57}$$

where q_f is again the parton density function.

Traditionally most semi-inclusive analyses use quark flavor tagging, where one relates the observed final state hadron to the struck quark. It is generally believed that in this kind of analyses a good separation between current and target jets is essential. However, there are arguments to believe that in fact this separation is not possible, even in principle, if one cares about a complete and coherent description of hadronic final states. In fact this is already reflected in the string and cluster hadronization pictures [65]. Since partons cannot exist as asymptotic states (free particles) due to confinement, one cannot expect to measure a primary parton, but instead one measures a degree of correlation between a detected hadron in either the current or the target region and the struck quark (see figure 2.20).

The current correlation $C(x_F, W)$ and target correlation $T(x_F, W)$ are defined in [66], suffice it to say here that they describe the degree of correlation between the observed hadron and the struck quark in either current or target jets. One can see that separation between the current and target region is never distinct. There is always a certain amount of correlation with the struck quark as well as with the target remnant. At small x_F **both** both target and current correlations decrease as W increases. This implies that indeed most of the hadrons come from the central rapidity region with increasing W. This is also found experimentally [67] in the EMC μp data, where the W dependence is in fact due to this central region. At higher x_F, the W dependence is suppressed in the current and target correlation functions (figure 2.20). Thus in summary, one has to take into account the central rapidity region to reproduce the correct kinematical dependencies.

[6] Or in other descriptions on the x_F variable

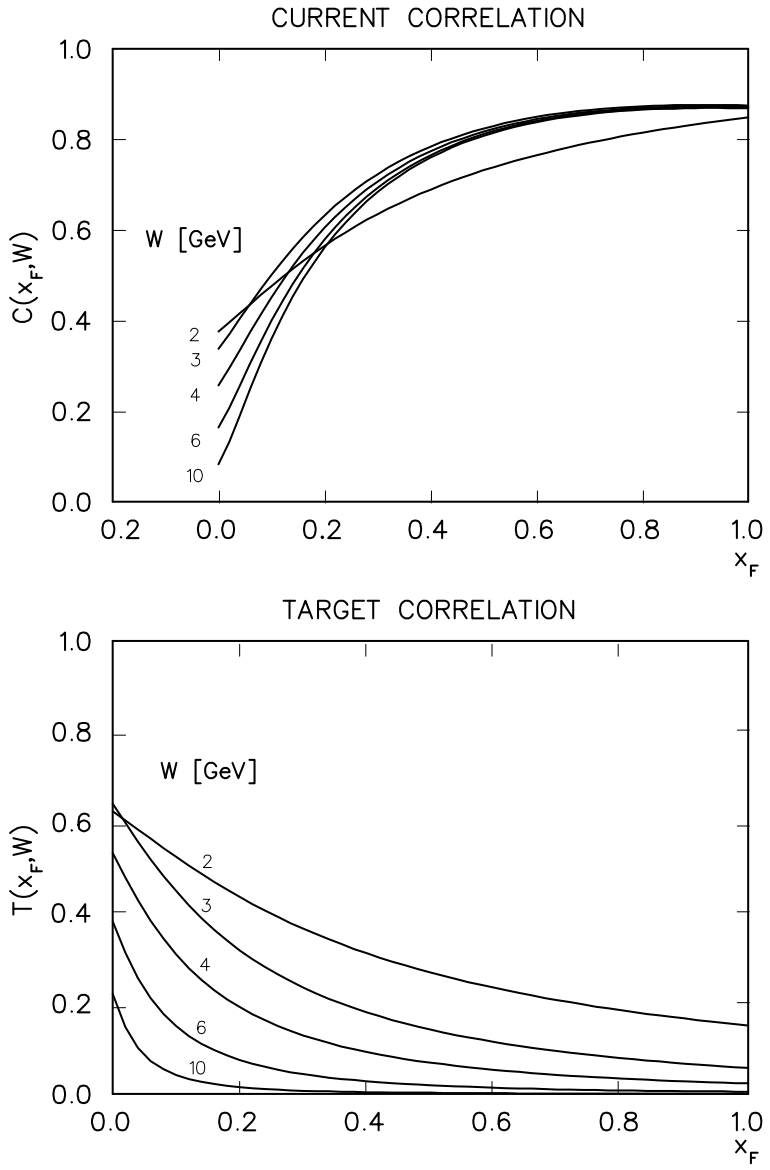


Figure 2.20: Quark hadron correlations for the current and target region, figures taken from [66]. These calculations were performed for pions.

Chapter 3

The HERMES experiment

"...Come here and decide which is the more excellent beauty of face, and to the fairer give this apple's lovely fruit..."

Hermes to Paris after bringing Hera, Athena and Aphrodite before him...

In this chapter the experimental setup used in this dissertation will be discussed in more detail. At first a general overview of the detector will be given after which we will go into more detail on the individual components. More time will be spent on the parts that are the most important to the presented analysis. This is mainly the identification of the different hadron types by a Ring Imaging Čerenkov detector. Attention will be given as well to the trigger hodoscopes since some hardware work was done with these detectors during preparation of this Ph.D. However, first we will start with a short description of the HERA storage ring, on which HERMES is situated.

3.1 The DESY facility and the HERA storage ring

The DESY[1] facility in Hamburg, Germany, was founded on the 18th December 1959 and has been a world leading particle physics and synchrotron laboratory ever since. Among others, it is the home of the HERA[2] storage ring. In this super-electron microscope, electrons or positrons are accelerated to an

[1] Deutsches Elektronen Synchrotron
[2] Hadron Elektron Ring Anlage

energy of 27.6 GeV. In the same tunnel, a superconducting storage ring for 820 GeV protons was constructed, which was upgraded in 1998 to 920 GeV. Both rings were fully operational in 1992 together with two collider experiments H1 (north hall) and ZEUS (south hall). In 1995 data taking started for the HERMES experiment which was constructed in the east hall. A fourth experiment, HERA-B was completed in the west hall in 1999 and took data until February 2003. The latter two experiments both have a fixed target, implying that they respectively only make use of the lepton and the proton beams. A schematic overview of the HERA ring can be seen in figure 3.1. This was the setup until the year 2000. Afterwards, spin rotators were also installed around the H1 and ZEUS experiments.

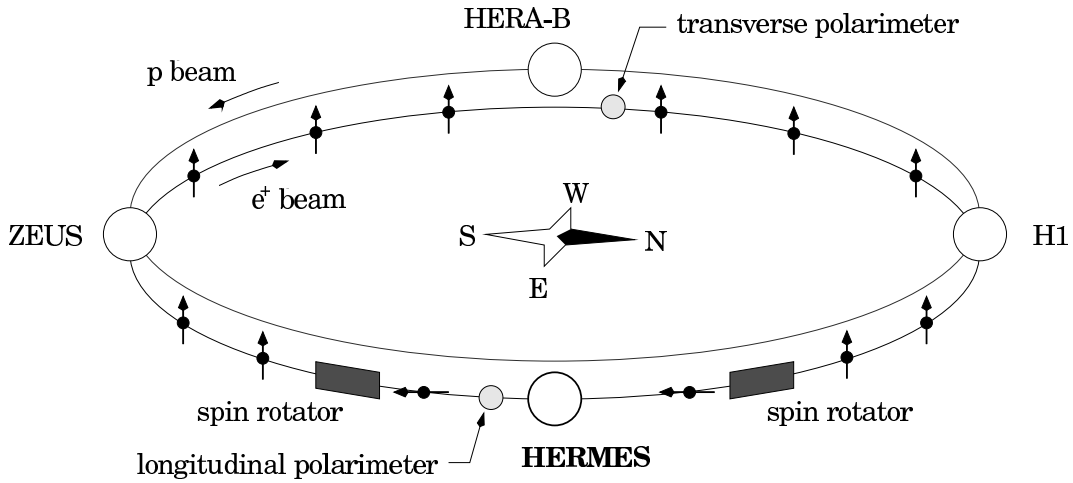


Figure 3.1: Schematic overview of the HERA ring with the 4 experiments and the spin rotators. The arrows denote the orientation of the lepton spin, so normally in the arcs a transverse polarization is realized, which is rotated to a longitudinal orientation by the spin rotators for use in the scattering reactions.

One distinct feature of the HERA lepton ring is indeed in fact its polarization. The emission of synchrotron radiation in the arcs of the ring involves a small asymmetry in the spin-flip amplitude so that gradually the spin states parallel[3] to the direction of the magnetic guiding field become more populated. This effect is named after A.A. Sokolov and I.M. Ternov who first described it in 1964 [68]. Thus a transverse polarization builds up gradually over time following an exponential law :

$$P_{ST}(t) = P_{ST}\left(1 - e^{-\frac{t}{\tau_{ST}}}\right) \tag{3.1}$$

The theoretical maximum polarization is $P_{ST} = 92.38\%$ with a rise time constant τ_{ST} of about 40 minutes, both depending on the energy of the beam, however. Of course in reality this is never achieved mainly due to small but unavoidable misalignments of the magnets. HERA is able to reach polarizations of up to 50 or even 60 %. Two spin rotators around the HERMES experiment flip the transverse spin orientation to a longitudinal one. The polarization is

[3]for positrons. For electrons the preferred state is antiparallel.

measured by Compton back - scattering with circularly polarized laser light in both transverse and longitudinal orientation as can be seen in figure 3.1.

The HERA ring is a bunched machine, with 220 bunch buckets of which 210 are filled. The buckets are separated in time by 96 ns. The circumference of the machine is about 6.3 km. HERA was designed to be able to store about 60 mA of positrons/electrons and 160 mA of protons. In practice HERA fills the beam pipes with currents of about 40 mA for the electrons and about 100 mA for the protons. When both beams are brought into collision at the H1 and ZEUS interaction points at the beginning of a so-called **fill**, a *luminosity run* is declared and the experiments can use the beams to take data. If the lepton beam current reaches a low enough value, beyond which data taking is no longer efficient, both beams are dumped and the fill is ended.

3.2 The target section

HERMES operates with gaseous target, a storage cell is used to locally increase the density of the gas. In figure 3.2 one sees a drawing of this storage cell. The cell is 40 cm long with an elliptical cross section of 9.8 $\times$ 21.0 mm^2 and is made out of 75 μm thin aluminum. It is cooled to about 70-100 K and is put inside the beam vacuum. For the polarized physics in HERMES an atomic

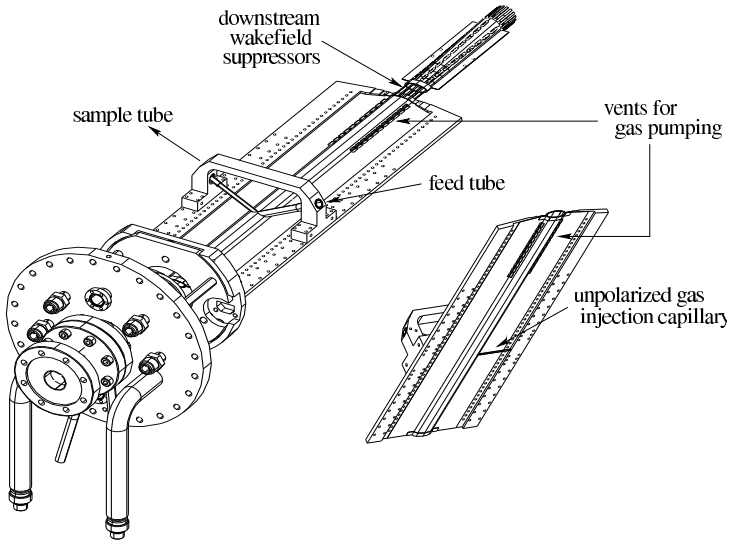


Figure 3.2: Technical drawing of the HERMES target storage cell.

beam source (ABS) was constructed, able to deliver polarized ^{1}H or ^{2}H. A dissociator atomizes molecular hydrogen or deuterium and nuclear polarization is then obtained via Stern - Gerlach separation. Cross section densities of 6.4 $\times$ 10^{16} (5.2 $\times$ 10^{16}) nucleons per cm^2 can be obtained for polarized hydrogen (deuterium). Polarization values are obtained up to about 85 %, even 90 % under ideal circumstances.

Next to having a polarized target, HERMES has also the possibility to inject

unpolarized high density ^{1}H, ^{2}H, ^{4}He and even some more complex nuclei in the storage cell using the *Unpolarized Gas Feed System* or UGFS. Data in high density mode was usually taken at the end of a HERA fill (see figure 3.3). This data is extremely important for research topics such as vector meson production, nuclear effects, exotic particles, light sea quark asymmetry and of course hadronization.

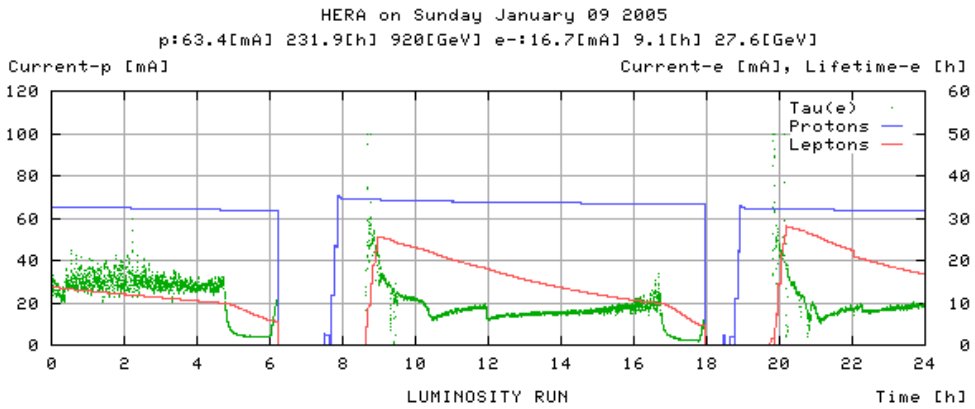


Figure 3.3: End of fill running, where HERMES injects high density unpolarized gas into the storage cell, causing the beam lifetime to reduce drastically. On this plot one can see the impact on the lepton beam around 05:00 h and 17:00 h. `Tau(e)` represents the electron lifetime in hours (scale on the right hand side).

3.3 The luminosity monitor

The luminosity for the HERMES experiment is measured by elastic Møller or Bhabha scattering in case of electrons or positrons respectively, off the electrons from the target atoms.

$$e^{-(+)} + e^- \rightarrow e^{-(+)} + e^- \tag{3.2}$$

In the case of a positron beam, also anihillation contributes

$$e^+ + e^- \rightarrow \gamma\gamma. \tag{3.3}$$

Both processes are exactly calculable in quantum electrodynamics (QED). The end products of these reactions are detected by two calorimeters which reside left and right of the beam pipe at the back of the HERMES experiment (see figure 3.4) and called the luminosity monitor [69]. Each of them consists of a matrix of 3×4 $NaBi(WO_4)_2$ crystals read out by a photomultiplier (PMT). The crystals are radiation resistant and measure $22 \times 22 \times 200$ mm^3. The ratio of the measured rate in the luminosity monitor and the exactly calculated Møller or Bhabha cross section provides the luminosity [70].

3.4 The HERMES spectrometer

In figure 3.4 a schematic overview of the HERMES setup is shown as it was in 2001. In total the spectrometer is about 10 m long and 5m high. It is symmetric with respect to the xz-plane [71]. The HERMES coordinate system

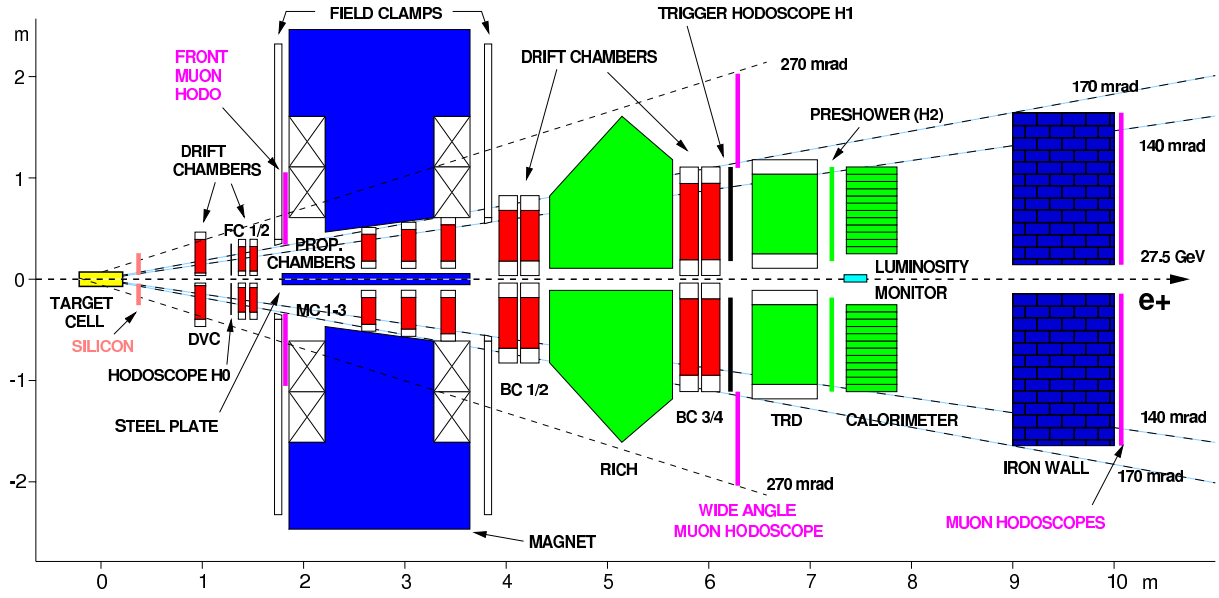


Figure 3.4: Schematic overview of the HERMES spectrometer as it was in 2001.

has its z axis along the beam pipe (along the lepton momentum), with 0 being the center of the target cell. The x axis points away from the center of the HERA ring, which is to the left if one looks downstream the beam pipe, and the y axis points up. The proton beam pipe passes undisturbed in between the two detector halves, about 70 cm left of the lepton beam. The red detectors in figure 3.4 represent the tracking detectors, whereas the green ones the particle identification system. We now go into more detail for each of the components, starting with the large dipole magnet.

3.4.1 The HERMES dipole magnet

A large dipole magnet is used to deflect the charged particles produced in the scattering reactions. It is operated at a field integral of $\oint B \cdot dl = 1.3$ T·m. Since the magnetic dipole field is oriented in vertical direction, the particles are deflected horizontally. The gap between the two pole shoes puts an upper limit to the angular acceptance of $\pm$ 140 mrad in the vertical and $\pm$ 170 mrad in the horizontal direction. The detector, however, adds another $\pm$ 100 mrad angular acceptance starting from the center of the magnet in the horizontal direction. This is to provide good detection for low momentum particles, which are deflected more. Inside the magnet, a septum plate is mounted to shield both HERA beams from the HERMES dipole field. This plate puts a lower limit on the vertical angular acceptance of $\pm$ 40 mrad. An additional correction coil

with a deflecting power of 0.08 T·m is accommodated inside the shielding of the lepton beam pipe. It is used to correct for fringe fields and imperfections of the magnetic shielding.

3.4.2 The tracking system

3.4.2.1 Detectors involved

The tracking system of HERMES consists of the detectors colored in red in figure 3.4. Except for the *Lambda Wheels* (LW) all the tracking detectors are wire chambers. About 1.1 m downstream, the Drift Vertex Chambers (DVC) are installed; then at z = 1.6 m we have the two sets of *Front Chambers* (FC 1/2). Behind the magnet, the *Back Chambers* [72] are installed: two sets before the RICH detector (see later), namely BC 1/2, and two sets after, BC 3/4. All of them are drift chambers, where each layer of drift cells consists of a plane of alternating anode and cathode wires between a pair of cathode foils. The chambers are constructed having six of layers of drift cells in three coordinate doublets (XX', UU' and VV'). For the X planes the wires are vertical, and give information on the x position. The U and V wires have a stereo angle of $\pm$ 30° with respect to the vertical and enable the reconstruction of the y position. The chambers are all filled with a mixture of Ar (90%), CO_2 (5%) and CF_4 (5%) which is both fast and non-flammable [73]. At a field $E = 800$ V/cm, the drift velocity has been measured to be about 7 cm/μs [71]. The FC resolution per plane is 225 μm, with an efficiency close to 99 %[4]. The resolutions for BC 1/2 are 210 μm and for BC 3/4 250 μm again with efficiency well above 99 %. It should be noted that the DVC pair is slightly larger then the standard HERMES acceptance. This detector extends from $\pm$ 35 mrad to $\pm$ 270 mrad vertically and covers $\pm$ 200 mrad horizontally. Its resolution is 220 μm per plane.

Inside the magnet, three sets of *Magnet Chambers* [74] (MC 1-3) are installed. These are proportional wire chambers. Their primary function is the momentum analysis of relatively low energy particles since most of them do not reach the back part of HERMES. The gas mixture of the MCs is the same as for the drift chambers, albeit in different proportions : Ar (65%), CO_2 (30%) and CF_4 (5%). They again consist of three submodules or planes (X,U and V), where each submodule consists of an anode plane at ground potential and two cathode planes with a common negative HV. They have a cell width of 2 mm, providing a resolution of 700 μm. Due to heavy radiation exposure aging effects have decreased the efficiency of the MCs [75]. The MCs however will not be used in this analysis.

Additionally, in 2002 a silicon detector called the Lambda Wheels (LW) was installed [76] directly behind the target cell. Its purpose is to increase the ac-

[4]Depending on the drift distance.

ceptance for decay products of long-lived particles, such as the Λ^0. It consists of two sets of 6 silicon strip modules arranged in a hexagonal shape, mounted 45 and 50 cm downstream of the HERMES interaction point. In order to safeguard the LW from too much radiation damage, a *Beam Loss Monitor* was installed [77] which measures the radiation intensity and triggers a magnet that kicks out the beam if the radiation level gets too high. As these directors were not yet fully commissioned, they will not be used here.

3.4.2.2 The tracking algorithm

The HERMES reconstruction program (HRC) [78] is based on a pattern tree search algorithm. The pattern made by the hits in the tracking chambers is matched against a database with possible patterns. This is done with increasing resolution as can be seen from figure 3.5. It is a well known technique,

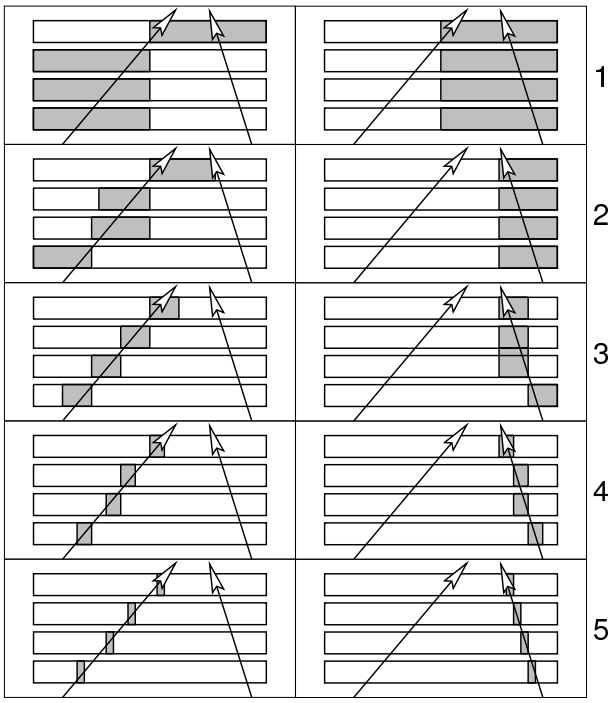


Figure 3.5: Tree search for two tracks. The detector pattern is matched against a pattern database with increasing resolution. Figure from [78].

however, before the HERMES experiment it was never used in such a large experiment and with such a high resolution. Again referring to [78], we can say that a special compression technique for storage of the patterns made this possible.

The momenta of the charged particles are measured by their deflection in the magnet, thus by the difference in slope between backtrack and fronttrack. The tracking resolution, determined from Monte Carlo is plotted in figure 3.6. Note that this is the resolution after installation of the RICH detector. The RICH material increased the total radiation length significantly, thus worsening the resolution by a factor up to 2.

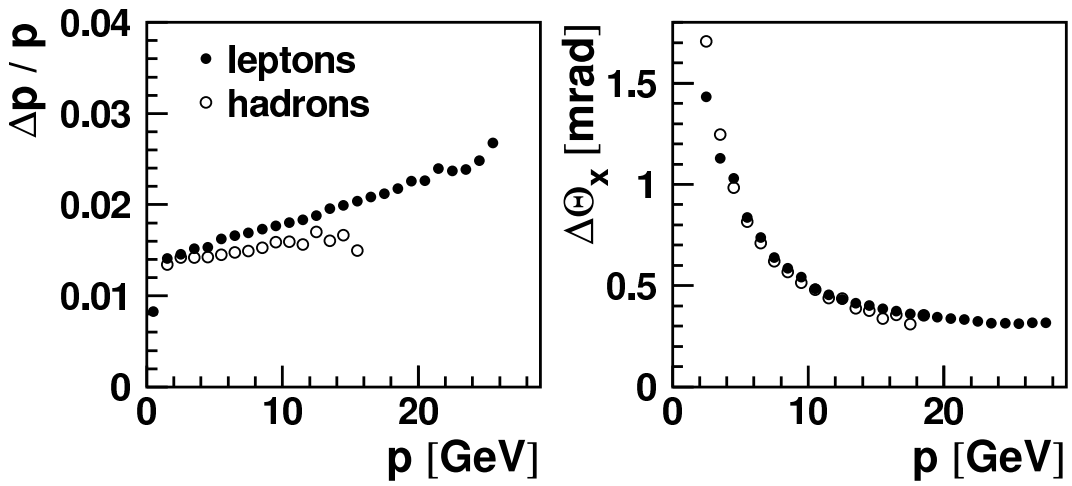


Figure 3.6: Momentum and angular θ_x resolution as a function of the momentum of the reconstructed particle.

3.4.3 Particle identification and energy measurement

In this section we will describe in somewhat more detail how different particles are identified in the HERMES experiment. At a first level a separation is made between leptons and hadrons. Since 1998 a Ring Imaging Čerenkov detector (RICH) is installed, which also provides identification of pions, kaons and protons. The particle identification (PID) detectors are marked green in figure 3.4.

3.4.3.1 The Transition Radiation Detector

When a particle crosses the boundary between 2 media with different electric properties, it will emit X-rays. In particular, the energy radiated when a particle with charge ze crosses a boundary between vacuum and a medium with plasma frequency ω_p is [6]

$$E = \frac{1}{3}\alpha z^2 \gamma \hbar \omega_p. \tag{3.4}$$

where one can see that it depends on the Lorentz factor γ of the incident particle. In this equation, α is the fine structure constant. The plasma frequency ω_p is proportional to the square root of the electron density $\sqrt{N_e}$ of the medium. The typical emission angle for the photons is $1/\gamma$. For a particle with $\gamma = E/m_0 = 10^3$, the radiated photons are in the soft X-ray range (2 to 20 keV). An easy calculation shows that the Lorentz factor γ for 5 GeV pions and electrons differs by almost 3 orders of magnitude[5].

The HERMES *Transition Radiation Detector* (TRD) consists of 6 modules in each detector half (see figure 3.7). The radiators are formed by layers of polyethylene/polypropylene fibers. Detection of the X-rays is done by proportional wire chambers filled with a Xe/CH_4 mixture (90/10 %). Most of the

[5] $\gamma_\pi \approx 35$ and $\gamma_e \approx 10^4$ for a total particle energy of 5 GeV.

transition radiation can only be detected from $\gamma > 500$, so hadrons will mostly deposit energy due to ionization.

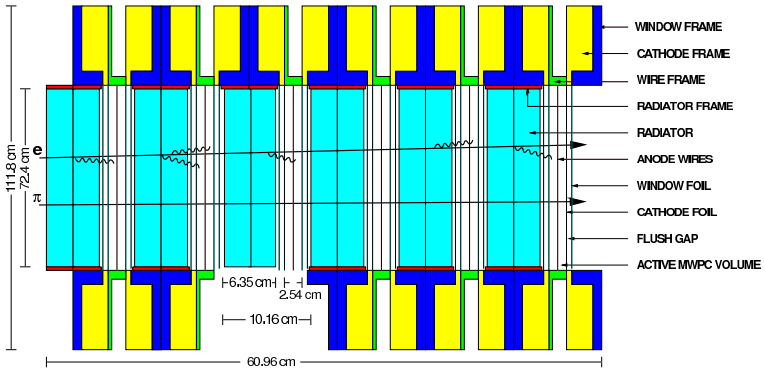


Figure 3.7: Schematic view of one half of the HERMES transition radiation detector.

To suppress the effects from delta electrons, the truncated mean of the six TRD modules is calculated, so the mean of all signals without the largest of the six responses. Figure 3.11 shows this for leptons and hadrons. The TRD misidentifies less than 1 hadron in 300 and has a 90% e^+/e^- detection efficiency.

3.4.3.2 The Electromagnetic Calorimeter

The purpose of the HERMES electromagnetic calorimeter [79] is actually four fold. Its functions are (i) providing a first level trigger for scattered positrons, (ii) separating positrons from pions with a rejection factor of more than 100 in the reconstruction, (iii) providing an energy measurement of the scattered positrons and the produced photons from radiative processes and e.g. π^0 decays and (iv) to give a position determination for the scattered leptons and produced photons. The detector consists of two times 420 radiation resistant F101 [80] lead-glass blocks[6] read out by photo-multipliers. The blocks measure $9 \times 9 \times 50$ cm^3 and are organized in a 42 by 10 matrix (figure 3.8). Both top and bottom detectors are mounted on a movable frame which can be moved away from the beam pipe in vertical direction during injection, beam tuning and beam dump. This is done to prevent too much radiation damage to the lead glass blocks during these times.

Including the effect of the pre-shower hodoscope (see later), the energy resolution of the calorimeter can be parameterized by [79] :

$$\frac{\sigma(E)}{E}(\%) = \frac{5.1 \pm 1.1}{\sqrt{E(\text{GeV})}} + (2.0 \pm 0.5) + \frac{10.0 \pm 2.0}{E(\text{GeV})} \tag{3.5}$$

Particle identification information from the calorimeter is obtained by making the ratio of the energy observed in the calorimeter E_{calo} and the reconstructed

[6]The Chemical composition of this material is Pb_3O_4 (51.2 %), SiO_2 (41.5 %), K_2O (7.0 %) and Ce (0.2 %). Its density is 3.86 g/cm^3 with a radiation length of 2.78 cm.

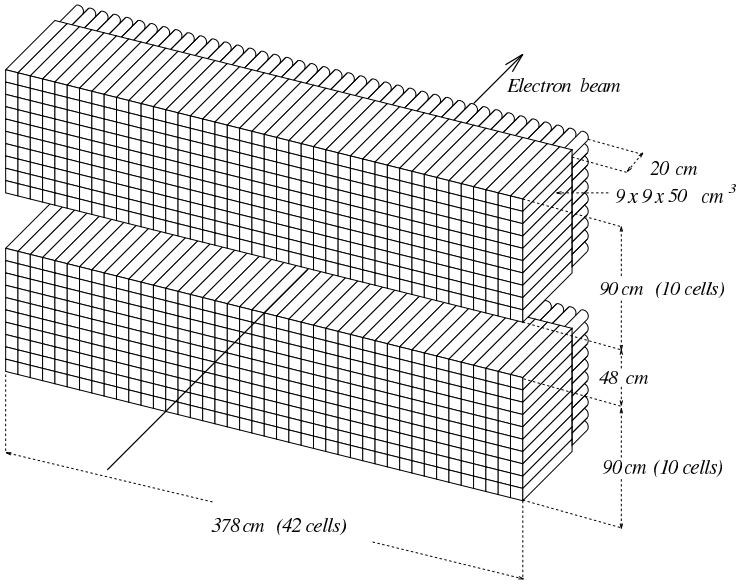


Figure 3.8: Schematic view of the HERMES electromagnetic calorimeter.

momentum p. The E_{calo}/p distribution for leptons is sharply peaked around 1, whereas for hadrons it is a lot broader and tends towards lower values. Figure 3.11 illustrates this.

3.4.3.3 The Hodoscopes

The hodoscopes are mainly responsible for generating the trigger signals for the experiment. To a lesser extent one of the hodoscope detectors also works as a particle identification detector. Hodoscopes consist basically of scintillator material read out by a photomultiplier (PM). In the default acceptance of HERMES, there are 4 hodoscopes H0, H1, H2 and the muon wall (HM) at the back (see further on). Figure (3.9) shows a technical drawing of H1 and H2. Per detector half, these two detectors are segmented in 42 paddles which overlap 3 mm. The paddles measure 9.3×91 cm^2 and are 1 cm thick [81]. H0 consists of one large 3.2 mm thick, 60×20 cm^2 scintillator panel for each detector half which is read out by two PMs using twisted light guides [82]. Figure (3.10) clarifies this. The material used for H1 and H2 scintillators is BC-412, whereas H0 is constructed from BC-400, both from Bicron Co[7]. The typical light attenuation length of these materials is about 200 cm. The H0,H1 and H2 detectors are read out by Thorn/EMI 2" 9954SB07 photomultiplier tubes. They have a rubidium bialkali photocathode with a quantum efficiency of 26 % at a wavelength of 430 nm and a rise time of about 2 ns.

In front of H2 a lead radiator is mounted in which electrons will initiate an electromagnetic shower, thus causing a larger signal in the H2 detector. Normally, minimum ionizing particles deposit about 2 MeV in 1 cm of scintillator

[7] http://www.bicron.com

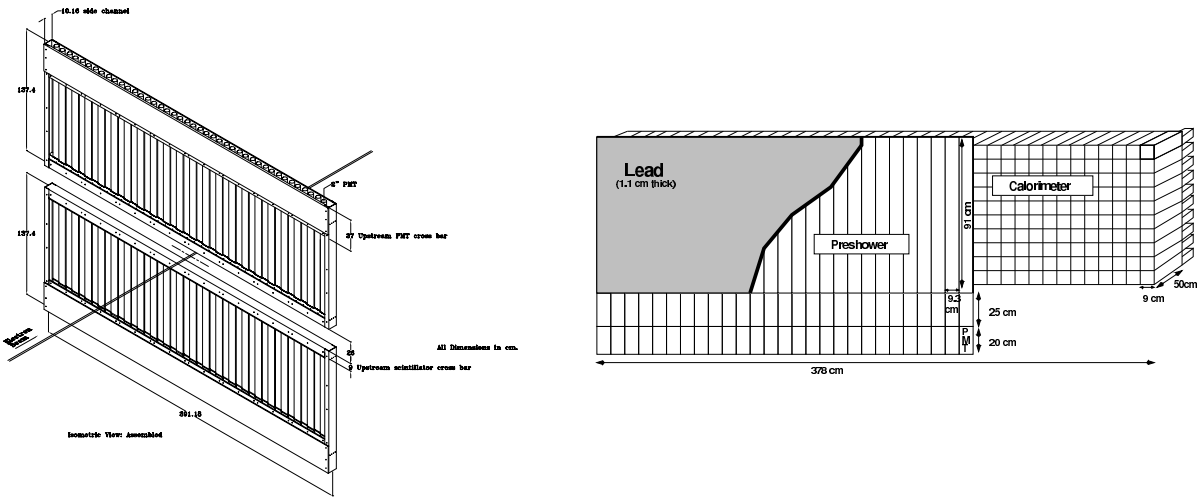


Figure 3.9: An isometric view of the H1 hodoscope and a more schematic view of H2 with the lead sheet in front of it.

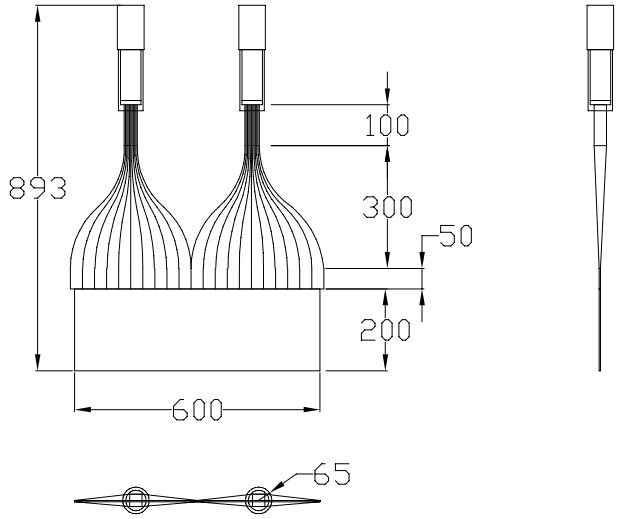


Figure 3.10: Technical drawing of a H0 detector. The 60 $\times$ 20 cm^2 scintillator is read out by 2 PMT's.

material. For electrons that have showered in the lead radiator, this can be up to 100 MeV. Figure 3.11 shows the response in MeV. Later on we will go into more detail on the role of these detectors in the HERMES trigger setup.

3.4.3.4 The PID algorithm

We have to somehow combine now the information from the individual PID detectors. This is done using Bayes' Theorem from statistics. If $\mathcal{P}(H_{l(h)}|E,p)$ is the probability for a track being a lepton (hypothesis H_l) or a hadron (hypothesis H_h) provided that the track had a momentum p and deposited energy E in the detector concerned, then we can write :

$$\mathcal{P}(H_{l(h)}|E,p) = \frac{\mathcal{P}(H_{l(h)}|p)\ \mathcal{P}(E|H_{l(h)},p)}{\mathcal{P}(E|p)} \tag{3.6}$$

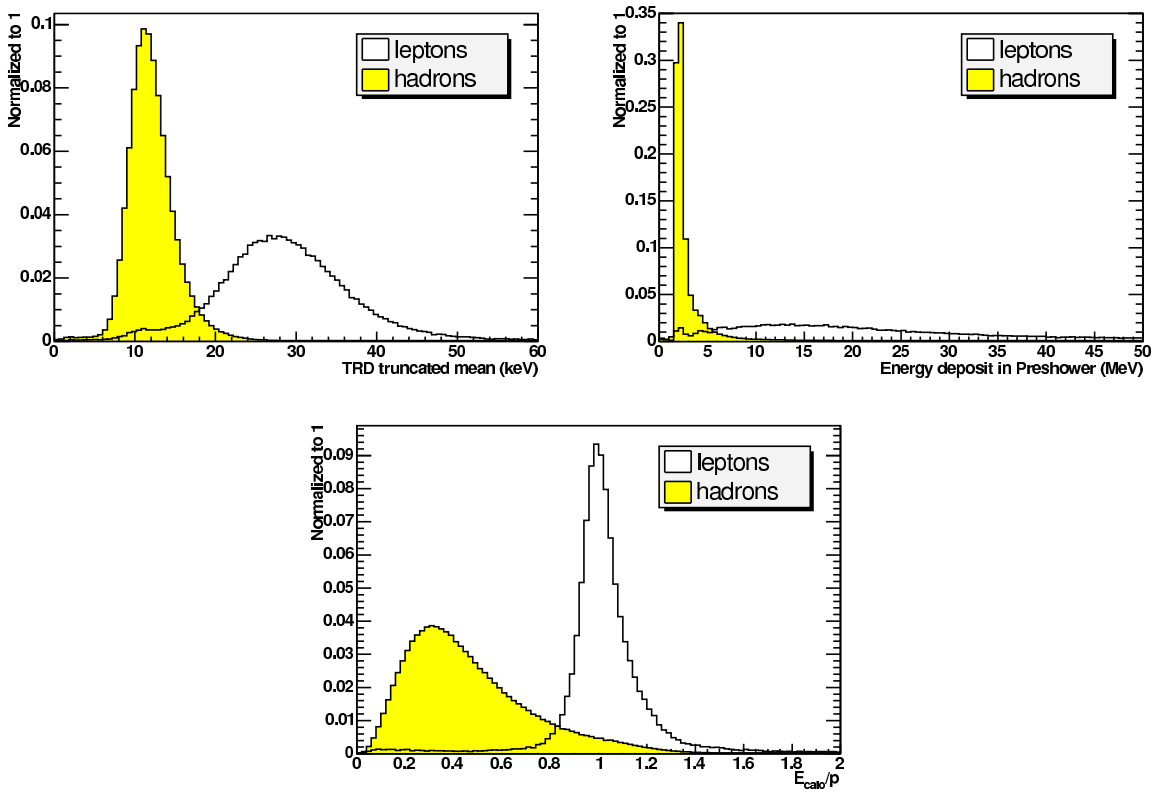


Figure 3.11: Responses of the above mentioned particle identification detectors. The separation between leptons and hadrons was made for each detector using hard cuts on the other PID detectors. The data used are from the HERMES 03c production. The distributions for both hadrons and leptons were normalized to unity.

So we have to calculate the probabilities $\mathcal{P}(E|H_{l(h)},p)$ that a lepton (hadron) with momentum p will deposit an energy E in the detector. These are called the **parent distributions** [83] and are extracted from HERMES data imposing hard cuts on the detector responses of the PID detectors other than the one considered. The probability $\mathcal{P}(E|p)$ that a particle with momentum p will deposit an energy E in the detector, can written as

$$\mathcal{P}(E|p) = \sum_{i=l,h} \mathcal{P}(H_i|p)\mathcal{P}(E|H_i,p) \tag{3.7}$$

This means that we now only have to know the *prior probabilities* $\mathcal{P}(H_{l(h)}|p)$. These are in fact the lepton (hadron) fluxes $\phi_{l(h)}$. Since particle identification is needed for these flux factors, we cannot determine them directly. However an iterative method is used [83]. By taking the logarithm of the ratio of the two particle hypotheses, we construct the PID numbers.

$$PID' = \log_{10}\frac{\mathcal{P}(E|H_l,p)\mathcal{P}(H_l|p)}{\mathcal{P}(E|H_h,p)\mathcal{P}(H_h|p)} = PID - \log_{10}\Phi \tag{3.8}$$

Where

$$\Phi = \frac{\phi_h}{\phi_l} = \frac{\mathcal{P}(H_h|p)}{\mathcal{P}(H_l|p)} \tag{3.9}$$

The fluxfactors are shown in figure 3.12 as a function of momentum and scattering angle θ for both positive and negative particles.

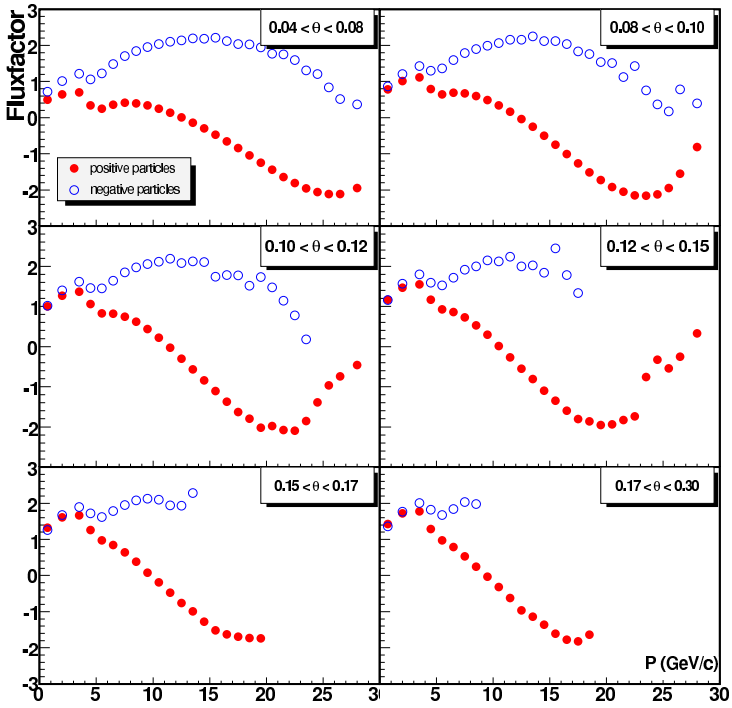


Figure 3.12: Flux factor ($\log_{10}\Phi$) calculated with the HERMES PIDLib v2.41 for the 00c1 data. The calorimeter threshold was 3.5 GeV for the data sample.

The influence of the flux factor correction can be seen from figure 3.13. By simple addition, we can now combine PID information from different combinations of detectors. In HERMES we define the following PID numbers :

- **PID2 =** $\mathrm{PID}_{cal} + \mathrm{PID}_{pre}$
- **PID3 =** $\mathrm{PID}_{cal} + \mathrm{PID}_{pre} + \mathrm{PID}_{cer}$
- **PID5 =** $\mathrm{PID}_{trd} = \sum_{i=1}^{6} \mathrm{PID}_{trd,i}$

Figure 3.14 shows PID3 versus PID5, where the disentanglement of hadrons and leptons is even clearer.

Up till now we only discussed detectors and algorithms which enable us to distinguish between hadrons and leptons in general. HERMES is also equipped with detectors which can tag individual particle types such as muons, pions, kaons and protons.

3.4.3.5 The muon detection system

Part of the HERMES Charm Upgrade proposal [84] was a muon detection system. Inside the standard HERMES acceptance there is the muon hodoscope

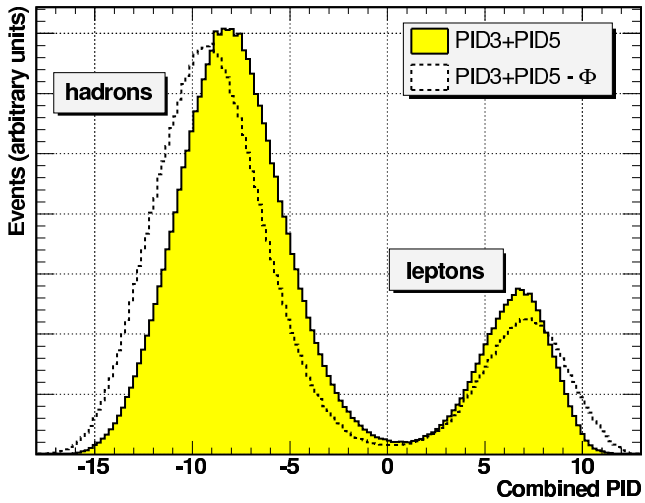


Figure 3.13: The PID3+PID5 distribution which is normally used for hadron/lepton separation. One can see the somewhat clearer separation between leptons and hadrons when using the flux factor correction. The sample of events is taken from the HERMES 00c1 unpolarized hydrogen dataset.

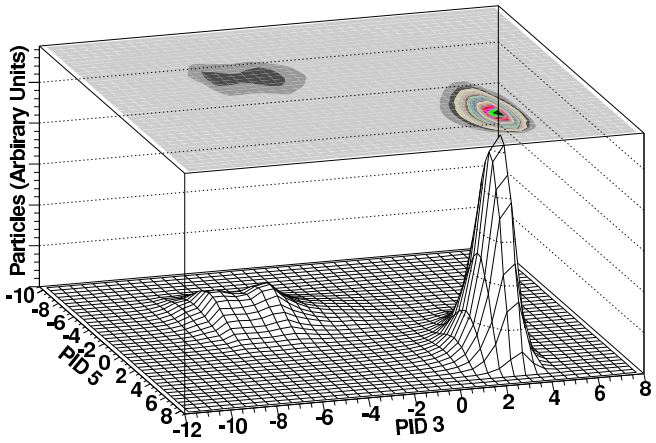


Figure 3.14: PID3 versus PID5. One can see a clear distinction between hadrons (negative PID) and leptons (positive PID).

(HM) which is located behind a 1m thick iron absorber (figure 3.4). From a GEANT Monte Carlo simulation the pion punch through probability[84] was determined to be about 5 to 10 %, depending on the pion energy. The muons are detected behind the iron wall using scintillator paddles organized in both x and y directions. They are 13 cm wide and organized in two layers with a 3 cm overlap between neighboring paddles. Figure 3.15 illustrates this. This way a spatial resolution for the muon detection can be obtained on the order of a couple of cm. A better resolution is not required since this corresponds also to the average deviation due to multiple scattering inside the iron absorber.

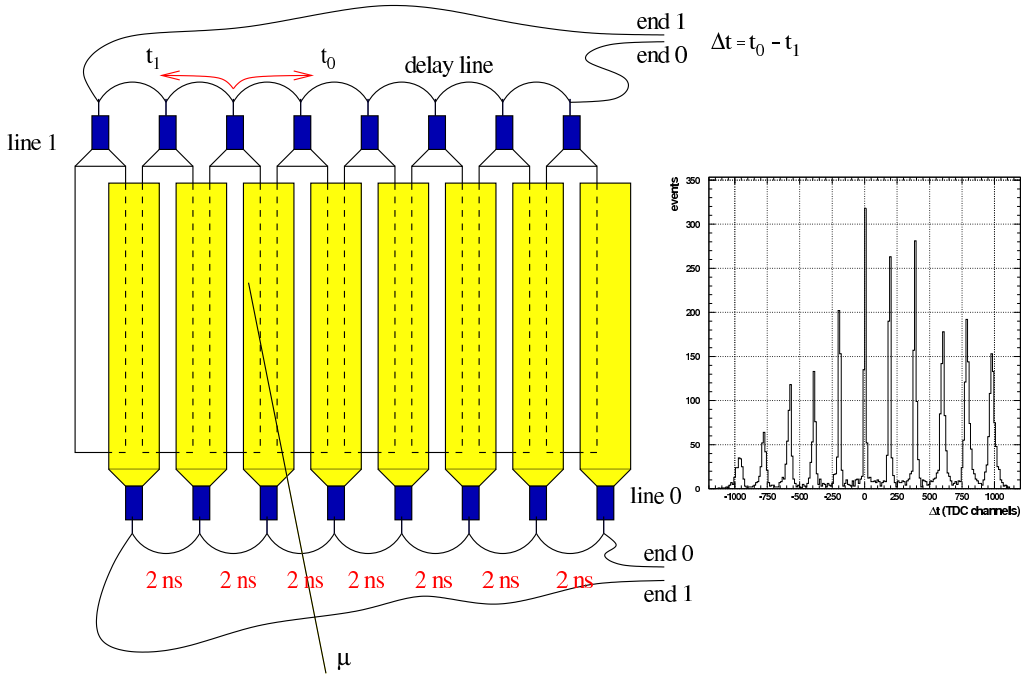


Figure 3.15: Schematic view of the layout for the muon hodoscopes with the delay line setup. The time difference spectrum is shown in raw TDC channels, set to 50 ps per channel.

Outside of the standard HERMES acceptance, a *Wide Angle Muon hodoscope* (WAM) was added as indicated in figure 3.4. It covers scattering angles between ± 170 and ± 270 mrad like the DVCs and the Lambda Wheels. For this detector muon filtering is accomplished by the iron of the spectrometer dipole magnet. The WAM only has paddles in the x direction, so no y coordinate determination can be done with it. Together with the front muon hodoscope[8] which is mounted just before the magnet it is used to trigger on muons coming from J/ψ decays under large angles.

As is visible in figure 3.15, the muon hodoscope and the WAM are read out by 1 simple delay line per detector 'layer'. The PM's are interconnected by 2 ns signal cables. Both ends of one delay line are then fed into a TDC (Time to Digital Convertor). This way the number of needed channels and hardware in general is heavily reduced. By taking the difference in the arrival time of the signals at both ends of a line, one can obtain the exact paddles which have fired (see spectrum in figure 3.15) and thus the hit position of the muon. In this way a very clean signal for J/ψ particles can be obtained as can be seen from figure 3.16.

3.4.3.6 The Ring Imaging Čerenkov Detector

Since 1998 the threshold Čerenkov detector was replaced by a Ring Imaging Čerenkov Detector. This enables HERMES to separate between different

[8]Also called J/ψ hodoscope.

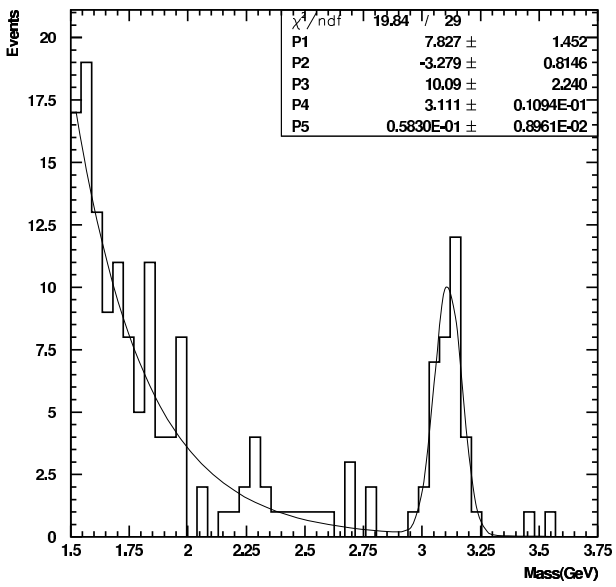


Figure 3.16: A clear J/ψ peak can be observed from Muon Hodoscope data. The resonance mass, 3.111 $\pm$ 0.011 GeV, is very close to the world average of 3.096 GeV.

hadron types, namely pions, kaons and protons over a momentum range of 2 to 15 GeV. When a particle enters a medium with a velocity greater than the

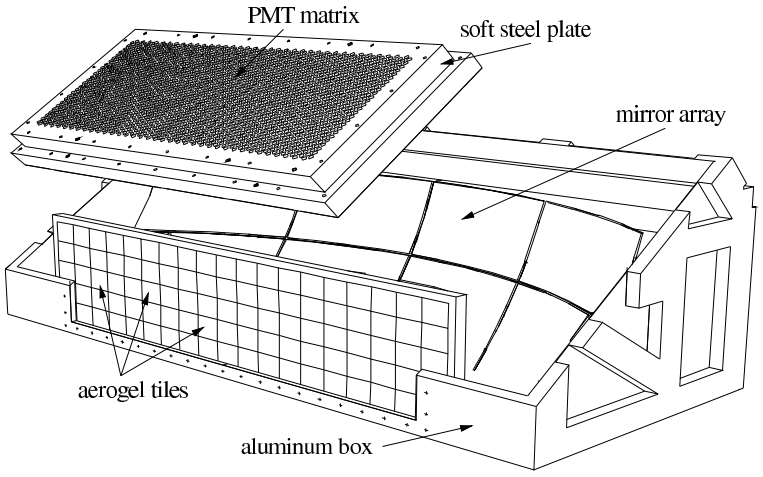


Figure 3.17: A schematic view of the RICH detector.

speed of light in that particular medium ($v > c/n$, with n the refractive index of the medium), it will emit Čerenkov light under an angle θ_c with the trajectory of the particle. This angle is given by

$$\cos\theta_c = \frac{c}{nv} = \frac{1}{\beta n}. \tag{3.10}$$

Depending on the medium (refractive index n) and the velocity and therefore mass of particle ($\beta = v/c$ or $\gamma = 1/\sqrt{1-\beta^2}$), the momentum dependence of θ_c will differ. This can clearly be seen from figure 3.18. The threshold velocity is given by $\beta_t = 1/n$ in which case $\theta_c = 0$. The HERMES RICH is the first one in the world that makes use of two radiators. Particles enter through a

wall of silica aerogel [85] tiles (see figure 3.17) and subsequently pass through a space filled with C_4F_{10}. Properties of these materials are displayed in the table in figure 3.18. The aerogel radiator is used for hadron separation at low

	Aerogel	C_4F_{10}
p_{thres} (GeV)		
π	0.6	2.7
K	2.0	9.4
p	3.8	17.9
n (λ = 633 nm)	1.0304	1.00137
$\beta_t\gamma_t$	4.03	19.10

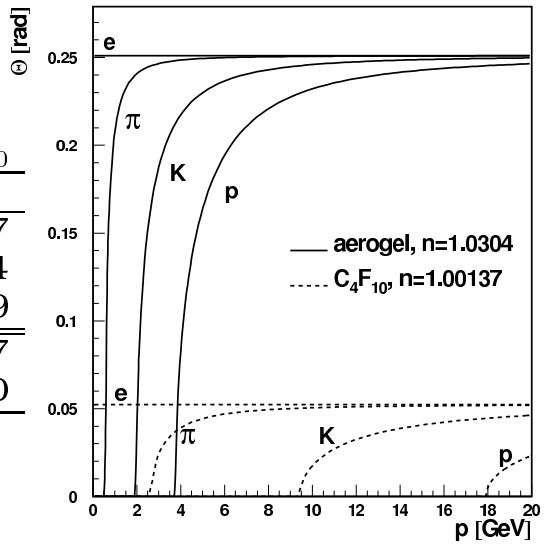


Figure 3.18: The Čerenkov angle versus momentum for different particle types and the two radiators used in the HERMES RICH: aerogel and C_4F_{10}. The table displays the values for the threshold momenta together with the refractive index for the two radiators and the threshold $\beta\gamma$ value for Čerenkov radiation.

momentum, whereas the C_4F_{10} enables identification at higher momentum. In this way the full range of 2 - 15 GeV is spanned. The Čerenkov photons are emitted in a cone around the particle track. This cone is reflected from a mirror and projected onto a matrix of 1934 photo-multipliers per detector half (figure 3.17). The projection of the cone is approximately elliptical around the 'virtual track hit point', this is the point where the track would hit the PM matrix if it were reflected from the mirror. Two different algorithms[86] are used to assign the hits to a Čerenkov cone of a certain track and thus reconstruct the Čerenkov angle θ_c [87], [88]. They are called *Indirect Ray Tracing* (IRT) and *Direct Ray Tracing* (DRT). For IRT, we start from a track and a hit in the RICH detector plane and seek to reconstruct the angle θ_c with respect to the track. For DRT we also start with the track parameters, however, we construct for each particle hypothesis the ring on the PM matrix. The ring which matches best the real data determines the hadron type assigned to the track. Although the RICH hadron tagging works very well, it is subject to misidentification and non-perfect efficiency. An unfolding method, based upon the GEANT MC description of the RICH provides means of dealing with these misidentifications. Since this topic plays a crucial role in the analysis presented further on in this thesis, we will postpone it until the subject is at hand.

3.4.3.7 RICH particle identification

Based on either one of the two methods, the RICH detector identifies a particle track as being either a electron, a pion, a kaon or a proton (or their respective anti particles). Next a quality parameter rQ_p is defined

$$rQ_p = \log_{10} \frac{L_1}{L_2} \tag{3.11}$$

for IRT and DRT where L_1 and L_2 are the highest and second highest likelihood values. For rQ_p equal to 0, both likelihoods are identical and the RICH has to leave the particle unidentified (X). Both IRT and DRT methods have their strengths and weaknesses, so a mechanism was developed to select the best method based on the track information and event topology. This *RICH PID Scheduler* or RPS is described in detail in [89].

3.5 Triggering

Several detectors go into the HERMES triggering system. However, the main components are the hodoscopes H0, H1 and H2 and the calorimeter. After discrimination and delay, the signals arrive at programmable lookup units (PLU) responsible for the logic combinations that define the triggers per detector half. Afterwards an **AND** or **OR** between both detector halves makes up the final trigger. These final triggers are then fed to scaler units and to the pre-scalers. These see to it that the rest of the data acquisition (DAQ) system only receives as much triggers as it is able to handle. The pre-scale factors as well as the logic combinations inside the PLU can be programmed. After this stage, the triggers are passed on to the gate generator to provide the ADCs (analogue to digital converters) with their trigger or gate signals. The main physics trigger (or *DIS trigger*), designated with number 21 is built up in the following way :

- The hodoscope detectors H1 and H2 give a signal above threshold, which is set to enable detection of minimum ionizing particles.
- H0 in the front region of the experiment has to give a signal above threshold. In this way the background coming from the proton beam is reduced. These protons travel in the opposite direction, and will have different flight times with respect to particles coming from the lepton beam.
- The total deposited energy in two adjacent columns of the calorimeter is above 1.4 GeV for polarized running and 3.5 GeV for unpolarized target operation.

This trigger 21 maximizes the probability that indeed a real deep inelastic scattering event took place. Typical rates are about 60 Hz at the beginning of a HERA e-fill ($\sim$ 40 mA) for polarized ^{1}H operation.

3.6 Data handling at HERMES

Data read out from the HERMES experiment is collected by the event builder program which runs on the online data acquisition computers. The information is written both to tape and hard disk in EPIO (Experimental Physics Input Output) format. Next, the HERMES DeCoder (HDC) converts these raw ADC (Analogue to Digital Convertor) and TDC values to pulse heights and positions. All this is stored in ADAMO/DAD[9] format. Determination of particle tracks, momenta and types is done by the HERMES reconstruction program (HRC). Next to particle tracks and other event level data, also data which is read out every couple of seconds is stored. The slow control data involves high-voltage values, target polarization, beam parameters and so on. The time period in the datastream for which one entry is made in the slow control tables is called a **burst**. In each burst one has several events. During online data taking, the data stream is written to a file untill it reaches about 500 MBytes, after which a new datafile is started. The corresponding period in the data handling is called a **run**. The HERMES μDST writer program combines and synchronizes the event tracks with the slow control information and summarizes everything in so-called μDSTs (DST stands for Data Summary Tape). The resulting runfiles are then used for physics analysis.

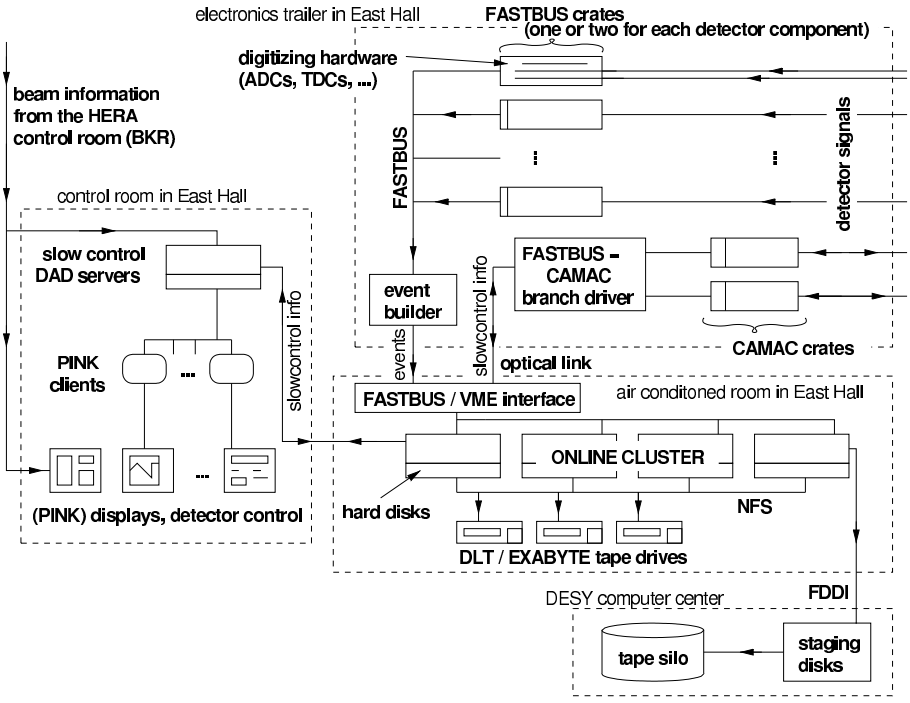


Figure 3.19: Diagram of the entire data acquisition flow. Figure is taken from Ph.D. thesis of F. Menden [91].

[9]ADAMO [90] stands for Aleph Data Model, a database format developed for the former LEP experiment. For the HERMES experiment, this format was extended to DAD (Distributed Adamo Database) enabling ADAMO data to be available between different processes and different computers on the network.

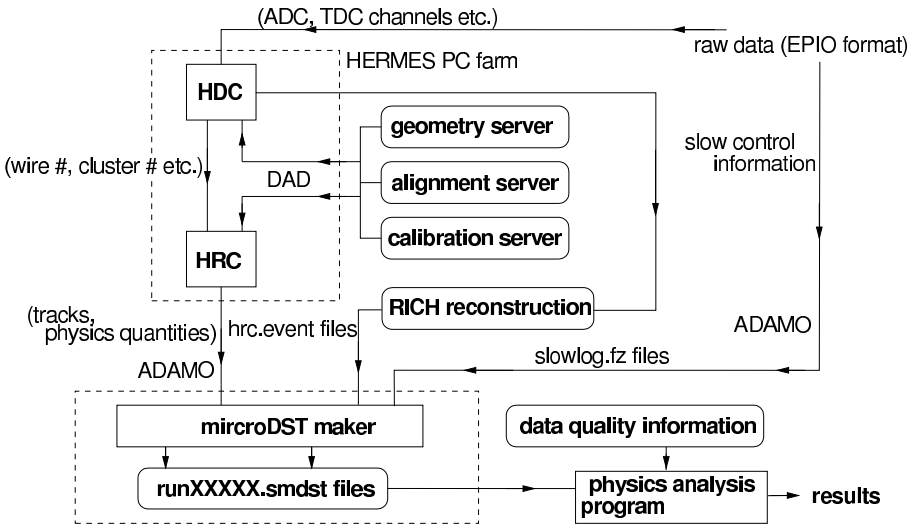


Figure 3.20: Diagram of the offline reconstruction software of the HERMES experiment. Figure is taken from Ph.D. thesis of F. Menden [91].

The mapping of the detectors, geometry and also the calibration are crucial at the decoding level. This information is stored in DAD servers for each of the data productions. During online data taking detector calibrations are given based upon previous running. The resulting online data production is called the 'a'-production. From the detector responses in this a-production, new calibration constants and detector efficiencies can be determined. This information subsequently serves as input for the 'b' production. This iterative process results in an improvement of the quality of the data.

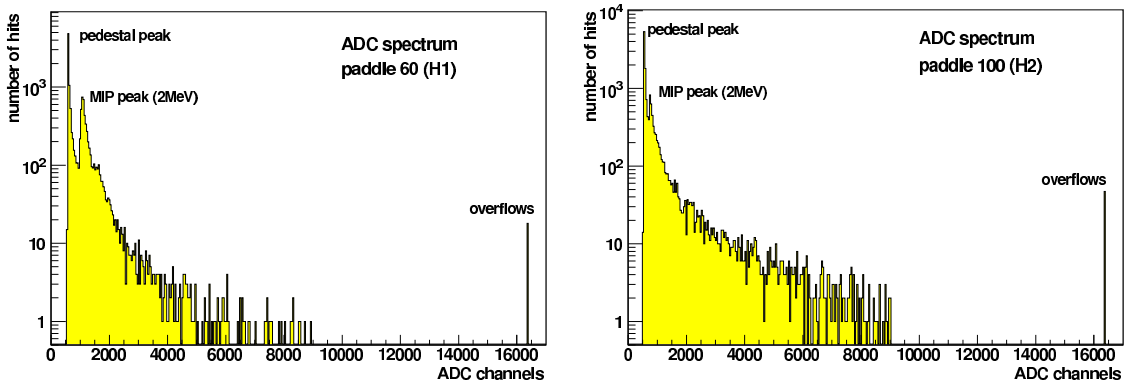
3.7 Calibration of the Hodoscopes

Since the preshower hodoscope is used in the HERMES PID scheme, it must be known what the energy deposition of particles in this detector is. Therefore a good calibration is necessary. Until 2001 a calibration program was in use that functioned well for the old computing setup, but was incompatible with the new setup. At the beginning of this PhD, a program was developed for this task.

3.7.1 Fitting procedure

In figure 3.21 typical decoded ADC spectra for an H1 and H2 paddle are shown. We clearly see in the case of H2 that the peak corresponding to energy deposition of minimum ionizing particles (MIPs) is less pronounced compared to the larger energy depositions. This is due to the presence of the lead preshower in front of the scintillator paddles. Light leptons induce electromagnetic showers depositing up to 100 MeV in H2. We also can see that the gain for the H2 detector is typically set lower than for H1, allowing for the

processing of the large energy depositions. Digitization of both the H1 and H2 signals is done by LeCroy FastBus 1881M ADCs with a 13 bit dynamic range and 50 fC least count above pedestal. As we know that MIPs deposit



(a) Typical H1 paddle ADC spectrum (b) Typical H2 paddle ADC spectrum

Figure 3.21: Data for this was taken from the 2000 running period. Clearly visible are the pedestal peak and the peak corresponding to the energy deposition of minimum ionizing particles in 1 cm of scintillator ($\sim$ 2 MeV). For H2 the MIP peak is less pronounced because of the lead sheet in front of the detector. We also see that the gain of the H2 PMTs is set a little lower. Also the overflows are clearly visible in channel 16383.

2 MeV in 1 cm of plastic scintillator, we can calibrate the spectra for each paddle by determining the pedestal and the MIP position. The pedestal for each paddle is never more than a couple of ADC channels wide. Therefore it suffices to take the spectrum maximum, which for the hodoscope spectra is always the pedestal, and determine the average within a couple of channels around it to fix the position. In the calibration program, these pedestals are determined for each run from the `hdc.hbook.gz` files which are output by the HRC program. They contain the raw decoded spectra as shown in figure 3.21 for each paddle. Next, the pedestal is subtracted from the track's ADC value for both hodoscope hits. Before determining the position of the MIP peak one has to apply a light attenuation correction on track level. For this correction we need the propagation distance of the scintillation light through the plastic. The multiplicative correction factor $\mathcal{A}(y,\lambda)$ takes the shape of an exponential with a 2nd order polynomial :

$$\mathcal{A}(y,\lambda) = \frac{a(\lambda)}{[a(\lambda) + b(\lambda)\, y^2] \cdot e^{-y/c(\lambda)}} \tag{3.12}$$

where y is the distance from the hit position to the PM. The $a(\lambda)$, $b(\lambda)$ and $c(\lambda)$ parameters were fit based on measurements of the attenuation versus position at Caltech and simulations of the signal propagation with attenuation and reflection from the end cap. They depend quadratically on the light attenuation length λ. Once the ADC spectrum is corrected for both the pedestal and the light attenuation, the MIP peak is determined [92]. The algorithm for determining the position of the peak is as follows:

1. To improve the fitting with low statistics, the spectra are rebinned, where

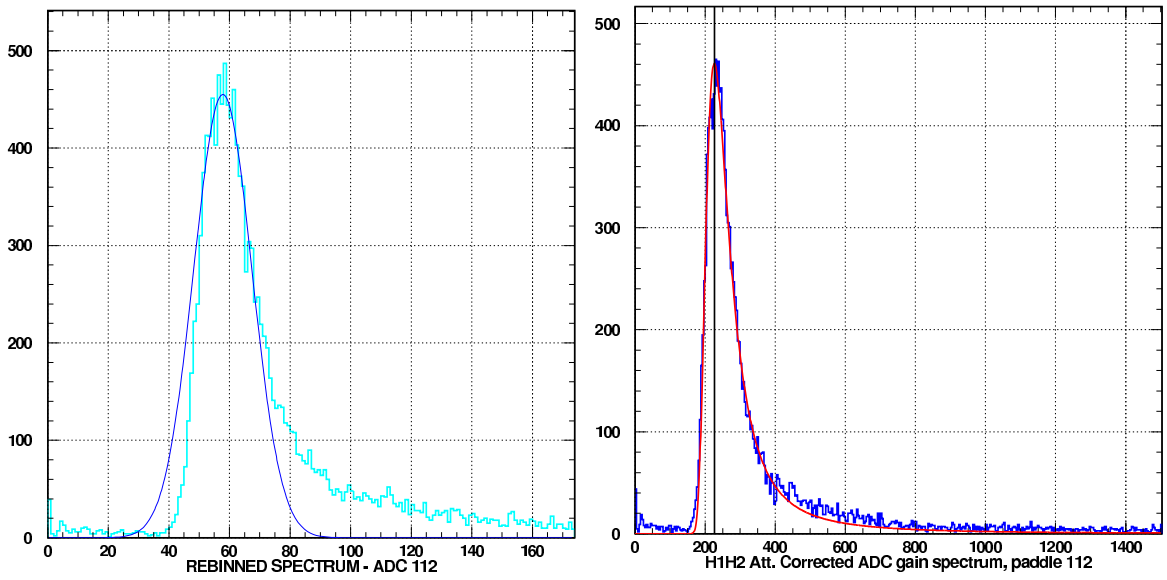
4 neighboring bins are added together, resulting in a spectrum as plotted in figure 3.22.

2. The data taken to determine the pedestals includes the ADC information in *unsparcified* mode at the beginning of each run. In this mode the entire 64-channel ADC is read out and its information put into the datastream, even though not all channels have seen a signal. This way we can determine the pedestals. This is done the first couple of bursts in each run. However, during normal data taking in the run itself, the ADCs are operated in *sparcified* mode in which case only channels which have seen a signal above threshold are put into the eventstream. This is the data that is used for calibration of the MIP peak position. Since the pedestal is already subtracted from the ADC values, the maximum in the sparcified spectrum corresponds roughly to the position of the MIP peak. The position of this maximum is given as a start parameter to a Gaussian fit procedure. The width of the Gaussian is initialized by running through the spectrum left and right until the height is reduced to 1/3 of the value at the maximum (the so-called valley points). The difference between the peak and these valley points is given as an initial value for the width of the Gaussian.

3. Next, a Gaussian is fitted twice using χ^2 minimization to determine the peak position. The start parameters of the second Gaussian are the fit parameters of the first one. Each time the fit is restricted around the maximum, the first time in between the two valley points. The second time it is fit within 1 σ of the Gaussian of the first iteration. This way we are able to determine quite precisely the position of the peak, even though the distribution itself is clearly not Gaussian.

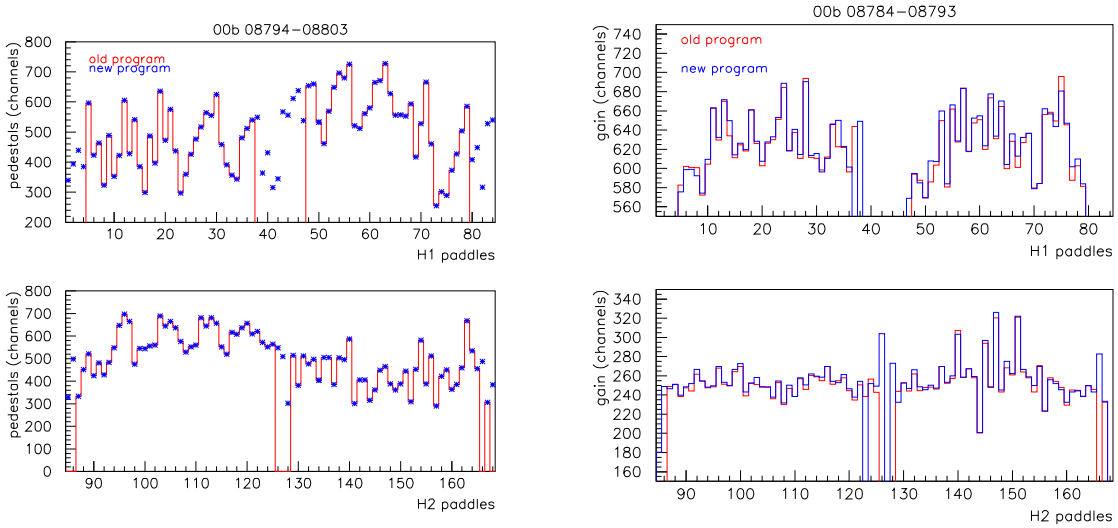
This algorithm is implemented in a program called `hoax`, or *HOdoscope Analysis and data eXtraction program*. Its results are shown in figure 3.22(a). It is in fact taken over from the old calibration program that was in use before 2001, with some minor adjustments to improve the fitting. A comparison for the same data can be seen in figure 3.23.

Later on a new calibration program (`hoaxNG`) was developed where the standalone fitting routine was replaced by the standard CERNLIB MINUIT minimization package [93]. Also the Gaussian fitting routines were abandoned and replaced by a Landau shape, although the Gaussian is still used as a fallback solution should the Landau fit fail. A similar algorithm is used to obtain initialization parameters as described above. The resulting fit for the same spectrum as in figure 3.22(a) can be seen from figure 3.22(b). The Landau fit does a far better job at describing the whole MIP peak and gives better determination of the peak position as well as its uncertainty (which results from MINUIT calculations). In figure 3.24 we see a comparison between the two implementations for the preshower detector where the determination of the exact gain is most important. It must be noted that for spectra with low statistics it is observed that the Landau algorithm is more likely to fail than



(a) fitted with Gaussian (b) fitted with Landau

Figure 3.22: Real examples of calibration spectra fitted with a Gaussian distribution as in the `hoax` program. On the right a fit with a Landau distribution is shown as in the new program (`hoaxNG`).



(a) comparison of pedestals (b) comparison of gains

Figure 3.23: Comparison between the old calibration program (pre 2001) and the new program for pedestal extraction of data and gain determination.

the Gaussian one. The PMT gains seem to be systematically lower using the Landau fitting procedure. However this difference is small (order of ~ 5 ADC channels). Furthermore, as one averages the hodoscope gains over a certain period in which the gains remain stable before loading the constants to the calibration server, this difference is well within acceptable limits.

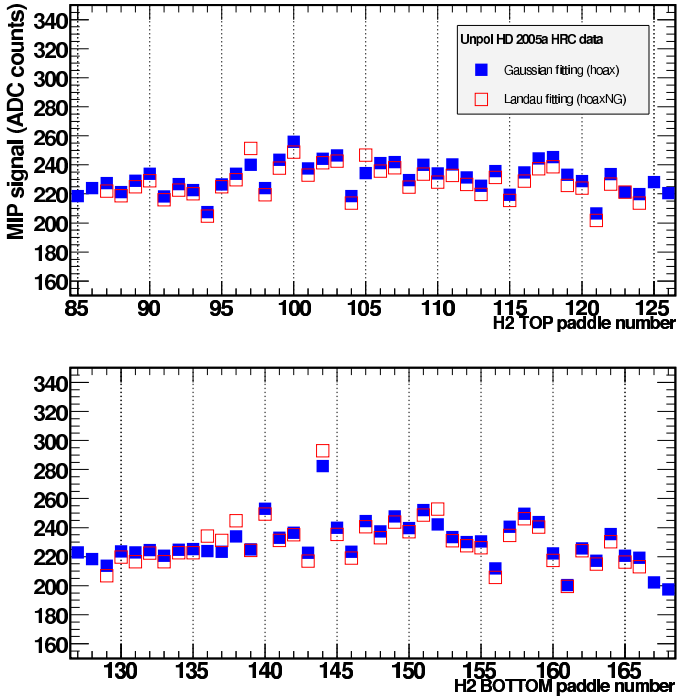


Figure 3.24: Comparison for the gain determining algorithms using Gaussian fitting (`hoax`) program and using Landau fitting (`hoaxNG`). The data for this plot was taking from the year 2005 for the H2 detector. One can see that the paddles are much better gain matched as compared to figure 3.23.

3.7.2 The calibration input

Once the size of the signal above pedestal is known in ADC counts, we can calculate the corresponding gain conversion factor since we know that minimum ionizing particles deposit 2 MeV in the scintillator material. These constants have to be loaded to the HERMES calibration servers (see figure 3.20). One defines the `Gain` as :

$$\texttt{MIP(MeV)} = \texttt{Gain} \cdot \texttt{MIP(ADC channels)} \tag{3.13}$$

For a number of historical reasons, however, what is loaded to the calibration server is not really these gain conversion factors but correction constants to a predefined gain. The reconstruction program HRC implements the following :

$$\texttt{PulsHodo} = \texttt{ADC value} \,/\, \texttt{Gain_Server} \tag{3.14}$$

$$\texttt{PulsPre} = \texttt{ADC value} \,/\, \texttt{Gain_Server} \tag{3.15}$$

where `Gain_Server` is defined by

$$\texttt{Gain_Server} = \texttt{1MeV_Channel} \cdot \texttt{GainCorr} \,/\, 0.002\ \text{GeV} \tag{3.16}$$

The `1MeV_Channel` corresponds to 250 for H1 and 100 for H2, these were chosen so that the MIP signal corresponds to 500 channels above pedestal for H1 and 200 channels above pedestal for H2 for nominal high voltage settings. The

GainCorr is exactly what the load program to the calibration server needs. So we somehow have to extract the GainCorr from the Gain defined in equation 3.13. From equations 3.13 and 3.14 and 3.15 we easily see

$$\texttt{Gain_Server} = 1/\texttt{Gain} \tag{3.17}$$

And therefore taking into account equation 3.16

$$\texttt{GainCorr} = 0.000008 \;/\; \texttt{Gain} \tag{3.18}$$

for H1 and

$$\texttt{GainCorr} = 0.00002 \;/\; \texttt{Gain} \tag{3.19}$$

for H2. These gain corrections are the values which are passed on to the server load program, together with the pedestal information.

Chapter 4

Experimental charge and flavor separated hadron multiplicities

This chapter is devoted to the extraction of hadron multiplicities within the HERMES acceptance. This means that no correction will be made to correct for the limited efficiency of the HERMES spectrometer or to extrapolate these results to a 4π geometry. For reasons that will be explained further on, we will postpone correcting the data for QED radiative effects as well. Nevertheless a number of issues besides obtaining the hadron spectra, will already be dealt with here: PID corrections, background corrections etc...

4.1 Data selection

4.1.1 Data Quality Cuts

In this thesis we make use of the unpolarized hydrogen data from the year 2000 dataset, more specifically the 00c1 production. The data quality (DQ) cuts that were applied are given below, together with the variables from the μDST production concerned. For reference, the DST database table names are also mentioned.Despite the fact that the main physics goal of HERMES is measuring polarized distributions and quantities, this analysis cares not about target nor beam polarization. The data quality cuts will therefore be a bit less stringent than in the latter case.

All selection criteria discussed here apply on burst level. As soon as a burst does not fulfill one of these restrictions, we discard the whole burst.

Experiment mode

First of all, the run has to be marked analyzable in the experimental logbook. This logbook is maintained by the shift crews during data taking and revised

afterwards to correct for errors.

```
g1Quality.iExpment = 1
```

The target has to be set up for unpolarized running, either normal or high density.

```
g1Upol.GasType = 1
g1Quality.iExpMode = 0 (normal density) or 16 (high density)
```

Experimental dead time

When HERMES accepts a trigger, it takes a finite amount of time to process, digitize and store the event. During this time, the data acquisition is not able to accept any new triggers. This amount of time is called **dead time**, or alternatively, the amount of time during which the DAQ *is* able to accept new triggers is called **live time**, essentially defined as

$$\tau_{live} = \frac{T_{acc}}{T_{gen}}, \tag{4.1}$$

being the ratio between the number of accepted and generated triggers within a given period of time. The requirement for a good live time in the burst considered was also put separately on trigger 21 as this is the DIS trigger for the experiment.

$$50\% < \tau_{live} < 100\%, \qquad 50\% < \tau_{live}^{21} < 100\% \tag{4.2}$$

At a certain moment in 1997, an *artificial dead time correction factor*, τ_{Art} was introduced to correct for events lost in the data stream. We selected only events for which this factor was 1 and no events were lost in the burst.

```
g1DAQ.rDeadCorr > 0.5 && g1DAQ.rDeadCorr <= 1.
g1DAQ.rDeadCorr21 > 0.5 && g1DAQ.rDeadCorr21 <= 1.
g1DAQ.rDeadCorrArt == 1.0
```

Burst length, beam current and luminosity rate

The restriction placed on burst length L_{burst}, beam current I_e and the gain corrected luminosity rate R_{lumi} are to eliminate errors in the reconstruction as well. They span the reasonable ranges of these variables. We require $L_{burst} < 11$ s, $5 < I_e < 50$ mA and $5 < R_{lumi} < 5000$ Hz.

```
g1DAQ.rLength > 0.  && g1DAQ.rLength <= 11.
g1Beam.rMdmCurr >= 5.  && g1Beam.rMdmCurr <= 50.
g1Beam.rLumiRate > 5.  && g1Beam.rLumiRate < 5000.
```

TRD data quality

As the TRD is very important for distinguishing between leptons and hadron, we require that the data for both top and bottom detector are found good by the detector experts. The information there is compiled into one variable :

```
g1Quality.iTrdDQ == 3
```

High Voltage trips, dead calorimeter and H2 blocks

Since the tracking chambers (both FC and BC) are essential to the track reconstruction, we want them to operate at maximum efficiency. This is not accomplished when one of the channels of either the FCs or BCs has tripped and therefore is not at its operating voltage. It takes a certain amount of time as well before a voltage trip in the tracking chambers is recovered. The table

```
g1HVTrip
```

contains all the channels within the current burst that have tripped. We demand here that none are present.

Besides this we will also require that information is available about, and no dead blocks appear in the H2 detector, the luminosity monitor or the calorimeter. These checks are done using the following bit masks in the 00c1 μDSTs.

```
( g1Quality.bCaloDead & 0x80000000 ) == 0 &&
( g1Quality.bCaloDead & 0x01FF01FF ) == 0 &&
( g1Quality.bH2LumiDead & 0x80000000 ) == 0 &&
( g1Quality.bH2LumiDead & 0x2F2F001F ) == 0 &&
```

RICH data quality

Since the RICH is crucial in separating hadron types we specifically want it to function properly for this analysis. This is accomplished by requiring

```
( g1Quality.bCereDQ & 0x00000001 ) == 0
( g1Quality.bCereDQ & 0x00010000 ) == 0
```

where the data quality information is again compiled by detector experts. Further on, some more criteria will follow on the level of an individual track.

μDST production quality issues

Also during the production of the μDST files things can go wrong. Issues like synchronization between the slow control information and the event stream, PID issues and so on... The first and last burst of each run will also be discarded since it can happen that the hardware is not yet fully initialized at the beginning of a run. This general μDST quality is summarized by the following bit masks :

```
!  ( g1Quality.iuDSTbad & 0x00040000 ) &&
!  ( g1Quality.iuDSTbad & 0x1E081FF9 ) &&
!  ( g1Quality.iuDSTbad2 & 0x0000000C )
```

Last we also check the calorimeter threshold to be set to 3.5 GeV since this is the default setting for unpolarized running !

```
g1Quality.rCaloThresh == 0
```

DQ cut	nr of bursts that fail	(%)
Bad Deadtime	116	0.33 %
Bad Beam Current	47	0.13 %
Bad Lumi Rate	10	0.03 %
Bad TRD Quality	123	0.35 %
HV Trips	444	1.27 %
Dead Calo Blocks	971	2.78 %
Dead H2/Lumi Blocks	268	0.77 %
Bad RICH DQ	407	1.17 %
Bad μDST Quality	2887	8.27 %
Total nr of bursts checked		34928

Table 4.1: Burst cuts for HERMES 2000, 00c1 unpolarized hydrogen

Table 4.1 summarizes what percentage of bursts does not fulfill each of the requirements discussed above. Note that this info is only valid for the running conditions in the year 2000 and the 00c1 production, for an unpolarized hydrogen target. So we see it is mainly due to μDST production quality issues that the most data gets thrown away. A further data selection is now done on track and event level.

4.1.2 Event and Track level Cuts

The starting point in this analysis is a sample of DIS events, in which we will subsequently look for hadrons. The following selection criteria were applied to the event kinematics:

Event level cuts

- $E_l^{calo} > 3.5$ GeV
- $P_{rc}^{tot} < 28.0$ GeV
- $Q^2 > 1$ GeV^2
- $W^2 > 10$ GeV^2
- $0.1 < y < 0.85$

First the reconstructed deposited energy in the calorimeter from the scattered DIS lepton is required to exceed 3.5 GeV. This is to match the calorimeter discriminator threshold settings during unpolarized data taking. This threshold is raised compared to unpolarized data taking[1] to avoid overload of the DAQ. In order to demand total momentum conservation in the tracking we require that the total reconstructed momentum does not exceed 28 GeV. The other 3 cuts

[1] where it is 1.4 GeV

that were applied are standard DIS kinematic cuts. With the Q^2 restriction the real deep-inelastic scattering regime is selected, in which the virtuality of the exchanged photon is large enough to see the quarks inside the nucleon. The W^2 cut assures that there is sufficient energy for the fragmentation process. The upper y cut sees to it that QED radiative corrections do not become too large. To illustrate this, figure 4.1 shows the ratio of the DIS Born level cross section to the full DIS cross section with the important correction diagrams included (see further on figure 5.5). One can clearly see that the impact of radiative effects rises steeply above $y = 0.85$.

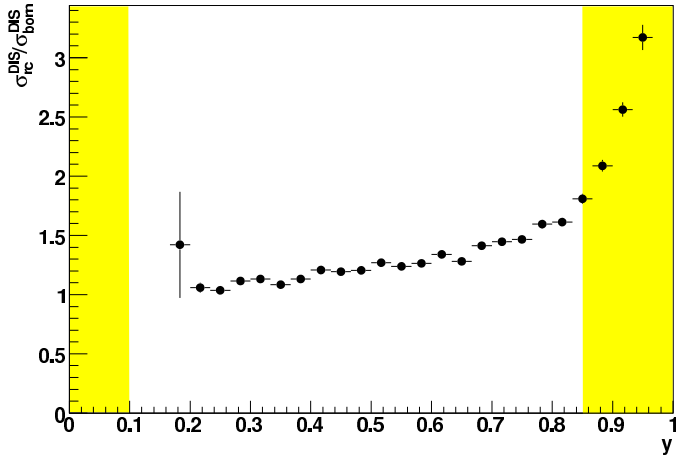


Figure 4.1: Ratio of the born level DIS cross section to the cross section with radiative corrections. This plot was generated using a HERMES DIS Monte Carlo generator equipped with RADGEN [94] to simulate radiative effects.

Track level cuts

In order to obtain optimal lepton - hadron identification, one needs to impose what is called a fiducial cut on the track impact position (x_{Calo}, y_{Calo}) on the calorimeter surface. The total energy deposited by a track is determined by summing the total energy deposition in a 3x3 matrix of calorimeter blocks around the impact point. Obviously at the edges of the detector this procedure loses efficiency. The fiducial cuts are listed below.

- -175.0 cm $< x_{Calo} <$ 175.0 cm
- 30.0 cm $< |y_{Calo}| <$ 108.0 cm

We also need to make sure that the track originated from the target scattering chamber, to cut away e.g. particles that originated from interactions in the collimator material. Figure 4.2 shows the z and $d = \sqrt{x^2 + y^2}$ vertex distributions. One can clearly see that the majority of the tracks indeed originates from within the target cell. However, also a substantial fraction comes from scattering off the fixed C2 collimator which is located directly in front of the

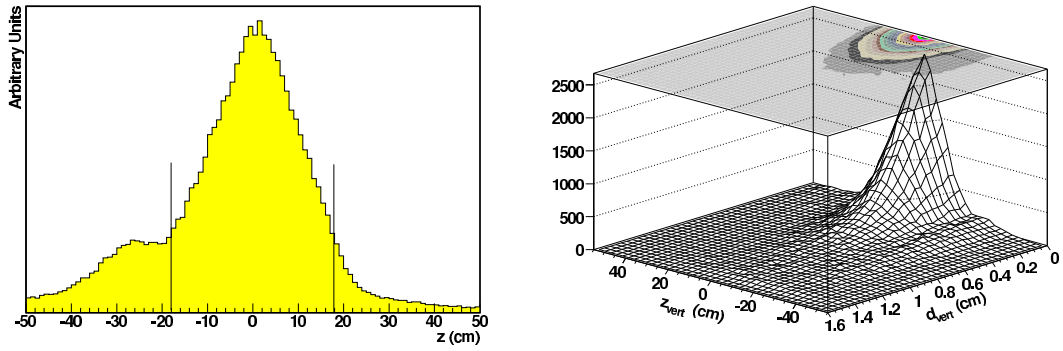


Figure 4.2: Vertex distributions of reconstructed particle tracks. Left the z_{vert} distribution, right a 2D plot of the z_{vert} vertex versus the $d = \sqrt{x_{vert}^2 + y_{vert}^2}$ distribution.

target cell. This collimator protects the cell from synchrotron radiation coming from the first collimator, C1, which sits 2 m upstream of the target cell and direct radiation during beam injection. At this time the C1 collimator, which is consists of 2 movable collimators (C1H, C1V) in horizontal and vertical direction, is opened by 0.5 m, exposing the target region to substantial radiation levels.

The need to cut away particle tracks originating from the C2 collimator imposes a lower cut on the track z vertex position. We also impose an upper cut on the vertex distance from the beam line (d). However, here we will make a distinction between lepton and hadron tracks. Leptons are required to originate from within the target cell. For hadron tracks, it is possible that they originated from a secondary vertex created by e.g. a Λ baryon traveling over some distance, even beyond the target cell, before decaying. To allow hadrons from those decays we open the upper z and d vertex cut. In summary we have :

- for leptons :
  -18 cm $< z_{vert} <$ 18 cm
  $d_{vert} = \sqrt{x_{vert}^2 + y_{vert}^2} <$ 0.75 cm
- for hadrons :
  -18 cm $< z_{vert}$

4.1.3 Lepton Hadron Separation

As already mentioned in section 3.4.3.4, HERMES makes use of a Bayesian algorithm that combines information from several detectors and returns so called PID numbers. In this analysis we use PID3 and PID5 corrected with the flux factor $\log_{10}\Phi$ (see figure 3.12). In figure 3.13 we already showed the corrected PID3+PID5 histogram on which we placed the lepton/hadron cut.

We used

- -100 $< PID3 + PID5 - \log_{10}\Phi <$ 0. : hadrons
- 0 $< PID3 + PID5 - \log_{10}\Phi <$ 100. : leptons

This way we obtain a lepton-hadron separation efficiency of more than 99 % [4] at contamination.

4.2 Hadron selection and PID correction

Efficient hadron detection and identification plays an essential role in this work, so correction for possible misidentifications in the RICH detector is needed. At first, the kinematical restrictions needed for this will be discussed briefly.

4.2.1 Hadron kinematical cuts

The RICH detector is only efficient (see section 3.4.3.6) for hadron momenta from 2 to 15 GeV, we therefore restrict ourselves to this range. For the pions one could in fact have gone down to 1 GeV/c. This however makes the RICH unfolding procedure discussed further on inconsistent. Furthermore the input matrices are only available starting from 2 GeV (see later on). So we also cut away pions between 1 and 2 GeV :

$2 < p_\pi, p_K, p_{prot} <$ 15 GeV

Additionally the RICH provides a quality parameter rQ_p which is defined in section 3.4.3.7, equation 3.11. If this quality parameter is 0, the RICH is unable to identify correctly the particle[2]. These are cut away by demanding

```
smRICH.rQp > 0
```

4.2.2 RICH unfolding

4.2.2.1 About $\mathcal{P}$ and $\mathcal{Q}$ matrices

The RICH hadron identification efficiency is subject to a certain amount of misidentification that varies with particle momentum. This is described by two complementary sets of probability matrices called the $\mathcal{P}$ and $\mathcal{Q}$ matrices. An element in the $\mathcal{P}$-matrix, $\mathcal{P}_t^i$, gives the probability that a particle of true type t is identified by the RICH as a type i, where $i, t = \pi, K, p$ and i can also be

[2] and is in fact a so-called X-particle

X, meaning that the RICH was unable to identify the particle. One can define $\mathcal{P}_t^i$ as

$$\mathcal{P}_t^i = \frac{N_t^i}{\sum_j N_t^j} \tag{4.3}$$

where N_t^i is the number of particles with true type t and identified type i, and write for the $\mathcal{P}$-matrix :

$$\mathcal{P}(p,\mathcal{T}) = \begin{pmatrix} \mathcal{P}_\pi^\pi & \mathcal{P}_K^\pi & \mathcal{P}_p^\pi \\ \mathcal{P}_\pi^K & \mathcal{P}_K^K & \mathcal{P}_p^K \\ \mathcal{P}_\pi^p & \mathcal{P}_K^p & \mathcal{P}_p^p \\ \mathcal{P}_\pi^X & \mathcal{P}_K^X & \mathcal{P}_p^X \end{pmatrix} \tag{4.4}$$

The matrix $\mathcal{P}(p,\mathcal{T})$ depends on the particle momentum p and the event topology $\mathcal{T}$, more specifically the number of tracks in the detector half where the particle tracks resides. This means there can be more than one Čerenkov cone image on the PMT-matrix. These overlapping rings cause difficulties in the reconstruction of the opening angles, therefore decreasing the identification efficiency. The HERMES RICH group has provided $\mathcal{P}$-matrices [95] in the case of a track multiplicity of 1, 2 and 'more than 2' tracks per detector half[3]. Note that the diagonal elements of this $\mathcal{P}$ matrix are the usual detection efficiencies (see equation 4.3). Due to its probabilistic interpretation, the columns of the $\mathcal{P}$ matrix are normalized to 1, i.e.:

$$\sum_i \mathcal{P}_t^i = 1 \tag{4.5}$$

It was mentioned before that two complementary sets of matrices can be used to describe hadron misidentifications. The second set is the so-called $\mathcal{Q}$ - matrix, where an element $\mathcal{Q}_t^i$ gives the probability that a particle identified as being of type i, in fact is of type t. With $\mathcal{P}_t^i$, we 'know' the particle is of true type t, and $\mathcal{P}_t^i$ gives us the probability that it is identified as type i, whereas with $\mathcal{Q}_t^i$ however we 'know' the particle is identified as i, and $\mathcal{Q}_t^i$ gives the probability that it is truly of type t. So,

$$\mathcal{Q}_t^i = \frac{N_t^i}{\sum_s N_s^i} \tag{4.6}$$

and we can write

$$\mathcal{Q} = \begin{pmatrix} \mathcal{Q}_\pi^\pi & \mathcal{Q}_\pi^K & \mathcal{Q}_\pi^p & \mathcal{Q}_\pi^X \\ \mathcal{Q}_K^\pi & \mathcal{Q}_K^K & \mathcal{Q}_K^p & \mathcal{Q}_K^X \\ \mathcal{Q}_p^\pi & \mathcal{Q}_p^K & \mathcal{Q}_p^p & \mathcal{Q}_p^X \end{pmatrix} \tag{4.7}$$

The $\mathcal{P}$ and $\mathcal{Q}$ matrix are related through Bayes' theorem:

$$\mathcal{Q}_t^i = \frac{\mathcal{P}_t^i \cdot \phi_t}{\sum_s \mathcal{P}_s^i \cdot \phi_s} \tag{4.8}$$

[3]Later RICH analysis has provided even more topology cases, distinguishing according to the type of the partner track in the same detector half [96]. However, in this analysis we have restricted ourselves to the 3 types of topologies mentioned, regardless of particle type.

In which ϕ_t is the relative flux of the hadron type t

$$\phi_t = \frac{N_t}{\sum_s N_s}, \qquad N_t = \sum_i N_t^i \tag{4.9}$$

Figure 4.3 shows these quantities versus z.

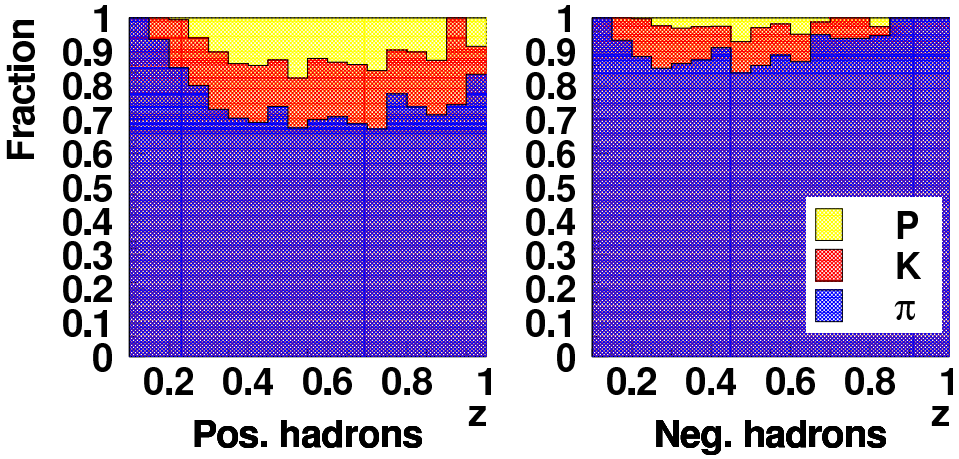


Figure 4.3: The relative particle fluxes of (identified) hadrons versus z. Note that these fluxes apply only for this analysis as they depend on the cuts imposed on the data sample.

It is important to note that the $\mathcal{Q}$ matrix depends on the event sample used, so in fact depends on the type of physics analysis. Its elements depend on the particle flux (4.8), which is obviously different for various physics analysis. Generation of $\mathcal{Q}$ matrices makes things rather complicated and increases the amount of overhead needed for an analysis. The $\mathcal{P}$ matrix, on the other hand, is a pure detector quantity and independent of the analysis. Obviously this is therefore the quantity made available to the collaboration by the RICH group. This choice, albeit very logical, does make things a little more complicated. Following the definition of the $\mathcal{Q}$ matrix, it directly gives us a vector with the true particle probabilities $\vec{N}$, starting from the identified particle type $\vec{I}$.

$$\vec{N} = \mathcal{Q} \cdot \vec{I} \tag{4.10}$$

where

$$\vec{N} = \begin{pmatrix} N_\pi \\ N_K \\ N_p \end{pmatrix}, \qquad \vec{I} = \begin{pmatrix} I_\pi \\ I_K \\ I_p \\ I_X \end{pmatrix} \tag{4.11}$$

For the $\mathcal{P}$ matrix this relation is:

$$\vec{I} = \mathcal{P} \cdot \vec{N}. \tag{4.12}$$

In order to obtain the true particle types $\vec{N}$ one has to '**unfold**' them from the identified particle type in $\vec{I}$. In this case this implies technically, the inversion of the $\mathcal{P}$ matrix.

4.2.2.2 The unfolding procedure

It is analytically not possible to invert a 3×4 matrix in a closed algorithm, and one would have to resort to minimization or iteration procedures [97]. However since the fraction of unidentified X-particles is small [95], we can neglect them and use what is called the **truncated $\mathcal{P}$ matrix**:

$$\mathcal{P}_{trunc}(p,\mathcal{T}) = \begin{pmatrix} \mathcal{P}_\pi^\pi & \mathcal{P}_K^\pi & \mathcal{P}_p^\pi \\ \mathcal{P}_\pi^K & \mathcal{P}_K^K & \mathcal{P}_p^K \\ \mathcal{P}_\pi^p & \mathcal{P}_K^p & \mathcal{P}_p^p \end{pmatrix} \tag{4.13}$$

which is a non-singular, square matrix. Therefore it has a unique inverse which enables us to solve equation 4.12 :

$$\vec{N} = \mathcal{P}_{trunc}^{-1} \cdot \vec{I}_{trunc} \qquad \text{, where} \quad \vec{I}_{trunc} = \begin{pmatrix} I_\pi \\ I_K \\ I_p \end{pmatrix} \tag{4.14}$$

So we can determine e.g. the real K flux by interpreting N_K as a weight. If we have a particle which is identified as a pion :

$$\vec{I}_{trunc} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \tag{4.15}$$

then we would also have to count $\left(\mathcal{P}_{trunc}^{-1}\right)_K^\pi$ in the true kaon spectrum since there was a small probability that the pion in fact was a kaon. Note that the elements in $\mathcal{P}_{trunc}^{-1}$ are not probabilities due to the fact the matrix was truncated. However, they do provide the correct true hadron fluxes because the so-called X-particles are cut out in the analysis anyway.

An important thing to note here is that this approach means that we have to take all 3 particle hypotheses into account when calculating hadron kinematics. If we want to extract an unfolded energy spectrum, we know that

$$E = \sqrt{p^2c^2 + m_0^2c^4}. \tag{4.16}$$

Since pions, kaons and protons have very different rest masses, we have to calculate for each of the true particle hypotheses for this identified pion track the correct energy, using the corresponding mass, and take this energy value to fill the spectrum. At HERMES energies, where the reconstructed hadron momenta range from 0.5 to about 18 GeV, the rest mass is not completely negligible when calculating the total hadron energy.

4.2.2.3 Generation of the $\mathcal{P}$ matrix and systematic uncertainties involved.

In order to evaluate the uncertainties associated with the $\mathcal{P}$ matrix, we have to look at the generation of these matrices using HERMES Monte Carlo data. The

RICH description is subject to a couple of parameters which were tuned to best fit the Čerenkov photon yield and angle distributions found in experimental data. The original tuning was done by B. Hommez [98] using electrons only. We will later on refer to this tune as the e-**tune**. The photon yields were tuned by varying the light attenuation parameters for the two radiators: the aerogel and the C_4F_{10} gas. In order to tune the Čerenkov angle distributions, a parameter called **Mirror Roughness (MR)** was introduced. It controls the distortion of the reflection of photons from the mirror. The MR parameter enables us to broaden the Čerenkov angle distribution in the Monte Carlo, as it had the tendency to be more narrow than the experimental distribution.

During a later study [99], however, that tried to extract the $\mathcal{P}$ matrix from experimental data using decaying particles like $\phi \to K^+ + K^-$ it was seen that the MR tune needed adjustment for this case. A new setting was obtained [96] using more decaying particle events like $K_S \to \pi^+ + \pi^-$, $\Lambda \to p + \pi^-$ and $\bar{\Lambda} \to \bar{p} + \pi^+$. Due to the limited acceptance of the experiment, and the nature and topology of decaying particle events, data is only available in a limited range. A minimum in χ^2 is found by varying the mirror roughness when comparing PYTHIA6 + HMC data. The HMC tune [99] will from now on be referred to as the ϕ **tune**.

In view of all of this, the center value (**MR** = 0.986667) between the e-tune (**MR** = 0.988333) and the ϕ-tune (**MR** = 0.985000) was used (**center-tune**) and the maximum difference between the e-tune of ϕ-tune as a measure for the systematic uncertainty on the $\mathcal{P}$ matrix. It is assumed that the $\mathcal{P}$ matrix value is uniformly distribution between the two tunes, and since for a uniform distribution between a and b :

$$\sigma^2 = \frac{1}{12}(b-a)^2, \tag{4.17}$$

we need to divide this difference by $\sqrt{3}$ to obtain one standard deviation. On figure 4.4 we have plotted the $\mathcal{P}$ matrix for the e- and ϕ-tune for 1, 2 and 3 (or more) tracks per detector half.

An immediate apparent feature in the plot is the strange jump around 10 GeV for kaons, this is the effect of the C_4F_{10} Čerenkov threshold of 9.4 GeV for kaons. For protons, the aerogel threshold is visible at 3.8 GeV/c. It should be noted also that all of this was only developed for the IRT reconstruction algorithm. So when selecting the information about the RICH one should only navigate the IRT link in the ADAMO database.

4.2.2.4 Propagation of the systematical uncertainties

A number of methods have been proposed to propagate the errors through the inversion of the truncated $\mathcal{P}$ matrix. One method would have been to generate independent random Gaussian matrix elements for each $\mathcal{P}$ matrix, taking its nominal value and the uncertainty as input. Each of these random matrices can than be inverted and the standard deviation we obtain for the inverted

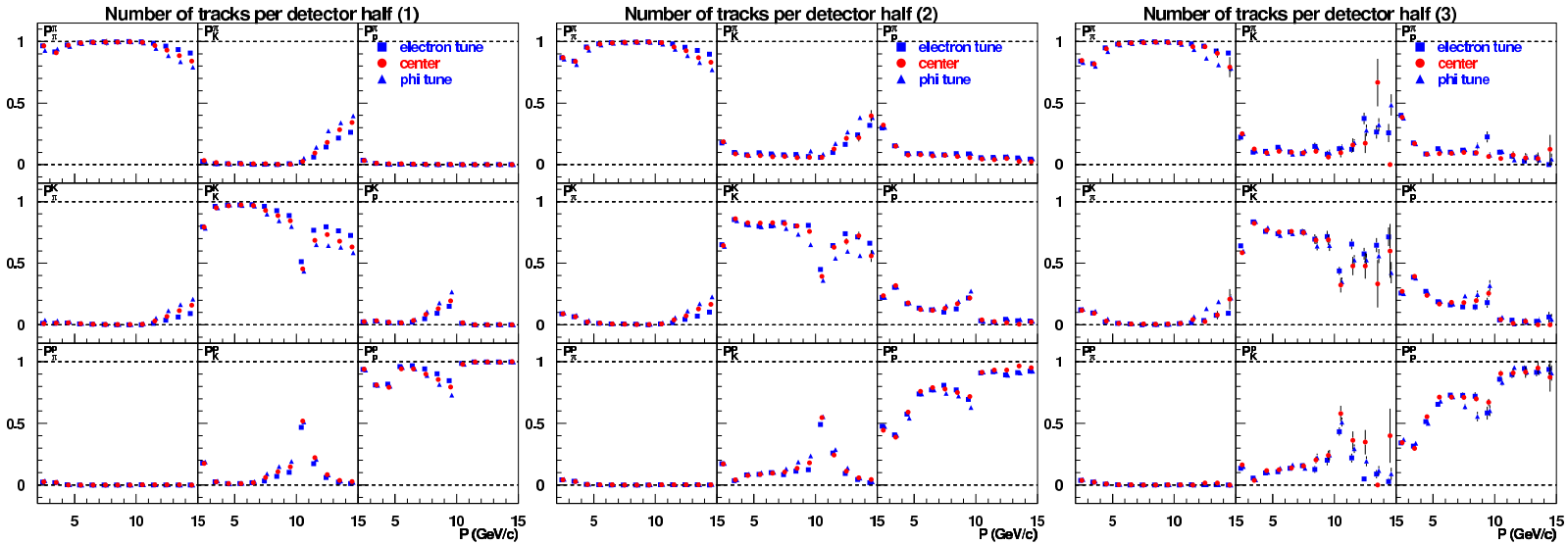


Figure 4.4: The truncated $\mathcal{P}$ matrix for 3 different event topologies. The matrices are shown for the two HMC tunes, the e tune (**MR** = 0.988333) and the ϕ-tune (**MR** = 0.985000) together with the center tune (**MR** = 0.986667). The data is provided by the RICH group.

matrix elements can than serve as the systematic error on the inverted matrices. However, this way we do not take into account any correlations between the truncated $\mathcal{P}$ matrix elements since we generate completely independent random numbers for each matrix element. Obviously the matrix elements are indeed highly correlated because of their normalization. The fourth identified particle probability $\mathcal{P}_t^i$ for a given true hadron type t within a single column of the $\mathcal{P}$ matrix is always uniquely defined when the 3 others are known since their sum has to be 1. Now, because we leave away the X particles, it is not clear anymore how these correlations behave when inverting the matrix.

A more correct way is just to redo the unfolding once with the e-tune $\mathcal{P}$ matrix and once with the ϕ-tune $\mathcal{P}$ matrix. The maximum difference of the unfolded hadron spectrum with either one of the *e-tune*/ϕ*-tune spectrum* is a measure for the systematic error due to the RICH on the hadron spectra. In this analysis, we have chosen this method to obtain the systematic errors of the unfolded hadron spectra.

4.3 Experimental hadron multiplicity distributions

Table 4.2 summarizes the DIS and hadron yields for the HERMES 2000 unpolarized hydrogen dataset using the cuts described in sections 4.1 and 4.2.1.

DIS leptons :	5 870 980
Total Hadrons :	2 126 799
Positive Hadrons :	1 318 609
Negative Hadrons :	808 190
π^+ :	959 059
π^- :	688 169
K^+ :	203 341
K^- :	87 858
p :	156 209
$\bar{p}$:	32 163
Integrated Luminosity $\mathcal{L}_{int}$ (pb^{-1}) :	144.68

Table 4.2: Experimental particle yields for the data set used in this thesis. These are the raw yields, without any correction or RICH unfolding applied.

We are mainly interested in the z dependence of the hadron spectrum. For protons and antiprotons, the statistics is rather low, so they are dropped from the rest of the analysis and we focus on obtaining charge and flavor separated pion and kaon multiplicities. Another reason to leave the proton data aside for the moment is that is not easy to have a clear physics interpretation. It is hard to disentangle protons that have formed in quark fragmentation from the nucleon remnant. Later on in this work we will also investigate other kinematical dependences for various reasons, but these will be dealt with in detail at that time. Note that the statistical uncertainty calculation on the

experimental multiplicity distributions will just be the fully correlated error between the hadron and the DIS sample as argumented in appendix A. .

In figure 4.5 the experimental hadron multiplicities are shown for π^+, π^-, K^+ and K^- separately. In each figure both the raw experimental data as well as the RICH PID corrected data is displayed. The uncertainty shown on each point without end bars is the quadratic sum of both the systematic and statistical uncertainty. The binning for the z spectrum is taken to be equidistant between 0.15 and 0.9 with 0.05 as bin width. The points are drawn at the center of each data bin. We can see that the RICH unfolding correction

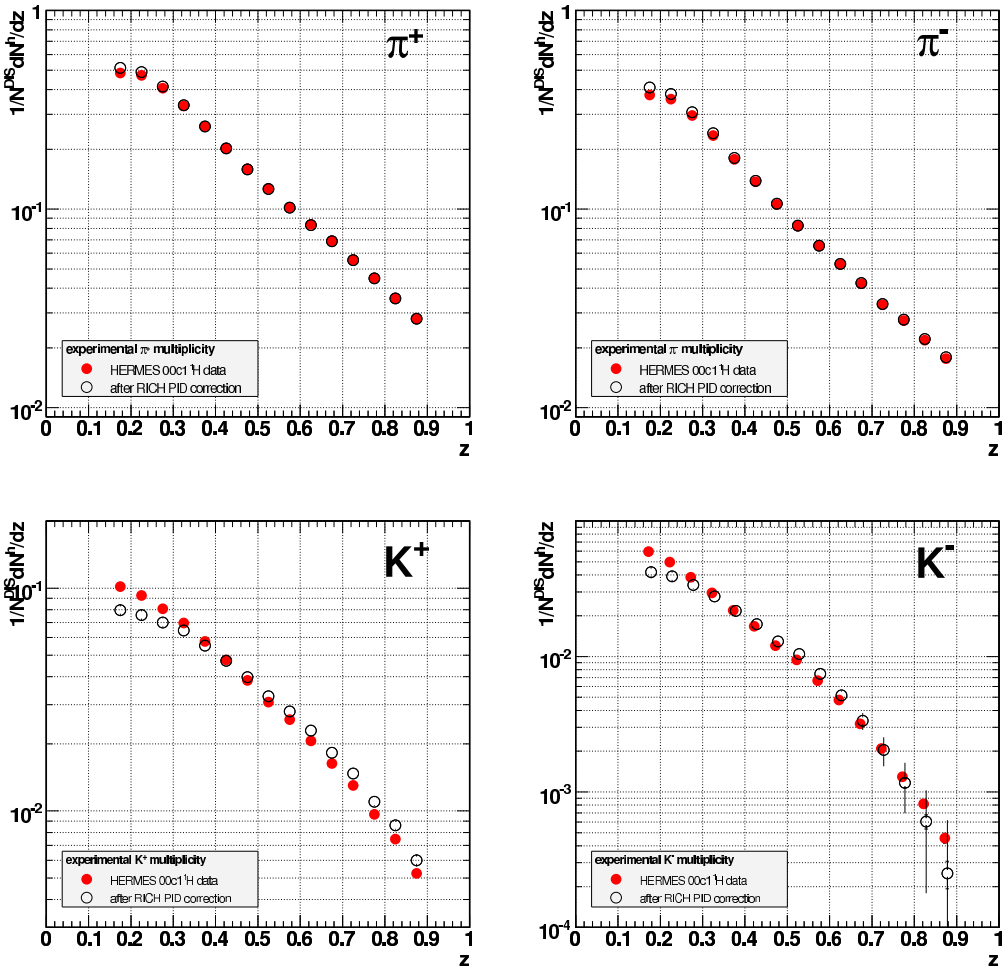


Figure 4.5: Experimental hadron multiplicities versus z for pions and kaons, charge and flavor separated. We show the obtained raw multiplicities (colored circles) and the ones after the RICH PID unfolding correction (open circles). Note that for the K^- spectrum the points were shifted a little to the left and right to make the error bars visible. The error flags with the end bars denote the statistical errors, on top of them are plotted the total uncertainty, so with the systematical uncertainty added in quadrature. See also appendix B.1.

affects the pions only very mildly and this only at low z. However, for the kaons the effect is rather large. This is expected from the RICH $\mathcal{P}$ matrix where $\mathcal{P}_K^K$ differs quite significantly from 1, especially at large momenta. The values for these plots are included in appendix B.1. For the pions, the statistical errors

are of the order of 0.5 % to 1 - 1.5 % at high z and lower at low z. The kaons show somewhat larger uncertainties: from 1.5 % at low z to about 3 % for K^+ and even larger than 10 % for K^- at high z. The systematic error for pions is of the same order of magnitude as the statistical error, whereas for kaons it is significantly larger, around 4-5 % maximum for K^+, and up to 10 % or even larger than 50 % at the highest z bins for the K^-.

4.4 Correction for pair creation $\gamma \rightarrow e^+ + e^-$

Normally we tag the candidate DIS events by detection of a positron within the DIS selection criteria mentioned before. However, a positron that ended up in the HERMES acceptance does not necessarily have to be the scattered lepton in a DIS event. It can also come from an e^+e^- pair creation process. As an example, photons with an energy of about 1 GeV and higher will, when interacting within a medium, almost always result in the creation of and e^+e^- pair [6]. So there is a distinct probability that these positrons wrongly end up in our DIS sample.

In order to ascertain this contribution, we note that most of the false leptons come from charge symmetric processes, like e^+e^- pair creation. Hence, statistically spoken, an equal number of oppositely charged leptons has to end up in our sample as well, thereby giving the number of correctly charged leptons which wrongly end up in the DIS sample. So by dropping the requirement that a DIS lepton has the same charge as the HERA beam particles and subtracting the wrongly charged events within the DIS cuts, we correct for these pair creation events. The influence of this correction is shown in figure 4.6.

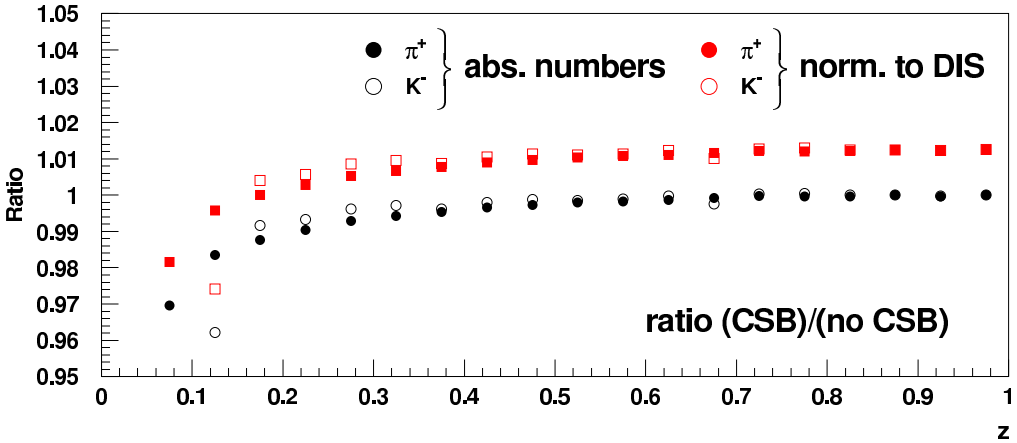


Figure 4.6: The charge symmetric background correction on the hadron yields, see also [100]. The black circles correspond to the absolute hadron yields, whereas the red squares to the yields, normalized to number of DIS events. To not over complicate the plot, only π^+ and K^- are displayed. As shown, they are very similar.

This correction has an effect on the total number of DIS events, lowering it by about 1.3 %. The influence on the hadron yields normalized to DIS is smaller, < 1 %, except for very low z. This is because these are in fact ratios where the correction in both nominator and denominator cancels to some extent. Overall, this correction results in a rise of the hadron multiplicities by 0.5 to 1.5 % (at low z) which is not visible in figure 4.6.

Chapter 5

Monte Carlo Simulations

"[...] Another change which I find disturbing is the rising tyranny of Carlo. No, I don't mean that fellow who runs CERN, but the other one, with first name Monte. The simultaneous increase in detector complexity and in computational power has made simulation techniques an essential feature of contemporary experimentation. The Monte Carlo simulation has become the major means of visualization of not only detector performance but also of physics phenomena. So far so good. But it often happens that the physics simulations provided by the Monte Carlo generators carry the authority of the data itself. They look like data and feel like data, and if one is not careful they are accepted as if they were data. All Monte Carlo codes come with a GIGO (garbage in, garbage out) warning label. But the GIGO warning label is just as easy for a physicist to ignore as that little message on a pack of cigarettes is for a chain smoker. [...]"

J.D. Bjorken

Extract from a talk given at the 75th anniversary celebration of the Max Planck Institute of Physics, Munich, Germany (Dec. 10th 1992). As quoted in: Beam Line, Winter 1992, vol 22., No. 4

5.1 Introduction

Now that we have a set of experimental multiplicity distributions, we need to apply a number of corrections for detector acceptance and efficiency or subtract background which we cannot resolve solely using the experimental data. For all of this, Monte Carlo (MC) simulations are needed. We will have a look in this chapter how the HERMES MC chain is built up, what physics generators are used and how these models were tuned to best describe the experimental data. Before we do, however, a small note on using weights in Monte Carlo is in order.

The standard accept/reject technique for the generation of events works

very well in practice as it enables us to sample basically any distribution. There are situations, however, where it can be rather time consuming. Suppose one wants to simulate diffractive processes. As the cross section follows a Q^{-6} law, it becomes very small at high Q^2 so one might want to revert to more efficient techniques for simulating this tail. One option is the use of weights. In weighted Monte Carlo, each generated event (in practice being set of kinematical variables, e.g. Q^2 and ν) comes with a certain weight w^i which is actually the value of the cross section at these kinematics. If we then want to calculate the number of events N within a certain bin of the spectrum, we just have to add the weights of the events for which the kinematics fall inside the bin:

$$N = \sum_{i=1}^{N_{gen}} w^i \tag{5.1}$$

where N_{gen} is the number of events generated. Note that for the calculation of the one standard deviation error on this value, we have to sum the squares of the weights

$$\sigma = \sqrt{\sum_{i=1}^{N_{gen}} (w^i)^2} \tag{5.2}$$

Weighted Monte Carlo samples can be very efficiently generated, but are a bit tricky to work with as far as filling spectra and error propagation are concerned.

5.2 Monte Carlo in HERMES

5.2.1 Overview

The HERMES Monte Carlo chain is depicted in figure 5.1. It starts with a physics generator, built inside a general framework called GMC based on the CERNLIB KUIP [101] language. Several generators were built for various physics analyses, like exclusive processes, deeply virtual Compton scattering (DVCS), transversity etc. The two generators that are important in this thesis are the deep inelastic scattering generator called `gmc_disNG` and the PYTHIA6 [102] generator, `gmc_pythia6`. `gmc_disNG` is based on LEPTO [103] to generate the DIS events themselves, RADGEN [94] for QED radiative corrections and JETSET [104] for fragmentation. `gmc_pythia6` takes the PYTHIA6 code and supplements it with radiative corrections done by RADGEN.

Once the physics events are generated and we know what particles are produced in the initial scattering interaction, we have to trace those particles though the detector. This is done by a program called HMC, or **Hermes Monte Carlo**. It contains a complete simulation of the experimental setup based on the GEANT3 library [105]. Figure 5.2 shows an example of a simulated event processed by HMC and tracked through the experimental setup. The program HRC, or **Hermes Reconstruction Code** is responsible for determining

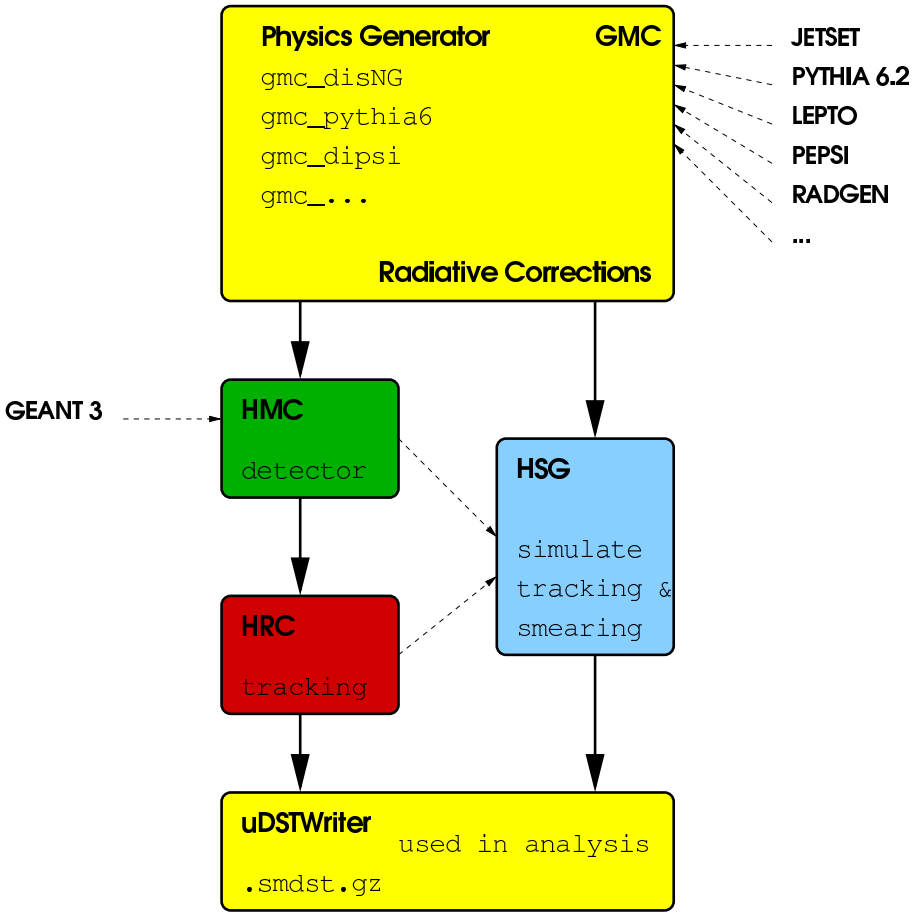


Figure 5.1: The structure of the HERMES Monte Carlo suite

the particle momenta from the deflection in the magnet by matching the back and front part of each track. Important to note that this is the exact same program which is used for reconstruction of the real experimental data. The HMC program prepares the generated events and simulated detector responses in such a way that they can be handled by HRC. Once the track information is available, a DST file is written which is very similar[1] to the DST files produced by the experimental data handling chain. The HMC + HRC step is the most time consuming of all. To give an idea, on a $\sim$ 1.4 GHz CPU it takes about 24 hours to produce 100'000 fully tracked events. Needless to say that even with sufficient CPU power available, producing HERMES Monte Carlo is a very time consuming occupation. Therefore a program was written to speed up the process a bit, or at least simulate it. This tool is called the **Hermes Smearing Generator** or HSG. It was originally conceived by F. Menden [91] and later on refined by A. Hillenbrand [106]. By taking momentum resolution $\left[\frac{\Delta p}{p}(p)\right]$ and θ_x, θ_y scattering angle resolution $\left[\frac{\Delta\theta_x}{\theta_x}(p)\right]$ and $\left[\frac{\Delta\theta_y}{\theta_y}(p)\right]$ obtained from a fully tracked Monte Carlo production, HSG smears the kinematics of the generated Monte Carlo tracks on a statistical basis. It also applies some acceptance cut

[1] Of course e.g. with the absence of slow control information, but with extra information on the 'real' event.

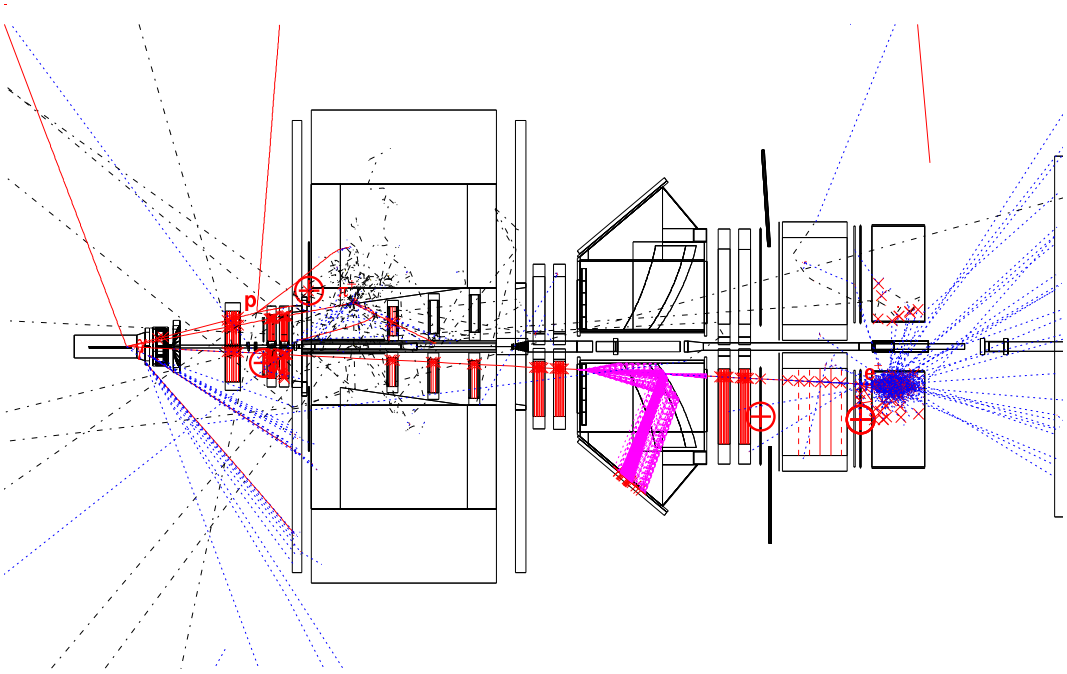


Figure 5.2: A nice picture of a simulated semi-inclusive event in HMC. The event is a normal semi-inclusive DIS event in which a π^+, a π^- and a π^0 (that decays into two γ's) are produced. The track in the bottom half of the spectrometer is the scattered positron.

based on lookup tables for the bending in the magnet. HSG Has proven to be a very reliable and convenient tool for Monte Carlo studies and especially for iterative tuning of Monte Carlo parameters. Figure 5.3 shows a comparison

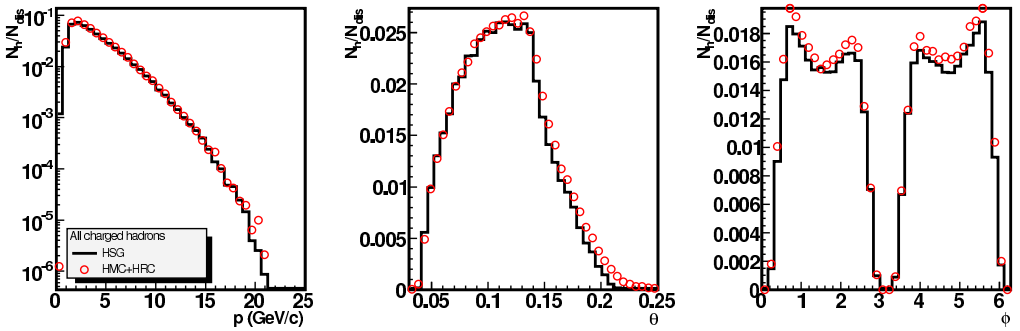


Figure 5.3: A comparison between HSG (histogram) and HMC + HRC (open circles) of total charged hadron yield (normalized to DIS events) versus momentum, scattering and azimuthal angle. This was done with the 2004c tune.

between HSG and HMC + HRC data done with a 2004c (see section 5.2.2.5) tune Monte Carlo production. One important thing to note with the smearing generator is that it is in fact only really suited to study hadrons yields. Correlations between different kinematical variables are distorted, and this will prove to be an issue when we will discuss the $\cos\phi_h$ moments (see chapter 7) and their influence on the HERMES acceptance correction.

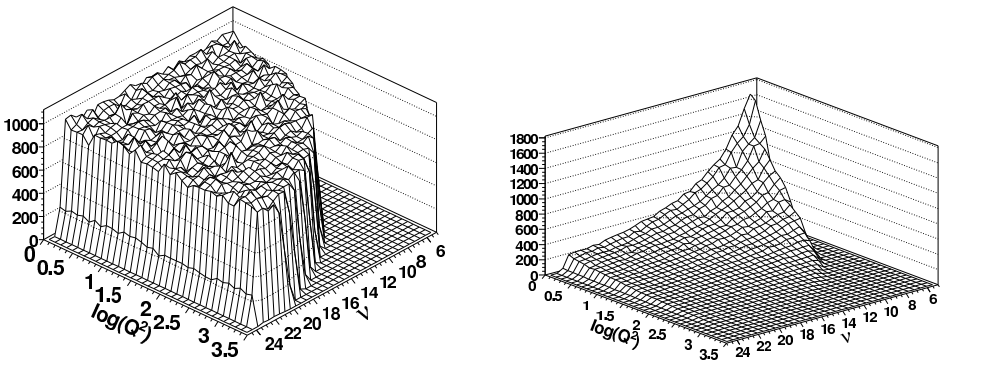
5.2.2 The gmc_disNG generator

The need for QED radiative corrections in a HERMES Monte Carlo implementation makes the implementation of the gmc_disNG generator somewhat complicated. Basically it starts by picking two independent kinematical variables uniformly in a box. In this analysis we have used ν and $\ln(Q^2)$ as independent variables (see figure 5.4). Note that of course e.g. for inelastic scattering $2M\nu - Q^2 > 0$. In the Monte Carlo initialization, this is set by

```
Set GenKine 1
```

The generation box (see figure 5.4(a)) is defined in the initialization by

```
Set Q2Min 0.10
Set Q2Max 30.
Set NuMin 0.1
Set NuMax 27.57
```



(a) Without event weights

(b) Taking into account the event weights

Figure 5.4: The box in which the kinematics is generated, uniform in $\ln(Q^2)$ and ν.

It is important to remember that these variables will be treated as **observed quantities** ! Next, gmc_disNG calculates the rest of the kinematics (x, y, W^2) and the Born cross section using an $F_2(x, Q^2)$ parameterization [8], [9]. Then the event is weighted with the Born cross section and a Jacobian factor which equals the size of the box in which one generates.

$$w^i = \sigma_{Born} \cdot \Delta \ln(Q^2) \Delta\nu \tag{5.3}$$

In figure 5.4(b) one can see the same generated distribution as before, but now taking into account the event weights. The real shape of the cross section becomes visible.

5.2.2.1 QED radiative effects

Subsequently, QED radiative corrections have to be applied before handing over the generated kinematics to LEPTO. This is done in a somewhat indirect

way. The generated kinematics by GMC itself are treated as observed quantities, so the radiative correction involves re-weighting the event and calculating back the true Born-level kinematics. This 'reverse-calculation' has the advantage that one always populates the kinematical phase space which is selected, and one does not have to generate outside of the region of interest to account for radiative smearing. This saves a lot of computing time.

RADGEN [94] proceeds by calculating a correction factor, which is the ratio between the Born and the observed cross sections at the generated observed kinematics

$$\Sigma_{RC} = \frac{\sigma_{Obs}(\nu, Q^2)}{\sigma_{Born}(\nu, Q^2)} \tag{5.4}$$

and applies this to the total event weight :

$$w^i \rightarrow w^i_{RC} = \sigma_{Born} \cdot \Sigma_{RC} \cdot \Delta \ln(Q^2) \Delta\nu \tag{5.5}$$

Figure 5.5 shows the diagrams which are taken into account when calculating the QED radiative corrections, where the first one shows the Born cross section. If a real photon has been radiated, as shown in diagrams b) and c),

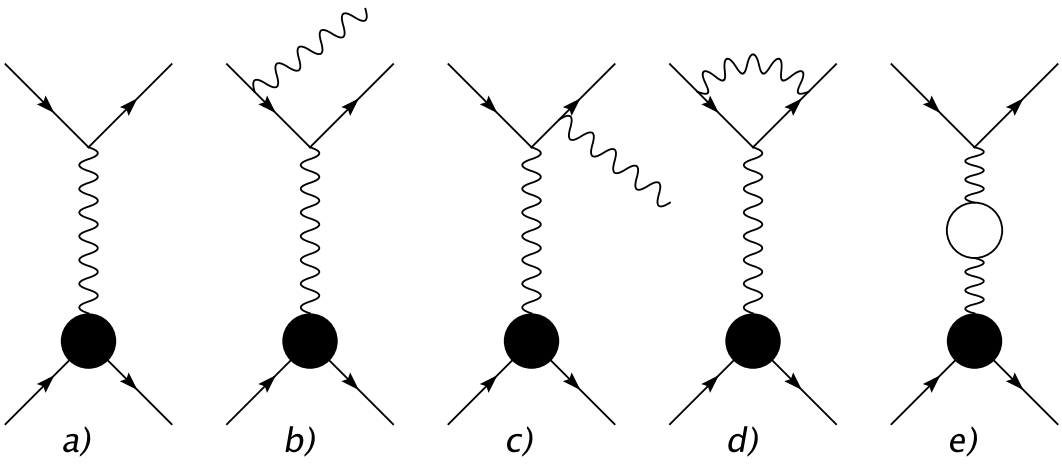


Figure 5.5: QED radiative correction diagrams taken into account by RADGEN.

the event kinematics are changed at the scattering vertex and thee additional variables are necessary to fix the kinematics. A possible choice [94] is E_γ, the energy of the real radiative photon, and the angles θ_γ and ϕ_γ. If $\vec{k}_\gamma$ represents the real photon momentum, and $\vec{q} = \vec{k} - \vec{k'}$ the momentum of the virtual photon, then θ_γ is the angle between $\vec{k}_\gamma$ and $\vec{q}$. ϕ_γ is then the angle between the $(\vec{k}, \vec{k'})$ and the $(\vec{k}_\gamma, \vec{q})$ planes. The true event kinematics are then calculated as

$$W_T^2 = W^2 - 2E_\gamma(\nu + M - \sqrt{\nu^2 + Q^2}\cos\theta_\gamma) \tag{5.6}$$

$$Q_T^2 = Q^2 + 2E_\gamma(\nu - \sqrt{\nu + Q^2}\cos\theta_\gamma) \tag{5.7}$$

$$\nu_T = \nu - E_\gamma \tag{5.8}$$

$$x_T = \frac{Q_T^2}{2M\nu_T} \tag{5.9}$$

$$y_T = \frac{\nu_T}{E} \tag{5.10}$$

It may happen that the energy carried away by the radiated photon is so high, that only an elastic scattering off the proton can happen. Thus one has to take into account the elastic scattering contribution as well as terms coming from loop corrections, like in figure 5.5 d) and e). The way in which the radiative corrections are applied (with or without radiating a real photon) is calculated from the respective contributing terms to the total observed cross section. For a more thorough discussion, we refer to [94].

Once the true kinematics are known, the event is passed on to LEPTO which selects a struck quark according to the parton distributions. Once a struck quark is selected, it is passed on to the Lund string model implementation in JETSET for fragmentation. Here we will discuss some of the important settings of LEPTO and JETSET. FORTRAN common blocks are used to hold the parameters and settings, the array-variables which make up this common block will be given below with their specific index for the settings we are using.

5.2.2.2 QCD effects

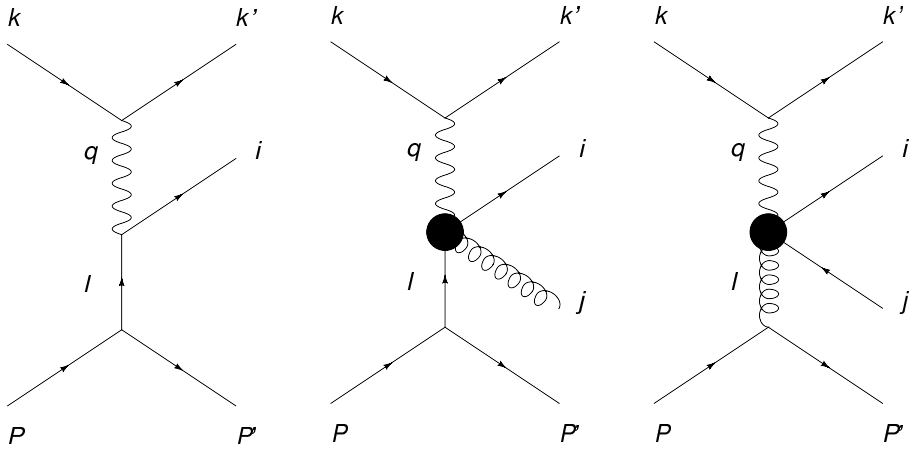


Figure 5.6: First order correction QCD processes. The Born diagram is on the left, QCD Compton $\gamma^* q \to qg$ (middle) and boson-gluon fusion $\gamma^* g \to q\bar{q}$ (right)

We use first order matrix elements for the simulation of QCD effects in the hadronic final state. The two contributing highest order QCD processes are **QCD Compton** scattering (σ_{qg}) and **boson/photon - gluon fusion** ($\sigma_{q\bar{q}}$). See also figure 5.6. These cross sections roughly behave [107] as

$$\sigma_{qg} \propto \frac{1}{(1-x_p)(1-z_p)} \tag{5.11}$$

$$\sigma_{q\bar{q}} \propto \frac{1}{z_p(1-z_p)}, \tag{5.12}$$

where $x_p = x/\xi$, ($l = \xi P$, so ξ is the fraction of the proton momentum carried by the struck quark, in the Breit frame one has $\xi \approx x$) and $z_p = P \cdot i / P \cdot q$. x_p and z_p are basically the counterparts at parton level of the variables x and z. We selected the W^2 cut-off scheme for handling the divergences. This means we

require $s_{ij} = (p_i + p_j)^2 > y_{cut} W^2$ for any pair ij of partons. This scheme is taken over from e^+e^- annihilation and is used in the so-called JADE algorithm for jet reconstruction [108], [109]. As settings we have used `PARL(8)` = $y_{cut} = 0.005$ and `PARL(9)` = $(s_{ij})_{min} = 1$ GeV^2. In regions of x_p and z_p where the contributions diverge rapidly, adding the parton showering mechanism (`LST(8) = 12`) is supposed to improve the treatment.

In a way, parton shower calculations are needed since the implementation of higher order QCD corrections in Monte Carlo is problematic due to the complexity and the number of diagrams and problems with the probabilistic interpretation. The emission of several partons is factorized into a series where individual terms are described by the basic QCD branching processes $q \to qg$, $g \to gg$ and $g \to q\bar{q}$. It is an iterative process which is halted when all the parton virtualities are below a cut off value of $m_0^2 = 1$ GeV^2. However at HERMES energies, turning on parton showering has been shown not to have a significant influence [110] and therefore we have not used this option.

```
LST(8) = 1
LST(20) = 1
PARL(8) = 0.005
PARL(9) = 1.0
```

5.2.2.3 The 'new' sea quark treatment and the nucleon remnant

```
LST(14) = 4
LST(35) = 1
PARL(14) = 0.35
```

In the case that LEPTO chooses a valence quark as struck quark there is not really a problem, according to the setting of `LST(14)`, the quark and diquark will be passed on to the JETSET routines for fragmentation. However, when a sea quark (q_s or $\bar{q}_s$) is chosen, the remnant is somewhat more complicated. We then have a four quark system $q_v q_v q_v q_s$ or $q_v q_v q_v \bar{q}_s$. In the old implementation of LEPTO, this was handled according to the following principle: if the sea quark in the four-quark system is a $\bar{u}_s$ or a $\bar{d}_s$, it is canceled with one of the valence quarks and a simple diquark is left. If it is a $\bar{s}$, it is joined also with a valence quark but in this case to form a strange meson. The meson is given a certain fraction z of the available longitudinal energy and a small transverse contribution according to a Gaussian with width $p_\perp$ = `PARL(14)` = 0.35 GeV.

A diquark is left over which will form a string with the struck quark and fragment according to `LST(14)`. When we are left with a sea-quark, it is combined with a random diquark in which case a baryon is formed $q_v q_v q_s$ and the left over quark will form a string with the struck quark and fragment.

In LEPTO 6.3 a new treatment was introduced in order to improve agreement with HERA data [111], [103]. The struck quark is assigned to be a valence or a sea quark in accordance with the relative size of the parton distribution functions $q_v(x, Q^2)$ or $q_s(x, Q^2)$. When it is a valence quark, the normal

treatment is used. For a sea quark however, the $q_s\bar{q}_s$ pair is treated dynamically. First, momentum is distributed between them according to the Altarelli-Parisi splitting function $P(g \rightarrow q\bar{q})$, next the three spectator valence quarks are split into a quark and a diquarks and two strings are build between the two sea quarks, the valence quark and the valence diquark. These strings are passed on to JETSET for fragmentation.

5.2.2.4 Intrinsic transverse momentum

```
PARL(3) = 0.44
```

This parameter of the LEPTO Monte Carlo defines the intrinsic primordial transverse momentum k_T of the partons in the nucleon. The world average and default value here is 0.44 GeV. In the past there were arguments [91], [112] to increase this value up to 0.8 GeV in order to compensate for the missing build-up of transverse hadron momenta since we have switched off parton showering. However, using this approach one completely loses all physical interpretation of this parameter. In this work we therefore chose to remain with the default intrinsic k_T value of 0.44 GeV and let the transverse momentum distribution follow in a natural way from tuning other fragmentation parameters.

5.2.2.5 Tuning the fragmentation model

As already mentioned the fragmentation model used in the `gmc_disNG` Monte Carlo is JETSET, which was tuned to LEP data to describe fragmentation in e^+e^- annihilation. So obviously one cannot just take the 'factory defaults' and use them for ep scattering at HERMES energy. A good hadronization description is of crucial importance not only to the multiplicity analysis, but also to the other semi-inclusive analyses in HERMES.

Table 5.1 summarizes the different tunes that were obtained, tested and used during the course of this work. We now discuss these parameters in somewhat more detail.

- `PARJ(1)` = $\mathcal{P}(qq)/\mathcal{P}(q)$ regulates the suppression of diquark-anti diquark formation as compared to quark-anti quark formation in the string field. This has a direct impact on baryon (p, $\bar{p}$, n, ...) production. Here, it was tuned to the antiproton to pion yields ($\mathcal{P}(\bar{p})/\mathcal{P}(\pi)$), since for this particle to emerge in fragmentation one needs a diquark in the string. An interesting fact to note is that in ep scattering we find that diquark-anti diquark production is more suppressed than at LEP (JETSET default). This is also confirmed by the H1 Collaboration [113].

- `PARJ(3)` = $\mathcal{P}(q^sq^s)/\mathcal{P}(q^s)$ is the extra suppression of strange diquarks. This parameter was tuned to the $\bar{\Lambda}$ yield.

parameter	JETSET	default	holger	2002a	2003a	2004a	2004c
qq_{supp}	PARJ(1)	0.10	0.10	0.02	0.03	0.025	0.029
q^s_{supp}	PARJ(2)	0.30	0.16	0.20	0.20625	0.120	0.283
$q^s q^s_{supp}$	PARJ(3)	0.40	0.40	0.40	0.25	0.25	0.40
$BM\bar{B}/B\bar{B}$	PARJ(5)	0.50	0.50	0.50	0.0	0.0	0.50
$s\bar{s}/BM\bar{B}$	PARJ(6)	0.50	0.50	0.50	0.0	0.0	0.50
$M_s/BM\bar{B}$	PARJ(7)	0.50	0.50	0.50	0.0	0.0	0.50
VM_{supp}	PARJ(11)	0.50	0.50	0.20	0.25	0.25	0.50
VM^s_{supp}	PARJ(12)	0.60	0.60	0.60	0.30	0.30	0.60
σ	PARJ(21)	0.36	0.33	0.37	0.382	0.382	0.381
f	PARJ(23)	0.01	0.01	0.03	0.03	0.03	0.01
P_T^f	PARJ(24)	2.00	2.00	2.50	2.50	2.50	2.00
a	PARJ(41)	0.30	0.89	1.74	1.1266	1.13	1.940
b	PARJ(42)	0.58	0.24	0.23	0.3672	0.37	0.544
a_{qq}	PARJ(45)	0.50	0.50	0.50	0.80	0.80	1.05
E_0	PARJ(33)	0.80	0.80	0.80	0.20	0.20	0.80
χ^2/ndf		226	235	163		74	43

Table 5.1: Tuning parameters for the JETSET fragmentation model. The meaning of the individual parameters is described in the text, and can also be found in the PYTHIA6/JETSET manual [104]

- PARJ(2) = $\mathcal{P}(q^s)/\mathcal{P}(q)$, suppression of strange quarks as compared to u and d production. This was tuned to the K multiplicity distribution and the K^0_S yield.

- PARJ(5), PARJ(6) and PARJ(7) are parameters belonging to the popcorn model for baryon production. (See section 2.4.2.2.) PARJ(5) regulates the relative occurrence of baryon production in $BM\bar{B}$ and $B\bar{B}$ configurations : $\mathcal{P}(BM\bar{B})/(\mathcal{P}(B\bar{B})+\mathcal{P}(BM\bar{B})) = \texttt{PARJ(5)}/(0.5+\texttt{PARJ(5)})$. PARJ(6) and PARJ(7) give an extra suppression of strange quarks and mesons in respectively the B and $\bar{B}$ on the one hand and the M in the $BM\bar{B}$ configuration on the other hand .

- PARJ(11) and PARJ(12) regulate the production of vector mesons (spin 1, like ρ^0, ϕ) compared to the production of scalar mesons like π. PARJ(12) is basically the same as PARJ(11), but respectively for strange vector mesons (e.g. K^{*0}). They were tuned to the ρ^0/π^0 and K^{*0}/K^0_S yield ratio.

- PARJ(21), PARJ(23) and PARJ(24) concern the transverse momentum buildup in the fragmentation. PARJ(21) gives the width of the Gaussian p_x and p_y transverse momentum distributions of the primary hadrons. This parameter is also closely related to the intrinsic transverse momentum (PARL(3)), (see above). PARJ(23) and PARJ(24) give a simple parameterization for deviations from this assumed Gaussian distribution. A fraction f (PARJ(23)) of the Gaussian transverse momentum distribution is taken to be a factor P_T^f (PARJ(24)) larger than was set in PARJ(21).

- `PARJ(41)` and `PARJ(42)` are the a and b parameters respectively in the Lund symmetric fragmentation function.

$$f(z) = z^{-1} \cdot (1-z)^{a+a_{q\bar{q}}} \cdot e^{\frac{-bm_T^2}{z}} \tag{5.13}$$

  These just have to be tuned to the hadron yields as a function of momentum p, longitudinal energy fraction z, transverse momentum p_T and rapidity η.

- `PARJ(45)`, or $a_{q\bar{q}}$ in equation 5.13 modifies the Lund fragmentation function with an extra a parameter in case of diquark production. Proton yields are very sensitive to it. Since in the JETSET model the sum $a + a_{q\bar{q}}$ is taken as the total 'a'-parameter for the Lund fragmentation function, we have to tune both parameters at the same time. Inclusion of this parameter greatly improved the proton yield description.

- `PARJ(33)` or E_0 regulates the energy below which the further fragmentation of the partons is stopped and the final hadrons are formed. At the default value of 0.80, one might think this would lead to a too low number of string breakings and resulting hadron multiplicity. It was suggested [114] to put this value somewhat lower to 0.2 GeV (2003a, 2004a tunes). However it was realized later on this is not needed.

In fact table 5.1 summarizes the different tunes in chronological order. The so-called 2002a tune was described in [91], however with intrinsic k_T set back to its default value of 0.44 GeV. Both 2003a and 2004a described in extensive detail in [115], [100] and [116]. For the 2004c tune we refer to [117] and [106].

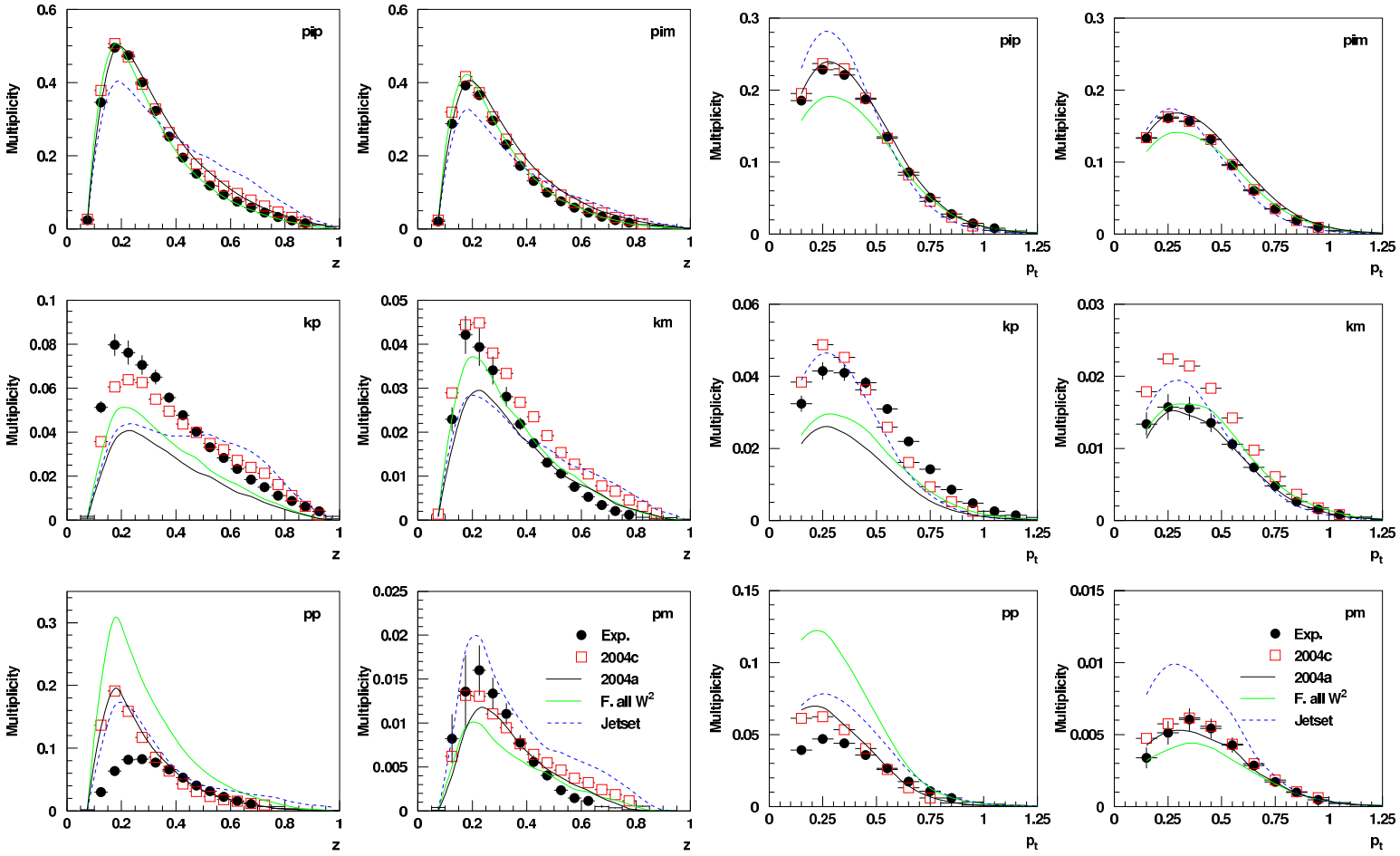
For the tuning process the aim was to minimize χ^2, defined as

$$\chi^2 = \frac{1}{N} \sum_{i=1}^{N} \frac{(m_{i,MC} - m_{i,EXP})^2}{\sigma_{i,MC}^2 + \sigma_{i,EXP}^2} \tag{5.14}$$

where $m_{i,MC}$ and $m_{i,EXP}$ are the simulated and experimental multiplicities in bin i and the corresponding σ values their uncertainty. Here the sum runs over all the bins of the different multiplicity spectra used. For the 2004c tune the spectra versus z, p_T and η were used for π^+, π^-, K^+, K^-, proton and antiproton. 18 Distributions in total.

Minimization was done using a *genetic algorithm*. It uses a group of parameter sets, sorted according to their χ^2. For every iteration, two sets are chosen, preferably better ones, and a new parameter set is produced by crossing the chosen ones according to specific rules. This continuously generates new, improved parameter sets in an inherently parallel way. This is what makes this algorithm excellent for running on parallel computing clusters. More details can be found in [106]. The ξ^2/ndf values for a number of tunes can be found in table 5.1 where one can see the improvement made.

In figure 5.7 we show a comparison of experimental multiplicities for π^+, π^-, K^+, K^-, proton and antiproton for 4 different tunes. One can see that the overall description of the 2004c tune is satisfactory.

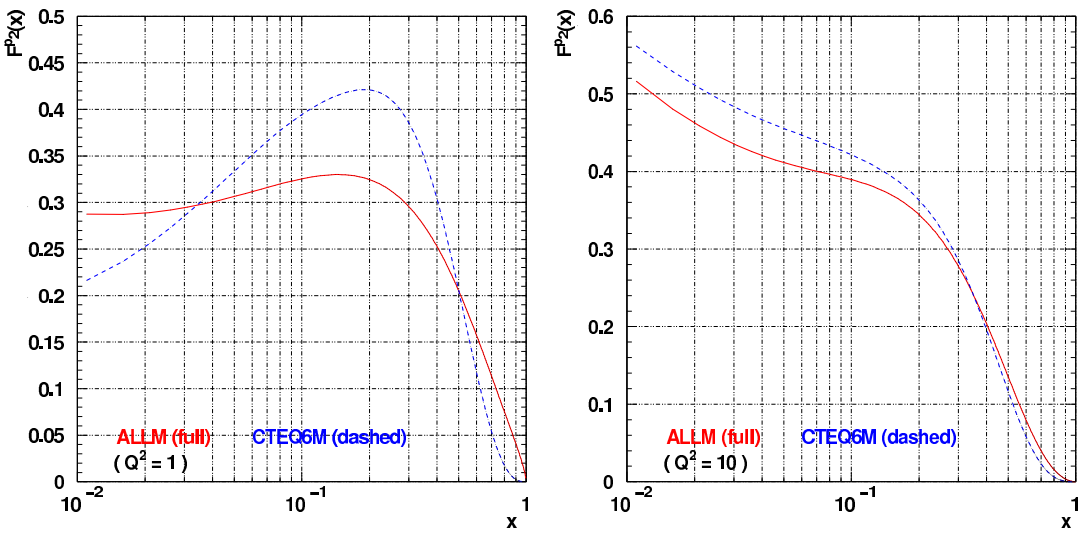


(a) Versus z (b) Versus p_T

Figure 5.7: Comparison of different JETSET tunes. In these plots the "F. All W^2" tune stands for the 2002a tune [91]. This figure is taken from [117]. One can clearly see the failure of default JETSET parameters in describing the multiplicities measured at HERMES. The overall agreement for the 2004c is much better than for the other tunes. Note that the abbreviations pip, pim, kp, km, pp and pm respectively stand for π^+, π^-, K^+, K^-, proton and antiproton.

5.2.2.6 Other parameterizations in the Monte Carlo

For completeness we mention here the other parameterizations used in the gmc_disNG Monte Carlo description. For the parton density function, we made use of the CTEQ6 parameterization [118] and for the proton structure function ($F_2(x,Q^2)$) the ALLM97 parameterization was used [119]. One important thing to remark here is the failure of the CTEQ6 parameterization to describe the total structure function at low Q^2. This can clearly be seen in figure 5.8 where



(a) At Q^2 = 1 GeV^2 (b) At Q^2 = 10 GeV^2

Figure 5.8: A comparison between the ALLM F_2 parameterization (blue line) and the sum $F_2 = x \cdot \sum e_f^2 q_f(x,Q^2)$ where $q_f(x,Q^2)$ (dashed line) is parameterized by the CTEQ6 parameterization.

we compare the ALLM $F_2(x,Q^2)$ parameterization and the F_2 structure function constructed from the individual quark contributions :

$$F_2(x,Q^2) = x \cdot \sum_f e_f^2 q_f(x,Q^2) \tag{5.15}$$

parameterized by CTEQ6. In order to correct for this, each of the parton distributions $q_f(x,Q^2)$ were reweighted with the same factor explicitly requesting that the in this way constructed new parton density functions $q'_f(x,Q^2)$ add up to the ALLM97 parameterized $F_2(x,Q^2)$ value.

$$x \cdot \sum_f e_f^2 \, q'_f(x,Q^2) = F_2^{\mathrm{ALLM}}(x,Q^2) \tag{5.16}$$

where

$$q'_f(x,Q^2) = w_{F_2}(x,Q^2) \cdot q_f(x,Q^2) \tag{5.17}$$

and

$$w_{F_2}(x,Q^2) = \frac{F_2^{\mathrm{ALLM}}(x,Q^2)}{x \cdot \sum_f e_f^2 \, q_f(x,Q^2)} \tag{5.18}$$

This admittedly leaves some room for discussion. However, the newly defined functions $q'_f(x, Q^2)$ do provide us with a correct normalization of the inclusive cross section and at the same time keep the same relative contributions of the different quark flavors to the total nucleon structure function. This technique is used in the Q^2 evolution of the obtained multiplicity distribution as well (See section 6.3).

5.2.2.7 Lepton box cut

As we want to use the Monte Carlo sample to extrapolate to a 4π geometry, we do not want the scattering angles θ and ϕ to be restricted to any specific range. This seems obvious, however, for efficient generation a lot of MC productions in HERMES are done with a box cut on the lepton scattering angle. This automatically restricts the lepton to the HERMES acceptance. This is not desirable in this analysis. We will, however, leave this cut not entirely open and still keep a restriction on the lepton scattering angle of $\theta >= 0.035$ rad.

Because of the box-like shape of acceptance, the Q^2 phasespace gets distorted. By imposing a lower θ cut we obtain a good equilibrium between conceptually correct and efficient event generation. Should we completely open the lower θ cut, then what would happen is that for most of the generated events the Q^2 would be to low. The cross section rises quickly with Q^4 as Q^2 approaches 0 and LEPTO would not be able to do something usefull with these events. Figure 5.9 illustrates the two ways of generating scattered leptons. Closing the lepton scattering box cut has got small but noticeable effects

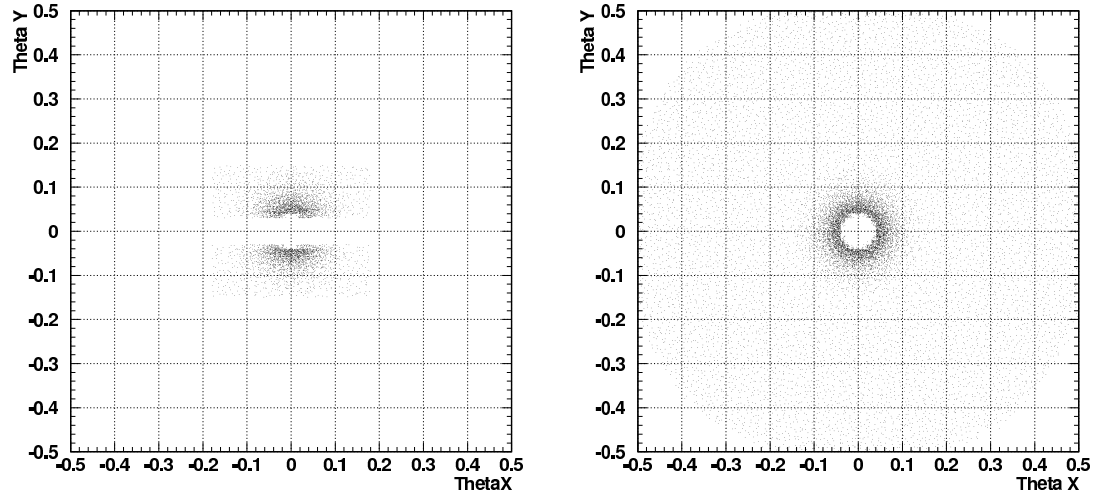


Figure 5.9: Illustration of the generation of the scattered DIS lepton for a MC production with box cuts (left) and with open cuts, leaving only the discussed lower θ cut in (right).

on the hadron acceptance function.

5.2.3 The gmc_pythia6 implementation

The main disadvantage of the gmc_disNG implementation is the lack of a VMD model to describe diffractive meson production. These contributions are present in the experimental data, and the decay products of these vector mesons cannot be easily disentangled from the pure fragmentation products. Use of the PYTHIA6 generator also has some drawbacks. The main one is that the implementation of QED radiative corrections is not so straightforward [120]. It is also worth noting that PYTHIA6 is an unweighted Monte Carlo and does not contain a model for elastic scattering. Therefore one cannot directly compare to inclusive data. If one normalizes a distribution with PYTHIA6 one should always do it using cross sections, so normalizing to luminosity and not e.g. to total number of DIS events.

The main interest we have in the PYTHIA6 implementation is whether or not it is suitable to be used to calculate the fractions of hadrons coming from diffractive processes. Since the main contribution comes from exclusive diffractive ρ^0 production, we want to make sure that gmc_pythia6 describes this process adequately. Figure 5.10 gives a comparison of the $\pi^+\pi^-$ invariant mass spectra for different regions of missing energy ΔE, preceded by a plot of the missing energy itself for an invariant mass around the ρ^0 resonance. The missing energy ΔE is the total energy of the undetected hadronic system apart from recoil proton. Around $\Delta E \approx 0$ one can observe the exclusivity peak. The comparison is at cross section level and one can clearly see that the version of gmc_pythia6 that was used, manages to describe very nicely the missing energy spectrum as well as the invariant mass spectra both for exclusive diffractive ρ^0 mesons ($\Delta E < 1.0$ GeV) and fragmentation ρ^0's ($\Delta E > 3.0$ GeV). Proof that this is not only the case for ρ^0 can be found in figure 5.11, where the K^0_S, $\pi^+\pi^-$-invariant mass spectra is compared on cross section level with equally satisfactory results.

In [100], one can also see that the PYTHIA6 simulation better describes the high z part of the hadron spectra since this is exactly where one would expect a large contamination of pions coming from diffractive ρ^0 decay.

5.3 Discussion on target fragmentation relying on Monte Carlo

Finally we will come back to the quote of J.D. Bjorken at the beginning of this chapter. Monte Carlo generators have indeed become an essential tool for feasiblility studies, simulation of anticipated background rates, acceptance corrections etc... However we must not blindly follow the models into their smallest of details because this is often where 'ad hoc' assumptions are made and they lose their real physical meaning. Let us have a look at an example triggered by a criticism in [121]. Inside the JETSET fragmentation model a switch is implemented

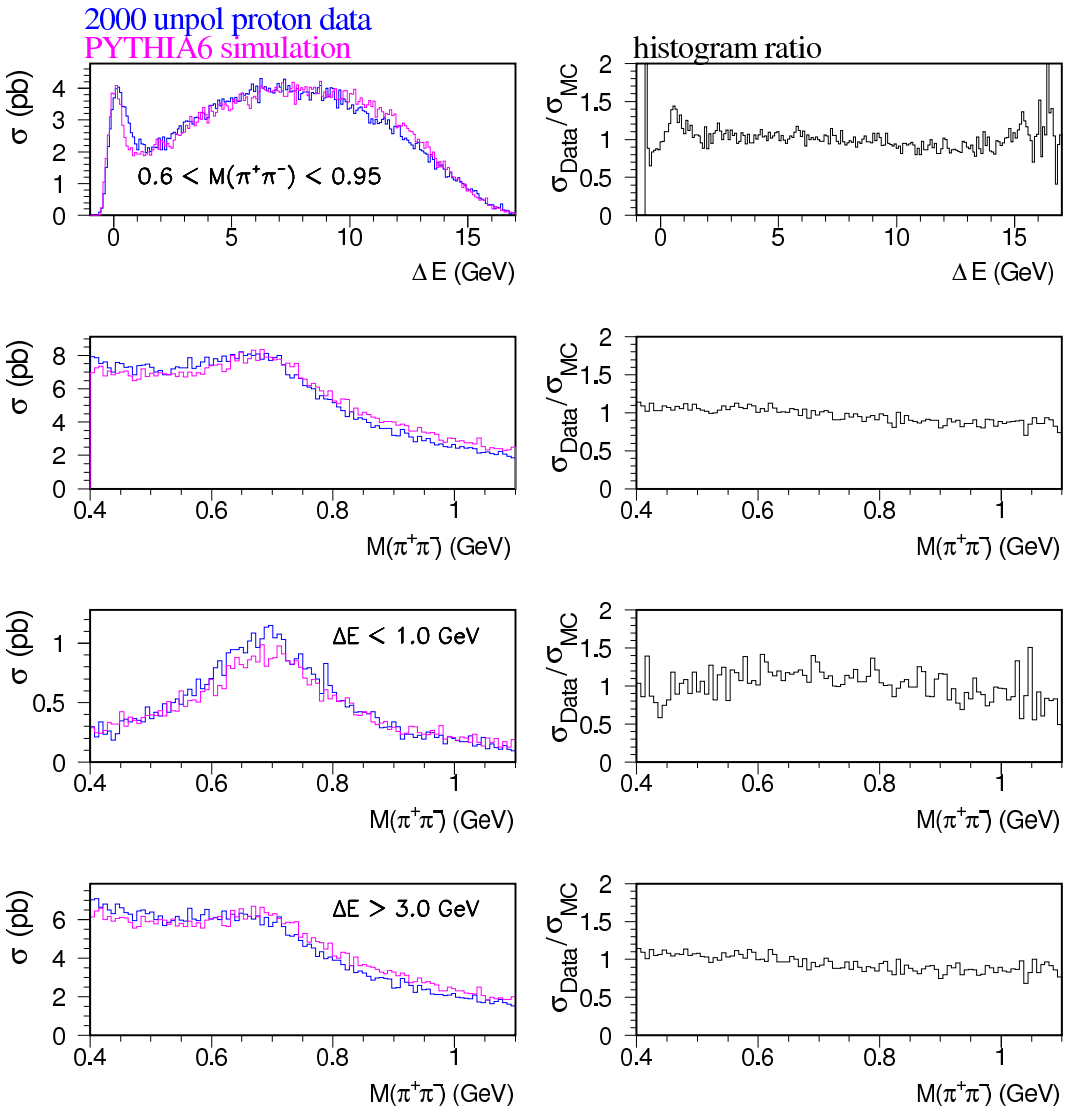


Figure 5.10: Comparison of the $\pi^+\pi^-$ invariant mass spectrum for different ΔE cuts. The comparison is on cross section level between HERMES unpolarized proton data (00c1) and the PYTHIA6 simulation including radiative corrections. Graph taken from [100]. On the right hand side of each spectra one sees the ratio of the experimental and simulated spectra.

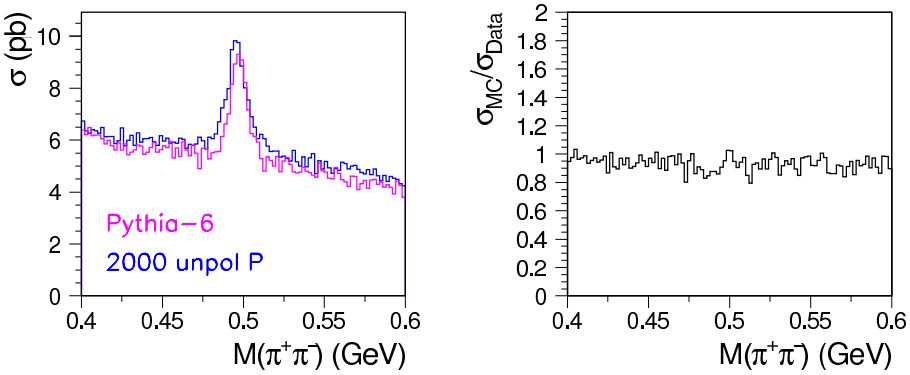
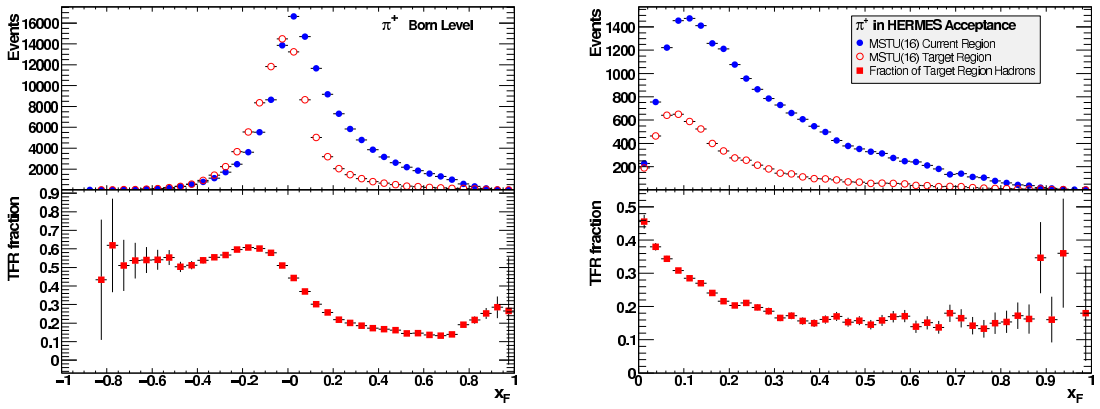


Figure 5.11: Cross section comparison for the K^0_S between the experimental data and the PYTHIA6 simulation.

```
MSTU(16) = 2
```

which enables tracing the hadrons back to the end point of the color string from which they originated, instead of just the string as a whole. In a naive approach one could think to use this information to distinguish between current and target fragmentation regions. A hadron belongs to the **current fragmentation** region if it traces back to the struck quark end of the string, and vice versa, a hadron belongs to the **target fragmentation** region if it traces back to the diquark or target end of the string. Should we trust this, we see



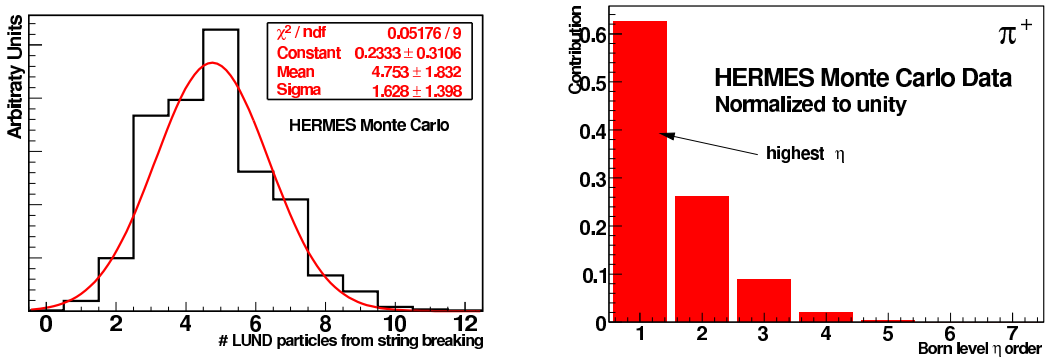
(a) in 4π geometry (b) in HERMES acceptance

Figure 5.12: x_F histograms for different hadrons distinguishing between the **current region (CFR)** and the **target region (TFR)** using the JETSET `MSTU(16)` switch. We've plotted this for both 4π geometry as well as inside the HERMES acceptance. Beneath each plot, the supposed fraction of TFR contamination is plotted (TFR events / all events).

that at HERMES energies a rather substantial contribution from target fragmentation (see fig 5.12), even though HERMES is only capable of detecting the forward region (x_F or $\eta > 0$). However, this kind of classification is taking the JETSET model one step too far [114]. In the JETSET model the fragmentation just starts from both ends of the string and moves inwards. Moreover, the determination of the pointer in the last string breaking is completely random. This, together with the fact that there are only a very limited number of string-breakings at HERMES energy (see figure 5.13(a)) leads to the conclusion that there is no real physical meaning in terms of current or target region when using this switch.

In the light of this discussion it might be interesting to look at this issue from a different perspective. Figure 5.13(a) shows the number of hadrons generated from the string. One can see that the average multiplicity is around 4.7, which is directly correlated to the number of string breakings. Figure 5.13(a) was fitted with a Gaussian to determine this number. From figure 5.13(b), we see that about 90 % of the pions ending up in the HERMES acceptance were either highest or second highest in rapidity at Born level. The rapidity order is defined by sorting the hadrons at Born level according to their rapidity, where we assign 1 to the the hadron with the highest rapidity.

This gives a somewhat clearer estimate of possible target fragmentation



(a) Hadron Multiplicity at Born level

(b) Rapidity ordering at Born level

Figure 5.13: On the left we show the hadron multiplicity at Born level. On the right hand side we have plotted the rapidity ordering at Born level for positive pions ending up in the HERMES acceptance.

contributions based upon our Monte Carlo. In any case it is difficult to estimate the contribution since the HERMES average W is only about 5 GeV. This is illustrated by the Berger criterion [61], see figure 2.19 in section 2.5. The real check for factorization and universality needs to come from comparing HERMES fragmentation data to higher energy and/or e^+e^- results.

We have repeated the exercise of figure 5.13(b) for π^-, K^+ and K^- as well. We also checked the effect of applying a z cut of 0.2 on the accepted hadrons as was used e.g. in the Δq analysis [4]. The results hereof are shown in figure 5.14. We can see quite some influence of the z cut. Selecting the lower z region, we clearly get more hadrons from the more central rapidity region. For $z > 0.2$ almost all hadrons in the acceptance are the highest rapidity hadrons. This is according to the Berger criterion indeed the value needed for the target and current jets to be separated [63]. There is also a striking difference between the hadron species. K^- appearantly get produced at much lower rapidity. Hadrons that contain the struck quark are found at high rapidity. At HERMES energy the sea quark contribution in the proton is still quite small compared to the valence quarks (as can be seen from figure 5.15). The contribution of K^- containing a struck quark, at high rapidiy, is therefore suppressed.

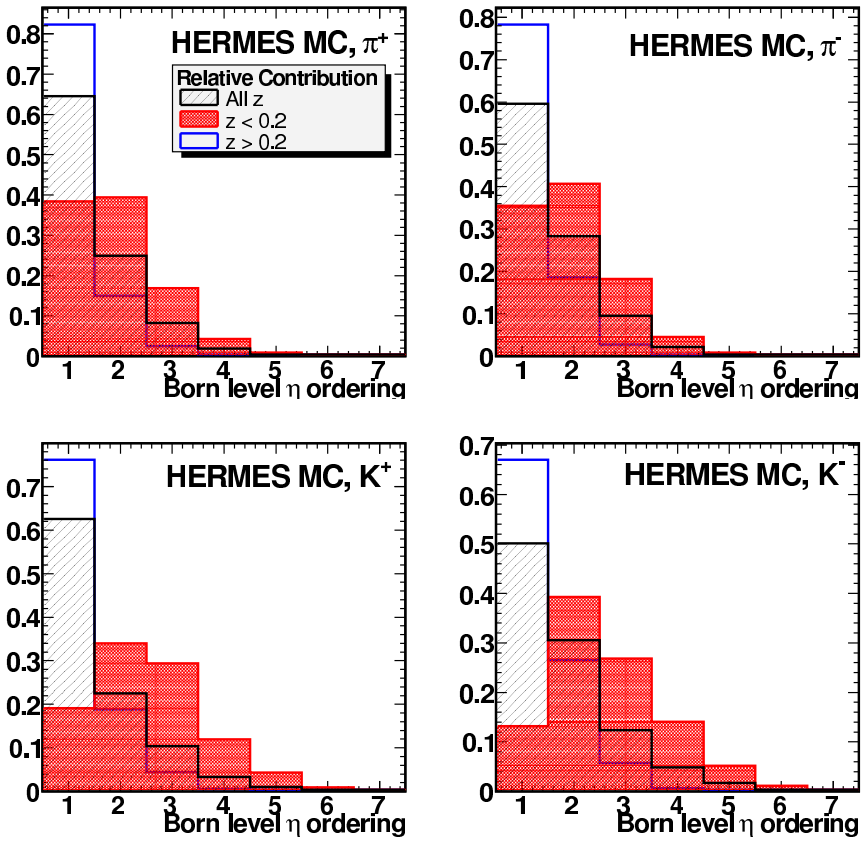


Figure 5.14: Rapidity ordering at Born level for hadrons ($\pi^{\pm}$ and $K^{\pm}$) in the acceptance. The shaded histogram represents the contributions for $z < 0.2$, the empty histogram $z > 0.2$ and the shaded all z.

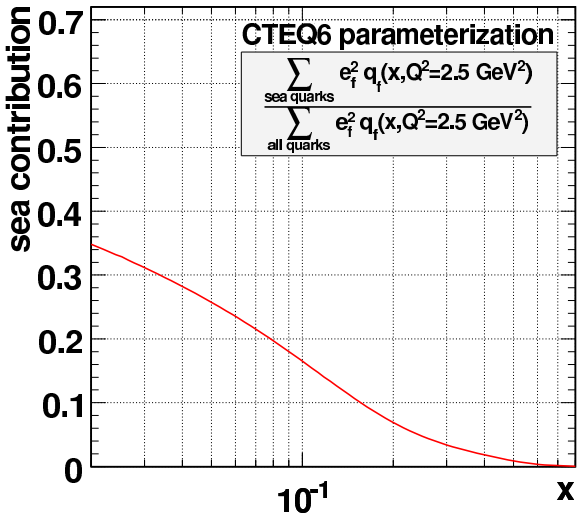


Figure 5.15: Contribution of the sea quarks tot the total structure function. CTEQ6 was used as parameterization.

Chapter 6

Corrected multiplicity distributions

6.1 Contamination from diffractive processes

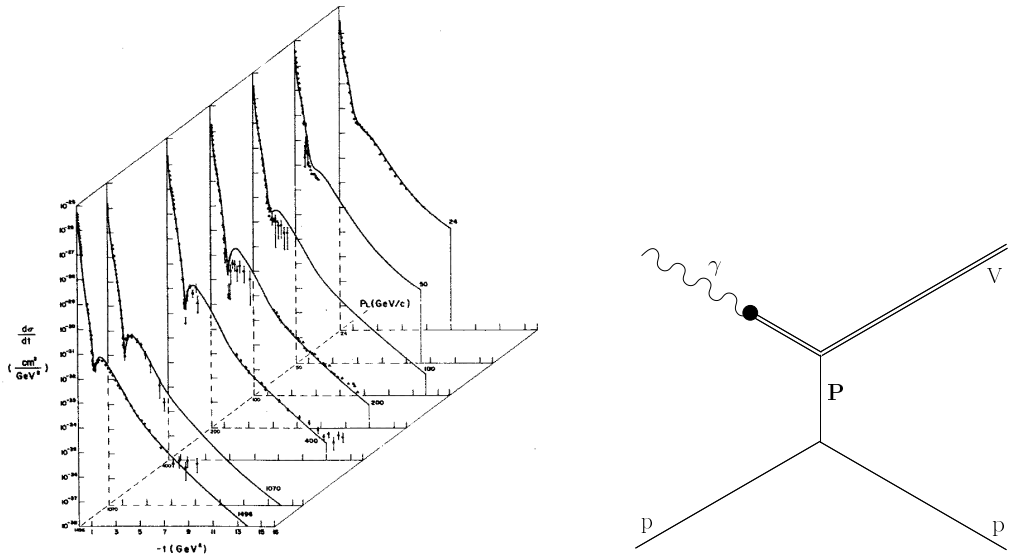
The phenomenon of diffraction has its origins in optics, where the typical example is the observation of a series of minima and maxima when light falls on a system of narrow slit openings. In elementary particle physics, a similar behavior was found in the $d\sigma/dt$ cross section of proton-proton scattering [122] with t the Mandelstam variable :

$$t = (p_1 - p_1')^2 = (p_2 - p_2')^2 \tag{6.1}$$

p_i and p_i' are the four momentum vectors of the two protons respectively before and after the interaction. This diffractive behavior can be seen in figure 6.1(a).

Diffractive interactions in deep inelastic lepton hadron scattering are understood in terms of a hadronic fluctuation of the virtual photon which interacts with the proton target by exchanging an object which has the quantum numbers of the vacuum. These objects, typically **pomerons** or **reggeons** are described in Regge theory and have been found to have a partonic structure [123]. Diffractive processes are typically described in *vector meson dominance* (VMD) models [124].

As will be shown in this section a substantial contribution to the total semi-inclusive scattering cross section comes from the decay of diffractively produced vector mesons [123], like the ρ^0, the ω and ϕ mesons. Especially the π^+ and π^- coming from exclusive diffractive ρ^0 decay constitute a big background fraction at high z. There is no experimental way to tag individual hadrons coming from this kind of events, so we have to rely on Monte Carlo to estimate the contribution. Obviously, a generator with both a diffraction as well as a deep-inelastic model is needed. PYTHIA6 [102], especially tuned to describe HERMES experimental data and include RADGEN radiative corrections [120] was used for this task. From the obtained level of agreement with the HERMES experimental data, as seen in section 5.2.3, we conclude that the Monte Carlo is in good condition to be used for determining this background.



(a) (b)

Figure 6.1: On the left $d\sigma/dt$ (pp) cross section showing diffractive behavior [122]. On the right: a schematic drawing of diffractive vector meson production in $\gamma^* p$ scattering in the VMD description showing the hadronic fluctuation of the virtual photon.

One could, however, argue that in fact the diffractive cross section is part of the total inclusive e^+p cross section. Nevertheless the fact remains that these hadrons are produced in a very different way. Moreover, their kinematical distributions follow totally different shapes as compared to fragmentation. Therefore it is at least desirable to see what the effect is of these contributions. The final results will be presented with and without this background subtracted.

6.1.1 Extraction of the background fraction

We now decompose the observed cross section in a non-diffractive part (σ_{nd}) and different diffractive contributions. One can distinguish the **elastic diffractive** contribution σ_{el} which is the main one: it has an elastic scattering between the proton that remains in its ground state and the partonic fluctuation of the virtual photon which is put on shell. The other contributions contain k diffractive subsystems σ_d^k [125]:

$$\sigma_{tot} = \sigma_{nd} + \sigma_{el} + \sum_k \sigma_d^k \tag{6.2}$$

where the two main contributions are the single and double diffractive where

$$\gamma^*_{q\bar{q}} + p \rightarrow X + p \tag{6.3}$$

for a single diffractive subsystem X and

$$\gamma^*_{q\bar{q}} + p \rightarrow X_1 + X_2 \tag{6.4}$$

for two (double) diffractive subsystems, showing a large rapidity gap between them. Here $\gamma^*_{q\bar{q}}$ represents the hadronic fluctuation of the virtual photon. Within PYTHIA6 the process numbers (ISUB) for these subprocesses are 91 for elastic diffractive, 92 for single diffractive and 93 for double diffractive [102]. These subprocesses can be seen in figure 6.2, where indeed we see that the elastic diffractive process constitutes the main contribution.

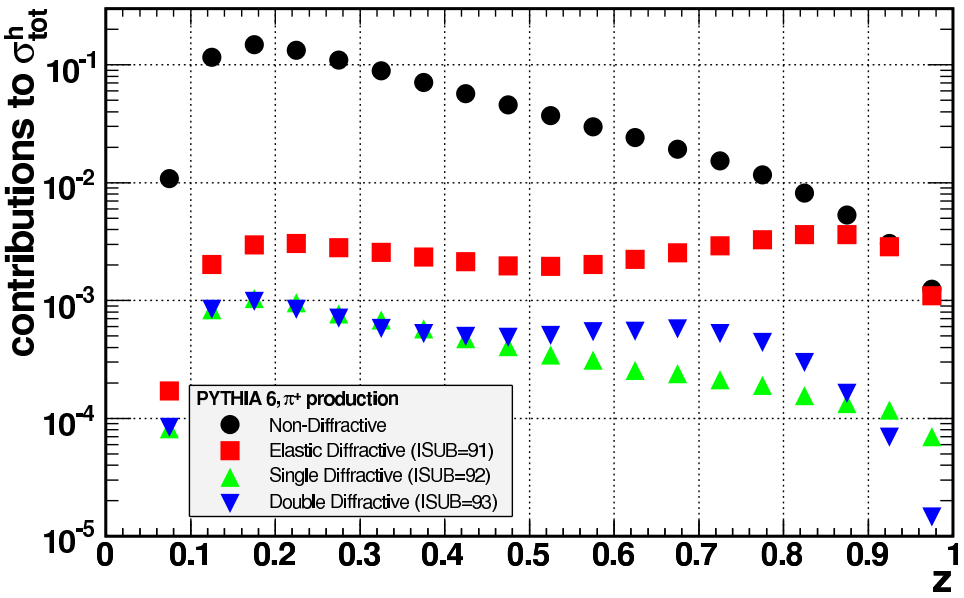
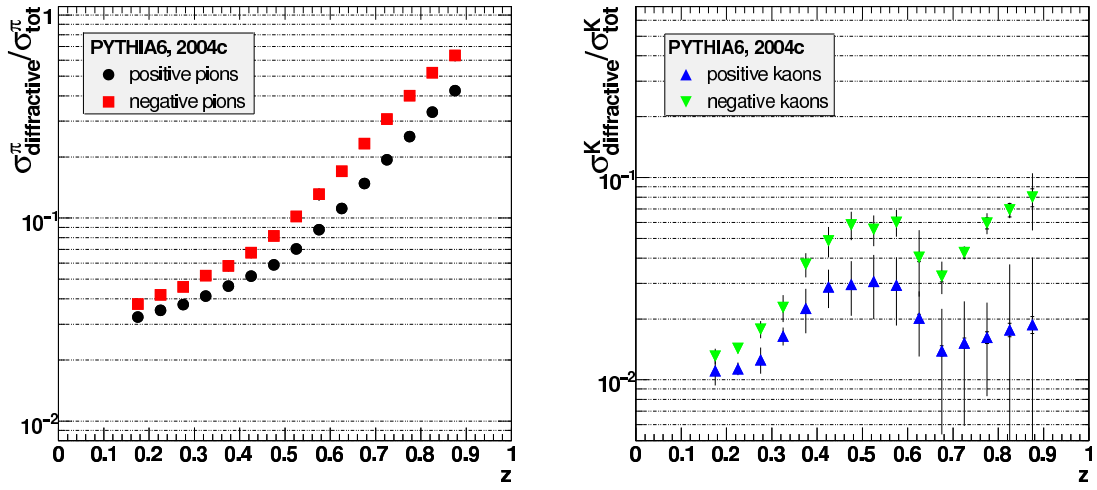


Figure 6.2: Contribution of the individual subprocesses to the total π^+ cross section versus z according to PYTHIA 6. The graphs were normalized to obtain unit integral for the total of the contributions.

To estimate the background we determine what fraction of the hadrons within the analysis criteria discussed above belongs to any one of these subprocesses. Figure 6.3 shows the diffractive background fraction as function of z for π^+, π^-, K^+ and K^-. One can see that indeed there is a substantial contribution at large z. Fragmentation hadrons will be most abundant at lower z. Because of the forward Lorentz boost we observe the ρ and ϕ vector mesons to decay very asymmetrically in the laboratory frame, yielding one hadron at high z and one at lower. The enhancement around $z = 0.5$ we see for the kaons is due to double diffractive events. In this case the proton will not stay in its ground state after the interaction and therefore an energy sharing has to occur between the diffractively produced ϕ meson and the proton subsystem. The Lorentz boost will not be so large anymore for the produced meson, therefore its decay is less asymmetric in the labframe, causing the enhancement at intermediate z. This behavior is also present for pions, however it is obscured by the magnitude of the elastic diffractive contribution. The inner errorbars are statistical, the outer ones the systematic errors with the statistical errorbars added in quadrature. This contribution is estimated very conservatively as the absolute difference between this analysis and the values used in [100] divided by $\sqrt{3}$ (assuming a uniform distribution between the two values).



(a) (b)

Figure 6.3: Diffractive contamination fraction using PYTHIA6 data.

It is striking that for π^- the diffractive contribution appears to be largest. This is due to the simple fact that there are less π^- produced in fragmentation, due to the dominance of the u quark in the nucleon. This can clearly be seen from equation 2.15 where because of its charge of $2/3$ the u quark carries much more weight than the d quark. In fact this would still hold for the neutron as well, even though there there are 2 valence d quarks present. So π^- have to come more from unfavored fragmentation of which we know the probability is lower. The background contributions shown in figure 6.3 we can subtract from the hadron yields (N_i^h) including the diffractive contribution in the following way:

$$(N_i^h)_{\text{excl diffr.}} = (N_i^h)_{\text{incl diffr.}} \times (1 - \frac{\sigma_{diffr}^{h,MC}}{\sigma_{tot}^{h,MC}}) \tag{6.5}$$

Figure 6.4 shows the effect of this background correction on the experimental multiplicity distributions. The corrected data are included in appendix (tables B.5 and B.6).

6.1.2 Error propagation of the diffractive background correction

A special note has to be made about the error propagation. In fact, what is done is that we subtract a background fraction from the spectra, so even though the correction factors are multiplicative, we still need to proceed as if the background is subtracted since this is how the two quantities are statistically correlated. Assume that we have a number of counts N in a certain kinematical bin. As said above, a correction factor, C, which is the ratio between the number of hadrons coming from diffractive events and the total

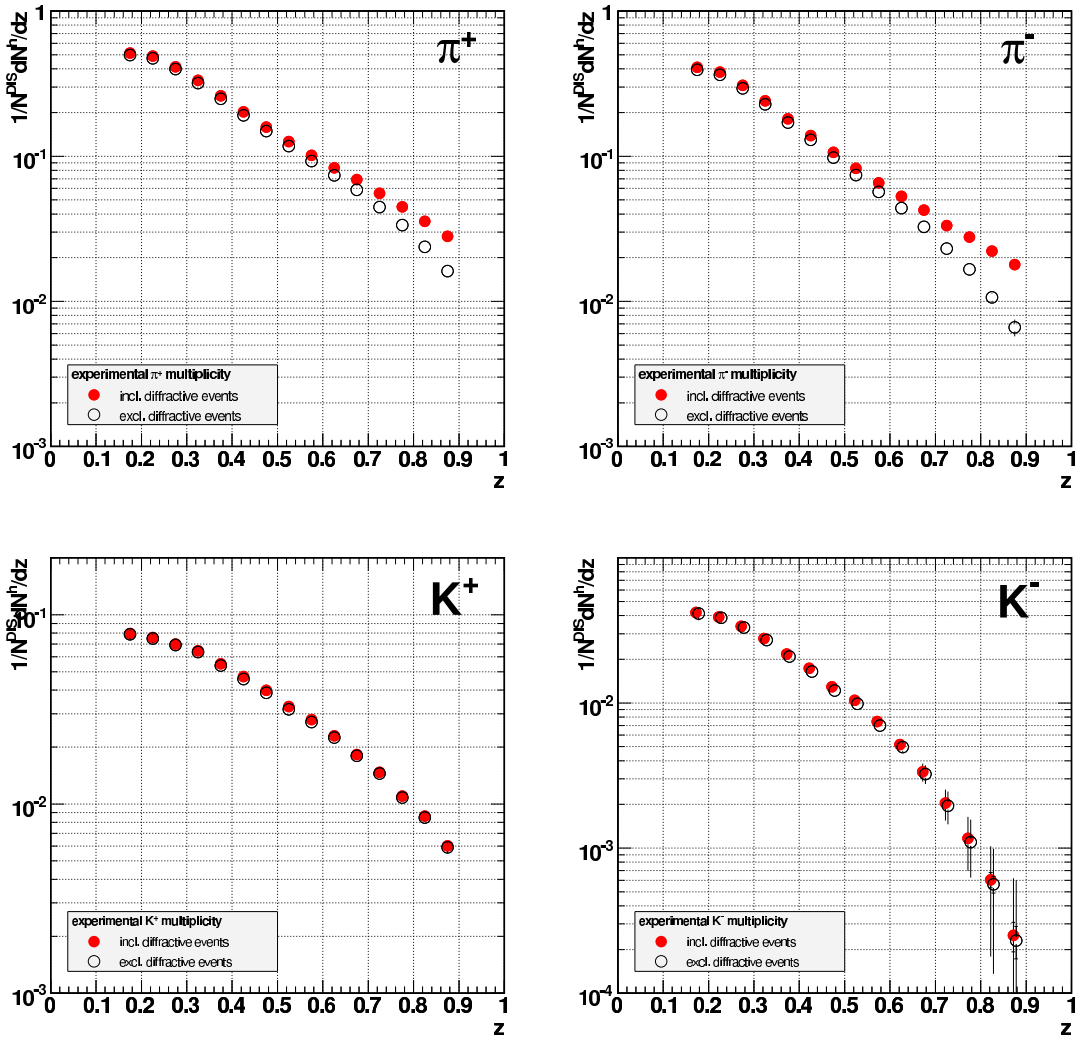


Figure 6.4: Experimental multiplicities, including (red closed circles) and excluding (white open circles) the diffractive background. The points for K^- have been shifted slighlty from the bincenter in z to the left and to the right a bit for clarity.

number of hadrons within a certain bin was calculated using PYTHIA6. Hence the corrected number of events is

$$N' = N(1 - C) \tag{6.6}$$

If we now assume that the correction is done in a multiplicative way, we treat the quantities N and C as being statistically independent. So we would calculate for the error propagation:

$$\sigma_{N'} = \sqrt{\sigma_N^2(1-C)^2 + N^2\sigma_C^2} \tag{6.7}$$

However, this is not correct since we subtract a background. Thus we have to assume that the total number of events and the background itself are statistically independent. Hence

$$N' = N - B \tag{6.8}$$

where $B = NC$, and we get

$$\sigma_{N'} = \sqrt{\sigma_N^2 + \sigma_B^2} \tag{6.9}$$

or

$$\sigma_{N'} = \sqrt{\sigma_N^2 + N^2\sigma_C^2 + C^2\sigma_N^2} \tag{6.10}$$

$$= \sqrt{N^2\sigma_C^2 + (1+C^2)\sigma_N^2} \tag{6.11}$$

We see that in this case we get an inflation of the error due to the $(1 + C^2)$ factor even if the uncertainty σ_C is very small (or even negligible). Figure 6.5 clearly shows the difference between a multiplicative error calculation and a subtractive one compared to the relative uncertainty on the experimental data as discussed in previous chapters.

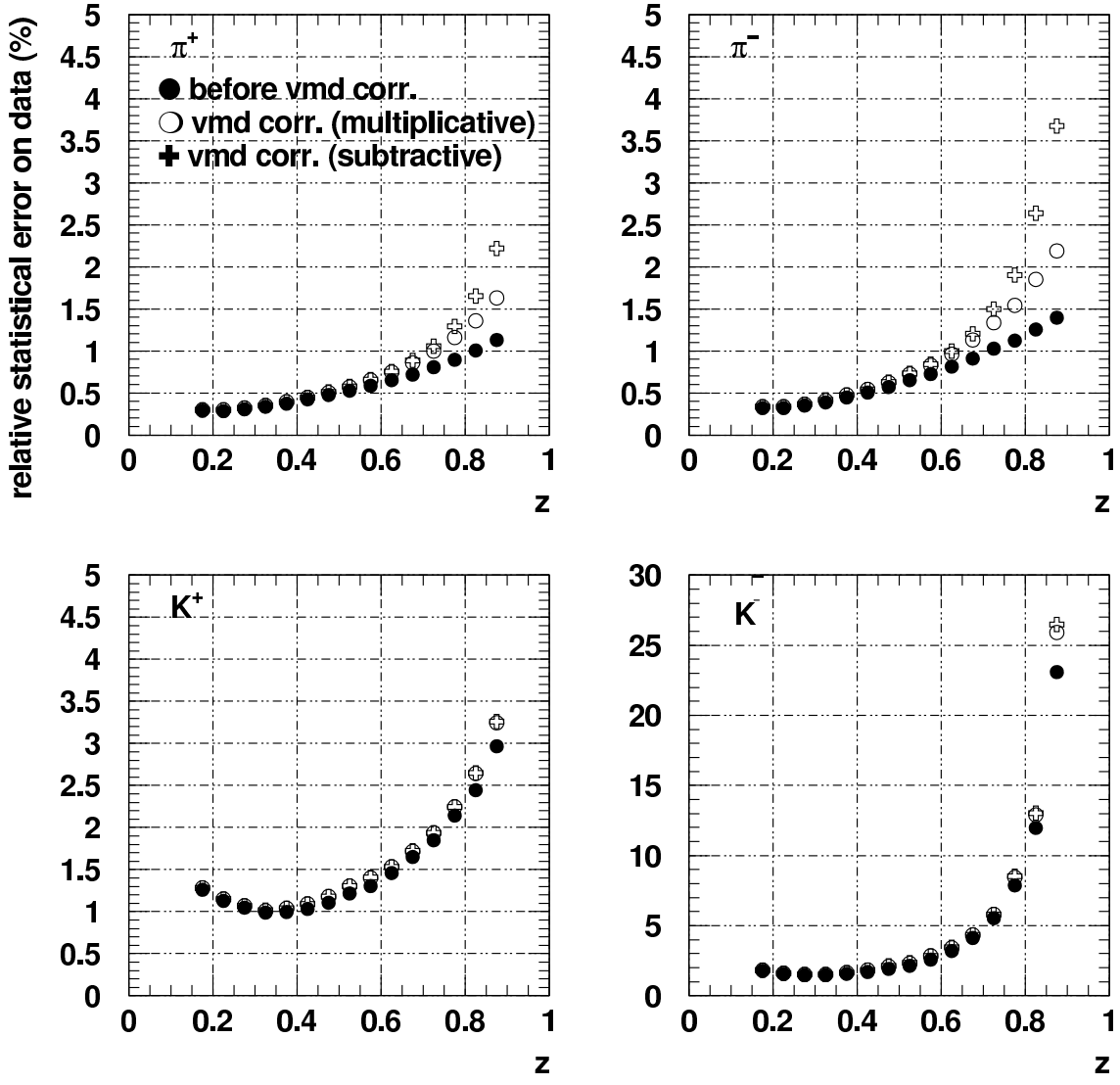


Figure 6.5: Comparison between the two error calculations. Full circles show the original statistical error before VMD correction, open circles the incorrect multiplicative propagation method. The crosses result from the correct subtractive error propagation.

6.2 Unfolding the radiative and experimental effects

As HERMES is a very forward spectrometer, its geometrical acceptance is limited. Results within the HERMES acceptance therefore have to be extrapolated using a Monte Carlo simulation. Obviously also detector efficiency and kinematical smearing due to the limited resolution of the setup have to be corrected for. Since the HERMES Monte Carlo is able to calculate QED radiative effects on an event level basis, we can reconstruct the actual radiative smearing as well. This smearing has a similar effect on measured quantities as kinematical smearing, namely migration of events between bins. It therefore makes sense to treat all these smearing effects in one step. For this we will use the `gmc_disNG` Monte Carlo simulation.

A classical method to correct an experimental quantity $\sigma_X(\Phi)$ for detector acceptance effects, efficiency and even QED radiation is to calculate an acceptance function. This is usually done using simulations where one calculates the quantity $\sigma_X^{MC}(\Phi)$ using Monte Carlo including all experimental and radiative effects and divide this by the same quantity using generator Born-level. Monte Carlo data $\sigma_B^{MC}(\Phi)$. One can then construct an acceptance function

$$\mathcal{A}(\Phi) = \frac{\partial\sigma_X^{MC}(\Phi)}{\partial\Phi} \Big/ \frac{\partial\sigma_B^{MC}(\Phi)}{\partial\Phi}. \tag{6.12}$$

Φ represents here a binning in arbitrary kinematics. The measured Born level cross section is trivially given by

$$\frac{\partial\sigma^B(\Phi)}{\partial\Phi} = \frac{1}{\mathcal{A}(\Phi)} \frac{\partial\sigma^X(\Phi)}{\partial\Phi} \tag{6.13}$$

Of course this relation only holds if the Monte Carlo simulation incorporates all detector inefficiencies and radiative effects. However, the problem lies exactly in the binning Φ. Because of QED radiation and inherent non-perfect resolution of the detector, the kinematics of each event are altered to some extent, leading to migration of the events in-between different bins. In case of radiative smearing this can in fact be rather substantial as will be seen later on. Obviously this means that different bins of the spectrum are highly correlated and these correlations should be taken into account when propagating any uncertainties through the procedure. When using equation 6.13 one completely neglects these correlations and over or underestimates the uncertainties in the final corrected spectrum. Therefore we can only apply a simple bin-by-bin acceptance correction if the amount of event migration is negligible.

A method to deal with this is to explicitely trace the migration of events by means of a migration matrix and to use this to calculate the correlations between the different bins. An **unfolding method** based on matrix inversion was described for HERMES by Andrew C. Miller in an internal note [126]. We use this method in a slightly adjusted form to correct our experimental spectra.

6.2.1 The formalism

Assume a Monte Carlo simulation which includes all experimental and radiative effects. Let for a certain process p, $n_p^X(i)$ be the number of Monte Carlo events that fall within a certain **experimental bin** i and $n_p^B(j)$ the ones within a **Born level bin** j. (Note that in this section we will always denote experimental bins using the index i and Born level bins using j.) Both i and j can represent an arbitrary binning in one **or more** kinematical variables. Next, we also define a matrix $n_p(i,j)$ where each element holds the number of events that fall in both Born level bin j and experimental bin i. In other words, each element holds the events that originated in Born bin j and were smeared into experimental bin i, by either detector or radiative effects. In this way we account for all smearing at the same time. If our experimental spectrum had n_X bins and the Born spectrum n_B, then the matrix $n_p(i,j)$ has dimensions $n_X \times (n_B + 1)$. The reason for the additional bin at Born level is the following: The experimental sample is defined by a number of kinematical cuts which we apply to the data, like $Q^2 > 1$ GeV^2 and $W^2 > 10$ GeV^2. The same kinematical cuts were also applied to the Born level Monte Carlo data as these are the general DIS kinematics for which we want to extract the quantities. Due to smearing, however, it is possible that events from outside these cuts are smeared into the experimental sample, hence the extra bin at Born level ($n_B + 1$), yielding an extra $j = 0$ column in the $n_p(i,j)$ matrix. In this way we can easily see that

$$n_p^X(i) = \sum_{j=0}^{n_B} n_p(i,j). \tag{6.14}$$

However, it is important to note that

$$n_p^B(j) \neq \sum_{i=1}^{n_X} n_p(i,j) \tag{6.15}$$

since radiative effects do not conserve the total DIS cross section, and events can migrate outside the initially generated box and are thus never detected. One therefore looses the normalization between the Born level and experimental Monte Carlo samples as also discussed in section 5.2.2. This problem is easily circumvented by generating two statistically independent Monte Carlo samples, one at Born level without radiative effects and one including both radiative and detector effects. From these $n(i,j)$ matrices we will define the so called **Smearing Matrix** $S_p(i,j)$ as :

$$S_p \equiv \frac{\partial \sigma_p^X(i)}{\partial \Phi_i} \Big/ \frac{\partial \sigma_p^B(j)}{\partial \Phi_j} = \frac{n_p^X(i)}{n_p^B(j)} \tag{6.16}$$

As mentioned in the last paragraph, radiative effects do not conserve the cross section so the S matrices are not unitary[1]. Therefore their elements do not

[1] Or since we are dealing with real matrices: orthogonal ($SS^T \neq 1$ or $S^{-1} \neq S^T$).

represent probabilities. We can write expression 6.16 in terms of our migration matrix :

$$S_p(i,j) = \frac{n_p(i,j)}{n_p^B(j)} \tag{6.17}$$

An important property for these smearing matrices is that the nominator and denominator scale together with the number of generated events in the Born bin, reducing model dependence. Let us now continue trying to extract a Born level distribution $B(j)$, from an experimental distribution $X(i)$. Note first that

$$n_p^X(i) = \sum_{j=0}^{n_B} S_p(i,j) \cdot n_p^B(j). \tag{6.18}$$

Following this equation and provided that our Monte Carlo simulation describes reality, we can write for the experimental distribution :

$$X_p(i) = L_p k_p(i) \sum_{j=0}^{n_B} S_p(i,j) \cdot B_p(j) \tag{6.19}$$

Where :

- $k_p(i)$: a vector with unknown normalization constants that incorporate unsimulated detector efficiencies.
- L_p : overall luminosity for the process p

This equation is in fact the analogue of 6.13. Starting from equation 6.19 one can extract the ratio $R_B(j) = B_2(j)/B_1(j)$ for two general processes $p = 1, 2$ [126], yielding :

$$R_B(j) = \frac{1}{n_1^B(j)} \times \sum_{i=1}^{n_X} [S_2']^{-1}(j,i) \left[\frac{R_X(i)}{R_L} n_1^X(i) - n_2(i,0) \right] \tag{6.20}$$

where $R_X(i) = X_2(i)/X_1(i)$. The matrix S_2' is in fact the trucated smearing matrix for process 2 in which we leave out the special $j = 0$ column. This only applies for equally binned spectra for both processes 1 and 2. However, we want to use this formalism for multiplicities which are ratios of a hadron kinematical distribution ($p = 2 = h$) and a number of DIS events ($p = 1 = DIS$). We will prove here that the formalism can still be used as described in [126]. Whether or not the model dependence is reduced here remains to be proven.

Since in our case the hadron sample is extracted from the DIS event sample, both cross sections have the same experimental luminosity, so we can equate L_1 and L_2. As already stated above we only have the z binning in the case of the hadrons (denominator). So we write equation 6.19 for hadrons and DIS leptons:

$$X_h(i) = L\, k(i) \sum_{j=0}^{n_B} S_h(i,j)\, B_h(j) \tag{6.21}$$

for the hadrons, and

$$X_{DIS} = L\,k\,\left(S_{DIS}\cdot B_{DIS} + S^0_{DIS}\cdot B^0_{DIS}\right) \tag{6.22}$$

for the DIS events. In this last B^0_{DIS} and S^0_{DIS} represent the Born DIS events that are smeared in to the acceptance together with its respective element in the smearing 'matrix'. We could actually rewrite this equation in the same form as 6.21 just by posing that in this case $S(i,j)$ is a 1×2 matrix. As $k(i)$ is an unknown vector of **unsimulated** inefficiencies we will have to assume that these are uniform over the kinematic range. The validity of this statement is supported by the fact that we obtained a good description of our experimental data using the Monte Carlo and therefore can assume that the contribution of any unknown effects is very small. Doing this we can write $k(i) = k$ and can solve equation (6.22) for k:

$$k = \frac{X_{DIS}}{L\,(S_{DIS}\cdot B_{DIS} + S^0_{DIS}\cdot B^0_{DIS})} \tag{6.23}$$

and write for the experimental hadron yield in bin i where we cancelled the luminosity in both nominator and denominator:

$$X_h(i) = \frac{X_{DIS}\,\sum_{j=0}^{n_B} S_h(i,j)\,B_h(j)}{(S_{DIS}\cdot B_{DIS} + S^0_{DIS}\cdot B^0_{DIS})} \tag{6.24}$$

Separating the row $j=0$ from the smearing matrix $S_h(i,j)$, we construct a square matrix, the truncated smearing matrix, which will be denoted as $S'_h(i,j)$

$$\frac{X_h(i)}{X_{DIS}} = \frac{\sum_{j=1}^{n_B} S'_h(i,j)\,B_h(j) + S(i,0)B_h(0)}{S_{DIS}\cdot B_{DIS} + S^0_{DIS}\cdot B^0_{DIS}} \tag{6.25}$$

Obviously the nominator $(S_{DIS}\cdot B_{DIS} + S^0_{DIS}\cdot B^0_{DIS})$ on the right hand side of the equation (6.25) is just the number of DIS events from the Monte Carlo at experimental level, n^X_{DIS}. So we have :

$$\sum_{j=1}^{n_B} S'_h(i,j)\,B_h(j) = \frac{X_h(i)}{X_{DIS}}\cdot n^X_{DIS} - S(i,0)\cdot B_h(0) \tag{6.26}$$

If the square matrix $S'_h(i,j)$has an inverse we can write

$$B_h(j) = \sum_{i=1}^{n_X}[S'_h]^{-1}(j,i)\left[R_X(i)\cdot n^X_{DIS} - S(i,0)\cdot B_h(0)\right] \tag{6.27}$$

where $R_X(i)$ is the experimentally observed ratio of number of hadrons within a bin i to the total number of DIS events[2]. Of course we want to have the ratio at Born level. So we put in again the Born DIS cross section from the Monte

[2]Note that we normalize this ratio R_X, as well as the R_B with the bin width, dz to obtain a real multiplicity.

Carlo, of which we know that it describes the experimental DIS cross section very well. This way we obtain:

$$R_B(j) = \frac{1}{n_{DIS}^B} \times \sum_{i=1}^{n_X} [S_h']^{-1}(j,i) \left[R_X(i) \cdot n_{DIS}^X - n_h(i,0) \right] \tag{6.28}$$

We introduced here $n_h(i,0)$, assuming that the Monte Carlo accounts for all the contaminating processes outside the acceptance. As discussed, the `gmc_disNG` Monte Carlo simulation does not take into account diffractive contributions, but we have corrected for these in the data.

6.2.2 Monte Carlo input to the correction procedure

Once the input matrices $n_h(i,j)$ and Born level distributions n_h^B are known, applying the formalism that was developed in the previous section is rather straightforward. One issue that nevertheless needs to be brought up here is the normalization of the Monte Carlo input matrices. As previously mentioned, two statistically independent Monte Carlo samples are required. However, one does not have to generate an equal number of events at Born and experimental level. If not, than normalizing the extracted migration matrices $n(i,j)$ and distributions $n_B(j)$ to the Monte Carlo luminosity is the way to go in this case. The Monte Carlo luminosity can be extracted in `gmc_disNG` simply by counting the total number of events generated, which is stored in `g1MEvent.iEvGen` :

$$\mathcal{L}_{MC} = \sum_{generated\ events} \text{g1MEvent.iEvGen} \tag{6.29}$$

6.2.2.1 Extraction of the migration matrix $n(i,j)$

As we are mainly interested in the z and x dependence of the hadron multiplicities, we will from now on label the bins with their correct kinematic variable as well. In order to fill the $n(i,j)$ matrix for the hadrons, one has to determine in which bin the experimental and the Born level z value fall. For this we of course use a Monte Carlo sample in the HERMES acceptance including all detector and radiative effects. However also the generated information is stored in the data files, so we can easily look up the z value at the Born level. In practice this means that one has to take the generated energy of the hadron and the true ν value of the event, which is stored in the `g1MEvent` table. Care must be taken when filling the special bin in which events reside that get smeared into the experimental sample from outside of the cuts. One has to repeat every selection criterion, imposed on the Monte Carlo at experimental level, with the Born-level kinematics. This incorporates implicit selection criteria such as the HERMES geometrical acceptance and fiducial volume cuts on the calorimeter as well. For this the track parameters after the bending in the dipole magnet are determined using momentum lookup tables. Of each track the offset and change in slope behind the magnet are searched in a lookup table. With

this information one can determine the impact positions of the tracks on the calorimeter if no radiative smearing had occurred[3].

The final matrices that were obtained in this way are shown in figure (6.6). The $n^{\pi,K}(i,j)$ matrices look very asymmetric. The kinematical smearing caused

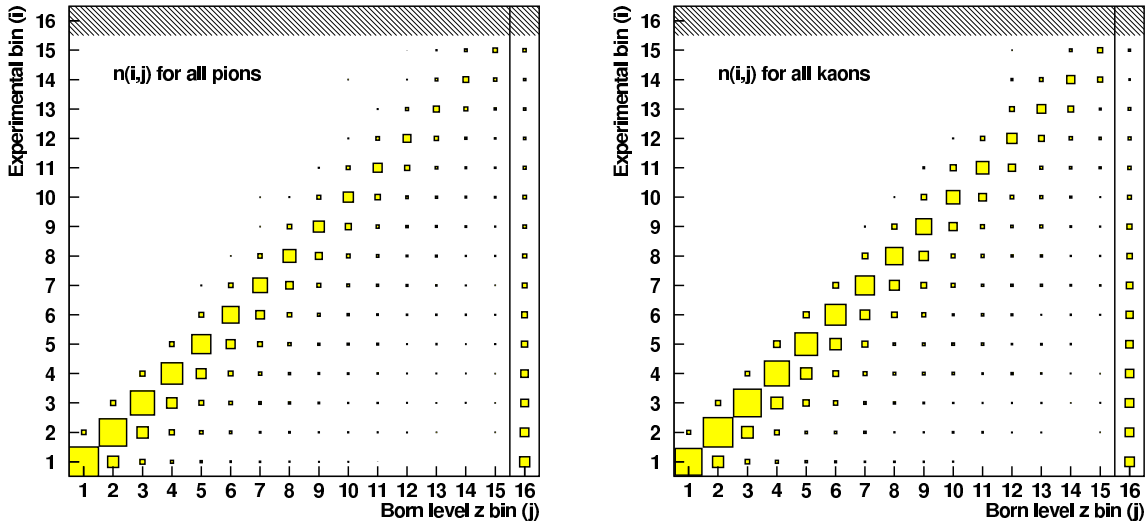


Figure 6.6: The $n(i,j)$ matrices. The extra bin $j=0$ at the Born level is here depicted as $j=16$. On the left side, the matrix for all pions together and on the right side for the kaons.

by the spectrometer and the track reconstruction is limited to at most one bin in z and works both way, so increasing or reducing z. However, radiative smearing is much larger, and works only in one direction. Figure 6.7 illustrates how both initial and final state radiation see to it that the Born level ν is always smaller than the reconstructed, or experimental ν. We know that

$$\nu = E - E' \qquad \text{and} \qquad \nu_{Born} = E_{Born} - E'_{Born} \tag{6.30}$$

For initial state radiation one has (figure 6.7)

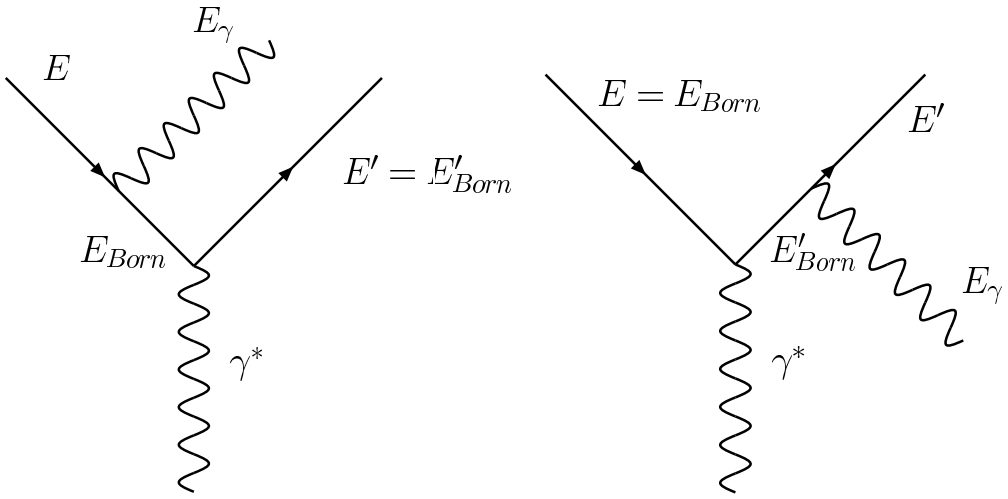
$$\nu_{Born} = (E - E_\gamma) - E' = \underbrace{\nu - E_\gamma}_{<\nu} \tag{6.31}$$

and for final state radiation

$$\nu_{Born} = E - (E' - E_\gamma) = \underbrace{\nu - E_\gamma}_{<\nu} \tag{6.32}$$

This means that the Born level z will be larger than the experimental z, as is observed in figure 6.6.

[3]In this particular case detector smearing obviously does not play a role



(a) Initial State Radiation (b) Final State Radiation

Figure 6.7: Schematic explanation of the asymmetric shape of the $n(i,j)$ matrices. Due to radiative effects the ν variable is increased by the amount of energy carried away by the Bremstrahlung photon.

6.2.2.2 Smearing of DIS events

In 6.22 we expressed the migration of DIS events as a 2×1 matrix and used this to obtain 6.28. However, the code that was developed for this was built in a more general way so that it can be used to tackle other unfolding problems within HERMES[4]. It is possible to do this if one constructs a migration matrix $n^{DIS}(i,j)$ for the DIS events in a special way. This is shown in equation 6.33

$$n^{DIS}(i,j) = \left(\begin{array}{ccccc|c} n^{DIS} & 0 & \cdots & 0 & 0 & n_0^{DIS} \\ 0 & n^{DIS} & \cdots & 0 & 0 & n_0^{DIS} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & n^{DIS} & 0 & n_0^{DIS} \\ 0 & 0 & \cdots & 0 & n^{DIS} & n_0^{DIS} \end{array}\right) \tag{6.33}$$

We simply have a diagonal matrix with the DIS events that remain inside the selection criteria, expanded with an extra column containing the events that are smeared into the experimental sample from outside the cuts.

6.2.3 Unfolding in multiple dimensions

This work is not restricted to the investigation of only differential z multiplicity distributions. In [29] a rather dramatic remaining x dependence of the 2-fold

[4]As was done e.g. by M. Contalbrigo and D. Reggiani [127] in the analysis of the g_1 structure function where it was also cross checked with code written by J. Wendland for the extraction of flavor separated polarized parton densities. This check provided the required assurance that the code that was developed in the framework of this thesis is fully correct.

multiplicity

$$\frac{d^2\sigma^h}{dzdx} \Big/ \frac{d\sigma_{DIS}}{dx} \tag{6.34}$$

was shown in various z bins which remained ununderstood. To be able to tackle such a problem one should be able to handle the smearing of events in between different x and z bins, or even more general in multiple dimensions. One could think of constructing a formalism in which the migration matrices indeed are multidimensional e.g. $n(i,j;k,l)$ where i,j would than be the z,x bin at experimental level and respectively k,l at Born level. However it is much easier to just alter the binning in the way that is shown in figure 6.8.

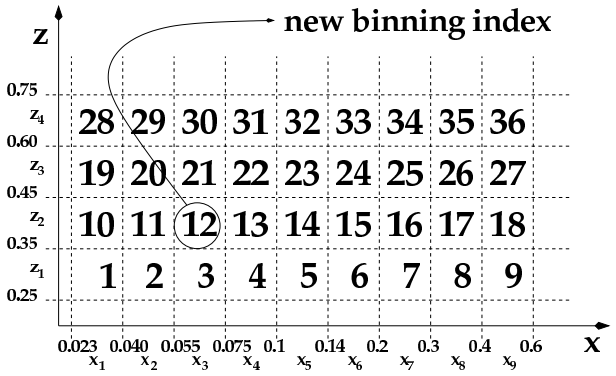


Figure 6.8: Schematic view of the new binning used to describe the 2D binning in x and z in 1D. The actual values represent also the real binning that is used in this analysis.

As each (z,x) coordinate is uniquely defined by a new bin number, also the smearing is uniquely defined. When applying multidimensional unfolding one has to ensure that every experimental bin is filled, excluding any null-rows from the smearing matrix. This will prove not to be a problem for this analysis. However, if one has to deal with it, inversion of the matrix becomes somewhat more complicated. As mentioned above, one can use the same formalism for the DIS events to construct a $n(i,j)$ matrix. However one has to take into account that one now *does* have a binning in x for the DIS events, but then in x alone (see equation 6.34). A normal smearing matrix was therefore extracted for the x binning and copied to the different z bins as shown in figure 6.9. The migration matrices for a binning in z and x obtained in this way for all pions and for DIS events are shown in figure 6.10.

6.2.4 Treatment of the uncertainties involved

One of the main motivations for using the unfolding method as outlined above, is a consistent treatment of the uncertainties involved. On the one hand this means the propagation of uncertainties on experimental input data, both statistical and systematic (RICH, diffractive background correction...). On the other hand a clear handling the systematic uncertainty contribution from the Monte Carlo statistics is aimed for as well. This is the topic of the next two sections.

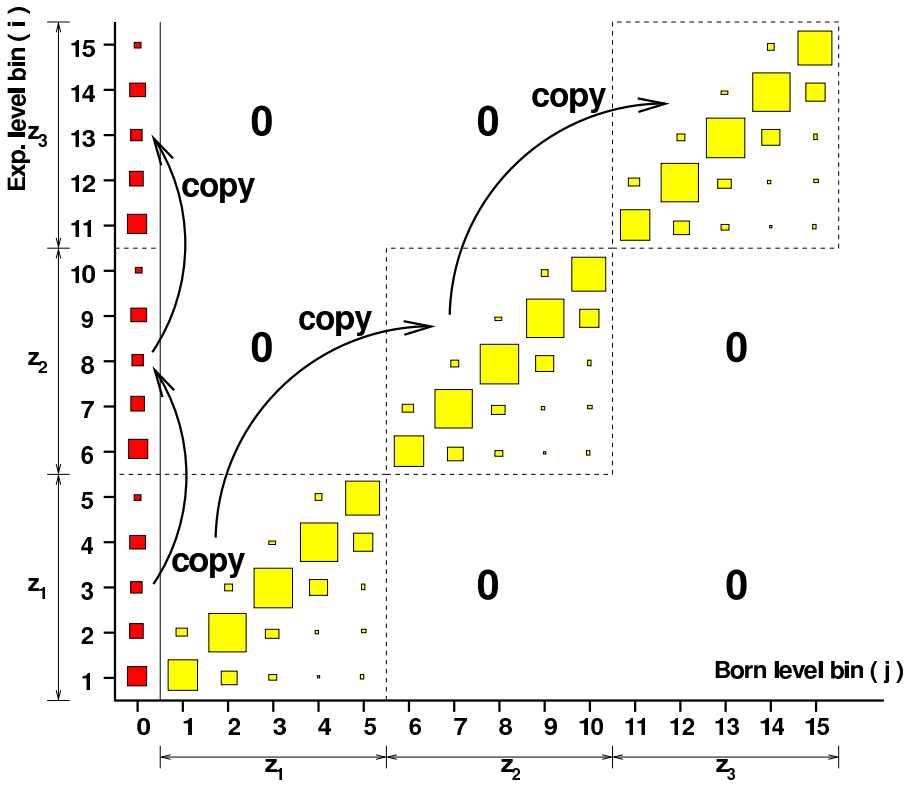
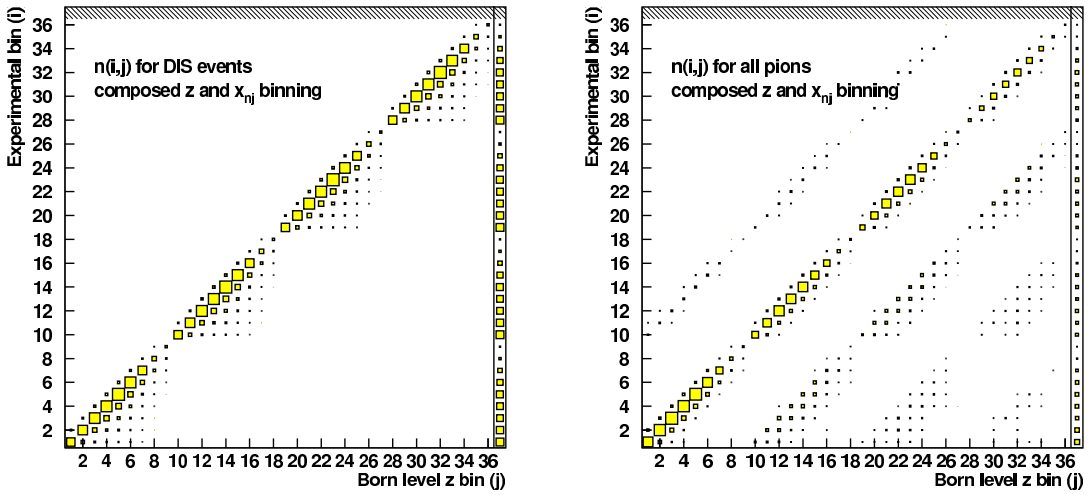


Figure 6.9: Construction of the $n(i,j)$ matrix for the DIS events using the binning explained in figure (6.8). In this figure we give the example of 3 z bins and 5 x bins.



(a) DIS events (b) All pions

Figure 6.10: The $n(i,j)$ matrices in the 2D case discussed. On the left the $n^{DIS}(i,j)$ matrix is drawn, on the right the one for the pions (both π^+ and π^-). The binning indexes are the ones defined in figure 6.8.

6.2.4.1 Propagation of uncertainties on the experimental input data

One can prove [126] that any uncertainties $\delta R_X(i)$ on the experimental ratio $R_X(i)$ can be propagated through the unfolding procedure using the following

equation

$$\delta^2(R_{Born}(j)) = \sum_{i=1}^{n_X} D^2(j,i)\delta^2(R_X(i)) \tag{6.35}$$

in which the error inflation due to the smearing of the events is governed by the so-called radiative dilution matrix $D(j,i)$. This matrix is based on the input parameterizations and the inverted truncated smearing matrix. It is given by :

$$D(j,i) = \frac{[S'_h]^{-1}(j,i)n^X_{DIS}(i)}{n^B_{DIS}(j)} \tag{6.36}$$

6.2.4.2 Intrinsic uncertainty due to limited Monte Carlo statistics

So far only the propagation of experimental uncertainties was considered. However, we do have an additional systematic contribution on the unfolded data points from the limited statistics used in the Monte Carlo production. It is not straightforward how this uncertainty is amplified through the unfolding procedure. Inversion of a matrix is a highly non-linear process and it is not obvious how this influences the correlations between the elements. Therefore this contribution was calculated using brute force. Consider the input from Monte Carlo: the $n^h(i,j)$, $n^{DIS}(i,j)$, $n^h_B(j)$ and $n^{DIS}_B(j)$ matrices and distributions and their uncertainties. All the elements of these objects are sampled simultaneously but independently by randomly generating random variates around the nominal value according to a Gaussian distribution. Obviously one cannot simply vary each element separately because of the correlations inside the migration matrix. In this way 100'000 smearing matrices are constructed for each spectrum and fed to the entire inversion procedure. Doing so, we obtain for the unfolded spectra a Gaussian-like distribution around the nominal value. Note that these obtained distributions are not necessarily Gaussian because of the nonlinearity of the matrix inversion. The standard deviation of this distribution is used as a the uncertainty on the unfolded elements coming from the Monte Carlo statistics.

From figure 6.11 one can see that this method is indeed stable around the unfolded values. We see that we more or less get back distributions which look like they are Gaussian. The centroid of these distributions agrees well with the nominal unfolded value within that bin, represented by the red line. Figure 6.12 shows the relative difference (in %) between the real unfolded value and the centroid of the distribution obtained by the generation of the random matrices. For very low statistics there is some noticeable difference mainly because the Monte Carlo accuracy for these bins is low. However, as can be seen from figure 6.12 these deviations are never more than 0.2 % except for the low statistics bins in the K^- spectra. This effect is well below the systematic uncertainties because of the RICH detector. We will nevertheless observe some instabilities because of this unfolding method, possibly due to the structure of the smearing matrix itself. Even though the individual datapoints are stable around their value, an entire unfolded spectrum may show some instability.

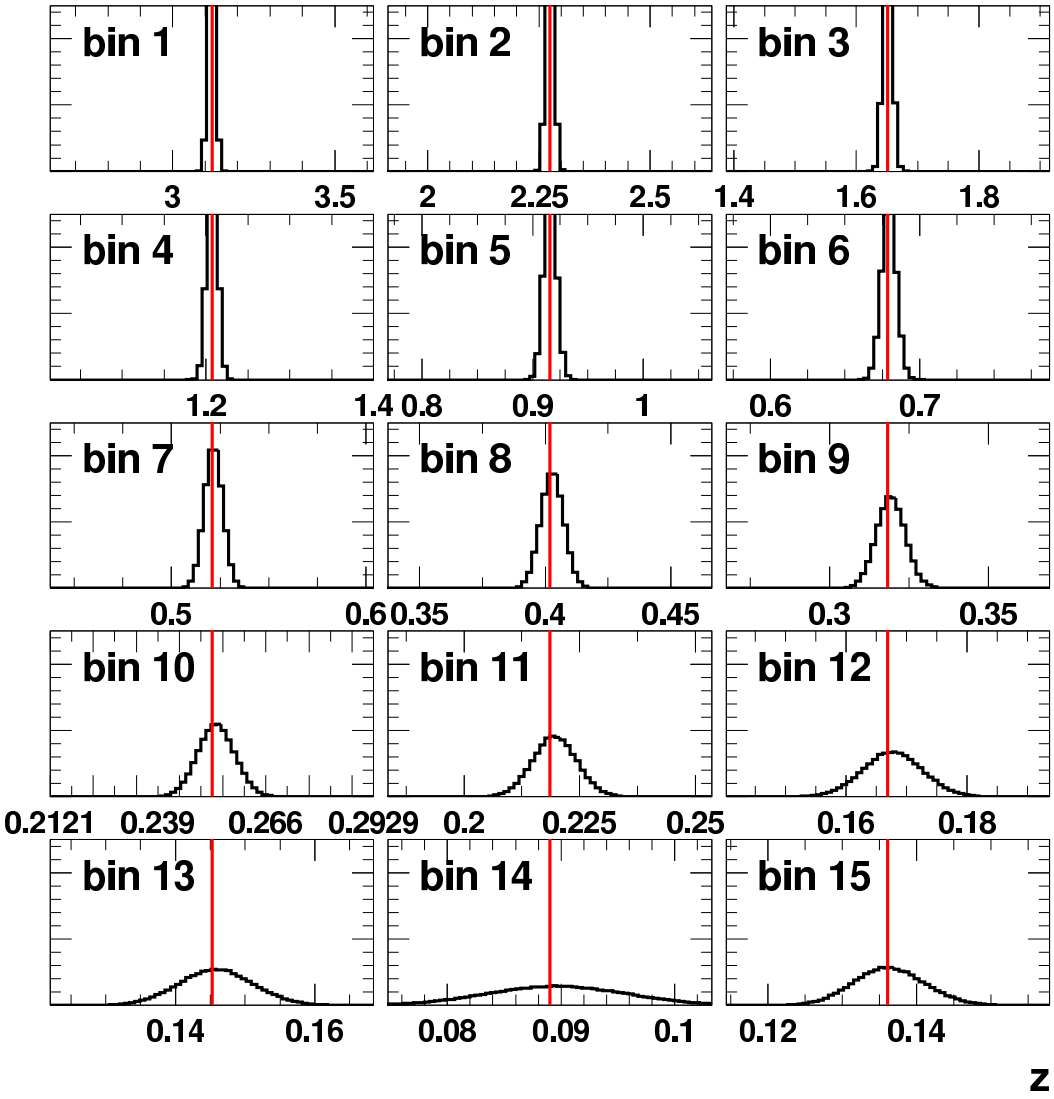


Figure 6.11: The obtained spectra for each element after unfolding the generated matrices. The units on the vertical axis are arbitrary.

6.3 Q^2 scaling of the multiplicities

As can be seen from figure 6.13, fragmentation functions show a mild but definite Q^2 dependence. It is therefore important to present the results at a fixed Q^2, since the average Q^2 varies between the different bins. Figure 6.14 shows the Q^2 variation in the binning used. This variation is very mild versus z: $\langle Q^2 \rangle$ only goes from about 2.4 to 2.6 GeV^2. However, due to the obvious correlation between x and Q^2 it is more pronounced in these spectra. Furthermore, we want to compare the obtained results to other experiments, therefore a tool is

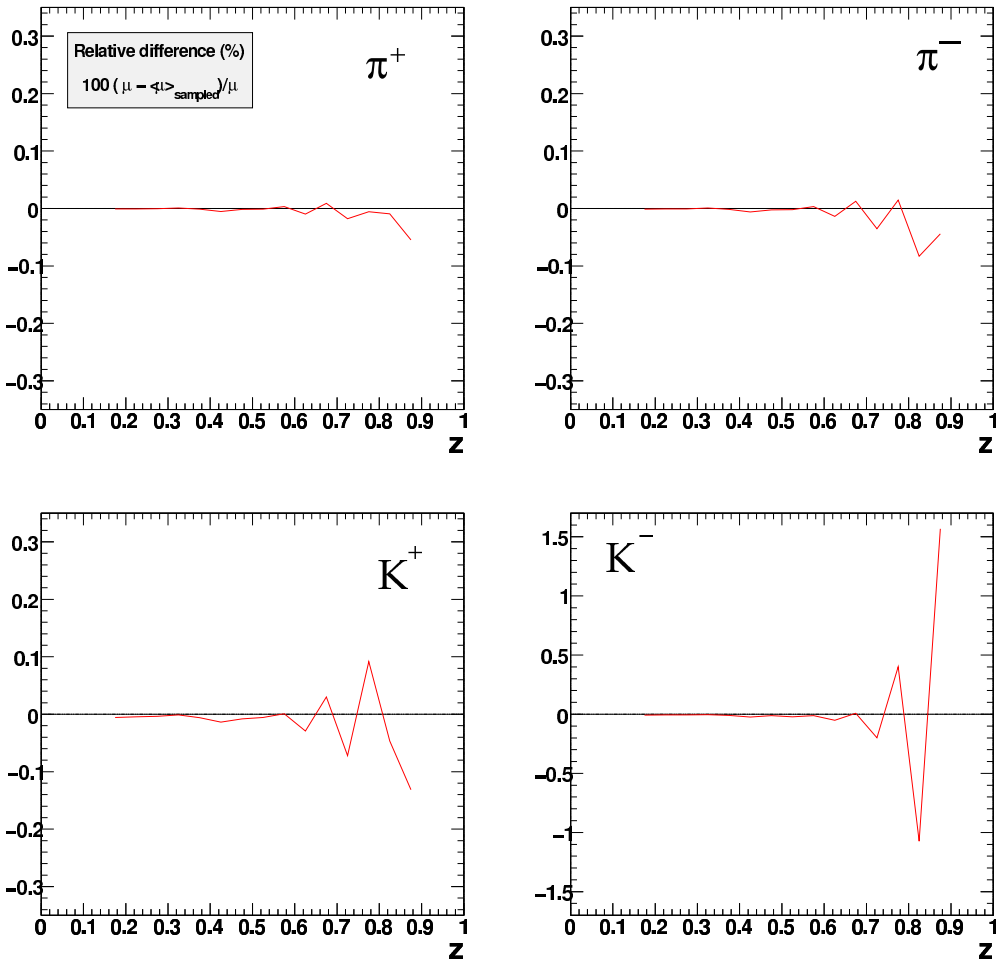


Figure 6.12: Relative difference in % between the unfolded value and the centroid of the distribution.

needed to handle the scaling. For example, the fragmentation functions published by the EMC collaboration [28] are at $Q_0^2 = 25$ GeV^2 and the correction factors for this will prove to be rather sizeable. This correction has to be determined at Born level since it is applied at this stage. The average Q^2 value in each bin has to be determined from a Monte Carlo simulation. The correction method is derived [29] from the expression for the multiplicity distribution :

$$\frac{1}{\sigma_{DIS}}\frac{d\sigma^h(z,Q^2)}{dz} = \frac{\sum_f e_f^2 \int_0^1 dx\, q_f(x,Q^2) D_f^h(z,Q^2)}{\sum_f e_f^2 \int_0^1 dx\, q_f(x,Q^2)} \tag{6.37}$$

However, where in previous HERMES results [29] only the fragmentation functions were used for the correction, we will use here the ratio of the full expression 6.37, evaluated at the average Q^2 in the bin and at a certain *fixed* value Q_0^2.

When presenting our results we will fix the Q^2 scale to more or less the average of the data, at $Q_0^2 = 2.5$ GeV^2. Since we also want to incorporate kaons into this analysis and separate the two charges, the simple approach where

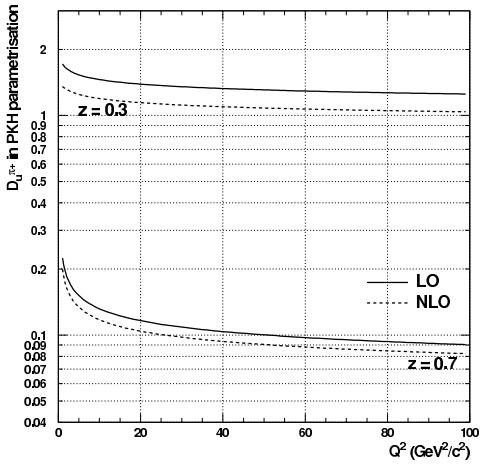
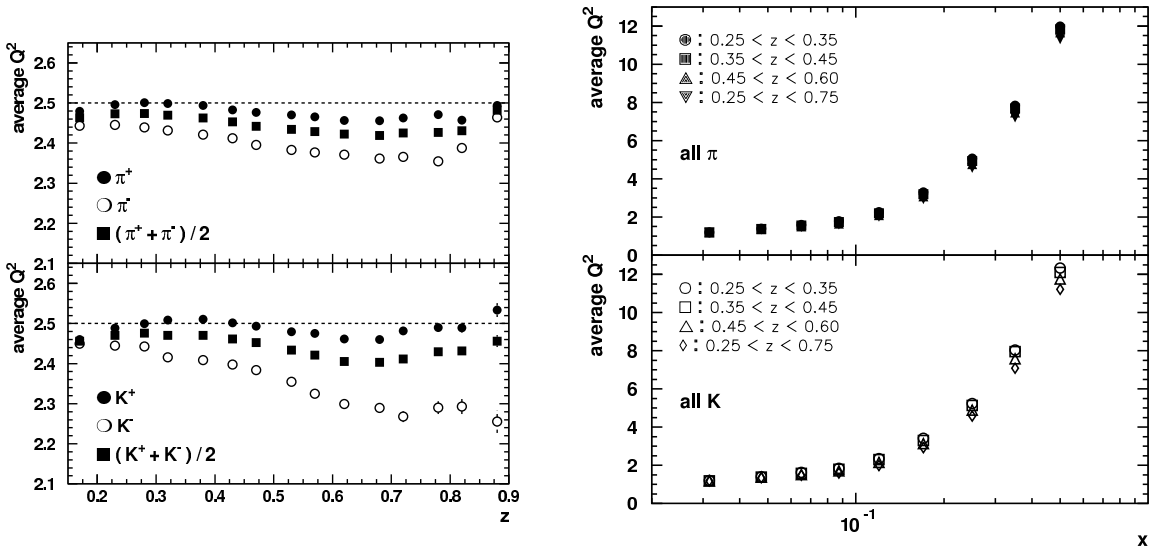


Figure 6.13: Scaling of the fragmentation function $D_u^{\pi^+}$ at two different fixed z values, and this for both LO and NLO parameterizations. The lines shown were calculated using the PKH model [52].



(a) Versus z (b) Versus x in 4 z bins

Figure 6.14: Average Q^2 values versus z and x at Born level.

one neglects strange quark contributions to the sum in equation (6.37) and assumes the symmetry relation $D_u^\pi = D_d^\pi$ [29], does not hold anymore. So, the parton density functions must be taken into account if we wish to obtain more general results.

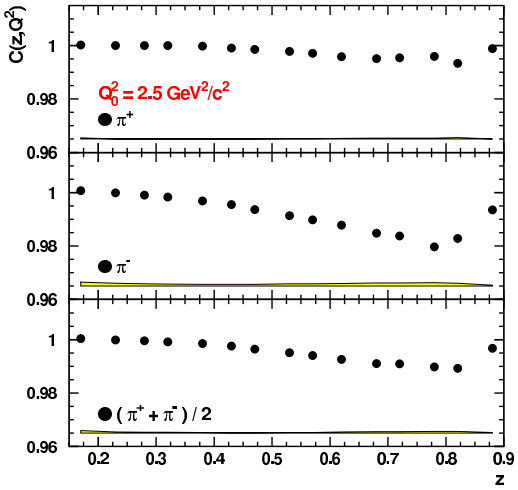
For the fragmentation functions we use the Kretzer parameterization [52]. This one has the advantage to be able to do a full flavor separation. Some more details were given in section 2.4.5. The parton densities were taken from the CTEQ6 parameterization [118] and the ALLM97 [9] parameterization for the structure function $F_2(x, Q^2)$ as it is known that the CTEQ6 parameterization fails to reproduce the total structure function at low Q^2, as demonstrated in

section 5.2.2.6.

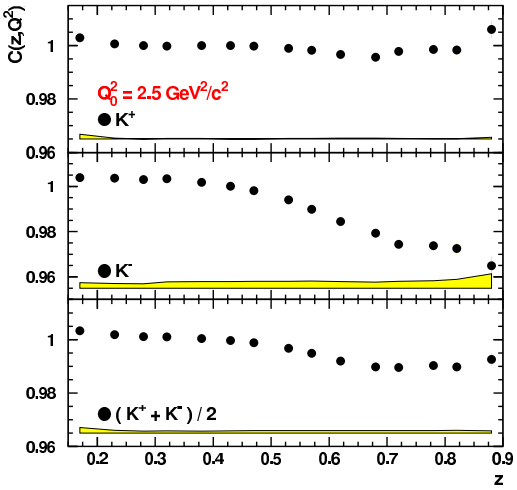
6.3.1 Uncertainty on the Q^2 correction

The only statistical input into the Q^2 scaling correction factors comes from the determination of the average Q^2 in each bin in the Monte Carlo data. The uncertainty contribution because of this was estimated and assigned as statistical uncertainty to this correction. However, this contribution is negligible. It is more difficult to assess systematic uncertainties to these correction factors. The Kretzer model is in fact the only one which we can use here as there are no other parameterizations which between charges and flavors. Therefore we cannot compare to other fragmentation models to assign an uncertainty coming from the parameterization. However, we have taken the difference between LO and NLO parameterizations as a conservative measure of the uncertainty due to the fragmentation model. Other contributions assigned come from the integration boundaries in x in equation 6.37 which were varied, the CTEQ6 parameterization and the discrepancy between CTEQ6 and ALLM97 in the structure function at low Q^2. The re-weighing of the parton density functions according to equation 5.18 in section 5.2.2.6 deals with this discrepancy and is thus used as a measure of the uncertainty. All these effects are added together and assigned as systematic error on the Q^2 scaling corrections. The resulting correction factors are included in appendix B.4 and are displayed below in figures 6.15 versus z and 6.16 versus x.

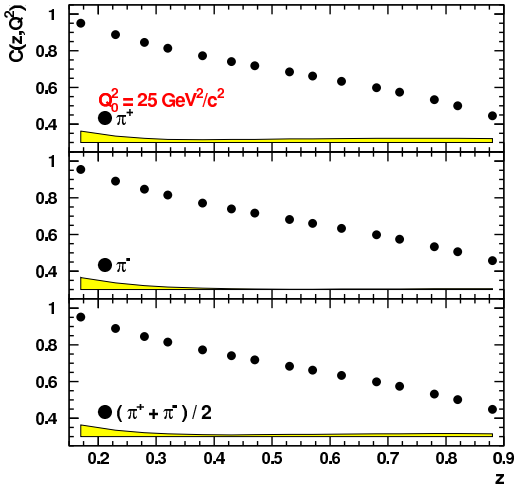
Since Q^2 does not change very much versus z the correction factor versus z at $Q_0^2 = 2.5$ GeV^2 is virtually 1 with a very small uncertainty. Evolving to 25 GeV^2 however, increases this uncertainty mainly due to the fragmentation parameterization. Since the difference is taken between LO and NLO. As Q^2 and x are strongly correlated quantities (see figure 6.14) the correction factor versus x is rather sizeable as can be seen from figure 6.16.



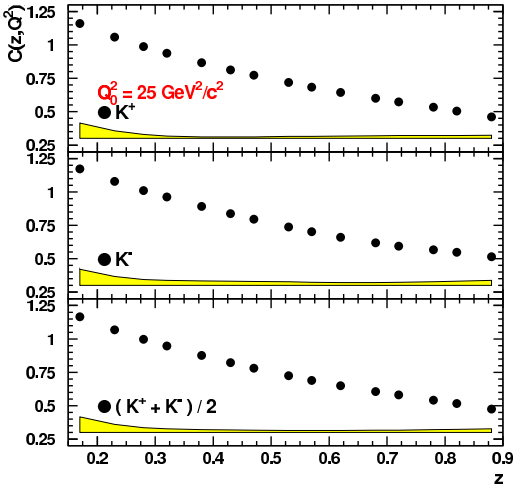
(a) Q^2 scaling correction: π, Q_0^2 = 2.5 GeV^2



(b) Q^2 scaling correction: K, Q_0^2 = 2.5 GeV^2

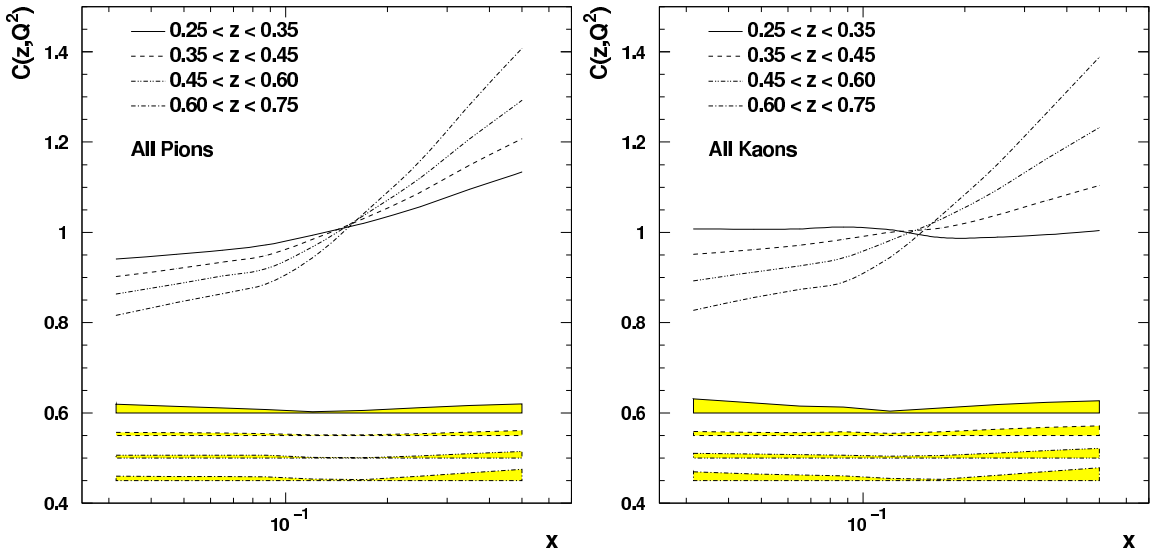


(c) Q^2 scaling correction: π, Q_0^2 = 25 GeV^2

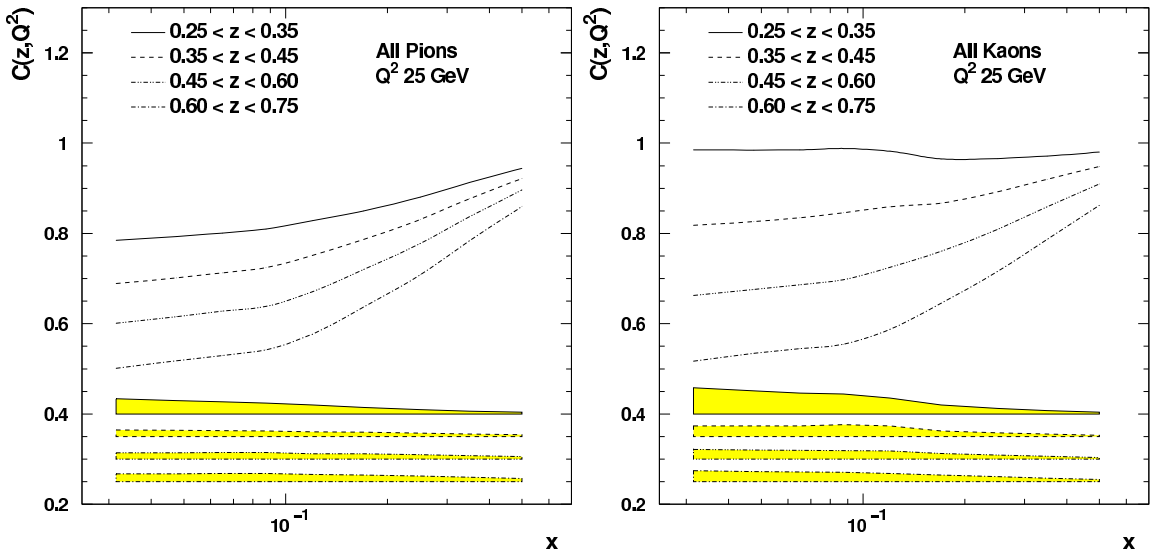


(d) Q^2 scaling correction: K, Q_0^2 = 25 GeV^2

Figure 6.15: Multiplicative Q^2 scaling correction factors versus z for $\pi^{\pm}$ and $K^{\pm}$. The first two plots provide the values at Q_0^2 = 2.5 GeV^2 (the average for HERMES) whereas the next two Q_0^2 = 25 GeV^2 in order to compare to EMC data [28].



(a) Q^2 scaling correction: π, Q_0^2 = 2.5 GeV^2 (b) Q^2 scaling correction: K, Q_0^2 = 2.5 GeV^2



(c) Q^2 scaling correction: π, Q_0^2 = 25 GeV^2 (d) Q^2 scaling correction: K, Q_0^2 = 25 GeV^2

Figure 6.16: Multiplicative Q^2 scaling correction factors versus x for pions and kaons. We only show these correction factors for the charge averaged case, so e.g. $(\pi^+ + \pi^-)/2$ not to overload this section with plots. All values are included in appendix. The first two again provide the values at Q_0^2 = 2.5 GeV^2, and the next two Q_0^2 = 25 GeV^2.

6.4 Multiplicity results and discussion

It is good recapitulate schematically the structure, as done in figure 6.17. We started of by obtaining raw multiplicities from the HERMES 00c1 data production using unpolarized hydrogen. A RICH hadron particle identification correction was performed using an unfolding method, and contributions from exclusive diffractive processes were subtracted from the data sample using a PYTHIA6 Monte Carlo production. The multiplicities were subsequently corrected for radiative and detector smearing as well as for geometrical acceptance, thus obtaining Born level distributions. Figure 6.17 clearly shows the contributing systematic effects in each step.

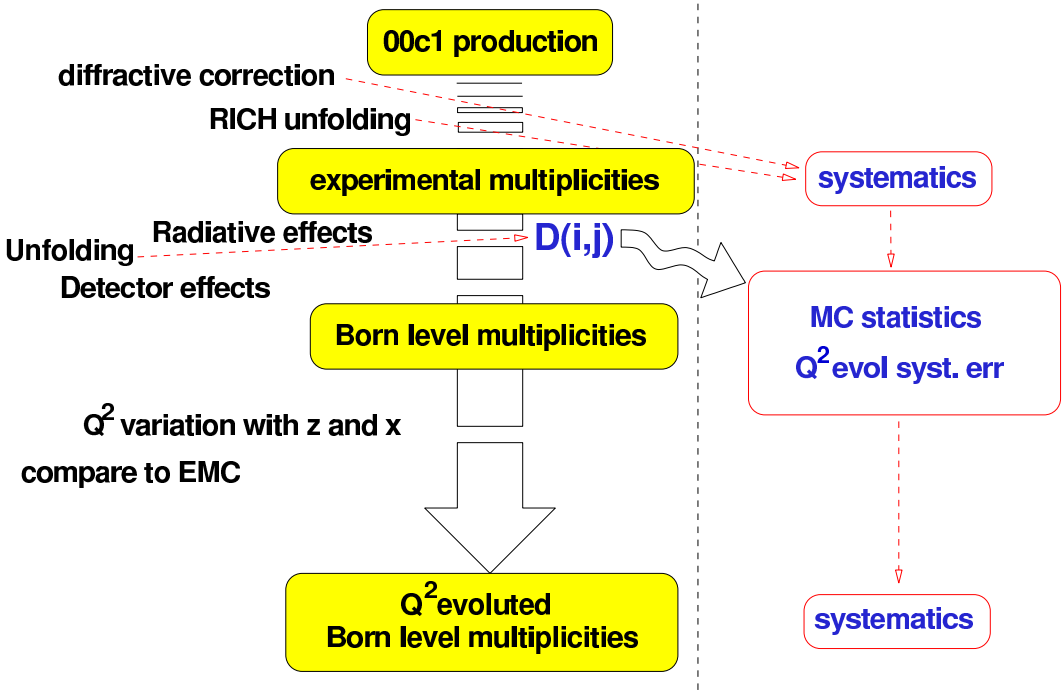


Figure 6.17: Schematic overview of the multiplicity analysis

Before we discuss the obtained results, we will group and highlight some of the features of the analysis presented here.

- The systematic error on the RICH unfolded experimental multiplicity distributions has been calculated using different sets of unfolding matrices (see section 4.2.2.3).
- A systematic uncertainty was assigned to the diffractive background correction as discussed previously in section 6.1.
- For the fragmentation tune in the `gmc_disNG` Monte Carlo production, this analysis has used the `2004c` tune (see table 5.1). In fact, compared to the `2003a` tune (used in [115]), the `2004a` version described the K^- distribution a lot better so it was generally decided upon to be more correct as there can be unsimulated K^+ contributions in the semi-inclusive

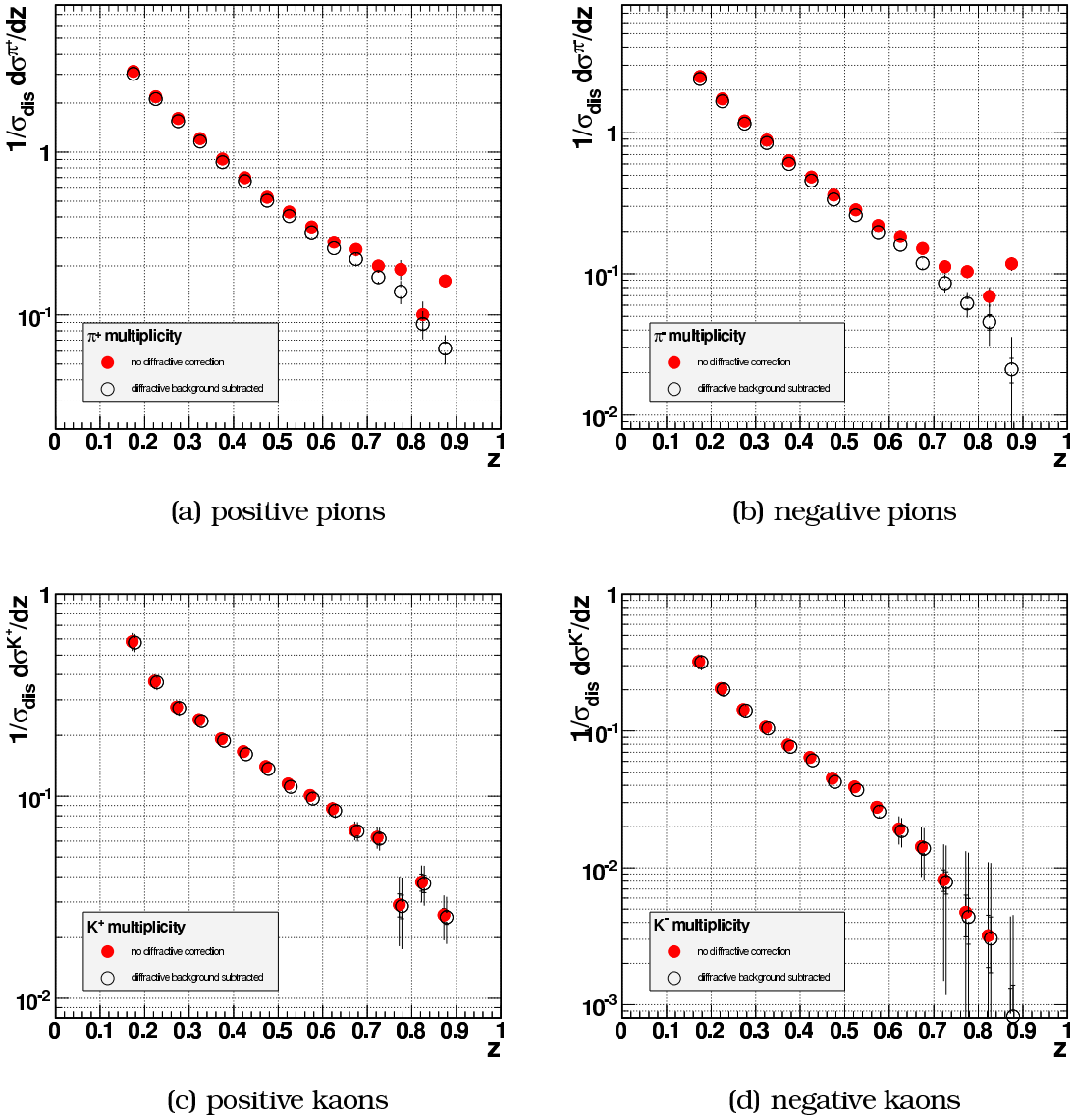
hadron sample from exclusive associated Λ production. However, the more recent `2004c` tune does a far better job in describing both the K^+ and K^- distributions. Therefore this tune has been chosen in this analysis. The difference between the `2004a` and `2004c` parameterizations has been treated as a systematic uncertainty contribution. Obtaining a complete scan of the parameter space in the LUND model is a very difficult and huge task. In the ideal case one would obtain a global minimum of χ^2 in the parameter space and define a 1 σ contour around this minimum in order to fix the uncertainty on the LUND parameters. This would then result in the correct systematic uncertainty. We believe, however, that the two tunes, `2004c` and `2004a` are satisfactory enough to be considered close to this global minimum. Hence assigning the discrepancy between the two assuming a uniform distribution between them can be considered as a reasonable estimation for this systematic effect.

- An elaborate study was performed of one other systematic effect, namely the influence of the azimuthal ϕ dependence of the semi-inclusive cross section, which was up till now completely ignored. This will be the topic of the next chapter.

6.4.1 Multiplicities versus z

After applying all the corrections that were discussed above we obtain the Born level multiplicity distributions as presented in figure 6.18. The data points are presented both with and without the diffractive background subtraction. The effect of the diffractive correction is clearly visible for the pion distributions where we see the impact at high z. As discussed above the impact on the kaon spectra is negligible. What is also striking is the instabilities that surface at the high z end of the spectrum. Especially the second but last z bin for the pions and the third last bin for the K^+ seem to be a bit unstable in this approach. This is due to the unfolding method with which we correct for smearing and geometrical acceptance. It is the price one has to pay for using this method. This in fact immediately raises the question whether an unfolding method reproduces the same results as applying an acceptance correction function as formulated in equations 6.12 and 6.13. Proof of this is given in figure 6.19. The two methods agree very well, except at the high z where the acceptance function method shows somewhat more smooth spectra. A smooth multiplicity spectrum, as obtained using acceptance functions, is indeed what one would naively expect. However this is not the correct method to account for the migration of events. Therefore one has to live with the instabilities which are inherent to unfolding.

Another issue here is the model dependence of the obtained results. The HERMES acceptance is rather low for hadrons, about 30 - 35 % at intermediate to high z and decreasing quickly at low z. Therefore it is but logical to raise questions about the model dependence of the extrapolation of experimental multiplicity distributions to 4π and Born level. Especially the Monte

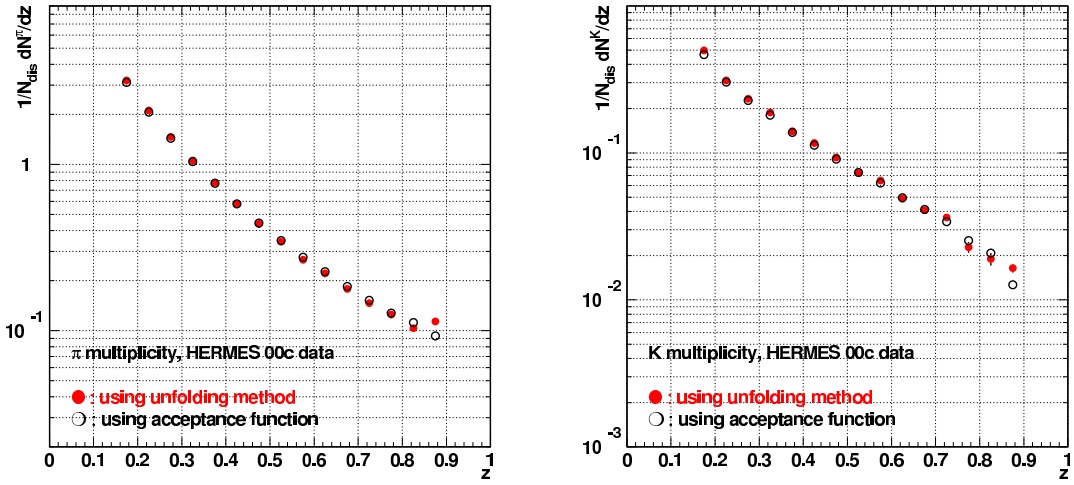


(a) positive pions (b) negative pions

(c) positive kaons (d) negative kaons

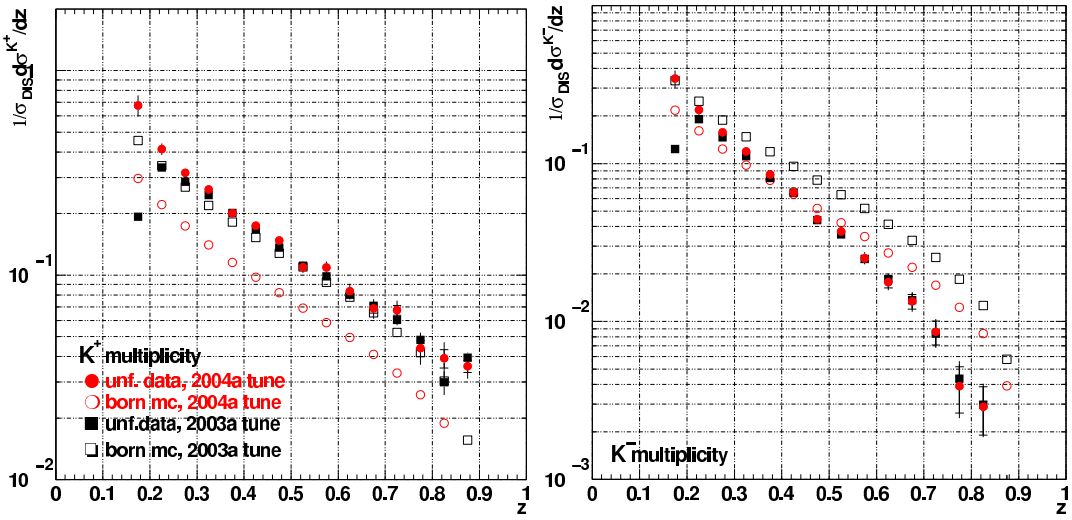
Figure 6.18: Born level multiplicity distributions versus z for π^+, π^-, K^+ and K^-. Data points are at HERMES average Q^2 of 2.5 GeV^2. The filled circles represent the datapoints without diffractive correction, the open circles with the diffractive background subtracted. The inner errorbars represent the statistical uncertainty, the outer errorbars the combined statistical and systematic uncertainty. These results are included in appendix in tables B.7 to B.14.

Carlo fragmentation model plays a very crucial role. In figure 6.20 one can see a comparison between unfolded kaon spectra (K^+ and K^-) using the `2003a` and `2004a` tunes for the JETSET model. The unfolded spectra are in the same plot compared to their pure Monte Carlo counterparts. It has already been discussed that the `2003a` and `2004a` tunes describe either K^+ or K^- spectra well, but never both at the same time. Although this problem has been largely overcome using the `2004c` tune, it nevertheless remains instructive to see what happens if we take these two extremes as input for the unfolding procedure. From figure 6.20 we see two very different Born level distributions obtained directly from the generated Monte Carlo sample (open symbols). The



(a) pions (b) kaons

Figure 6.19: Comparison of pion and kaon multiplicities using the unfolding method and using an acceptance correction method. Note that these plots show charge averaged distributions for pions and kaons.



(a) positive kaons (b) negative kaons

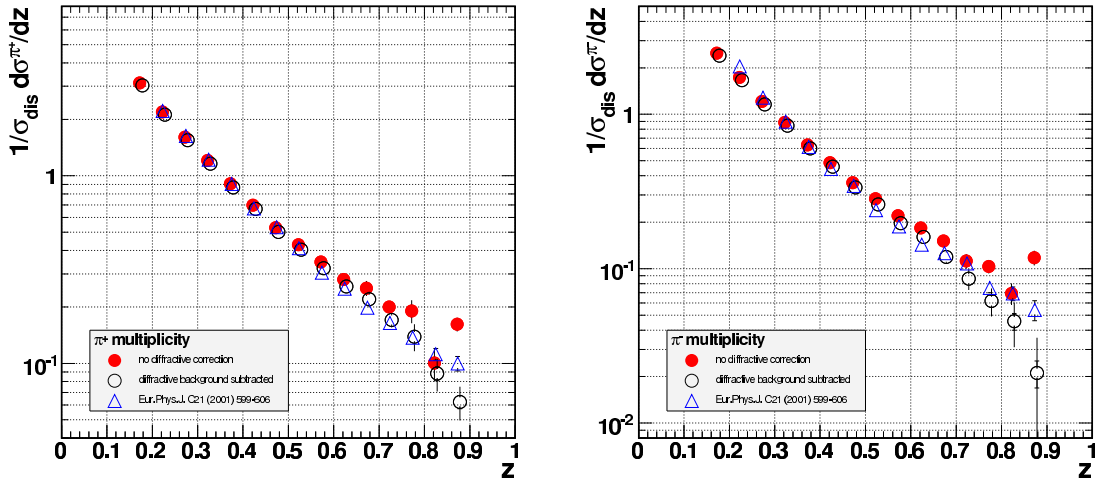
Figure 6.20: Comparison between 2003a and 2004a fragmentation tunes for kaons. Data represented by the black square symbols is obtained using the 2003a tune either directly in the Monte Carlo (open symbols) or using such a Monte Carlo production to unfold experimental data (closed symbols). The circular symbols represent the same, but using the 2004a tune.

2003a tune compares best to the K^+ spectrum, whereas 2004a reproduces the K^- spectrum better. Still, using these two tunes as input to the unfolding procedure, the obtained data spectra are very much the same, proving that the applied correction makes sense. As discussed above the uncertainty due to the fragmentation tune is included into a systematic error and is of the same order as the systematic error due to the RICH. As the Monte Carlo determines

the shape of the smearing matrices and the input distributions, this systematic uncertainty in a way implicitly also accounts for the instabilities in the unfolding procedure. The smearing matrices based on different fragmentation tunes will lead to larger variations as one would expect just based upon the difference in fragmentation model. This is because in the unfolding process, small differences in the smearing matrices can have large effects due to instabilities for bins where these input matrices and distributions are not smooth enough because of limited Monte Carlo statistics.

6.4.2 Comparison of z dependence to published data

Based upon the HERMES 96/97 data, π^+ and π^- multiplicity distributions were extracted and published in [29]. In figure 6.21 one can see a comparison between the presented and the previously published HERMES data. We see a



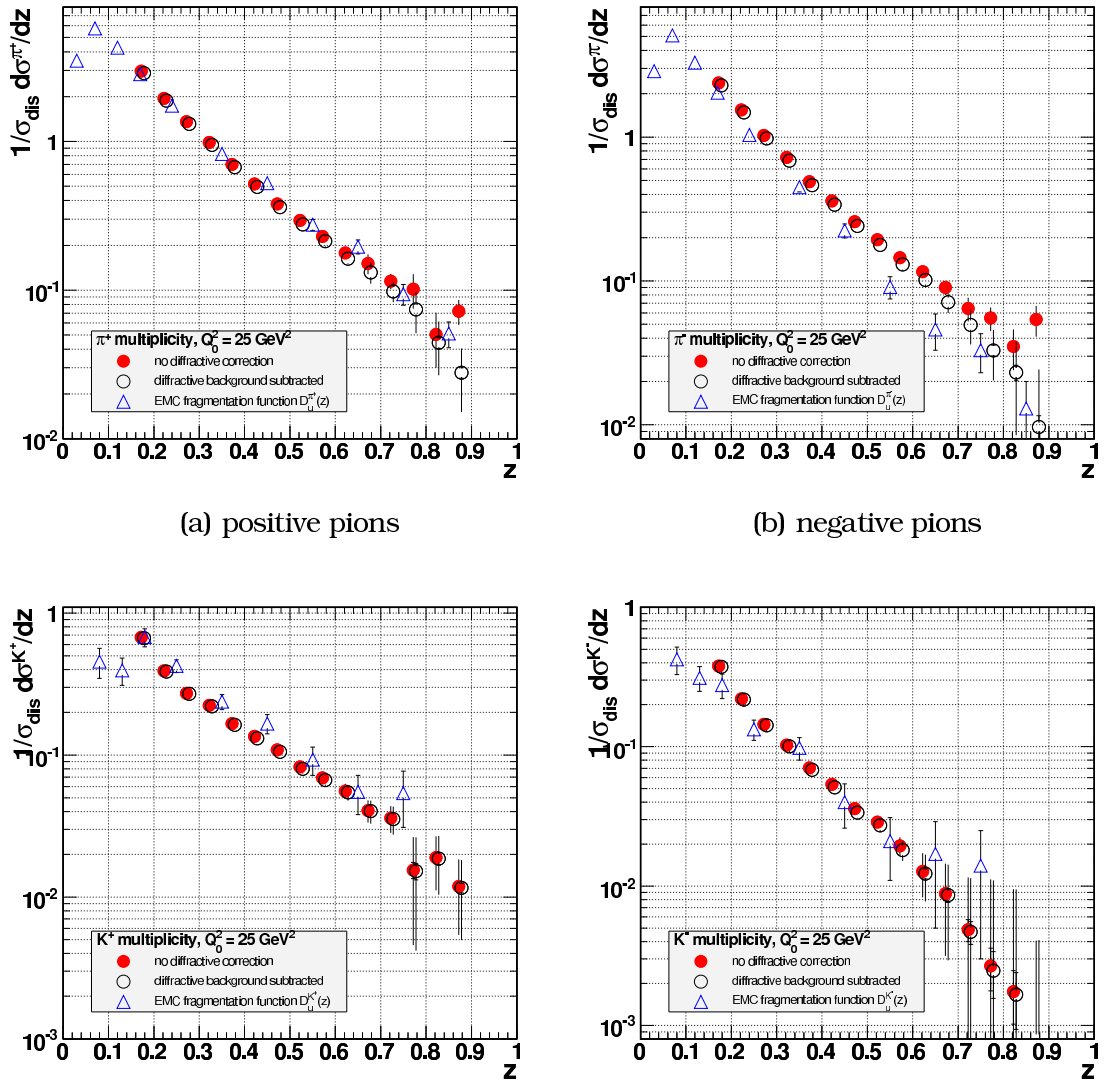
(a) positive pions (b) negative pions

Figure 6.21: Comparison of the z dependence of the π^+ and π^- multiplicity distributions with (open circles) and without (filled circles) diffractive correction between this analysis and the HERMES 96/97 results published in [29] (triangles).

very good agreement up till $z \sim 0.6$, while for higher z the two analyses start to disagree. What is in fact rather amazing is that the published data seems to agree better with the data from this analysis where the diffractive background was subtracted. However, no such correction was made in [29]. A number of other different approaches form the reason of this discrepancy. We will come back to this point when we discuss the residual x dependence of the multiplicity distributions. As we will see it is most pronounced there.

Another, very interesting, comparison is with fragmentation data published by the EMC experiment. This data was obtained from a deep inelastic muon-proton scattering experiment at 280 GeV [28]. Obviously, the average Q^2 of

this experiment is much higher, so in order to compare we have scaled the HERMES data up to $Q_0^2 = 25$ GeV^2 using the evolution factors from section 6.3. The comparison is seen in figure 6.22. We must, however, note that the quan-



(a) positive pions

(b) negative pions

(c) positive kaons

(d) negative kaons

Figure 6.22: Comparison between the data points obtained in this analysis, scaled up to $\langle Q^2 \rangle$ = 25 GeV^2 and the u-quark fragmentation functions $D_u^{\pi^+}$, $D_u^{\pi^-}$, $D_u^{K^+}$ and $D_u^{K^-}$ published by the EMC collaboration in [28].

tities provided by the EMC collaboration are in fact u quark fragmentation functions, and therefore the comparison is not completely accurate. However, since we know that the u quark is very dominant in the proton, we still can extract some physical meaning from this comparison. Furthermore, neglecting the strange quark content of the proton and assuming isospin symmetry in the fragmentation function reduces equation 6.37 to a simple fragmentation function. For π^+ which is formed in favored fragmentation, we observe a strikingly good agreement given the fact that the data has been taken at

very different energies. At EMC energy, applicability of perturbative QCD and factorization are generally well accepted, whereas at HERMES energies there has been some debate about how well one can interpret the semi-inclusive data in terms of factorization and fragmentation. This comparison shows that at HERMES, one still obtains the hadron spectra measured at higher energy DIS. Exactly for this reason the comparison for π^- is a bit troublesome as well. EMC data is clearly below the HERMES multiplicity. The u quark is not a valence quark in the π^- meson, therefore the EMC data only provides the contribution from unfavored fragmentation $u \to \pi^-$ which is suppressed, whereas a multiplicity distribution sums over all quarks. Of course in the kaon case, statistics is rather poor for the EMC data points as compared to HERMES, so it remains difficult to make any clear statement, especially for K^-. Nevertheless the distributions seem to show a very similar behavior.

Another very interesting item is the multiplicity versus Q^2 compared to the Q^2 evolution of the fragmentation function parameterization provided by S. Kretzer [128] that was integrated over the 4 z ranges. Figure 6.23 shows this comparison. One can see that there is a reasonable match between the

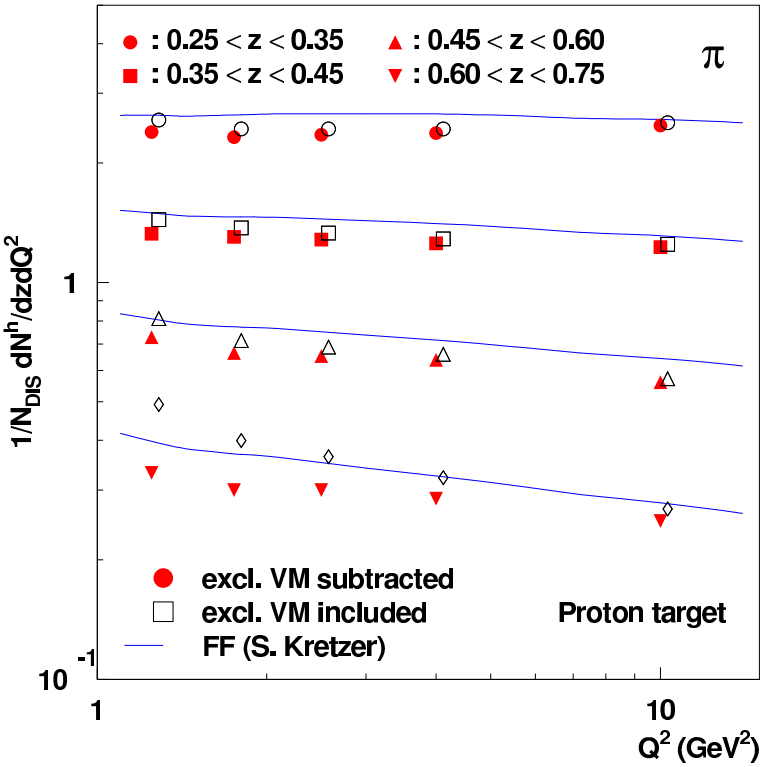


Figure 6.23: Comparison of $\pi^+ + \pi^-$ multiplicities versus Q^2 with a parameterization provided by S. Kretzer.

HERMES data and the curves. Especially the Q^2 scaling is reproduced rather well. Once again we see that a parameterization which uses experimental input at much higher energy still holds quite well at a moderate Q^2 of 2.5 GeV^2.

6.4.3 The residual x dependence of the pion multiplicities

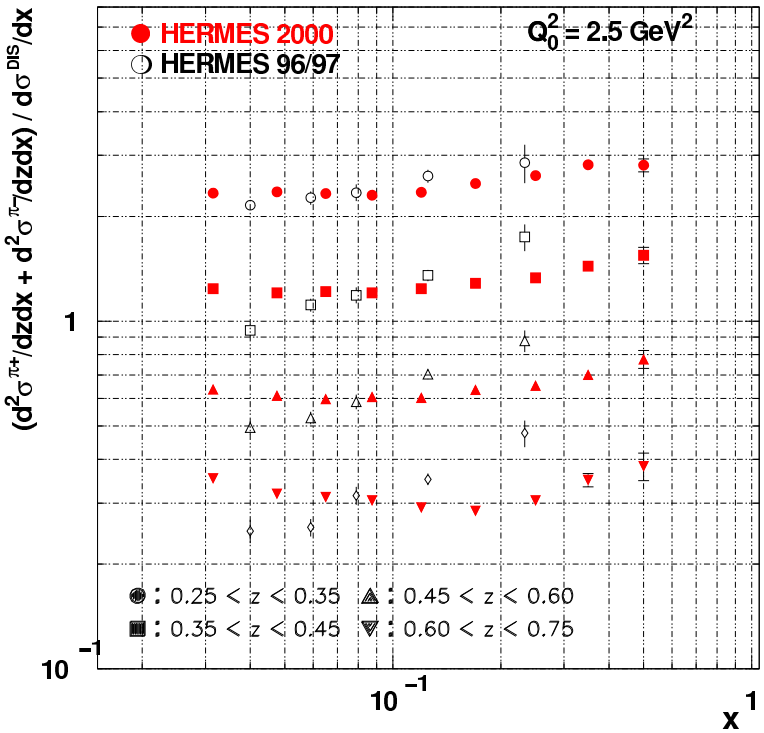


Figure 6.24: Comparison between the x dependence from this analysis and the data from [29]. Note that this plot was obtained using the `2004a` tune to correct for acceptance and radiative effects. Figure taken from [100]. The HERMES 96/97 data presented comes from [29].

At the time of the publication of the HERMES 96/97 data [29], some excitement was stirred about possible factorization breaking effects which remained unexplained. These can be seen from figure 6.25. The pion multiplicities show a pronounced residual x dependence after fixing the Q^2 value and binning in z. This dependence seems to get stronger with increasing z as well. Even more surprising, is that the EMC data displays such a behavior as well. Let us have a look what we can learn from an analysis performed using the methods explained in this work. In figure 6.24 we can see the x dependence of the multiplicity in 4 z bins. The 96/97 data points indeed show the steep slope whereas the newly obtained data points display a drastically different x behavior. There is still a rising tendency of the multiplicity distributions, but much less pronounced. We must note that a certain x dependence is expected. All too often the argument is given that fragmentation functions should be independent of x. This is of course true, but first of all in this case the quantities plotted are not fragmentation functions, but multiplicities. Neglecting strange quark contributions and assuming symmetry under isospin rotation, the expression for the multiplicity does indeed reduce to a pure u quark fragmentation function if one sums over the pions. However, this is only approximate. One has to include contributions from all the parton density functions, which obviously do depend on x and this behavior does not simply cancel out in the expression 6.37. Furthermore, one has to take into account that the fragmentation process in lepton nucleon scattering is somewhat more complicated as

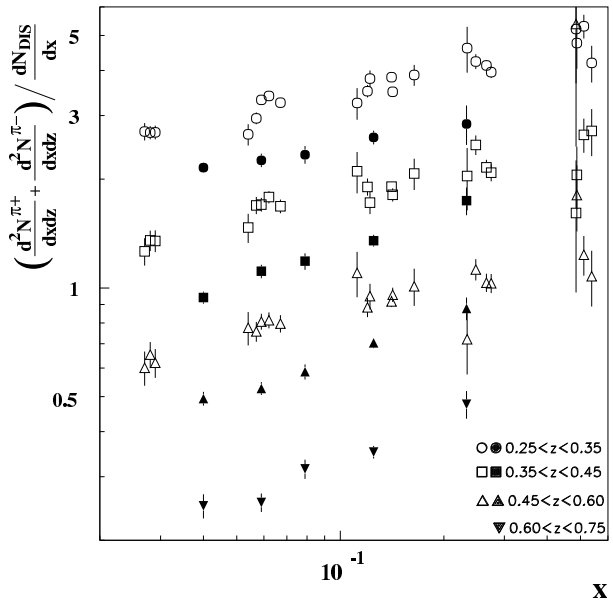


Figure 6.25: Residual x dependence of the pion multiplicity for HERMES 96/97 data. Figure taken from [29].

compared to e^+e^-. The struck quark comes from within a hadron and due to possible gauge links between the final[5] and initial quark propagator one can have factorization breaking effects that play a role (see figure 2.8(b)).

Nevertheless, the x dependence observed from this analysis is very mild, as opposed to what was observed in [29]. For a number of historic reasons, a direct comparison between the two analysis has proven not to be possible so here we will outline some in depth checks performed to explain this difference. We believe the method of correcting for detector geometry and efficiency plays a crucial role in this. The experimental data between the different data taking years has proven to be consistent over the data taking periods considered [100], so in any case we have to look to the Monte Carlo input, and there, a lot has happened [116].

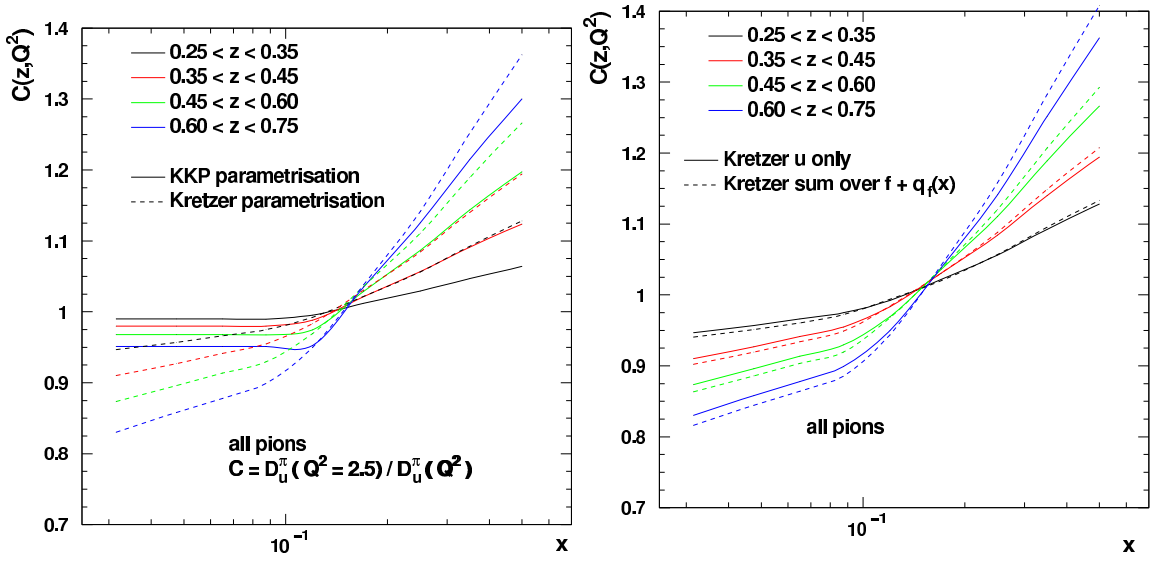
First of all the Q^2 **scaling correction** did change dramatically. As already stated, in [29] the Q^2 evolution was performed by simple ratios of u-quark fragmentation functions provided by the Kniehl, Kramer and Pötter (KKP) parameterization [51].

$$C_{Q^2}(z, \langle Q^2 \rangle, Q_0^2) = \frac{D_u^\pi(z, Q_0^2)}{D_u^\pi(z, \langle Q^2 \rangle)} \tag{6.38}$$

This parameterization, however, does not handle the low Q^2 region (correspondingly low x region) very well as can be seen from figure 6.26. We also see that the Kretzer parameterization shows a drastically stronger correction. However, as this is a multiplicative correction factor, the Q^2 scaling does not account for the difference in x slopes observed between the two analyses.

Another drastic change of course is the **JETSET tune**. The Monte Carlo production used by [29] contained the so-called Holger-tune (see table 5.1), for

[5]Right before and after the hard scattering.

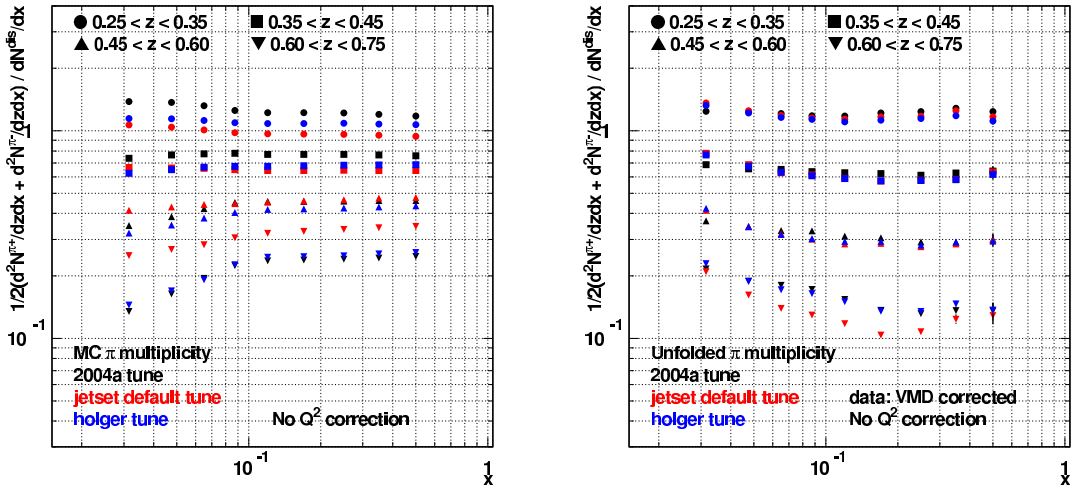


(a) Kretzer vs. KKP (b) Simple method vs. full pdf

Figure 6.26: Q^2 scaling correction checks performed. On the left side a comparison between the KKP parameterization and the Kretzer parameterization (respectively [51], [52]) is shown using just equation 6.38. On the right hand side we see the difference between the two methods and the effect of folding in the parton density functions.

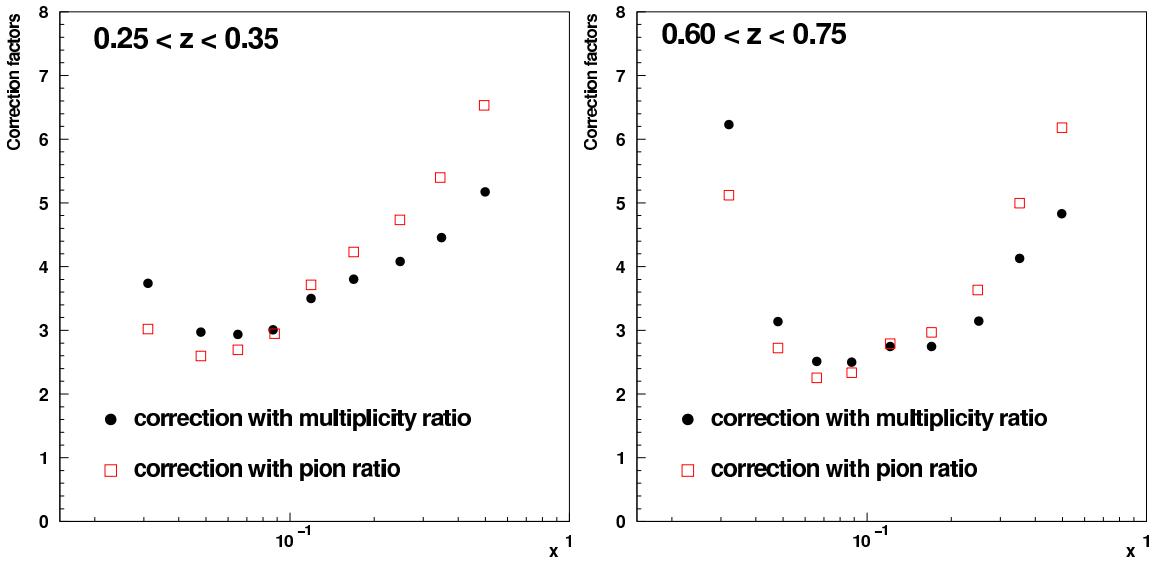
which the χ^2 lies far from the minimum obtained by the 2004c tune. Therefore in figure 6.27 the x dependence of both a Born level Monte Carlo simulation and unfolded data was plotted for the JETSET default tune, the Holger-tune and the more recent 2004a tune. For the pure Monte Carlo multiplicities one clearly sees big discrepancies between the different tunes, which largely cancel the experimental unfolded spectra, albeit not completely. However we can observe no different slope behavior for the x dependence and therefore the JETSET tune alone cannot be responsible for what is seen in figure 6.24. Note that in the right panel of figure 6.27 we see some disagreement between the JETSET default tune and the Holger/2004a tune in the unfolded spectra. This effect is clearly outside other systematic uncertainty contributions. However, it is important to state here that in fact it hardly makes sense to use the JETSET default tune to describe HERMES data. The default settings have been tuned to describe much higher energy e^+e^- data. It completely fails at describing HERMES experimental data and therefore using it to assign a systematic uncertainty as well would be far too conservative.

Now, the reason for the disagreement shown in figure 6.24 and is believed to be in the fact that the lepton detection efficiency was not taken into account as argumented in [29]. The acceptance correction method presented in this work does not make any assumptions upon the lepton detection efficiency. The old analysis simply used Monte Carlo pion yields in the acceptance and at generated level. Figure 6.28 shows the acceptance plotted as a multiplicative correction factor. We see that simply using pion yield ratios will result in a much steeper x dependence than using multiplicity ratios. So this might



(a) Monte Carlo Multiplicities (b) Unfolded Multiplicities

Figure 6.27: Comparison between different fragmentation tunes for the $\pi^+ + \pi^-$ multiplicity. Details for the tunes can be found in table 5.1.



(a) (b)

Figure 6.28: Comparison between two acceptance corrections, simply using pions yield ratios (open squares) and using multiplicity ratios (closed dots). The acceptance is displayed as a multiplicative correction factor.

to a large extent explain the different behavior in figure 6.24. However we must note here that this is probably not the only contribution. QED radiative corrections were applied using a standalone code, whereas here these are treated on event level together with detector smearing, which is a more elegant and also conceptually more correct approach.

Chapter 7

Influence of $\langle \cos\phi \rangle_{UU}$ moments on the acceptance correction

Here we investigate whether or not the extracted multiplicity distributions are influenced by unsimulated effects coming from the intrinsic transverse motion of the partons inside the nucleon. Since the importance of this discussion extends widely beyond the multiplicity analysis, it was placed in a separate chapter.

7.1 Introduction, problem statement

7.1.1 Some theoretical considerations

So far in the entire analysis transverse degrees of freedom have been neglected, or in other words silently integrated over. Nevertheless, in its full form the semi-inclusive deep-inelastic scattering cross section is 5-fold differential (as compared to equation 2.29) :

$$\frac{d^5\sigma^{lp\to lhX}}{dxdQ^2dzd^2\vec{p}_T} \tag{7.1}$$

and depends apart from the traditional x, Q^2 (or y) and z also on two transverse degrees of freedom, usually taken as the magnitude of the transverse momentum of the hadron p_T^2 and the azimuthal angle ϕ, formed between the lepton scattering plane and the plane formed by the hadron and the virtual photon. This is visualized in figure 7.1 below. The azimuthal ϕ angle is calculated as :

$$\cos\phi = \frac{(\vec{q}\times\vec{k})\cdot(\vec{q}\times\vec{p}_h)}{|\vec{q}\times\vec{k}|\cdot|\vec{q}\times\vec{p}_h|} \tag{7.2}$$

where $\vec{q}$, $\vec{k}$ and $\vec{p}_h$ respectively are the (3-) momentum vectors of the virtual photon, the incoming lepton and the produced hadron.

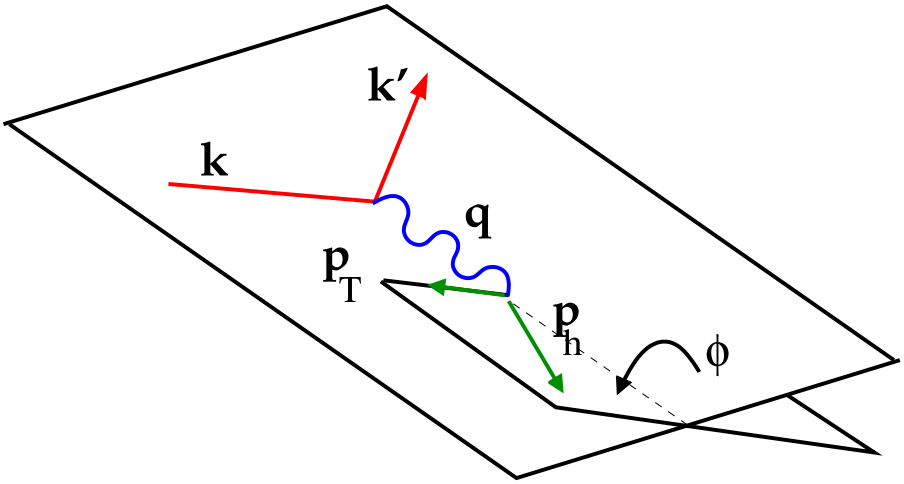


Figure 7.1: Definition of the azimuthal ϕ angle. k and k' are the momenta of the incoming and outgoing lepton, q of the exchanged virtual photon and p_h of the produced hadron.

It was in 1978 that Robert N. Cahn criticized [129], [130] basic parton model calculations in terms of the azimuthal ϕ distribution. By explicitely adding the intrinsic transverse parton momentum $k_\perp$ to the calculation, the total parton momentum is given by :

$$p = xP + k_\perp \tag{7.3}$$

From a simple derivation he obtained the following results for the moments in the azimuthal distribution for unpolarized electron-proton semi-inclusive deep-inelastic scattering (see [131] and [132] as well).

$$\begin{aligned} \langle \cos\phi \rangle_{UU} &= -\left[\frac{2k_\perp}{Q}\right] \frac{(2-y)\sqrt{1-y}}{1+(1-y)^2} \\ \langle \cos 2\phi \rangle_{UU} &= \frac{2k_\perp^2}{Q^2} \frac{1-y}{1+(1-y)^2} \end{aligned} \tag{7.4}$$

where the UU denotes the fact that both the beam and the target are unpolarized.

These results were confirmed later on and refined by J. Levelt and P. Mulders [133]. Recently another approach appeared [134], where a transverse momentum dependence is introduced as a Gaussian factor in the parton distribution $q_f(x)$ and fragmentation functions $D_q^h(z)$:

$$\begin{aligned} q_f(x,k_\perp) &= q_f(x) \frac{1}{\pi \langle k_\perp^2 \rangle} e^{-k_\perp^2/\langle k_\perp^2 \rangle} \\ D_q^h(z,p_\perp) &= D_q^h(z) \frac{1}{\pi \langle p_\perp^2 \rangle} e^{-p_\perp^2/\langle p_\perp^2 \rangle} \end{aligned} \tag{7.5}$$

such that

$$\begin{aligned} \int d^2\mathbf{k}_\perp q_f(x,k_\perp) &= q_f(x) \\ \int d^2\mathbf{p}_\perp D_q^h(z,p_\perp) &= D_q^h(z) \end{aligned} \tag{7.6}$$

When integrating the 5 fold differential cross section (equation 7.1) over the intrinsic transverse quark momenta (assuming $\vec{p}_\perp \approx \vec{p}_T - z\vec{k}_\perp$, with $\vec{p}_T$ the final hadron transverse momentum), a $\cos\phi$ - like dependence for the hadron azimuthal angle distribution is obtained (up till $\mathcal{O}(1/Q)$) [135]. One can think of $\vec{p}_\perp$ as the 'intrinsic' transverse momentum picked up in the fragmentation process.

If one does *not* allow for time-reversal odd structure functions [136], a $\cos(2\phi)$ term appears only at order $\mathcal{O}(1/Q^2)$. However, if one *does* allow for T-odd effects, an azimuthal $\cos(2\phi)$ asymmetry appears at leading order as well. The $\cos(2\phi)$ term is related to the quark transversity distribution $h_1^\perp(x)$ and Collins fragmentation function $H_1^\perp(z)$ [137], [136]. One can parameterize the azimuthal dependence of the semi-inclusive cross section simply as follows:

$$\frac{d\sigma}{d\phi} = A + B\cos\phi + C\cos 2\phi \tag{7.7}$$

where the resulting moments of this distribution are straightforwardly given by :

$$\langle\cos\phi\rangle = \frac{\int_0^{2\pi}\cos\phi\frac{d\sigma}{d\phi}}{\int_0^{2\pi}\frac{d\sigma}{d\phi}} = \frac{B}{2A} \tag{7.8}$$

$$\langle\cos 2\phi\rangle = \frac{\int_0^{2\pi}\cos 2\phi\frac{d\sigma}{d\phi}}{\int_0^{2\pi}\frac{d\sigma}{d\phi}} = \frac{C}{2A} \tag{7.9}$$

Experiments have measured the $\cos\phi$ moment to be negative on the order of -0.1 and the $\cos(2\phi)$ to be small but positive [138] (ZEUS), [139] (E665), [140] (EMC). Figure 7.2 shows experimental results obtained by the ZEUS collaboration. In fact this is yet another experimental proof [19] of the partonic structure of the nucleon. Should the nucleon be a continuous distribution of charge, one would only expect a $\cos(2\phi)$ modulation of the angular distribution. However, the Fermi motion of the struck point-like particle inside the nucleon gives rise to an intrinsic transverse momentum k_T of the parton and this leads to important $\cos\phi$ contributions to the cross section.

Recent work done by K. Oganessyan [137] has shown that, considering the shape of the HERMES acceptance, the azimuthal effects cannot be simply neglected by the standard hand-waiving argument that one has integrated over this angle. In fact, the HERMES acceptance already shows a rather dramatic $\cos\phi$ dependence in itself, strongly depending also on x as this variable is highly correlated with Q^2, which is in its turn related to the scattering angle. Expanding the acceptance function versus ϕ into a Fourier series yields

$$\varepsilon(\phi) = C_0 + \sum_{m=1}(C_m\cos(m\phi) + D_m\sin(m\phi)) \tag{7.10}$$

Applying this to a measured asymmetry [137], shows that even after integration over ϕ, contributions related to the $\cos\phi$ moment still remain.

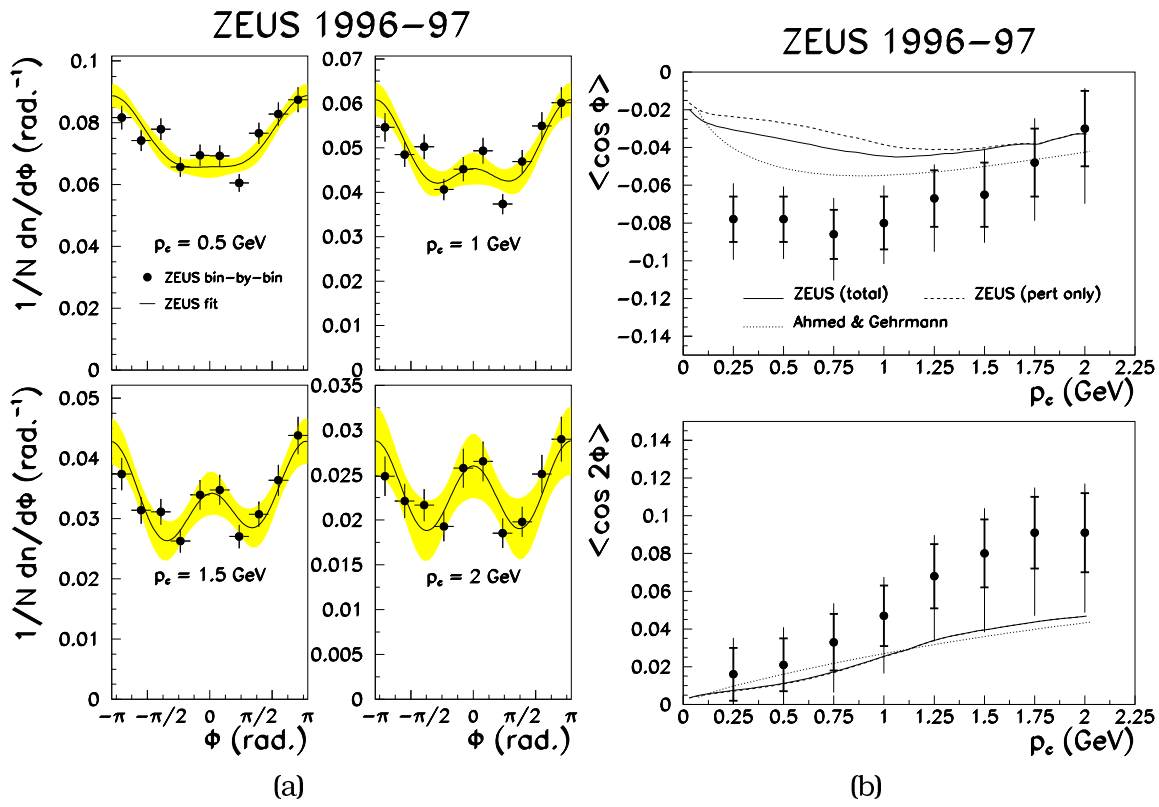


Figure 7.2: Results for the azimuthal ϕ dependence of the semi-inclusive cross section and the dependence of the first and second moment on the p_T cutoff parameter. Figures taken from [138].

7.1.2 Description of the azimuthal ϕ distribution in the Monte Carlo

The basic question then is how well the LEPTO Monte Carlo program, which was used for the acceptance correction, reproduces the azimuthal angle dependence, which should be something like

$$\frac{d\sigma}{d\phi} \approx 1 - 0.2\cos\phi \tag{7.11}$$

As it turns out however, no azimuthal dependence is generated by LEPTO at all in its default configuration, see figure 7.3. There is a switch, `LST(6)` which influences the azimuthal angle of the lepton scattering plane. As this is the only one which has an impact on the azimuthal distribution of the hadrons, it was tested. However, activation of this switch did not provide a satisfactory azimuthal angle dependence as can be inferred from figure 7.3. Furthermore, other kinematical distributions were completely distorted by it. To make things even worse, the shape of the HERMES acceptance function shows a dramatic variation in ϕ as illustrated in figure 7.4. One can first of all see a strong dependence of the acceptance function on ϕ. At low x, more hadrons are accepted near $\phi = \pi$, whereas at large x the situation is reversed and more hadrons are accepted at $\phi = 0$, or 2π. This means that if $\langle \cos\phi \rangle_{UU}$

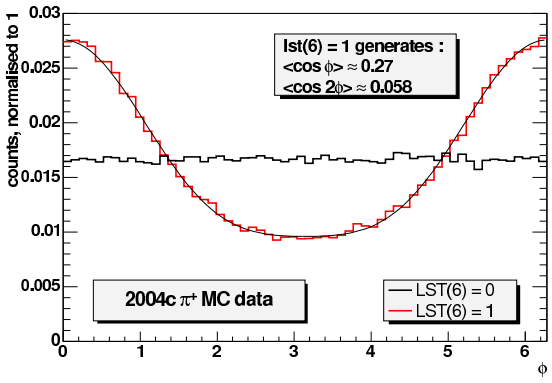


Figure 7.3: Azimuthal angle dependence generated by the LEPTO Monte Carlo. Note that the histograms both have been normalized to unity. The figure is made from 2 different setting of the `LST(6)` parameter of the LEPTO model.

is sizeable, e.g. at small x we would detect more pions than we would expect based upon a uniform $d\sigma/d\phi$ distribution.

Thus we need to know how these $\langle\cos\phi\rangle_{UU}$ moments affect the acceptance function and possibly correct for it. So first of all we must obtain a rough estimate of these moments. As the HERMES kinematical range is far away from previously published experimental results [138], [139], [140], we choose not to rely on them but rather extract the moments from HERMES data. In our opinion, it is reasonable to assume, since the used Monte Carlo production has a uniform ϕ distribution, that the difference in shape between the data and the reconstructed Monte Carlo is due to the $\langle\cos\phi\rangle_{UU}$ moment left over in the data. This will be our starting point, later on we will extract weights to reweight our Monte Carlo production **at the Born level** and apply a shape like equation (7.7) to the otherwise uniform $d\sigma/d\phi$. Of course we will have to show that this reweighting works. We can then investigate the effect on the acceptance function and thus construct a correction or systematic error for the multiplicities.

7.2 Extraction of the $\cos\phi$ moments

Correlations between different kinematic variables are very important in the extraction of the moments. Therefore we need to make the extraction using fully tracked Monte Carlo data. Since the smearing generator (HSG) uses random smearing it does not account for any correlations between kinematical variables in the way that the HMC + HRC chain does. It is therefore safer to use the latter combination. The production used, had the `2004c` JETSET tune, used CTEQ6 LO pdf set and radiative effects switched on by RADGEN. For the data we used the same dataset from the `00c1` production as was used for the multiplicity extraction. Obviously we need to impose identical cuts on the data and the Monte Carlo in the acceptance. This includes the hadron momentum cut $2 < p < 15$ GeV, a point which was ignored in [141]. This will also explain

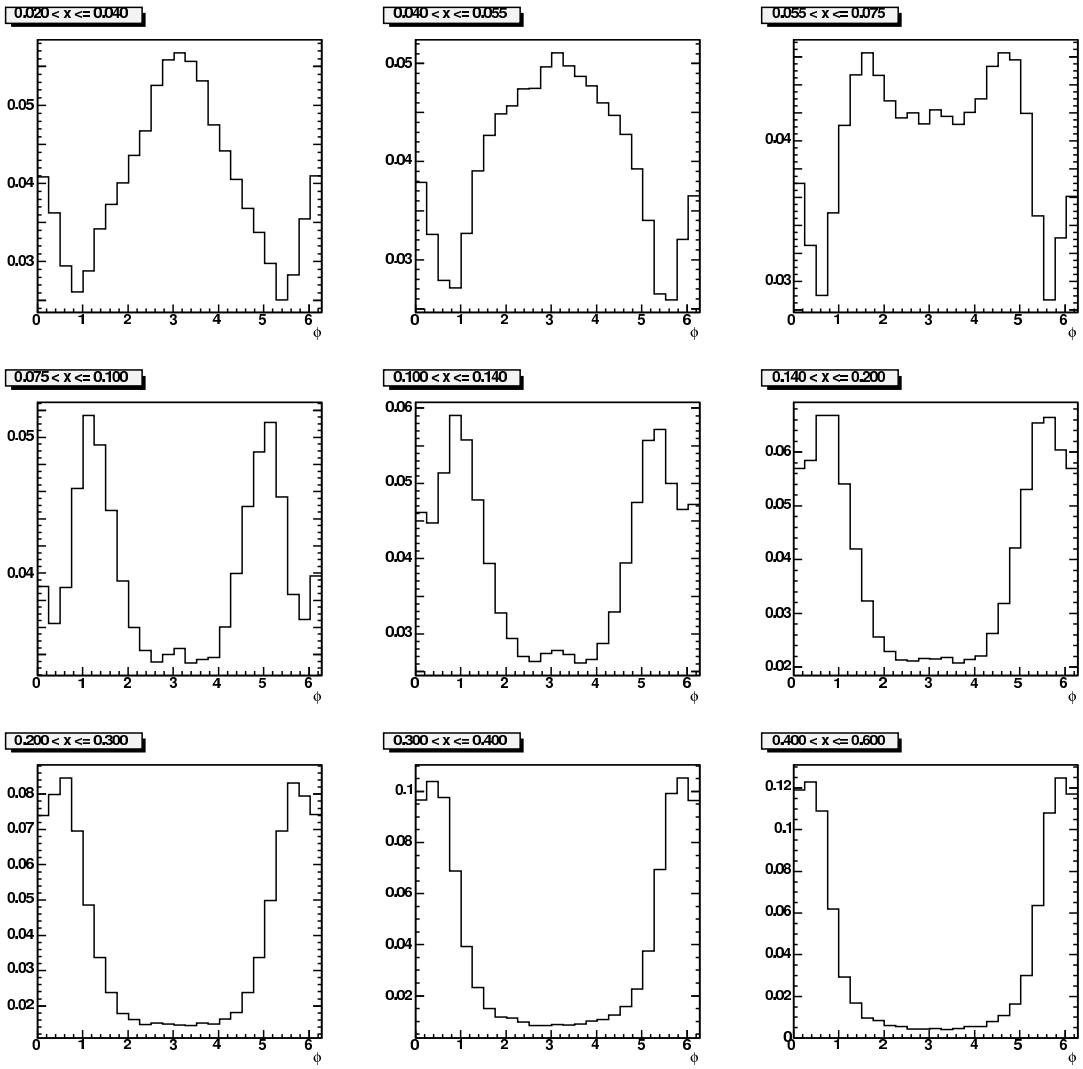


Figure 7.4: Shape of the pion acceptance function versus ϕ in various ranges of x. The histograms were normalized to unity and thus show only the shape of the acceptance, not the magnitude.

the discrepancy between the moments found in this work and the moments presented in [141].

7.2.1 The methods compared

Let us first explore how the $\langle \cos\phi \rangle_{UU}$ can be extracted from Monte Carlo. In figure 7.5 one can see a shape comparison of the azimuthal ϕ distribution in 4 different x bins. The data clearly overshoots the Monte Carlo around $\phi \sim \pi$ and undershoots it at $\phi \sim 0$ or 2π. As discussed, we can attribute this behavior to a $\cos\phi$ modulation which is present in the data but unsimulated in the Monte Carlo. This becomes clearer when we make the ratio data / Monte Carlo. Note that for now we have used 4 x bins, but integrated over all other kinematical variables. This will be refined later on as this exploratory exercise is done to

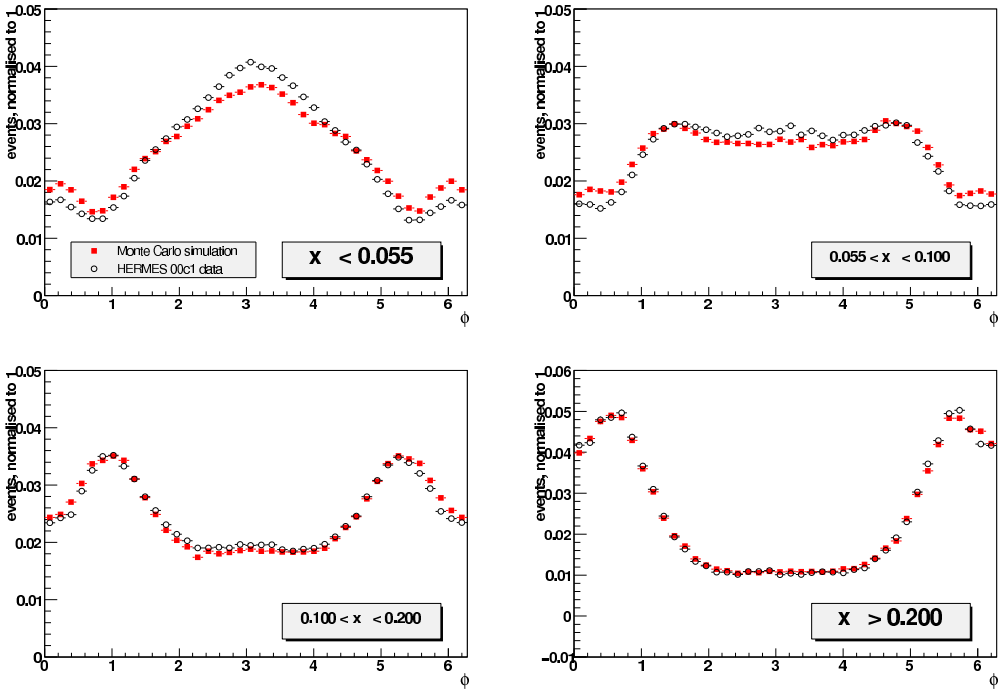


Figure 7.5: Shape comparison for semi-inclusive π^+ production between the ϕ distribution for the HERMES experimental data (open circles) and the Monte Carlo simulation (closed circles). Both histograms were normalized to unity.

fix the framework for a more complete extraction.

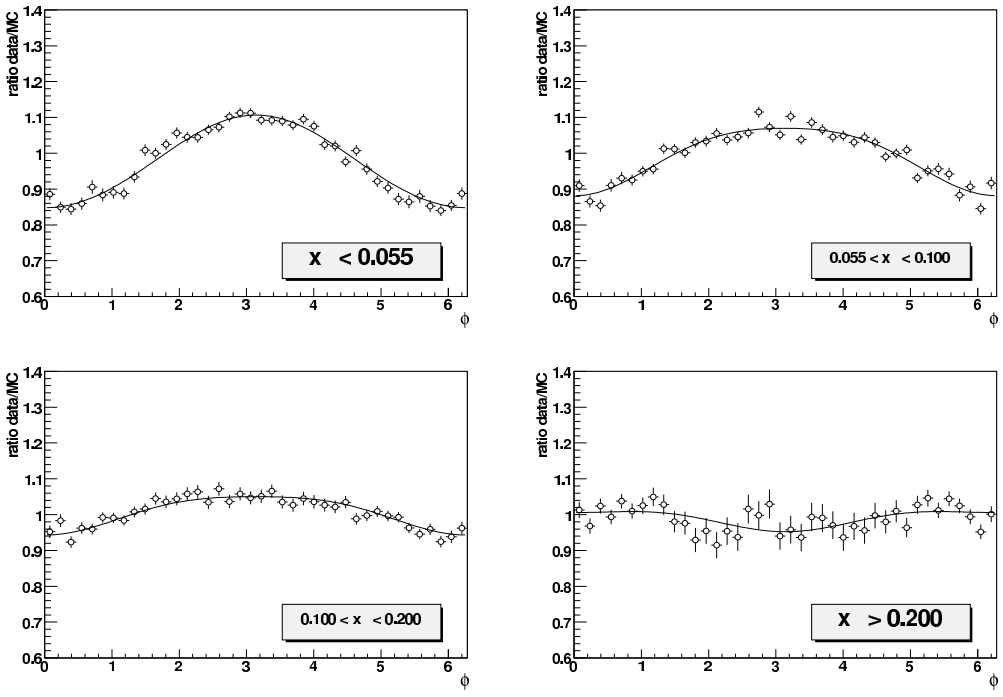


Figure 7.6: Fit of equation 7.7 to the ratio data/Monte Carlo in the HERMES acceptance. The result is displayed on the figure.

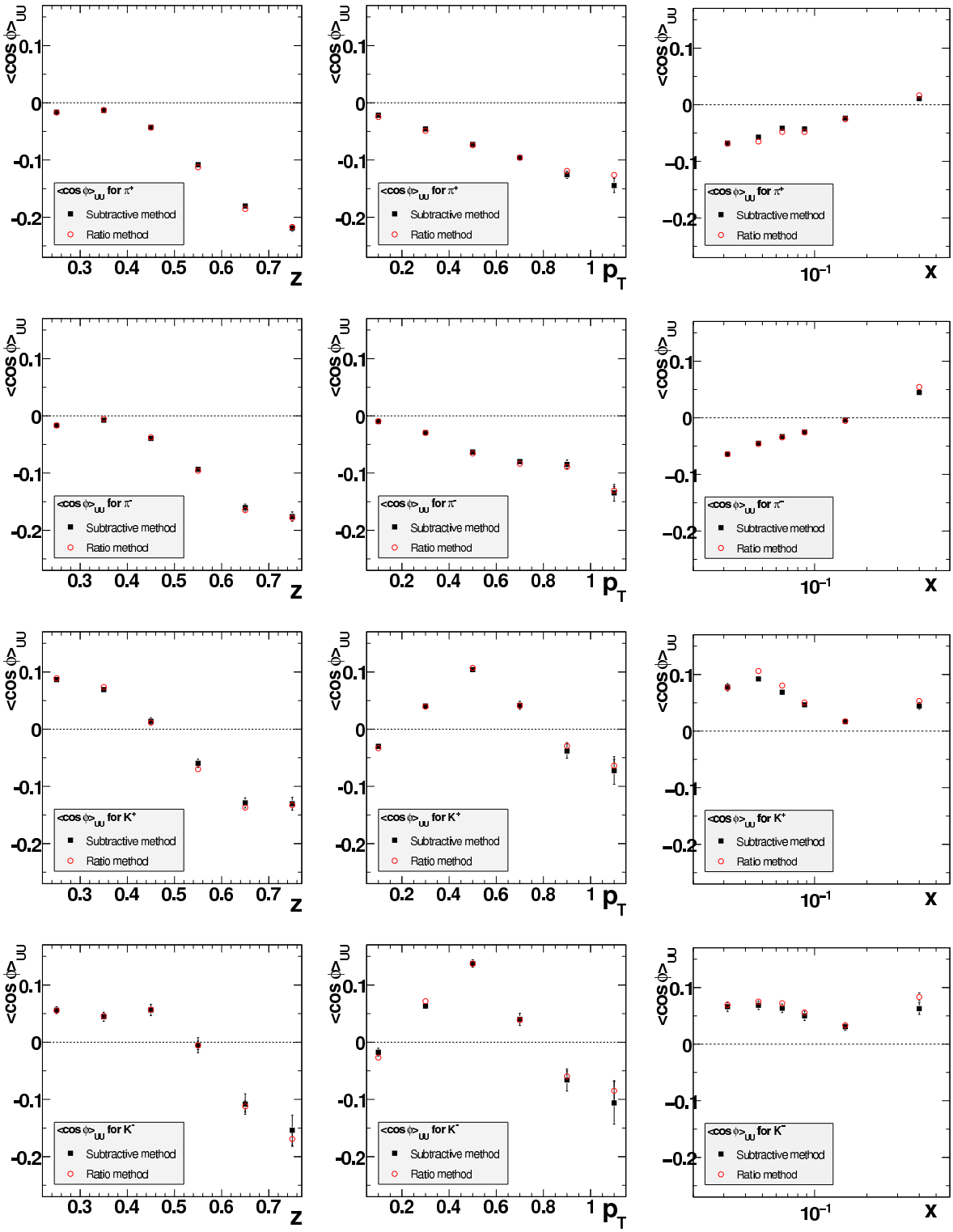


Figure 7.7: Extracted $\langle \cos\phi \rangle_{UU}$ moments for π^+ (top row), π^-, K^+ and K^- (bottom row) versus z, p_T and x. The two methods discussed in the text are compared, where the closed black squares represent the subtractive method and the open red circles the ratio method. Each time the results were obtained integrated over the other kinematical variables. The data points are included in appendix C

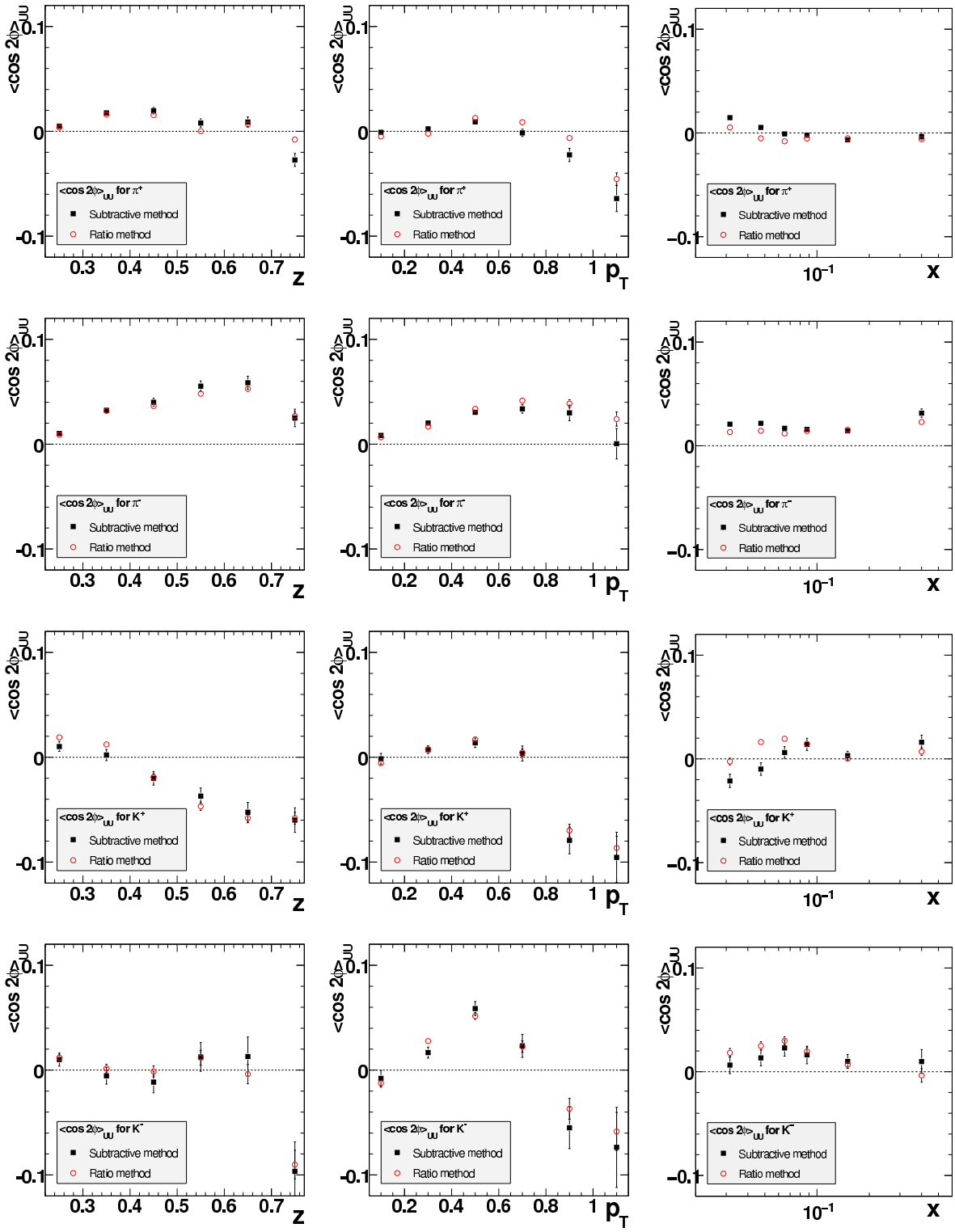


Figure 7.8: Extracted $\langle\cos 2\phi\rangle_{UU}$ moments for π^+ (top row), π^-, K^+ and K^- (bottom row) versus z, p_T and x. The two methods discussed in the text are compared, where the closed black squares represent the subtractive method and the open red circles the ratio method. Each time the results were obtained integrated over the other kinematical variables. The data points are included in appendix C

As can be seen from figure 7.6, the difference observed between the data and Monte Carlo does indeed have the shape as proposed by equation 7.7.

The method to extract the $\langle\cos\phi\rangle$ and $\langle\cos 2\phi\rangle$ moments from making the ratio between data and Monte Carlo we will call the **ratio method**. It was criticized, however, [142] in favor of a **subtractive method** using the moments alone. One can express the measured cross section $\sigma_{meas}(\phi, x, y, z)$ as follows [143] :

$$\sigma_{meas}(\phi, x, y, z) = \tag{7.12}$$
$$\sigma_0(x, y, z) \cdot (1 + 2\langle\cos\phi\rangle_{UU} \cdot \cos\phi + 2\langle\cos 2\phi\rangle_{UU} \cdot \cos 2\phi) \cdot \epsilon_{acc}(\phi, x, y, z) \cdot \epsilon_{rad}(\phi, x, y, z)$$

The acceptance function ϵ_{acc} can be expanded into a Fourier series where

$$\epsilon_{acc} = a_0 + \sum_{n=1}^{\infty} a_n \cos(n\phi) + \sum_{m=1}^{\infty} b_m sin(m\phi). \tag{7.13}$$

Radiative effects change the total cross section and also exhibit $\cos\phi$ modulations. The real photon is radiated approximately collinear with the lepton momentum, therefore the virtual photon momentum is changed within the scattering plane, symmetrically in ϕ about the incident lepton beam. Therefore one can write $\epsilon_{rad} = r_0 + \sum_{k=1}^{\infty} r_k \cos(k\phi)$. If we determine the $\cos\phi$ and $\cos 2\phi$ moments from the measured cross section (eq. 7.12) we can easily see :

$$2\langle\cos\phi\rangle_{meas} = 2\langle\cos\phi\rangle_{UU} + \frac{a_1}{a_0} + \frac{r_1}{r_0} \tag{7.14}$$

$$2\langle\cos 2\phi\rangle_{meas} = 2\langle\cos 2\phi\rangle_{UU} + \frac{a_2}{a_0} + \frac{r_2}{r_0} \tag{7.15}$$

In the Monte Carlo, no $\langle\cos\phi\rangle_{UU}$ modulation is present at Born level, so the reconstructed Monte Carlo ϕ spectrum only contains modulations from acceptance and radiative effects :

$$2\langle\cos\phi\rangle_{MC} = \frac{a_1}{a_0} + \frac{r_1}{r_0} \tag{7.16}$$

$$2\langle\cos 2\phi\rangle_{MC} = \frac{a_2}{a_0} + \frac{r_2}{r_0} \tag{7.17}$$

Therefore by subtracting the Monte Carlo moments from the measured moments in the data we obtain [143] the true moments $2\langle\cos\phi\rangle_{UU} = 2\langle\cos\phi\rangle_{meas} - 2\langle\cos\phi\rangle_{MC}$ and $2\langle\cos 2\phi\rangle_{UU} = 2\langle\cos 2\phi\rangle_{meas} - 2\langle\cos 2\phi\rangle_{MC}$.

Note that in order to fit the measured and Monte Carlo distributions we have to take into account more than 2 terms in the Fourier expansion. We have used

$$p_0 + \sum_{j=1}^{4} p_j \cos(j\phi) \tag{7.18}$$

as shape. Using more slightly improves the goodness of fits, but the extracted cosine amplitudes remain unchanged [143]. Figures 7.7 and 7.8 show respectively the $\langle\cos\phi\rangle_{UU}$ and $\langle\cos 2\phi\rangle_{UU}$ moments from both the ratio and the

subtractive method versus z, p_T and x where only a 1-dimensional binning was employed for each of the variables. The error bars are statistical and include the correlation between the fit parameters. Inside each bin the ϕ spectrum was binned in 15 equidistant bins between 0 and 2π. A systematic check was performed where this number of bins was doubled. The impact on the extracted moments was marginal and well within the statistical error. Data from these figures is included in appendix C for reference.

First of all we observe a good agreement between the two methods used. Especially versus z and p_T. Versus x we do note a small systematic discrepancy at low x for the $\langle\cos 2\phi\rangle_{UU}$ moments. The results obtained are in agreement with [143]. The $\langle\cos\phi\rangle_{UU}$ moments are negative for the pions and increase with increasing z, which is also predicted by theory [134]. The same behavior is shown for p_T, which is intuitively clear as these moments originate from the transverse motion of the quarks in the nucleon. The $\langle\cos 2\phi\rangle_{UU}$ moments are interesting as they are directly related in leading order to the Collins fragmentation function $H_1^\perp(z)$. The rising behavior and the falloff back to 0 for π^+ and π^- are in agreement with theoretical predictions [144]. We, however, must make the note here that the realistic reproduction of the detector efficiency and geometry is crucial. Especially misalignment effects will play an important role in a more complete and systematic extraction. In [141] the $\langle\cos\phi\rangle_{UU}$ moments introduced by misalignment were measured to be up to -0.02.

We choose the subtractive method to proceed as this method appears to be analytically most correct. Integrating over other kinematical variables introduces large systematic effects upon the moments as well [145], [146]. We therefore need a finer binning in multiple variables to do a good extraction. This will be attempted next. However the remark should be made here that this analysis will not strive to obtain final $\langle\cos\phi\rangle_{UU}$ and $\langle\cos 2\phi\rangle_{UU}$ moments. They will merely be used to estimate and correct for the systematic effect in the multiplicity analysis.

7.2.2 Binning and extraction

To extract a set of moments, to be used in reweighting the Monte Carlo, we have to take into account the kinematical dependence of the $\langle\cos\phi\rangle_{UU}$ and $\langle\cos 2\phi\rangle_{UU}$ moments. In a world with unlimited statistics, one would of course be able to bin completely in all 4 remaining degrees of freedom. However, this is not the case. It was decided to choose a binning in function of the analysis, which means rather fine in z and x, where we chose 4 bins in x and only 3 in z. This way, however, we did integrate over Q^2 (or y). This does introduce some unavoidable systematic effects [145] on the extracted moments. We also decided to split the data into just 2 p_T bins, this is because transverse momentum is a key component to the shape of the ϕ distribution. We defined one low p_T (< 0.33 GeV) and one high p_T (> 0.33 GeV) bin. The binning is displayed in table 7.1. We have a total of 24 kinematical bins and in each of them the ϕ spectrum was histogrammed again in 15 equidistant ϕ

bins between 0 and 2π.

$p_T < 0.33$		$p_T < 0.33$	
$0.00 < z < 0.30$	$0.30 < z < 0.55$	$0.55 < z < 1.00$	
$x < 0.055$	$0.055 < x < 0.1$	$0.1 < x < 0.2$	$x > 0.2$

Table 7.1: The binning used for the extraction of the $\langle\cos\phi\rangle_{UU}$ and $\langle\cos 2\phi\rangle_{UU}$ moments.

The results of the multidimensional extraction is included in appendix C as well. When calculating the moments, the correlation between the fit parameters p_j in equation 7.18 was taking into account in the calculation of the uncertainty on the resulting moment. The results are presented versus x in each of the p_T and z bins in appendix C. (See figures C.1, C.2, C.3 and C.4 with the data points in tables C.7 for the pions and C.8 for the kaons.) Each plot was fitted with a second degree logarithmic polynomial.

$$f(x; a_0, a_1, a_2) = a_0 + a_1 \ln(x) + a_2 \ln^2(x) \tag{7.19}$$

We then used a reweighting method to introduce these moments in the Monte Carlo.

7.3 Reweighting the Monte Carlo

The extracted moments were thus used to construct additional weights for the hadron tracks. This way we explicitely introduced the cosine modulations at the **Born level**. For smoothness we used the fitted function values from equation 7.19 instead of the binned data. The weights assigned to each hadron track are then calculated as :

$$w^p(\phi; p_T, z, x) = 1 + 2\,\langle\cos\phi\rangle^p_{p_T,z,x}\,\cos(\phi) + 2\,\langle\cos 2\phi\rangle^p_{p_T,z,x}\,\cos(2\phi) \tag{7.20}$$

Where the index p denotes the hadron type. The resulting Born level spectra can be seen in figure 7.9, where we plotted the ϕ, z, p_T and x dependence for π^+ production. The reweighted Monte Carlo shows a Born distribution in ϕ which has a indeed a clear cosine modulation. Integrating over ϕ, we observe as well that the distributions versus p_T, x and z remain identical at the Born level between the normal Monte Carlo simulation and the reweighted one, as desired. Note that the two histograms in these cases are really on top of each other, therefore only one shows up in the figure.

As a first consistency check for the reweighting procedure, it is a good thing to see whether or not we can extract again the same moments from the Born level ϕ distribution as were put into the Monte Carlo. Figure 7.10 shows the agreement between what was inserted in the Monte Carlo (fit lines) and the re-extracted Born level moments to be satisfactory.

Another consistency check is to assess whether or not the reweighted, reconstructed ϕ distribution in the Monte Carlo now matches the shape of the

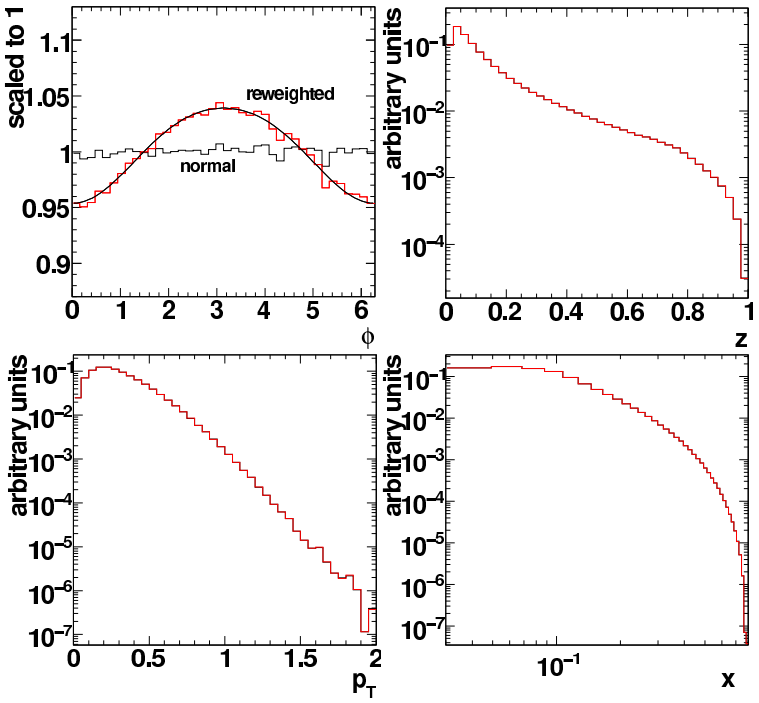


Figure 7.9: Born level distributions, the reweighted Monte Carlo production is shown by the fitted red line. The ϕ histograms have been rescaled to match the unweighted, flat distribution. The other histograms were normalized to unit area. Note that the agreement versus p_T, x and z is so well that the two curves are on top of each other.

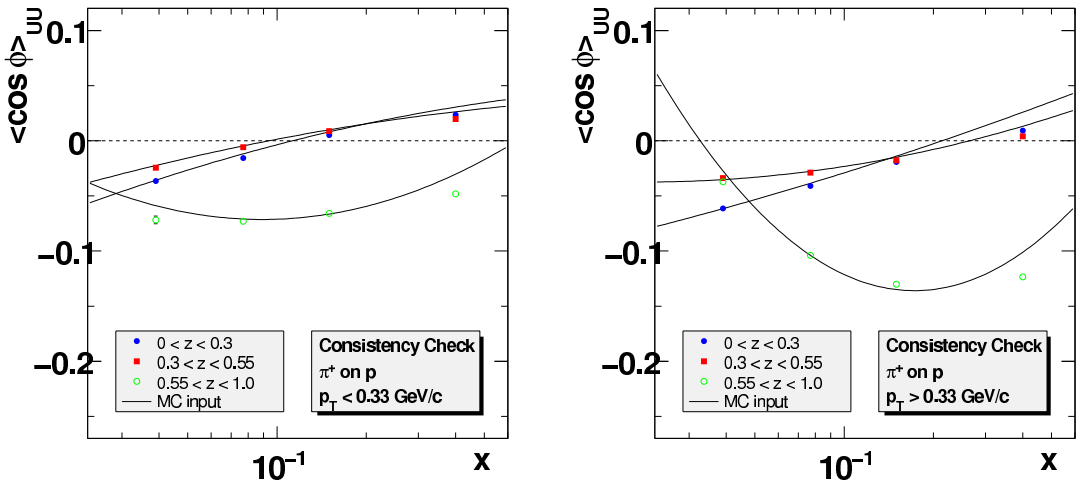


Figure 7.10: Consistency check for the Monte Carlo. The full line represents the fitted functions that were used as input to reweight the Monte Carlo at born level. The data points represent the moments that were extracted back from the reweighted Born level Monte Carlo.

experimentally measured ϕ distribution better. The ratio of the reconstructed HERMES data and the reconstructed, reweighted Monte Carlo should contain no ϕ moments anymore. Again we focus this check on the π^+ versus x. In figure 7.11 actually the same data/Monte Carlo ratio is plotted as in figure 7.6, however, this time we also reweighted the hadron tracks. So a direct compar-

ison is possible between the two figures. We can clearly see that the moments

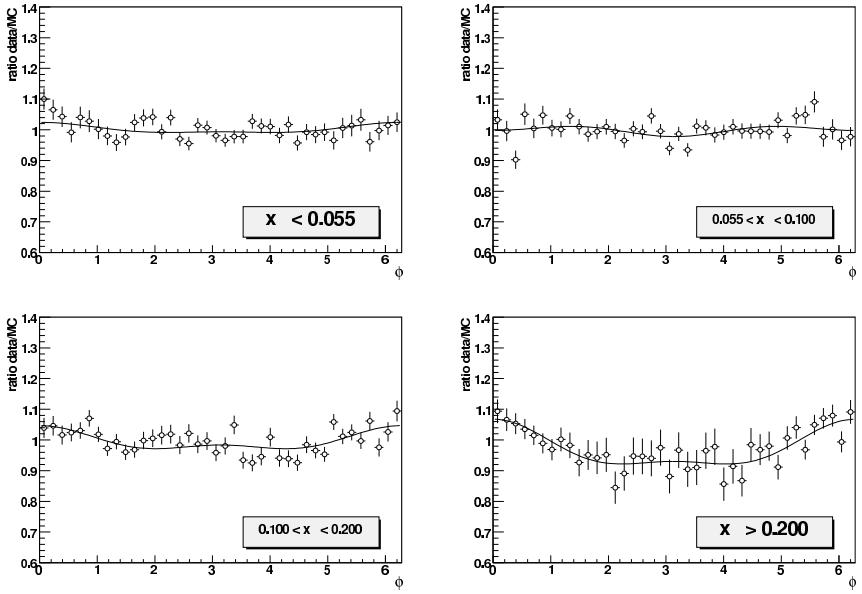


Figure 7.11: Analogous to figure 7.6, however, now using the additional hadron weights in the Monte Carlo.

virtually disappear. At high x, however, we over-correct a little bit. This is probably due to the fact that we used fitted functions which do not completely match the extracted moments. Also we have not used the same binning in figures 7.6 and 7.11 as we had for the multidimensional extraction. Anyway, we learn from this and the consistency check above that the reweighting is a good approach leading to the results that were expected. We can now proceed in calculating the acceptance functions from the normal Monte Carlo and the reweighted sample.

7.4 Influence on the acceptance correction

The acceptance function was calculated versus z itself and versus x in the 4 z bins as shown in equation 6.12. We define the acceptance correction function as being the ratio between the hadron multiplicity at Born level and the multiplicity for the fully tracked Monte Carlo[1]. Explicitely written versus x and z this looks like :

$$\mathcal{A}(x_i, z_j) = \frac{N^h_{trk}(x_i, z_j)/N^{DIS}_{trk}(x_i)}{N^h_{Born}(x_i, z_j)/N^{DIS}_{Born}(x_i)} \tag{7.21}$$

The complete unfolding formalism will not be applied here as the idea is to simply determine a systematic effect. The result of the acceptance comparison is shown in figure 7.12 versus z and figure 7.13 versus x in 4 z bins.

[1]Of course also with radiative effects turned on.

Versus z we only show the π^+ and K^+ particles to present an idea of the effect. Versus x we restrict ourselves to the positive pion only. However, the same conclusions apply for all 4 particles. One can see that in general the effect of

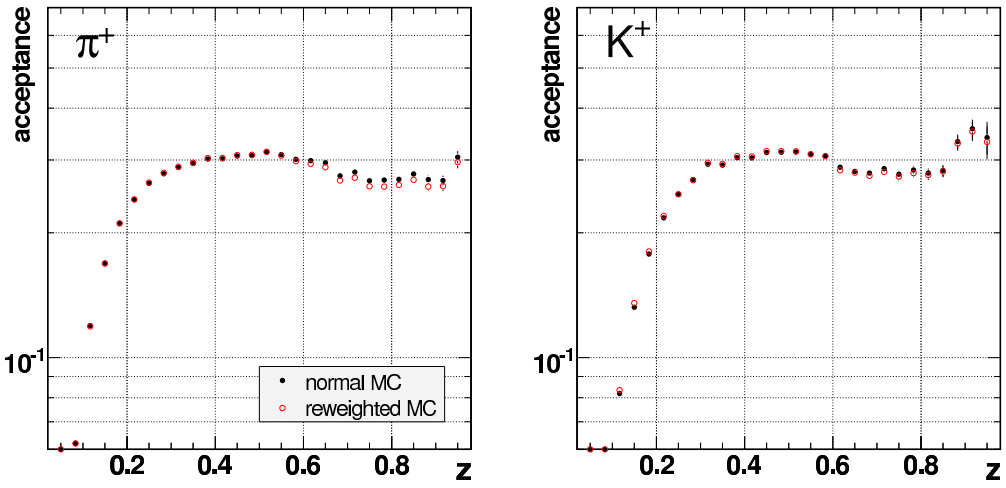


Figure 7.12: Comparison in the hadron acceptance function versus z for standard fully tracked Monte Carlo (2004c tune) and Monte Carlo reweighted according to ϕ.

the reweighting is rather small. At high z we notice some discrepancy in both the one dimensional acceptance function (figure 7.12) and also in the highest z bins versus x. As we noticed also the highest $\langle \cos\phi \rangle_{UU}$ moments at high z this was to be expected. Surprisingly enough the effect is larger at high x within this highest z bin, even though in the 1-dimensional case the $\langle \cos\phi \rangle_{UU}$ moment vanished at high x. This is a clear example of how binning can influence this analysis.

We must give here a word of caution. It was observed that misalignment effects do play some role in this discussion. In figure 7.14 we see possible $\langle \cos\phi \rangle_{UU}$ and $\langle \cos 2\phi \rangle_{UU}$ moments introduced by misalignment. For this check a Monte Carlo sample was used where both the top and bottom detector were rotated in θ_x and θ_y by a small amount. Also an offset was introduced in x and y. In fact the very same production was used in misalignment studies by the transversity group [143]. A similar approach was adopted as compared to what we discussed before in order to extract the cosine moments. We can see that the moments introduced through misalignment are of similar order of magnitude as the moments extracted in figure 7.7. However an elaborate study of alignment effects on the extraction of the cosine moments falls outside the scope of this work. It was therefore decided to assign the variations observed in the acceptance function as a systematic uncertainty to the multiplicity distributions. This uncertainty versus z is no more than 1.5 % for all hadrons except for π^+ where it reaches up till 3 % at high z (> 0.7). Versus x the situation is somewhat worse. At high x one sees relative differences between the reweighted and the ordinary acceptance function that can reach up to 10 %. This must clearly be the effect of the strong coupling of the intrin-

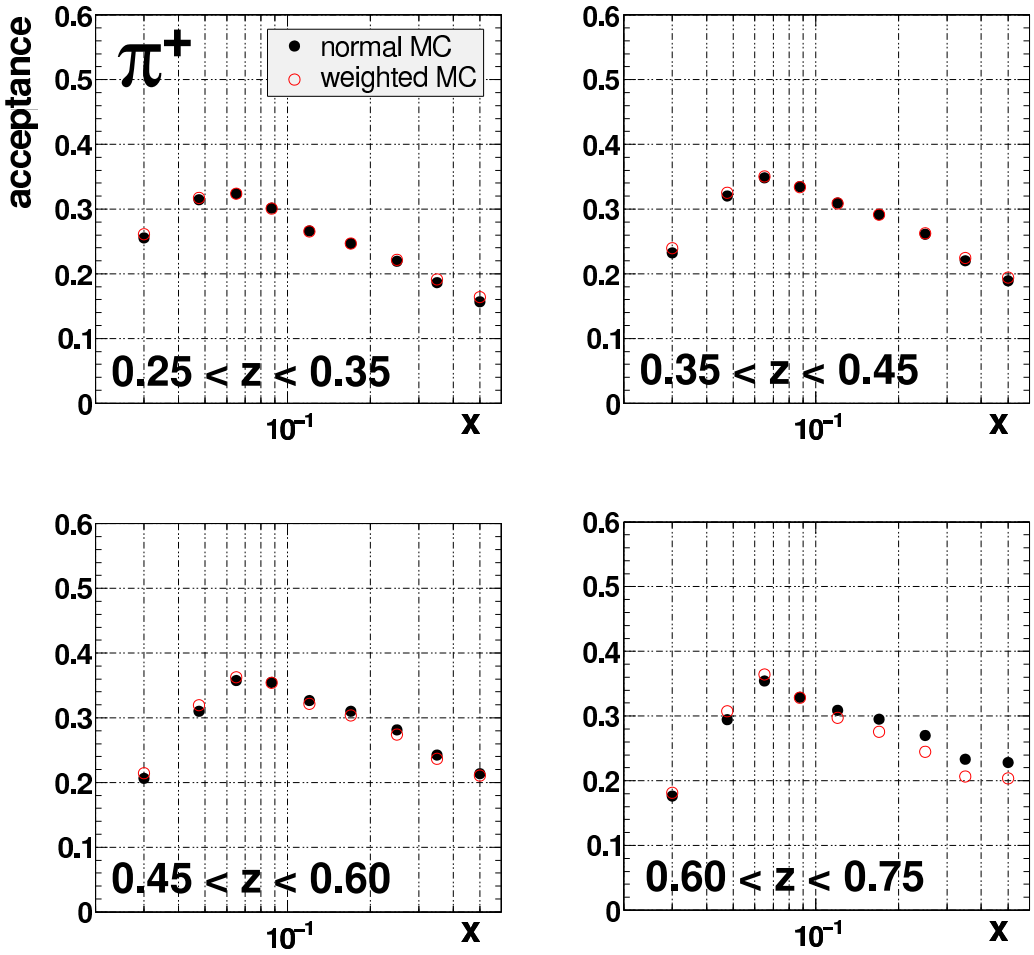


Figure 7.13: Comparison in the hadron acceptance function versus x in 4 z bins for standard fully tracked Monte Carlo (2004c tune) and Monte Carlo reweighted according to ϕ. Only π^+ is shown.

sic cosine modulations and the HERMES acceptance. Figures 7.15 and 7.16 show this behavior resp. versus z and versus x. As the additional systematic contribution versus z is only in the order of a couple of %, it does not change the results obtained and presented in the previous chapter.

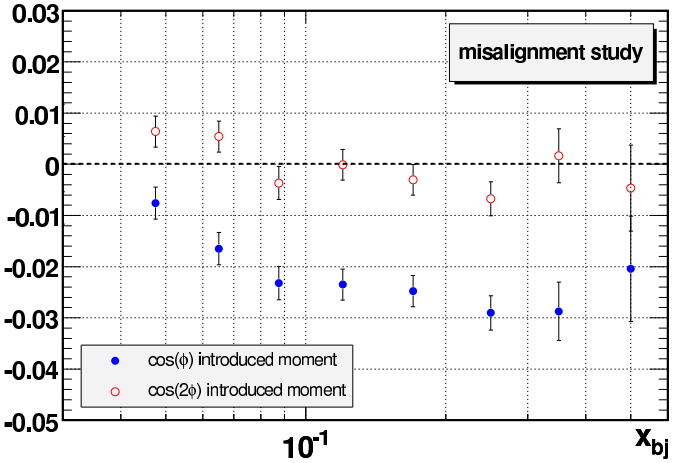


Figure 7.14: Possible cos moments introduced by misalignment of the detector. Extracted for positive pions. [141].

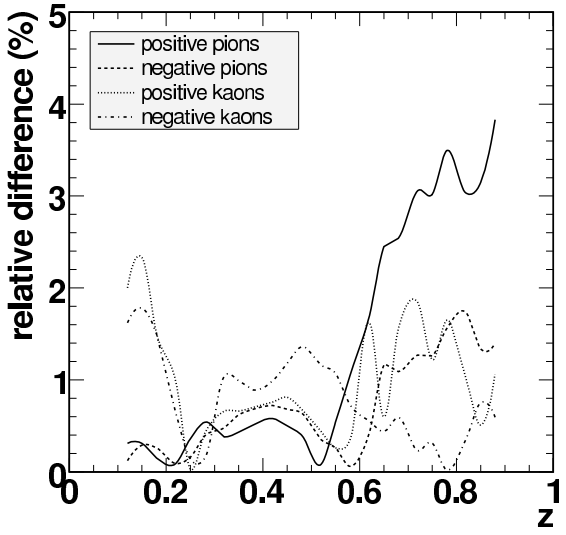


Figure 7.15: Relative differences in % for the acceptance functions versus z. Data in appendix C.9.

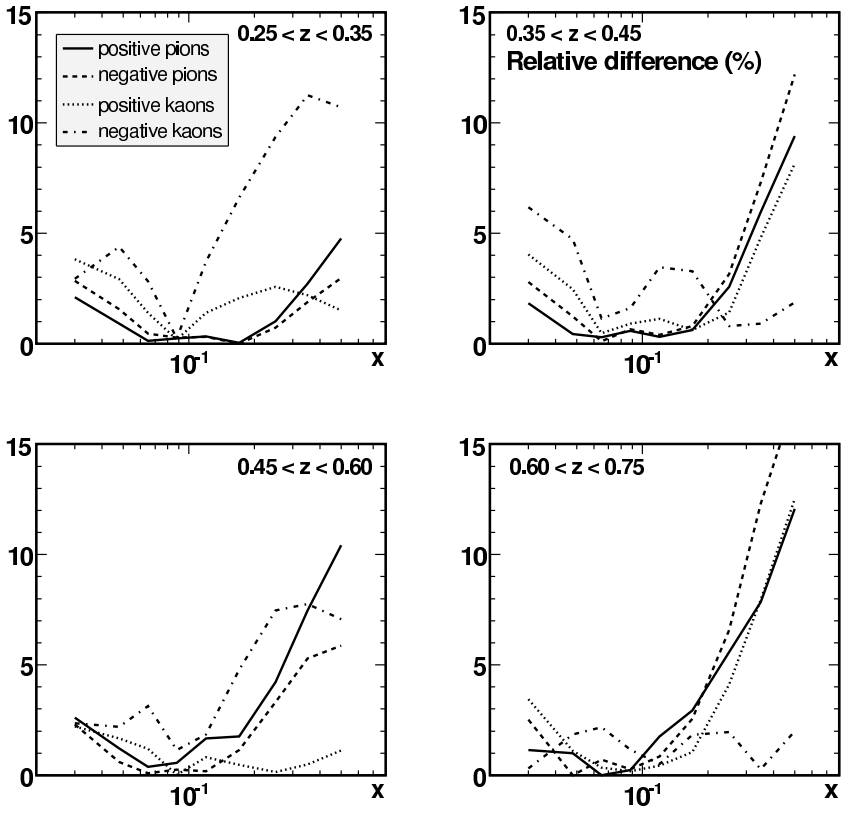


Figure 7.16: Relative differences in % for the acceptance functions versus x. Data in appendix C.10.

Chapter 8

Summary and discussion

We have come to the end of this thesis so a small summary and discussion of the results obtained are in order. The aim of this work was to extract Born level multiplicity distributions for the new HERMES 2000 unpolarized hydrogen data set for π^+, π^-, K^+ and K^- separately. The emphasis during this study was put on understanding and tuning the fragmentation model in the Monte Carlo simulation, a clear treatment of the instrumental and QED radiative smearing as well as an extensive study of systematic effects involved.

The interpretation of hadron multiplicities (σ^h/σ^{DIS}) in the quark-parton model is given by the following expression:

$$\frac{1}{\sigma_{DIS}}\frac{d\sigma^h(z,Q^2)}{dz} = \frac{\sum_f e_f^2 \int_0^1 dx\; q_f(x,Q^2) D_f^h(z,Q^2)}{\sum_f e_f^2 \int_0^1 dx\; q_f(x,Q^2)} \tag{8.1}$$

One can think of them as fragmentation functions $D_f^h(z,Q^2)$ summed over the quark flavors f. These fragmentation functions parameterize the hadronization process, which cannot be determined from QCD because a perturbative expansion in α_S, the strong coupling constant, does not converge because of confinement. Assuming some symmetry relations like isospin rotation and neglecting strange quark contributions, the pion multiplicity can even be set equal to the u quark fragmentation function D_u^π.

At first experimental multiplicity distributions were obtained for the 4 given hadron types and charges. Good separation between pions and kaons is possible in HERMES because of the presence of a Ring Imaging Čerenkov detector. This detector is very efficient over a momentum range from 2 to 15 GeV. However, tagged hadrons are subject to a certain amount of misidentifications. These misidentifications were corrected for using an unfolding algorithm based on a purity matrix $\mathcal{P}$. Each element $\mathcal{P}_t^i$ in this matrix gives the probability that a particle of true type t is identified by the RICH as being of type i. Systematic uncertainties because of the RICH unfolding remain small for pions but can be as large as 4 - 5 % at high z for K^+ and even more than 50 % for the K^- spectrum at high z.

The HERA beam consisted of positrons during the year 2000 data taking period. Positrons however coming from e^+e^- pair production that ended up in the acceptance form a source of background. They were corrected for by subtracting the oppositely charged leptons in the DIS sample. This correction, however, was determined to be small and affects the multiplicities only by 1 - 1.5 %. A second identified source of background came from the decay of diffractively produced vector mesons like the ρ^0 or the ϕ mesons. For this the PYTHIA6 generator, adapted to HERMES kinematics and equipped with QED radiative corrections was taken because it has models in it for both fragmentation as well as diffraction. It was first checked that the PYTHIA6 generator adequately described the fragmentation data as well as the exclusive diffractive ρ^0 data. At low Q^2 and high z contaminations can reach up to 40 - 50 %, consisting mainly of pions coming from the decay of exclusive diffractive ρ^0 mesons. The contribution for kaons here remains well below 10 %. This correction was applied following the philosophy that the production process for these hadrons is totally different from quark fragmentation.

The obtained experimental hadron distributions were used as input for tuning a LEPTO + JETSET Monte Carlo production which was used to account for detector acceptance as well as experimental and radiative smearing. Significant progress was made in understanding the JETSET fragmentation model and applying it to HERMES kinematics. The obtained JETSET setting, in the form of the `2004c` tune, describes very well both the shape and the magnitude of the experimental multiplicity distributions for all individual hadron spectra. QED radiative corrections in this Monte Carlo simulation were handled by the RADGEN program. Each event was reweighted and true kinematics were calculated. In order to have a clear treatment of the uncertainties involved, we kept track of the smearing between the different kinematical bins by means of a smearing matrix $S(i,j) = n(i,j)/n^B(j)$, of which each element represents the flow of events from a Born level bin j to an experimental bin i. It is defined as the migration matrix $n(i,j)$ normalized to the Born level distribution $n^B(j)$. Since both of these quantities are extracted from a Monte Carlo simulation having the same model, the dependence on the input parameterization is reduced. A special bin ($j = 0$) represents events smeared from outside the Born level sample into the experimental sample. A clear asymmetric shape of the migration matrix due to QED radiative corrections was observed. Radiative smearing causes large migrations and works only in one direction of ν (and thus z). The change in kinematics due to detector smearing works in both directions, but is much smaller in size. The model dependence of this approach was checked by using different tunes for the JETSET fragmentation model in the Monte Carlo. As a complete scan of the LUND parameter space has not yet been accomplished to explore the vicinity of the χ^2 minimum, the variation between results based on the `2004a` and `2004c` tunes was taken as a measure for the systematic uncertainty coming from the fragmentation model in the Monte Carlo. After all, these two tunes are the two best descriptions available for HERMES. Normalizing to a 4π Born distribution $n^B(j)$, also the HERMES geometrical acceptance is taken into account.

The inherent Q^2 variation of the resulting multiplicity distributions was fixed at $Q_0^2 = 2.5$ GeV^2. Scaling up the obtained spectra to a Q_0^2 of 25 GeV^2, we were able to compare these results to previously published fragmentation functions by the EMC experiment. Although the energy scales of EMC and HERMES are very different, the obtained distributions match very well.

The obtained distributions were compared to previously extracted and published pion distributions from HERMES 96/97 data as well. Apart from the fact that this analysis provides a much higher statistical accuracy and a more extensive determination of the systematical effects involved, a small discrepancy between the two is visible at high z. This discrepancy becomes very obvious when examining the residual x dependence of these quantities. The 96/97 results showed a rather steep and unexplained slope versus x. The analysis presented in this work shows only a very mild dependence. First of all one should note that one in fact does expect a mild dependence as the equality between multiplicities and fragmentation functions does not completely hold because of the parton density functions. Possible QCD effects like gauge links between the initial and final quark propagator after the hard scattering interaction can play a role in this as well. The observed difference was thoroughly investigated. The causes for this discrepancy lie in the way radiative corrections were applied and the fact that the lepton detection efficiency was not properly taken into account in the old analysis.

The Q^2 dependence of the multiplicities in different z bins was compared to predictions provided by S. Kretzer [128] coming from the integration of the fragmentation functions and the parton density functions. Taken into account the fact that probably the theoretical uncertainties on the fragmentation function parameterizations are rather large at this low Q^2, the comparison works rather well. No evidence from this analysis could be found for any strong factorization violating effects. It is therefore a good assumption for the HERMES kinematical regime.

In the last chapter a systematic effect was investigated which comes from the $\cos\phi$ modulation of the semi-inclusive DIS cross section. This modulation was not modeled in the Monte Carlo used for the acceptance correction. Taking into account the transverse motion of quarks in the nucleon, sizeable $\langle\cos\phi\rangle_{UU}$ and $\langle\cos 2\phi\rangle_{UU}$ moments are expected. As the HERMES acceptance shape versus ϕ shows large cosine moments, which vary strongly versus x strong coupling between the acceptance and the intrinsic ϕ modulation was expected. By comparing ϕ distributions within the acceptance from the HERMES experimental data and a fully tracked Monte Carlo simulation without any moments at Born level, we extracted a set of $\langle\cos\phi\rangle_{UU}$ and $\langle\cos 2\phi\rangle_{UU}$ moments versus z, p_T and x for π^+, π^-, K^+ and K^-. These were in agreement with previous studies and were conceptually consistent with theory. The extraction in itself is also very important as the $\cos 2\phi$ moment is directly related to the Collins fragmentation function $H_1^\perp$. As binning is a crucial systematic effect in the extraction of these quantities, we used a set of extracted moments applying a multidimensional z, p_T and x binning to reweight a Monte Carlo simulation. This way a $\cos\phi$ and $\cos 2\phi$ modulation was explicitely introduced

at Born level. From the difference between the acceptance functions with and without this modulation in, we estimated the systematic effect on the multiplicity distributions. This effect was found to be very small versus z and well within the systematic errors assigned. Versus x is could can reach up to 10 % because of the strong coupling of the acceptance and the ϕ spectrum in x.

Chapter 9

Nederlandstalige samenvatting

Tot op heden hebben natuurkundigen quarks en leptonen geïdentificeerd als zijnde de kleinste, ondeelbare entiteiten waaruit alle gekende materie is opgebouwd. Het bekendste voorbeeld van een lepton is uiteraard het elektron, terwijl quarks de fundamentele bouwsteentjes zijn van protonen en neutronen waaruit atoomkernen zijn opgebouwd. Er zijn drie generaties quarks, elk bestaande uit een doublet. Het up (u) en down (d) quark zijn de lichtste met respectievelijk fractionele elektrische ladingen van $+\frac{2}{3}$ en $-\frac{1}{3}$ van de elementaire lading e en halftallige intrinsieke spin $\frac{1}{2}\hbar$. Dan volgen de charm (c), strange (s), top (t) en bottom (b) quarks. Samen met de 3 generaties leptonen (e^-, μ^- en τ^- en hun respectieve neutrinos) vormen de quarks het **standaard model** van de fysica der elementaire deeltjes. De zwaartekracht buiten beschouwing gelaten kunnen deze 12 deeltjes[1] interageren via een aantal wisselwerkingen die gedragen worden door zogenaamde ijkbosonen van de onderliggende kwantumveldentheorieën. Voor de elektromagnetische wisselwerking die ageert tussen elektrisch geladen deeltjes worden fotonen (γ), of 'lichtdeeltjes' uitgewisseld. Voor de zwakke kracht zijn dit de massieve Z^0 en $W^\pm$ ijkbosonen. De sterke wisselwerking wordt gedragen door gluonen (g), die in tegenstelling tot fotonen, wel de lading dragen waaraan ze koppelen. Gluonen zijn met andere woorden in staat te interageren met andere gluonen. Deze eigenschap heeft vérstrekkende gevolgen in de kwantumchromodynamica (QCD), de theorie die de wisselwerking tussen quarks en gluonen beschrijft. Quarks kunnen niet als vrije en ongebonden entiteiten in de natuur voorkomen. Ze zijn immer opgesloten in tripletten (baryonen) met elk een verschillende kleurlading of in doubletten van een quark en een antiquark. Dit is het zogenaamde *lange afstandsgedrag* van de quarks. Het gekke is dat het omgekeerde ook geldt. Naarmate de energieschaal toeneemt waarmee men quarks bestudeert, des te ongebondener ze worden. Men spreekt van *asymptotische vrijheid*. Deze eigenschappen worden in feite geparametriseerd door de variatie van de koppelingsconstante $\alpha_S(Q^2)$ van de sterke wisselwerking. De energieschaal Q^2 geeft een maat voor het oplossend vermogen waarmee we de interne structuur van protonen beschouwen.

[1] en natuurlijk ook hun anti-deeltjes.

Een onderdeel van het bestuderen van de dynamica van quarks en gluonen is het kijken naar de spinstructuur van het proton. Sinds de ontdekking van de quarks begin de jaren '70 van vorige eeuw, weet men dat het proton niet langer een elementair deeltje is. Zijn intrinsiek baanimpulsmoment, of *spin*, $\frac{1}{2}$ moet bijgevolg opgebouwd zijn uit de spins van de samenstellende delen. Dankzij het EMC experiment weet men dat de spin van het proton slechts voor een heel klein stukje gedragen wordt door de 3 valentiequarks (uud) die de overige kwantumgetallen van het proton bepalen. De aanwezige gluonen (met intrinsieke spin 1), de zeequarks[2] en het baanimpulsmoment van deze constituenten maken de totale spinpuzzel van het proton een complex gegeven. In 1995 zag HERMES als tweede generatie-experiment het levenslicht aan de HERA opslagring te DESY, Hamburg. Elektronen of positronen worden er versneld tot een energie van 27.6 GeV en ondergaan diep-inelastische verstrooiingsreacties met het trefgas in de targetcel van HERMES. Door het optreden van spontane polarisatie van de elektronbundel en de mogelijkheid tot polarisatie van het trefgas zijn de HERMES-fysici erin geslaagd stukjes van deze spinpuzzel bij elkaar te zoeken. Hoewel het ontrafelen van de nucleon spinstructuur het hoofddoel van het HERMES experiment is, staan uiteraard nog andere natuurkundige fenomenen in het daglicht.

We vermeldden reeds het lange afstandsgedrag van quarks en hun eigenschap dat ze niet als vrije deeltjes in de natuur kunnen voorkomen. Door de hoge energie waarmee een elektron interageert met de quarks in het proton wordt het getroffen quark echter weggekatapulteerd. Er ontstaat met andere woorden een zeer sterk veld tussen het getroffen quark en de rest van het proton. De sterkte van dit veld neemt lineair toe met de afstand waarmee het quark zich verwijdert ($\sim$ 1 GeV/fm). Omwille echter van *confinement* zullen quark-antiquark paren ontstaan in dit veld die de vrije kleurlading van het getroffen quark gaan neutraliseren. Er worden met andere woorden mesonen en baryonen gevormd. Dit proces noemt men **hadronizatie** of **fragmentatie**. Doordat de koppelingsconstante α_S groot wordt voor het lange afstandsgedrag van quarks is een perturbatieve aanpak van dit process, zoals gebruikelijk in kwantumveldentheorie, niet langer mogelijk. De reeksontwikkelingen in α_S divergeren. Daarom kunnen we het fragmentatieproces op dit moment enkel via fenomenologische modellen benaderen. Dit gebeurt enerzijds door Monte Carlo modellen waarbij het LUND *string-model* [104] en het *clusterfragmentatie-model* [46] de meest succesvolle en meest gebruikte zijn op dit ogenblik. Anderzijds kent men parametrisaties in de vorm van fragmentatie functies $D_f^h(z, Q^2)$ die een soort waarschijnlijkheid uitdrukken dat een quark van het type f zal hadronizeren tot een hadron h. De meest bekende parametrisaties hier zijn die van S. Kretzer [52] en de zogenaamde KKP [51], [53] parametrisatie. Fragmentatie functies zijn in eerste instantie afhankelijk van z, de longitudinale energie fractie die weggedragen wordt door het hadron. Net zoals de dichtheidsverdelingen $q_f(x, Q^2)$ van quarks in het proton, waarin x de impulsfractie, vertonen fragmentatie functies een zekere Q^2 afhankelijk-

[2] quark-antiquark paren die voortdurend gecreëerd en geannihileerd worden in het sterke kleurveld dat het proton samenhoudt.

heid. Q^2 is hier tevens de virtualiteit van het foton dat wordt uitgewisseld tussen het inkomend elektron en het getroffen quark.

Beschouwen we nu de uitdrukking

$$\frac{1}{\sigma_{DIS}}\frac{d\sigma^h(z,Q^2)}{dz} = \frac{\sum_f e_f^2 \int_0^1 dx\; q_f(x,Q^2) D_f^h(z,Q^2)}{\sum_f e_f^2 \int_0^1 dx\; q_f(x,Q^2)} \tag{9.1}$$

die we een hadron *multipliciteit* zullen noemen. De bepaling van deze multipliciteiten zal ons iets leren over het fragmentatieproces voor het HERMES kinematisch regime. Het is namelijk zo dat het HERMES experiment, met een interactie energie van $\sqrt{s} = 7.2$ GeV in het massacentrumstelsel en een gemiddelde virtualiteit van $Q^2 = 2.5$ GeV^2, zich bevindt in het gebied waar een perturbatieve QCD aanpak van de diep inelastische processen niet algemeen aanvaard is. Dit geldt ook voor de factorizatie tussen fragmentatie functies en parton dichtheidsfuncties van de semi-inclusieve werkzame doorsnede die vervat zit in uitdrukking 9.1. Een uitgebreide systematische studie van deze hadron multipliciteiten is daarom zeker aan de orde binnen HERMES en is dan ook het onderwerp van dit werk.

In eerste instantie werden experimentele hadron multipliciteitsdistributies uit de HERMES data bekomen. Daarvoor werd gebruik gemaakt van de ongepolariseerde dataset met waterstof als trefgas. Deze data werd genomen tijdens het jaar 2000. De HERMES experimentele opstelling bestaat zoals reeds vermeld uit een opslagcel waarin zich het trefgas bevindt. De voorwaartse spectrometer die na deze trefcel is opgesteld bestaat ondermeer uit een aantal driftkamers vóór en na een grote dipoolmagneet. De driftkamers zorgen voor het bepalen van ruimtecoördinaten om de sporen van de deeltjes doorheen de opstelling te kunnen reconstrueren. Uit de afbuiging in de dipoolmagneet kan men de impuls en de elektrische lading van de deeltjes achterhalen. Een transitie-stralingsdetector, een preshower-scintillatie detector en een calorimeter zorgen voor een goed onderscheid tussen de verstrooide leptonen en de geproduceerde hadronen. De calorimeter is verder verantwoordelijk voor het meten van de totale energie van de verstrooide elektronen. Een uniek kenmerk van het HERMES experiment is zijn mogelijkheid met grote precisie én over een groot impulsbereik een onderscheid te maken tussen pionen, kaonen en (anti-)protonen. Dit gebeurt met een zogenaamde *Ring Imaging Čerenkov*- of RICH-detector die aan de hand van kegelprojecties van Čerenkovlicht op een fotomultipliermatrix. Door een raytracing algoritme kan men de openingshoek van de Çerenkov kegel en dus de massa of het type van de deeltjes kan achterhalen.

De experimentele data werd eerst en vooral aan een uitgebreide kwaliteitscontrole onderworpen. Het is uiteraard belangrijk er ons van te vergewissen dat de opstelling naar behoren heeft gefunctioneerd voor de data die gebruikt wordt in de analyse. Er werden criteria aangewend in verband met dode tijd, luminositeit, doorslag in de hoogspanningsmodules enz... Niet alleen de kwaliteit van de detectoren zelf werd nagegaan, maar ook de reconstructie software is aan uitgebreide toetsen onderworpen.

Naast de datakwaliteitscontrole zijn ook een aantal kinematische selectiecriteria aan de orde. We vermelden hier als belangrijkste : $Q^2 > 1$ GeV^2 wat ons de zekerheid geeft dat we ons in een diep-inelastisch verstrooiingsregime bevinden. De virtualiteit van het uitgewisselde foton, onze probe, moet m.a.w. hoog genoeg zijn om de substructuur van het proton te ontwaren. $W^2 > 10$ GeV^2 zorgt ervoor dat er voldoende energie (W) is in de hadronische finale toestand voor fragmentatie en door $0.1 < y < 0.85$ blijven de kwantumelektrodynamische stralingscorrecties binnen de perken. y wordt in het laboratoriumstelsel gedefiniëerd als zijnde de fractie van de energie weggedragen door het virtueel foton t.o.v. het inkomend elektron (of positron). Een totaal van ongeveer 5.8 miljoen diep inelastische verstrooiingsevents werden uiteindelijk opgenomen in de analyse.

Het onderscheid tussen leptonen (het verstrooide elektron of positron) en hadronen (de producten van het fragmentatieproces) wordt gemaakt in HERMES door een Bayesiaans algoritme dat informatie van de 3 eerder vernoemde detectoren combineert. De analyse van de RICH-informatie gebeurt additioneel aan de hand van een ontvouwingsalgoritme dat corrigeert voor eventuele foute identificaties. Op basis van een detectorsimulatie van de RICH werd hiertoe een zogenaamde $\mathcal{P}$ matrix berekend waarin elk element $\mathcal{P}_t^i$ de waarschijnlijkheid weergeeft dat een deeltje van het werkelijke type t door de RICH geïdentificeerd wordt als zijnde van het type i. De elementen in deze matrix zijn sterk afhankelijk van de impuls van het hadron en de topologie van het evenement. Vooral voor de kaon spectra heeft deze correctie wel degelijk een niet te onderschatten impact. De systematische fouten die optreden door een niet perfecte beschrijving van de RICH detector in de Monte Carlo simulatie dragen voor de experimentele pion spectra bij tot ongeveer 4 - 5 % voor hoge z waarden. Voor het K^- meson zorgen deze systematische onzekerheden ervoor dat zijn spectrum bij hoge z veel van zijn significantie verliest.

Positronen afkomstig van e^+e^- paarproductie zijn uiteraard een vorm van achtergrond in het DIS ensemble. Rechtstreeks is er geen manier om deze achtergrond te bepalen. Tijdens het jaar 2000 was de HERA ring met positronen gevuld. Daar paarproductie symmetrisch is, weten we dat het aantal elektronen dat ons sample DIS events binnendringt een maat is voor het aantal positronen dat hetzelfde doet. Door de tegengesteld geladen DIS events m.a.w. negatief te tellen, kunnen we deze achtergrond bepalen. De correctie voor de multipliciteiten is echter heel klein ($\sim$ 1 - 1.5 %). Een tweede vorm van achtergrond vormen exclusieve diffractief geproduceerde vectormesonen. Daar de PYTHIA Monte Carlo generator zowel een diffractief model als het JETSET string-fragmentatie model omvat is het mogelijk via een Monte Carlo simulatie deze achtergrond te bepalen. Eerst werd hiertoe geverifiëerd dat de PYTHIA simulatie de diffractieve productie van vectormesonen wel degelijk goed beschrijft. Het is hiervoor nodig gebleken de PYTHIA simulatie te voorzien van QED stralingscorrecties. Bij lage Q^2 en hoge z kan de contributie van hadronen afkomstig van verval van dergelijke vectormesonen oplopen tot 40 - 50 %. In hoofdzaak pionen afkomstig van het ρ^0 verval dragen bij. Voor kaonen blijft de achtergrond ver beneden de 10 %.

De aldus verkregen hadrondistributies dienen vervolgens gecorrigeerd te worden voor migratie van events door QED straling en de niet perfecte resolutie van de experimentele opstelling. Ook moeten we de resultaten extrapoleren naar een 4π ruimtehoek en dus corrigeren voor de beperkte voorwaartse geometrie van de HERMES detector. Hiertoe werd een Monte Carlo simulatie aangewend die bestaat uit het LEPTO model voor de diep-inelastische verstrooiing en JETSET voor de fragmentatie. De experimentele opstelling werd door een GEANT3 simulatie beschreven. QED stralingscorrecties werden opnieuw door het RADGEN programma berekend. Het gebruikte fragmentatiemodel is van cruciaal belang in deze discussie. Na een intensieve afstemmingsprocedure werd een minimum in de parameterruimte van het LUND model gevonden in de vorm van de `2004c` tune die de experimentele HERMES multipliciteitsverdelingen heel dicht benadert. Om de consequenties van de migratie van events tussen verschillende kinematische intervallen ten volle in rekening te brengen, is ook hier een ontvouwingsalgoritme ter correctie toegepast. De migratie tussen de bins wordt geparametriseerd door een matrix $S(i,j)$ waarin elk element de migratie van events van een Born - niveau interval j naar een experimenteel interval i voorstelt. Deze migratiematrix wordt gedefiniëerd als de verhouding van twee grootheden komende uit dezelfde Monte Carlo simulatie, waardoor een zekere graad van modelafhankelijkheid wordt weggedeeld. Er is ook rekening gehouden met events die van buiten het Born ensemble in de acceptantie migreren. Doorheen de ontvouwing wordt deze intrinsieke 'achtergrond' apart behandeld (interval $j = 0$). De extractie van de migratiematrix $S(i,j)$ uit een Monte Carlo simulatie leert ons dat event migratie veroorzaakt door detectorresolutie heel klein is en in beide richtingen werkt, terwijl QED stralingscorrecties de migratiematrix een duidelijk asymmetrische vorm geven.

Door het aanwenden van verschillende JETSET parametrisaties die in meerdere of mindere mate de HERMES experimentele data beschrijven, werd de afhankelijkheid van de bekomen resultaten van het fragmentatiemodel in de Monte Carlo nagegaan. Een complete scan van de JETSET parameterruimte rond het χ^2 minimum is tot op heden nog niet mogelijk gebleken. De variatie van de resultaten ten gevolge van het verschil tussen de `2004c` en de `2004a` parametersets werd daarom als maat voor de systematische onzekerheid genomen. De ontvouwingscorrectie corrigeert zoals vermeld ook voor de detectorgeometrie. Met een opening van 80 mrad symmetrisch rond de bundelpijp is HERMES beperkt in zijn geometrische acceptantie. Zoals hierboven echter beschreven wordt de faseruimte voldoende bedekt om fysisch significante resultaten te bekomen. Uiteraard worden de statistische en systematische onzekerheden gepropageerd doorheen de ontvouwingsprocedure. Dit gebeurt aan de hand van de zogenaamde *radiative dilution* matrix die volledig bepaald is door kennis van de migratiematrix $S(i,j)$ en de Monte Carlo distributies. Additionele onzekerheden ten gevolge van de beperkte Monte Carlo statistiek zijn ook in rekening gebracht en zijn met brute kracht doorheen de ontvouwingsprocedure gepropageerd. Een Gaussisch ensemble van migratie matrices waarin elk element tegelijkertijd en onafhankelijk van de rest gesam-

pled werd, is hiertoe gegenereerd.

De inherente Q^2 variatie van de resulterende distributies werd gefixeerd op $Q_0^2 = 2.5$ GeV^2. De correctie hiertoe werd berekend aan de hand van de $D_f^h(z, Q^2)$ parametrisaties van S. Kretzer en de CTEQ6 parton dichtheidsverdelingen $q_f(x, Q^2)$. Deze correctie werd ook berekend voor een energie schaal van $Q_0^2 = 25$ GeV^2. We willen immers onze bekomen resultaten vergelijken met resultaten van experimenten bij hogere energie. Zo bestudeerde bijvoorbeeld het EMC experiment muon-proton verstrooing bij een muon bundelenergie van 280 GeV, m.a.w. een factor 10 hoger dan HERMES. De gepubliceerde u quark fragmentatie functies van het EMC experiment kunnen de vergelijking doorstaan met HERMES multipliciteitsverdelingen omwille van het feit dat het u quark heel dominant aanwezig is in het proton. De distributie functie wordt immers gewogen met de kwadratische lading van het u quark in rekening gebracht. De overeenkomst tussen de herschaalde HERMES resultaten en de gepubliceerde EMC resultaten is bijzonder goed.

Verder werden de verkregen resultaten ook vergeleken met eerder gepubliceerde pion distributies uit de HERMES 96/97 data. Behalve het feit dat de analyse die hier werd voorgesteld, een veel grotere statistische nauwkeurigheid vertoont samen met een uitgebreidere studie van systematische effecten, werd bij hoge z een kleine discrepantie tussen de twee resultaten vastgesteld. Het onderscheid werd pas echt duidelijk wanneer de data verdeeld werd in 4 z intervallen en in elk van deze intervallen de x afhankelijkheid bij vaste Q^2 werd nagegaan. Waar men in de 96/97 data een sterke x afhankelijkheid vaststelde, vertoonde deze analyse slechts een heel milde variatie. Residuele variaties van de fragmentatie functies versus x kunnen duiden op het niet voldaan zijn van de factorizatie hypothese die stelt dat de semi-inclusieve werkzame doorsnede kan geschreven worden als direct product van de fragmentatie functies die enkel van z en quark dichtheidsverdelingen die enkel van x afhankelijk zijn[3]. Hoewel we geen exact schaalgedrag van de hadron multipliciteitsverdelingen versus x verwachten aangezien de parton dichtheidsverdelingen niet volledig wegdelen, kan men toch stellen dat dit gedrag bij benadering vervuld moet zijn. Na uitgebreide studie werden dan ook enkele oorzaken van de discrepantie tussen de 96/97 en 2000 resultaten gevonden. We stelden vast dat de lepton detectie efficiëntie en het aanbrengen van QED stralingscorrecties cruciaal zijn en aan de basis liggen van het verschil.

De Q^2 afhankelijkheid zelf van de hadron distributies werd ook nagegaan en vergeleken met parametrisaties geleverd door S. Kretzer. De curves zijn geïntegreerd over de gebruikte z intervallen. De vergelijking gaat relatief goed op. We moeten echter de bedenking maken dat de gebruikte parametrisaties gefit werden aan data afkomstig van e^+e^- verstrooiing bij veel hogere energie en dat de evolutie naar lage Q^2 onderworpen is aan een zekere graad van theoretische onzekerheid. Toch kunnen we stellen dat uit deze analyse geen aanwijzingen gevonden werden voor een sterke schending van de factorizatie hypothese in

[3]Het schaalgedrag in Q^2 ten gevolge van de aanwezigheid van gluonen eventjes buiten beschouwing gelaten.

het HERMES energiebereik.

Additioneel werd in een laatste hoofdstuk een systematisch effect bestudeerd dat afkomstig is van de transversale beweging van quarks in het nucleon. Men verwacht namelijk een intrinsieke $\cos\phi$ en $\cos 2\phi$ modulatie van de semi-inclusieve werkzame doorsnede, waarbij ϕ de azimutale hoek is tussen het lepton verstrooiingsvlak en het vlak gevormd tussen het virtuele foton en het geproduceerde hadron. In de gebruikte Monte Carlo simulatie voor de 4π correctie was dit zogenaamde Cahn-effect echter niet in rekening gebracht. Daar de HERMES geometrische acceptantie grote cosinus momenten vertoont die nog eens sterk variëren met x, verwachten we een sterke koppeling tussen de intrinsieke $\langle\cos\phi\rangle_{UU}$ en $\langle\cos 2\phi\rangle_{UU}$ momenten en de momenten geïntroduceerd door de vorm van de acceptantie. Om de invloed hiervan na te gaan werd in eerste instantie een extractie gedaan op basis van data - Monte Carlo discrepantie van de $\cos\phi$ en $\cos 2\phi$ amplitudes in de Fourier reeks. Een 1-dimensionale extractie werd met succes vergeleken met resultaten verkregen in de *transversity* analyse in HERMES en conceptueel vergeleken met theoretische verwachtingen. Een vergelijking tussen twee onafhankelijke extractie methodes leverde compatibele resultaten op. Een multi-dimensionale analyse in z, p_T en x leverde ons vervolgens een meer precieze set momenten op waarmee extra gewichtsfactoren voor de Monte Carlo simulatie berekend werden. Op die manier werd op Born niveau expliciet een $\cos\phi$ en $\cos 2\phi$ modulatie ingevoerd. Na extensieve controles dat dit herschalingsproces van de hadron gewichten in de Monte Carlo correct werkt, werden de acceptantie functies met elkaar vergeleken. Uit de discrepantie haalden we een systematisch onzekerheid voor de multipliciteitsdistributies tengevolge van het gebrek aan een cosinus modulatie op Born niveau. Er werd niet gekozen voor een correctie factor daar de invloed van detectoruitlijning geacht wordt een belangrijke invloed uit te oefenen op deze cosinus momenten. De additionele systematisch onzekerheid ten gevolge van dit Cahn effect is echter heel klein als functie van z. Versus x kan het relatieve verschil in acceptantie functie tot 10 % bedragen om wille van de sterk correlatie tussen de cosinus modulaties en de acceptantie als functie van x.

Bibliography

[1] M.J. Alguard et al. Phys. Rev. Lett. **37**, 1261 (1976).

[2] J. Ashman et al. Phys. Lett. B **206**, 364 (1988).

[3] A. Airapetian et al. Phys. Lett. B **442**, 484–492 (1998).

[4] A. Airapetian et al. Phys. Rev. D **71**, 012003 (2005).

[5] W. Greiner and A. Schäfer. Springer-Verlag Institut für Theoretische Physik der Johann Wolfgang Goethe-Universität Frankfurt.

[6] Particle Data Group. *Review of Particle Physics*. Phys. Rev. D **66** (2002).

[7] C.G. Gallan Jr. and D.J. Gross. Phys. Rev. Lett. **22**, 156 (1969).

[8] A. Levy H. Abramowicz, E. Levin and U. Maor. *Phys. Lett. B.* **269**, 465 (1991).

[9] H. Abramowicz and A. Levy. *The ALLM parametrization of* $\sigma_{tot}(\gamma^* p)$ *an update*. **hep-ph/9712415** (1997).

[10] H.W. Kendall M. Breidenbach, J.I. Friedman. Phys. Rev. Lett. **23**, 935–939 (1969).

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

[12] J.D. Bjorken. Phys. Rev. **179**, 1547 (1969).

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

[14] M. Gell-Mann. Phys. Lett. **8**, 214 (1964).

[15] M. Goncharoc et al. Phys. Rev. D **64**, 112006 (2001).

[16] D. Mason. *New Strange Asymmetry Results from NuTeV*. hep-ph/0405037. for the NuTeV Collaboration.

[17] K. Gottfried. Phys. Rev. D **18**, 1174 (1967).

[18] G.T. Garvey and J.C. Peng. Prog. Part. Nucl. Phys. **47**, 203 (2001).

[19] R. Voss T. Sloan, G. Smadja. Phys. Rep. **162**, 45–167 (1988).

[20] Y.L. Dokshitzer. Sov. Phys.¬JETP **46**, 641 (1997).

[21] L.N. Lipatov V.N. Gribov. Sov. J. Nucl. Phys. **15**, 438 (1972).

[22] L.N. Lipatov. Sov. J. Nucl. Phys. **20**, 94 (1975).

[23] G. Parisi G. Altarelli. Nucl. Phys. B **126**, 298 (1977).

[24] G. Altarelli. Phys. Rep. **81**, 1–129 (1982).

[25] A.H. Mueller. Phys. Rev. D **9**, 963 (1974).

[26] R.K Ellis et al. Nucl. Phys. B **152**, 285–329 (1979).

[27] G. Sterman J.C. Collins, D.E. Soper. Adv. Ser. Direct. High Energy Phys. **5**, 1–91 (1988). hep-ph/0409313.

[28] J.J. Aubert et al. Phys. Lett. B **114**, 373–377 (1982).

[29] A.Airapetian et al. Eur.Phys.J **C21**, 599–606 (2001).

[30] R. D. Field and R. P. Feynman. Nucl. Phys. **B136**, 1 (1978).

[31] P. Hoyer, P. Osland, H. G. Sander, T. F. Walsh and P. M. Zerwas. Nucl. Phys. **B161**, 349 (1979).

[32] Ahmed Ali, J. G. Korner, G. Kramer and J. Willrodt. Nucl. Phys. **B168**, 409 (1980).

[33] Ahmed Ali, E. Pietarinen, G. Kramer and J. Willrodt. Phys. Lett. **B93**, 155 (1980).

[34] Frank E. Paige, Serban D. Protopescu, Howard Baer and Xerxes Tata. *ISAJET 7.69: A Monte Carlo event generator for p p, anti-p p, and e+ e- reactions*. (2003).

[35] X. Artru and G. Mennessier. Nucl. Phys. **B70**, 93–115 (1974).

[36] Bo Andersson, G. Gustafson and C. Peterson. Z. Phys. **C1**, 105 (1979).

[37] W. Bartel et al. JADE Collaboration. Phys. Lett. B **101**, 129 (1981).

[38] W. Bartel et al. JADE Collaboration. Phys. Lett. B **134**, 275 (1984).

[39] G. Gustafson B. Andersson and B. Söderberg. Z. f. Physik C **20**, 317–329 (1983).

[40] G. Gustafson B. Andersson and T. Sjöstrand. Phys. Scripta **32**, 574 (1985).

[41] M. Althoff et al. Z. f. Physik C **26**, 157–174 (1984).

[42] S. Bethke. Z. f. Physik C **29**, 175–181 (1985).

[43] J. Chrin. Z. f. Physik C **36**, 163–171 (1987).

[44] I. Schmitt C. Peterson, D. Schlatter and P.M. Zerwas. Phys. Rev. D **27**, 105–111 (1983).

[45] T. Sjöstrand. Z. f. Physik C **26**, 93–103 (1984).

[46] G. Marchesini et al. Computer Phys. Commun. **67**, 465 (1992).

[47] B.R. Webber. Nucl. Phys. B **238**, 492 (1984).

[48] B.R. Webber G. Marchesini. Nucl. Phys. B **238**, 1 (1984).

[49] D.A. Morris T.D. Gottschalk. Nucl. Phys. B **288**, 729 (1987).

[50] G. Kramer J. Binnewies, B.A. Kniehl. Phys. Rev. D **52**, 4947–4960 (1995).

[51] B. Pötter B.A. Kniehl, G. Kramer. Nucl. Phys. B **582**, 514–536 (2000).

[52] S. Kretzer. Phys. Rev. **D62**, 054001 (2000).

[53] G. Kramer S. Albino, B.A. Kniehl. Nucl. Phys. **B725**, 181–206 (2005).

[54] SLD Collaboration. Phys. Rev. D **59**, 052001 (1999).

[55] ALEPH Collaboration. Phys. Lett. B **357**, 487 (1995).

[56] ALEPH Collaboration. Phys. Lett. B **364**, 247 (1995).

[57] H. Aihara et.al. Phys. Rev. Lett. **61**, 1263–1266 (1988).

[58] OPAL Collaboration. Eur.Phys.J. C **16**, 407 (2000).

[59] P. Nason and B.R. Webber. Nucl. Phys. B **421**, 473 (1994).

[60] M.Arneodo et al. Nucl. Phys. B **321**, 541–560 (1989).

[61] E.L. Berger. Nucl. Phys. B **85**, 61 (1975).

[62] P. Mulders. AIP Conf. Proc. **588**, 75–88 (2001). ePrint: hep-ph/0010199, Talk given at 2nd Workshop on Physics with an Electron Polarized Light Ion Collider (EPIC 2000), Cambridge, Massachusetts, 14-16 Sep 2000.

[63] Mike Vetterli. *Interpretation of the HERMES semi-inclusive data.* 1995. Talk given at Collaboration Meeting.

[64] L. Trentadue and G. Veneziano. Phys. Lett. B **323**, 210–211 (1994).

[65] J.M. Niczyporuk and E.E.W. Bruins. Phys. Rev. D **58**, 091501 (1998).

[66] J.M. Niczyporuk and E.E.W. Bruins. *Flavor tagging at HERMES.* (1996). HERMES Internal Note 96-060.

[67] M. Arneodo et al. Z. f. Physik C **31**, 1 (1986).

[68] I.M Ternov A.A. Sokolov. Sov. Phys.–Doklady **8**, 1203 (1964).

[69] T. Benisch et al. Nucl. Inst. & Meth. **A471**, 314 (2001). Hermes Internal 01-078.

[70] U. Elschenbroich. *Analysis of Luminosity Monitor Data for Different Years.* Hermes Internal Note **02-013**.

[71] K. Ackerstaff et al. Nucl. Inst. & Meth. **A417**, 230–265 (1998).

[72] S. Bernreuther et al. Nucl. Inst. & Meth. **A416**, 45–58 (1998). Hermes Internal 98-006.

[73] J.M. Butler et al. Nucl. Inst. & Meth. **A290**, 122 (1990).

[74] A. Andreev et al. Nucl. Inst. & Meth. **A465**, 482 (2001).

[75] V.Vikhrov S.Belostotski, G.Gavrilov. *Magnet Chambers Status on fall of 2003 year.* Hermes Internal Note **03-043**.

[76] M.G. van Beuzekom et al. Nucl. Inst. & Meth. **A461**, 247–250 (2001).

[77] M.G. van Beuzekom et al. Nucl. Inst. & Meth. **A512**, 44–51 (2003).

[78] W. Wander. *Reconstruction of High Energy Scattering Events in the HERMES Experiment.* PhD thesis Friedrich Alexander Universitaet Erlangen-Nuernberg April 1996. HERMES Internal note 97-031.

[79] H. Avakian et al. Nucl. Inst. & Meth. **A417**, 69–78 (1998).

[80] H. Avakian et al. Nucl. Inst. & Meth. **A378**, 155 (1996). Hermes Internal 96-024.

[81] B. Maiheu and U. Elschenbroich. *The HERMES Trigger Hodoscope (H1) and Preshower detector (H2) : Experimenters Guide.* Hermes Internal.

[82] S.F. Pate. *Test of the H0 Forward Trigger Scintillators for HERMES.* Hermes Internal.

[83] J. Wendland. *Particle Identification for HERMES Run I.* Hermes Internal Note **01-067**.

[84] M. Amarian et al. *The HERMES Charm Upgrade Program: A Measurement of the Double Spin Asymmetry in Charm Leptoproduction.* Hermes Internal Note **97-004**. Proposal to the DESY PRC.

[85] E.C. Aschenauer et al. Nucl. Inst. & Meth. **A440**, 338 (2000). Hermes Internal 99-046.

[86] R. Kaiser D. De Schepper and E. Cisbani. *Particle Identification with the HERMES RICH Detector : Description of the Different Approaches*. Hermes Internal Note **98-008**.

[87] N. Akopov et al. Nucl. Inst. & Meth. pages 511–530 (2002).

[88] B. Hommmez. *A Study of Fragmentation Processes in the HERMES experiment using a Ring Imaging Cerenkov Detector*. PhD thesis Universiteit Gent July 2003. HERMES Internal 03-031.

[89] A. Maas. *The RICH PID Scheduler*. Hermes Internal Note **00-029**.

[90] ECP Division CERN PTG, Programming Techniques Group. *ADAMO Entity-Relationship Programming System, Version 3.3* 1993. CERN.

[91] F.M. Menden. *Determination of the Gluon Polarization in the Nucleon*. PhD thesis Albert-Ludwigs-Universität Freiburg December 2001. DESY THESIS 2001-060.

[92] Brad Filipone. *Hodoscope Logbook* 1999.

[93] F. James. *MINUIT, Function Minimization and Error Analysis*. Part of the CERNLIB libraries.

[94] H. Boettcher I. Akushevich and D. Ryckbosch. *RADGEN 1.0, Monte Carlo Generator for Radiative Events in DIS on Polarized and Unpolarized Targets*. hep-ph/9906408. Proceedings of the 'Monte Carlo Generators for HERA Physics' Workshop, Hamburg, Germany, 1999.

[95] R. Kaiser B. Hommez, H. Jackson and Y. Miyachi. *Using the RICH Detector for Physics Analysis*. Hermes Internal. See HERMES RICH group website.

[96] T. Hasegawa and Y. Miyachi. *Estimation of RICH particle identification accuracy using decaying particles*. Hermes Internal Note.

[97] R. Kaiser. *Unfolding of Hadron Distributions and Asymmetries*. Hermes Internal Note **00-027**.

[98] B. Hommez. *Tuning the RICH parameters in the Monte Carlo*. Hermes Internal Note **02-003**.

[99] Y. Miyachi. *Systematic uncertainties on RICH P-matrix*. Hermes Internal Note.

[100] B. Maiheu A. Hillenbrand and E.C. Aschenauer. *Extraction of pion and kaon multiplicities from HERMES 2000 data*. Hermes Release Report **March 2004**.

[101] `http://cernlib.web.cern.ch/cernlib/`.

[102] T. Sjöstrand et al. *PYTHIA 6.2 Physics and Manual*. Comp. Phys. Comm. **135**, 238 (2001).

[103] G. Ingelman et al. *LEPTO 6.5: A Monte Carlo generator for deep inelastic lepton - nucleon scattering*. Comp. Phys. Comm. **101**, 108–134 (1997).

[104] T. Sjöstrand et al. *PYTHIA 5.7 And JETSET 7.4*. Comp. Phys. Comm. **82**, 74 (1994).

[105] `http://wwwasd.web.cern.ch/wwwasd/geant/`.

[106] A. Hillenbrand. *Measurement and Simulation of the Fragmentation Process at HERMES*. PhD thesis Friedrich Alexander Universität Erlangen-Nuernberg.

[107] G. Ingelman A. Edin and J. Rathsman. *Unified description of rapidity gaps and energy flows in DIS final states*. Z. f. Physik C **75**, 57–70 (1997).

[108] W. Bartel et al. JADE Collab. Z. f. Physik C **33**, 23 (1986).

[109] S. Bethke et al. JADE Collab. Phys. Lett. B **213**, 235 (1988).

[110] E.C. Aschenauer G. Ingelman and A. Hillenbrand. *Private communication, phone conference 20.01.2003*.

[111] J. Rathsmann A. Edin, G. Ingelman. Phys. Lett. B **366**, 371 (1996).

[112] M. Arneodo et al. Phys. Lett. B **149**, 415 (1984).

[113] K. Long. *QCD at High Energy (Experiments)*. 2002. Plenary talk at the International Conference on High Energy Physics (ICHEP02), Amsterdam.

[114] T. Sjöstrand. *Private communication*.

[115] B. Maiheu A. Hillenbrand and E.C. Aschenauer. *Extraction of hadron multiplicities from the 2000 unpolarized data and Monte Carlo*. Hermes Release Report **April 2003**.

[116] B. Maiheu A. Hillenbrand and E.C. Aschenauer. *Extraction of pion and kaon multiplicities from HERMES 2000 data, Additional Studies en Cross Checks*. Hermes Release Report **June 2004**.

[117] A. Hillenbrand. *Update on Lund tuning and multiplicty extraction*. Talk given at the HERMES Collaboration Meeting, April 18th 2005.

[118] J. Pumplin et al. *New Generation of Parton Distributions with Uncertainties from Global QCD Analysis*. J. High Energy Phys. **JHEP07**, 012 (2002). Preprint: hep-ph/0201195.

[119] H. Abramowicz and A. Levy. *The ALLM parametrization of $\sigma_{tot}(\gamma^* p)$ - an update*. **hep-ph/9712415**. DESY 97-251.

[120] P. Liebing. *Can the gluon polarization in the nucleon be extracted from HERMES data on single high p_T hadrons ?* PhD thesis Universität Hamburg June 2004. HERMES Internal 04-025.

[121] A. Kotzinian. Phys. Lett. B **552**, 172–176 (2003).

[122] G. Alberi and G. Goggi. Phys. Rev. **74**, 1 (1981).

[123] M. Tytgat. *Diffractive Production of ρ^0 and ω Vector Mesons at HERMES.* PhD thesis Universiteit Gent March 2000. HERMES Internal 01-014.

[124] D.R. Yennie T.H. Bauer, R.D. Spital and F.M. Pipkin. Rev. Mod. Phys. **50, No.2** (1978).

[125] T. Sjöstrand G.A. Schuler. Phys. Rev. D **49**, 2257 (1994).

[126] A.C. Miller. *Applying Radiative Corrections to Ratios of Cross Sections for Deeply Inelastic SCattering.* HERMES Internal.

[127] M. Contalbrigo and D. Reggiani. *First g_1 unfolding results and cross-checks: input parametrizations and e.m. background (deuteron).* Talk given at the HERMES Analysis Week, December 2003.

[128] S. Kretzer and A. Hillenbrand. *private communication.*

[129] R.N. Cahn. Phys. Lett. B **78**, 269 (1978).

[130] R.N. Cahn. Phys. Rev. D **40**, 3107 (1989).

[131] P.M. Zerwas G. Köpp, R. Maciejko. Nucl. Phys. B **144**, 123 (1978).

[132] A. Mendez. Nucl. Phys. B **145**, 199 (1978).

[133] J. Levelt and P.J. Mulders. Phys. Rev. D **49**, 96–113 (1994).

[134] M. Anselmino et.al. Phys. Rev. D **71**, 074006 (2005).

[135] A. Prokudin. *The role of Cahn and Sivers effects in Deep Inelastic Scattering.* 2005. HERMES Collaboration Week, Hamburg 6.12.2004 - 10.12.2004.

[136] D. Boer and P. Mulders. Phys. Rev. D **57**, 5780 (1988).

[137] Oganessyan et.al. Phys. Lett. B **564**, 60 (2003).

[138] ZEUS Collaboration. Phys. Lett. B **481**, 199 (2000).

[139] E665 Collaboration. Phys. Rev. D **48**, 5057 (1993).

[140] M. Arneodo et al. (EMC Collaboration). Z. f. Physik C **34**, 277 (1987).

[141] B. Maiheu. *Monte Carlo study of the influence of possible moments on the HERMES acceptance function for hadrons.* (2005). HERMES Internal Note 05-004.

[142] U. Elschenbroich. *private communication*.

[143] U. Elschenbroich. *Transverse spin structure of the Proton Studied in Semi Inclusive DIS*. PhD thesis Universiteit Gent February 2006.

[144] A. Bacchetta. *Probing the transverse spin of quarks in deep inelastic scattering*. PhD thesis Vrije Universiteit Amsterdam Oktober 2002.

[145] A. Jgoun. *private communication*.

[146] A. Jgoun. *Measurement of Hadronic Azimuthal Distributions in Deep Inelastic Scattering*. (2003). HERMES Internal Note 03-001.

Acknowledgments
Dankwoordje

Vooreerst wens ik mijn promotor, Dirk Ryckbosch, van harte te danken voor de kans die hij me gegeven heeft. Gedurende mijn doctoraat heeft hij het perfecte kader gecreëerd om m'n eigen weg te kunnen volgen binnen HERMES. Ook om wille van zijn niet aflatend enthousiasme voor natuurkunde was het echt heel fijn samenwerken aan ondermeer de activiteiten rond het *World Year of Physics*. Op zijn verzoek bracht ik ook de eerste twee jaar van mijn doctoraat door in Hamburg. Deze internationale ervaring heeft me sterk beïnvloed op persoonlijk vlak en heeft me laten inzien dat de wereld veel groter is dan onze eigen achtertuin. Echt bedankt !

Arne en Michael, excuseer... Missel :-), de laatste twee jaar met jullie in het bureau waren echt fantastisch. Ik heb je echt wel gemist toen je weg was Michael, en dat zal nu niet anders zijn ! En Arne, wat wij tweetjes niet al beleefd hebben in naam der wetenschap (en ook daarbuiten)... Echt merci ! Ook de rest van de groep hier in Gent : Peter, Tiko, Yves, Brecht, Charlotte, Freija, Junior ... enorm bedankt voor de fijne sfeer op 't labo ! Ook jullie Klaas, Stijn, Veerle en alle andere mensen op het INW, van harte bedankt ! Robert, ik denk niet dat ik er ooit in zal slagen een derdemachtswortel te trekken uit een getal van 5 cijfers, maar in elk geval .... toch bedankt ;-). En luc... live long and prosper, resistance is futile !

The one person at DESY in Hamburg, who I am very much indebted to is Elke Aschenauer. I don't think I ever met anyone quite as spectacular as she. Be it leak-chasing the beam pipe at 5 AM on a Saturday morning, preparing the final plots for the release report or discussing in some meeting, passion is the word that always applies to her. Thanks for being there for me, Elke ! I'm also very grateful to Achim Hillenbrand with whom I spent countless evenings debugging and cross checking before we finally managed to put our release report together. Uli... Schätzchen ;-) ... we made the Blasenkammer rock ! Watching "Sesame street" will never be the same again. Thanks ! Gunar, you made me really appreciate hardware and electronics, I've learned a lot from you during my early Hamburg time. And Maurice... I will never forget one of the first times we sat down together over an HM problem. You simply

said "Oh, well, let's just change this and that in the source and recompile the whole bunch." That moment really impressed me quite a bit and learned me to just go ahead and do things and not be frightened by the magnitude of a task. Also, you were rather pleasant company... coming from Holland (couldn't resist... sorry ;-p). Thanks a lot ! Many thanks as well to the rest of the Hamburg crew for the 2 splendid years I had at DESY. Andreas, Larry, Beni, Ralf, Frank, Jim, Marc and all you other guyes I'm forgetting... let me say it this way... there's no meeting like off-site meeting !

Tanja, obwohl unsere Wege ganz unterschiedliche Wendungen genommen haben und sich wahrscheinlich nie wieder treffen werden, möchte ich dir für die schöne Zeit, die wir in Hamburg gehabt haben, danken.

Ma en pa, zussie, onnodig te zeggen dat zonder jullie ik hier nu geen dankwoord zou schrijven. Maar de steun die je me altijd gegeven hebt, het gevoel dat ik ondanks alles altijd welkom ben thuis... weet dat ik dit zeker niet vanzelfsprekend vind. Ik ben echt blij dat ik uit jullie nestje kom, enorm bedankt ! Zelfs al bevind ik me aan de andere kant van de wereld, dat gevoel draag ik altijd bij me.

Arne, Sarah, Phil, Val, Johanna en Lieven... Maak es iemand wijs dat 7 mensen die in Rome, Berlijn, Bridge of Allan, York, Hamburg en Gent leefden en elkaar kennen van tijdens hun natuurkundestudies, reeds 8 jaar beste maatjes zijn... Ondanks de afstand soms hecht ik enorm veel belang aan die vriendschap. Jullie weten ook als geen ander wat er zo allemaal door je heen gaat tijdens een doctoraat. Michiel, jouw muziektipjes hebben de heel vele uurtjes achter de pc van een fijne achtergrond voorzien !

Tom, enorm bedankt om me 'in te wijden' in de wereld van de djembé. Dankzij jou heb ik in Afrikaanse percussie echt wel mijn ding gevonden ! Ik hoop echt dat we nog veel eelt en blaren op onze handen zullen krijgen ! Lieve, Pascal, Ann, Wim, Fille, Nancy, Barbara, Bettina, Imka en Bert. Ik vergeet nooit de eerste keer dat ik met jullie op het podium gestaan heb, fantastisch ! Bedankt voor de vriendschap en bedankt om me erbij te nemen ! Het ganse djembé gebeuren is meer dan eens een enorme steun en uitlaatklep geweest tijdens mijn doctoraat.

Ellen, jouw spontaneïteit heeft me bij sommige beslissingen echt wel over de schreef geholpen. Elke, Hilde, jullie zijn in staat enorm veel liefde te geven, bedankt om daar iets van te delen met mij. En natuurlijk jij Nele... want de wereld zou zonder zo'n fantastische krulliekop eigenlijk maar een beetje saai zijn hé :-). Griet, Bart... ik kan me eigenlijk geen leven meer voorstellen zonder jullie. In al die jaren dat ik je ken ben je echt als een broer voor me geworden, Bart ! Het is fantastisch dat om te ervaren dat ik op eender welk moment op jullie kan rekenen ! En jij Eva, wat jij voor me betekende het afgelopen jaar, daar zijn gewoonweg geen woorden voor. Vriendschap heeft een gans nieuwe betekenis gekregen sinds ik jou ken. Bedankt dat je er bent !

Appendix A

Calculation of statistical uncertainties on multiplicity distributions

We want to calculate the uncertainty on a hadron multiplicity, which is essentially the number of hadrons found *within* a sample of DIS events divided by that total number of DIS events[1]. These two quantities are clearly not statistically independent, so we have to take into account their correlation. It is not so easy just to take the formula for fully correlated variables, because that would mean that the hadron sample is simply a subset of the DIS events. One DIS event can however contain more than one hadrons when we do not just simply take the leading hadron in each event. So how do we solve this problem ?

The fraction has to be written in terms of uncorrelated variables, we do this in the following way :

$$f = \frac{N_h}{N_{DIS}} = \frac{1 \cdot N_{DIS}^{1h} + 2 \cdot N_{DIS}^{2h} + 3 \cdot N_{DIS}^{3h} + 4 \cdot N_{DIS}^{4h} + ...}{N_{DIS}^{0h} + N_{DIS}^{1h} + N_{DIS}^{2h} + N_{DIS}^{3h} + N_{DIS}^{4h} + ...} = \frac{\sum_{i=0}^{\infty} i \cdot N_{DIS}^{ih}}{\sum_{i=0}^{\infty} N_{DIS}^{ih}} \qquad \text{(A.1)}$$

Where N_{DIS}^{ih} is simply the number of DIS events containing i hadrons. So the total number of hadrons in the DIS sample is just given by the nominator of the last equation. These variables are all uncorrelated, so we can write the uncertainty as

$$\sigma^2(f) = \sum_{j=0}^{\infty} \left(\frac{\partial f}{\partial N_{DIS}^{jh}} \right)^2 \cdot \sigma^2(N_{DIS}^{jh}) = \sum_{j=0}^{\infty} \left(\frac{\partial f}{\partial N_{DIS}^{jh}} \right)^2 \cdot N_{DIS}^{jh} \qquad \text{(A.2)}$$

One can easily calculate that

[1] Off course in reality one has to normalize with the bin with as well.

$$\frac{\partial f}{\partial N_{DIS}^{jh}} = \frac{j \cdot \sum_{i=0}^{\infty} N_{DIS}^{ih} - 1 \cdot \sum_{i=0}^{\infty} i \cdot N_{DIS}^{ih}}{\left(\sum_{i=0}^{\infty} N_{DIS}^{ih}\right)^2} \tag{A.3}$$

So that with replacing $\sum_{i=0}^{\infty} i \cdot N_{DIS}^{ih}$ by N_h and $\sum_{i=0}^{\infty} N_{DIS}^{ih}$ by N_{DIS} and putting some common terms to the front we get

$$\sigma^2(f) = \frac{N_h^2}{N_{DIS}^4} \sum_{j=0}^{\infty} \left(j \cdot \frac{N_{DIS}}{N_h} - 1 \right)^2 \cdot N_{DIS}^{jh} \tag{A.4}$$

If we work out the square of the difference and put the index - independent factors in front of the sum we get :

$$\sigma^2(f) = \frac{N_h^2}{N_{DIS}^4} \left[\left(\frac{N_{DIS}}{N_h} \right)^2 \sum_{j=0}^{\infty} j^2 \cdot N_{DIS}^{jh} - 2 \cdot \frac{N_{DIS}}{N_h} \sum_{j=0}^{\infty} j \cdot N_{DIS}^{jh} + \sum_{j=0}^{\infty} N_{DIS}^{jh} \right] \tag{A.5}$$

Then we see that we again can replace the sums in the two last terms and simplify the expression to (with taking the square root)

$$\sigma\left(\frac{N_h}{N_{DIS}}\right) = \frac{N_h}{N_{DIS}^2} \sqrt{\frac{N_{DIS}^2}{N_h^2} \sum_{j=0}^{\infty} j^2 \cdot N_{DIS}^{jh} - N_{DIS}} \tag{A.6}$$

As a final step we rewrite this a little just bringing one N_{DIS} out of the square root.

$$\boxed{\sigma\left(\frac{N_h}{N_{DIS}}\right) = \frac{N_h}{N_{DIS}} \sqrt{\frac{1}{N_h^2} \sum_{j=0}^{\infty} j^2 \cdot N_{DIS}^{jh} - \frac{1}{N_{DIS}}}} \tag{A.7}$$

Then one clearly sees that should we only take the leading hadron of each event, we get the fully correlated error formula back as expected. As an approximation we have just used this formula, this should not be too far off for two reasons. First one is that the number of pions per event found in HERMES is for the large majority of the events not more than 1. And secondly if we e.g. do a binning in z than it is rather unlikely you will have more than one pion from the same event in the same z bin. Hence we can safely use the formula for the fully correlated error to not over complicate things. Still we did need this exercise to come to this conclusion as it is not so obviously clear.

Appendix B

Data tables for multiplicity analysis

B.1 Experimental Data Multiplicities

Note that in the following tables we will denote a multiplicity with the symbol μ. Statistical errors will be written as $d\mu$ and systematic errors as $\delta\mu$. This section contains the experimental multiplicities obtained solely from the HERMES data. The values with and without RICH unfolding are presented.

$\langle z \rangle$	$\mu^{\pi^+}_{raw}$	$d\mu^{\pi^+}_{raw}$	$\mu^{\pi^+}_{unf}$	$d\mu^{\pi^+}_{unf}$	$d\mu^{\pi^+}_{unf}$ (%)	$\delta\mu^{\pi^+}_{unf}$	$\delta\mu^{\pi^+}_{unf}$ (%)
0.175	0.4846	0.0013	0.5137	0.0015	0.3	0.0063	1.2
0.225	0.4719	0.0013	0.4894	0.0014	0.3	0.0047	1.0
0.275	0.4067	0.0012	0.4137	0.0013	0.3	0.0027	0.7
0.325	0.3311	0.0011	0.3334	0.0011	0.3	0.0014	0.4
0.375	0.2608	0.0009	0.2609	0.0010	0.4	0.0006	0.2
0.425	0.2028	0.0008	0.2021	0.0009	0.4	0.0004	0.2
0.475	0.1593	0.0007	0.1585	0.0008	0.5	0.0003	0.2
0.525	0.1269	0.0007	0.1262	0.0007	0.5	0.0003	0.2
0.575	0.1023	0.0006	0.1016	0.0006	0.6	0.0003	0.2
0.625	0.0836	0.0005	0.0830	0.0005	0.7	0.0004	0.4
0.675	0.0692	0.0005	0.0689	0.0005	0.7	0.0005	0.7
0.725	0.0556	0.0004	0.0554	0.0004	0.8	0.0005	0.9
0.775	0.0449	0.0004	0.0448	0.0004	0.9	0.0004	1.0
0.825	0.0356	0.0004	0.0355	0.0003	1.0	0.0004	1.0
0.875	0.0280	0.0003	0.0280	0.0003	1.1	0.0004	1.3

Table B.1: Experimental π^+ multiplicities, both raw and after RICH unfoling correction. Also the the RICH systematical error is included in the last columns

$\langle z \rangle$	$\mu^{\pi^-}_{raw}$	$d\mu^{\pi^-}_{raw}$	$\mu^{\pi^-}_{unf}$	$d\mu^{\pi^-}_{unf}$	$d\mu^{\pi^-}_{unf}$ (%)	$\delta\mu^{\pi^-}_{unf}$	$\delta\mu^{\pi^-}_{unf}$ (%)
0.175	0.3762	0.0011	0.4097	0.0013	0.3	0.0050	1.2
0.225	0.3579	0.0011	0.3797	0.0012	0.3	0.0037	1.0
0.275	0.2960	0.0010	0.3078	0.0011	0.3	0.0021	0.7
0.325	0.2345	0.0009	0.2408	0.0009	0.4	0.0011	0.5
0.375	0.1779	0.0008	0.1811	0.0008	0.4	0.0006	0.3
0.425	0.1373	0.0007	0.1388	0.0007	0.5	0.0003	0.2
0.475	0.1057	0.0006	0.1066	0.0006	0.6	0.0002	0.2
0.525	0.0821	0.0005	0.0825	0.0005	0.6	0.0002	0.2
0.575	0.0652	0.0005	0.0655	0.0005	0.7	0.0003	0.4
0.625	0.0527	0.0004	0.0530	0.0004	0.8	0.0004	0.8
0.675	0.0421	0.0004	0.0424	0.0004	0.9	0.0005	1.1
0.725	0.0330	0.0003	0.0333	0.0003	1.0	0.0005	1.4
0.775	0.0274	0.0003	0.0277	0.0003	1.1	0.0004	1.6
0.825	0.0218	0.0003	0.0221	0.0003	1.3	0.0004	1.8
0.875	0.0176	0.0003	0.0179	0.0003	1.4	0.0003	1.9

Table B.2: Experimental π^- multiplicities, both raw and after RICH unfoling correction. Also the the RICH systematical error is included in the last columns

$\langle z \rangle$	$\mu^{K^+}_{raw}$	$d\mu^{K^+}_{raw}$	$\mu^{K^+}_{unf}$	$d\mu^{K^+}_{unf}$	$d\mu^{K^+}_{unf}$ (%)	$\delta\mu^{K^+}_{unf}$	$\delta\mu^{K^+}_{unf}$ (%)
0.175	0.1015	0.0006	0.0796	0.0010	1.3	0.0048	6.1
0.225	0.0927	0.0006	0.0757	0.0009	1.1	0.0041	5.5
0.275	0.0808	0.0005	0.0700	0.0007	1.0	0.0028	4.1
0.325	0.0696	0.0005	0.0645	0.0006	1.0	0.0018	2.9
0.375	0.0575	0.0004	0.0551	0.0006	1.0	0.0011	2.1
0.425	0.0476	0.0004	0.0471	0.0005	1.0	0.0005	1.2
0.475	0.0385	0.0004	0.0398	0.0004	1.1	0.0002	0.7
0.525	0.0307	0.0003	0.0327	0.0004	1.2	0.0002	0.8
0.575	0.0256	0.0003	0.0279	0.0004	1.3	0.0002	0.8
0.625	0.0206	0.0003	0.0229	0.0003	1.5	0.0003	1.5
0.675	0.0163	0.0002	0.0182	0.0003	1.6	0.0004	2.4
0.725	0.0130	0.0002	0.0147	0.0003	1.8	0.0004	3.0
0.775	0.00965	0.00018	0.0110	0.0002	2.1	0.0004	3.6
0.825	0.00746	0.00016	0.0086	0.0002	2.4	0.0003	4.0
0.875	0.00524	0.00013	0.00601	0.00018	3.0	0.00034	5.7

Table B.3: Experimental K^+ multiplicities, both raw and after RICH unfoling correction. Also the the RICH systematical error is included in the last columns

$\langle z \rangle$	$\mu^{K^-}_{raw}$	$d\mu^{K^-}_{raw}$	$\mu^{K^-}_{unf}$	$d\mu^{K^-}_{unf}$	$d\mu^{K^-}_{unf}$ (%)	$\delta\mu^{K^-}_{unf}$	$\delta\mu^{K^-}_{unf}$ (%)
0.175	0.0595	0.0005	0.0419	0.0008	1.8	0.0042	10.0
0.225	0.0497	0.0004	0.0391	0.0006	1.6	0.0034	8.8
0.275	0.0384	0.0004	0.0337	0.0005	1.5	0.0022	6.8
0.325	0.0294	0.0003	0.0278	0.0004	1.5	0.0014	5.1
0.375	0.0219	0.0003	0.0217	0.0003	1.6	0.0007	3.6
0.425	0.0167	0.0002	0.0173	0.0003	1.7	0.0003	2.2
0.475	0.0120	0.0002	0.0129	0.0003	1.9	0.0002	1.7
0.525	0.00944	0.00018	0.0104	0.0002	2.1	0.0001	1.7
0.575	0.00663	0.00015	0.00744	0.00019	2.6	0.00027	3.7
0.625	0.00477	0.00013	0.00518	0.00016	3.2	0.00040	7.6
0.675	0.00319	0.00010	0.00335	0.00014	4.1	0.00046	13.6
0.725	0.00209	0.00008	0.00204	0.00011	5.5	0.00048	23.5
0.775	0.00129	0.00007	0.00117	0.00009	7.9	0.00046	39.5
0.825	0.00082	0.00005	0.00060	0.00007	12.0	0.00042	69.4
0.875	0.00046	0.00004	0.00025	0.00006	23.1	0.00036	145.6

Table B.4: Experimental K^- multiplicities, both raw and after RICH unfoling correction. Also the the RICH systematical error is included in the last columns

B.2 Diffractive background $\sigma^h_{diff}/\sigma^h_{tot}$

In this section the diffractive background ratios for π^+, π^-, K^+ and K^- that were obtained from the PYTHIA6 Monte Carlo simulation are presented as a function of z.

$\langle z \rangle$	$C^{\pi^+}_{vmd}$	$dC^{\pi^+}_{vmd}$	$\delta C^{\pi^+}_{vmd}$	$C^{\pi^-}_{vmd}$	$dC^{\pi^-}_{vmd}$	$\delta C^{\pi^-}_{vmd}$
0.175	0.0325	0.0002	0.0003	0.0377	0.0003	0.0007
0.225	0.0350	0.0003	0.0014	0.0418	0.0003	0.0012
0.275	0.0375	0.0003	0.0012	0.0458	0.0004	0.0016
0.325	0.0412	0.0004	0.0019	0.0520	0.0005	0.0030
0.375	0.0461	0.0004	0.0026	0.0580	0.0005	0.0033
0.425	0.0517	0.0005	0.0030	0.0676	0.0007	0.0042
0.475	0.0587	0.0006	0.0023	0.0816	0.0008	0.0045
0.525	0.0703	0.0007	0.0018	0.1017	0.0010	0.0055
0.575	0.0877	0.0009	0.0004	0.1308	0.0013	0.0093
0.625	0.1114	0.0011	0.0004	0.1701	0.0017	0.0079
0.675	0.1479	0.0014	0.0051	0.232	0.002	0.012
0.725	0.1933	0.0018	0.0071	0.307	0.003	0.015
0.775	0.251	0.002	0.012	0.401	0.004	0.017
0.825	0.332	0.003	0.013	0.519	0.005	0.026
0.875	0.424	0.004	0.010	0.631	0.007	0.039

Table B.5: Diffractive background fractions $C^h_{vmd} = \sigma^h_{diffractive}/\sigma^h_{tot}$ for pions versus z. dC Represents the error due to limited Monte Carlo statistics, δC represents the estimation of the systematics. Results obtained from PYTHIA using `ISUB`=91,92 and 93.

$\langle z \rangle$	$C^{K^+}_{vmd}$	$dC^{K^+}_{vmd}$	$\delta C^{K^+}_{vmd}$	$C^{K^-}_{vmd}$	$dC^{K^-}_{vmd}$	$\delta C^{K^-}_{vmd}$
0.175	0.0110	0.0004	0.0016	0.0130	0.0005	0.0010
0.225	0.0113	0.0004	0.0007	0.0143	0.0005	0.0005
0.275	0.0125	0.0004	0.0018	0.0177	0.0006	0.0016
0.325	0.0164	0.0005	0.0015	0.0227	0.0008	0.0033
0.375	0.0225	0.0007	0.0055	0.0371	0.0011	0.0050
0.425	0.0288	0.0008	0.0061	0.0486	0.0014	0.0081
0.475	0.0296	0.0009	0.0088	0.0583	0.0017	0.0092
0.525	0.0306	0.0010	0.0106	0.0553	0.0018	0.0094
0.575	0.0294	0.0010	0.0108	0.060	0.002	0.008
0.625	0.0202	0.0009	0.0071	0.0403	0.0019	0.0144
0.675	0.0139	0.0008	0.0085	0.032	0.002	0.005
0.725	0.0151	0.0009	0.0092	0.042	0.003	0.001
0.775	0.0162	0.0011	0.0078	0.059	0.004	0.005
0.825	0.0176	0.0014	0.0194	0.069	0.005	0.003
0.875	0.0187	0.0018	0.0214	0.079	0.008	0.023

Table B.6: Diffractive background fractions $C^h_{vmd} = \sigma^h_{diffractive}/\sigma^h_{tot}$ for kaons versus z. dC Represents the error due to limited Monte Carlo statistics, δC represents the estimation of the systematics. Results obtained from PYTHIA using `ISUB`=91,92 and 93.

B.3 Born level multiplicity distributions versus z without Q^2 correction

π^+ **multiplicity versus** z ***(no diffr. correction)***					
$\langle z \rangle$	μ^{π^+}	$d\mu^{\pi^+}$	$\delta\mu^{\pi^+}_{RICH}$	$\delta\mu^{\pi^+}_{MC}$	$\delta\mu^{\pi^+}_{FF}$
0.175	3.130	0.012	0.048	0.009	0.026
0.225	2.192	0.008	0.028	0.007	0.013
0.275	1.607	0.007	0.014	0.006	0.004
0.325	1.208	0.006	0.007	0.006	0.008
0.375	0.906	0.006	0.003	0.005	0.004
0.425	0.697	0.005	0.002	0.005	0.017
0.475	0.528	0.005	0.002	0.005	0.011
0.525	0.428	0.005	0.002	0.004	0.013
0.575	0.346	0.005	0.002	0.005	0.015
0.625	0.280	0.005	0.003	0.005	0.009
0.675	0.251	0.006	0.005	0.006	0.020
0.725	0.200	0.006	0.006	0.007	0.006
0.775	0.190	0.007	0.007	0.008	0.023
0.825	0.100	0.007	0.007	0.008	0.015
0.875	0.161	0.004	0.004	0.006	0.010

Table B.7: Born level π^+ multiplicity, no VMD correcion applied. The systematic uncertainties respectively stand for the contribution from the RICH, propagated through the smearing and acceptance correction ($\delta\mu_{RICH}$), the systematic contribution due to the limited statistics in the Monte Carlo used for this smearing correction ($\delta\mu_{MC}$) and the assigned uncertainty due to the fragmentation model in the Monte Carlo ($\delta\mu_{FF}$).

π^+ **multiplicity versus** z ***(diffr. background subtracted)***					
$\langle z \rangle$	μ^{π^+}	$d\mu^{\pi^+}$	$\delta\mu^{\pi^+}_{EXP}$	$\delta\mu^{\pi^+}_{MC}$	$\delta\mu^{\pi^+}_{FF}$
0.175	3.027	0.012	0.048	0.008	0.023
0.225	2.116	0.008	0.028	0.007	0.011
0.275	1.549	0.007	0.015	0.006	0.005
0.325	1.161	0.006	0.008	0.005	0.009
0.375	0.867	0.006	0.005	0.005	0.005
0.425	0.664	0.005	0.004	0.005	0.017
0.475	0.502	0.005	0.003	0.004	0.011
0.525	0.403	0.005	0.002	0.004	0.012
0.575	0.321	0.005	0.002	0.004	0.015
0.625	0.256	0.005	0.004	0.004	0.009
0.675	0.219	0.006	0.007	0.005	0.018
0.725	0.170	0.007	0.009	0.005	0.007
0.775	0.138	0.007	0.011	0.005	0.016
0.825	0.088	0.007	0.011	0.005	0.009
0.875	0.062	0.005	0.006	0.003	0.008

Table B.8: Born level π^+ multiplicity, with diffractive background subtracted. The uncertainties have the same meaning as in the non-vmd corrected case, with the execption that $\delta\mu_{EXP}$ stands for the propagated total uncertainty on the experimental multiplicities, so due to the RICH **as well as** the diffractive background correction.

π^- **multiplicity versus** z ***(no diffr. correction)***					
$\langle z \rangle$	μ^{π^-}	$d\mu^{\pi^-}$	$\delta\mu^{\pi^-}_{RICH}$	$\delta\mu^{\pi^-}_{MC}$	$\delta\mu^{\pi^-}_{FF}$
0.175	2.490	0.010	0.038	0.008	0.028
0.225	1.735	0.007	0.022	0.006	0.001
0.275	1.210	0.006	0.011	0.005	0.000
0.325	0.886	0.005	0.006	0.005	0.005
0.375	0.633	0.005	0.003	0.004	0.000
0.425	0.485	0.004	0.001	0.004	0.008
0.475	0.360	0.004	0.001	0.004	0.001
0.525	0.283	0.004	0.001	0.004	0.006
0.575	0.220	0.004	0.002	0.004	0.005
0.625	0.183	0.004	0.003	0.004	0.006
0.675	0.150	0.004	0.005	0.004	0.006
0.725	0.112	0.005	0.006	0.005	0.007
0.775	0.103	0.005	0.006	0.006	0.000
0.825	0.069	0.005	0.006	0.007	0.001
0.875	0.117	0.003	0.004	0.005	0.010

Table B.9: Born level π^- multiplicity, no VMD correcion applied. The systematic uncertainties respectively stand for the contribution from the RICH, propagated through the smearing and acceptance correction ($\delta\mu_{RICH}$), the systematic contribution due to the limited statistics in the Monte Carlo used for this smearing correction ($\delta\mu_{MC}$) and the assigned uncertainty due to the fragmentation model in the Monte Carlo ($\delta\mu_{FF}$).

π^- **multiplicity versus** z ***(diffr. background subtracted)***					
$\langle z \rangle$	μ^{π^-}	$d\mu^{\pi^-}$	$\delta\mu^{\pi^-}_{EXP}$	$\delta\mu^{\pi^-}_{MC}$	$\delta\mu^{\pi^-}_{FF}$
0.175	2.396	0.010	0.038	0.007	0.026
0.225	1.664	0.007	0.022	0.006	0.000
0.275	1.157	0.006	0.012	0.005	0.001
0.325	0.842	0.005	0.007	0.004	0.005
0.375	0.600	0.005	0.004	0.004	0.001
0.425	0.457	0.004	0.003	0.004	0.008
0.475	0.336	0.004	0.003	0.003	0.002
0.525	0.260	0.004	0.003	0.003	0.006
0.575	0.197	0.004	0.005	0.003	0.006
0.625	0.160	0.004	0.005	0.003	0.006
0.675	0.118	0.004	0.007	0.003	0.005
0.725	0.085	0.005	0.009	0.003	0.006
0.775	0.061	0.005	0.010	0.003	0.000
0.825	0.045	0.006	0.013	0.003	0.001
0.875	0.021	0.004	0.010	0.002	0.008

Table B.10: Born level π^- multiplicity, with diffractive background subtracted. The uncertainties have the same meaning as in the non-vmd corrected case, with the execption that $\delta\mu_{EXP}$ stands for the propagated total uncertainty on the experimental multiplicities, so due to the RICH **as well as** the diffractive background correction.

K^+ **multiplicity versus** z ***(no diffr. correction)***					
$\langle z \rangle$	μ^{K^+}	$d\mu^{K^+}$	$\delta\mu^{K^+}_{RICH}$	$\delta\mu^{K^+}_{MC}$	$\delta\mu^{K^+}_{FF}$
0.175	0.582	0.009	0.043	0.004	0.039
0.225	0.370	0.006	0.027	0.003	0.014
0.275	0.276	0.004	0.016	0.002	0.015
0.325	0.239	0.004	0.010	0.002	0.007
0.375	0.192	0.003	0.006	0.002	0.000
0.425	0.166	0.003	0.003	0.002	0.001
0.475	0.140	0.003	0.001	0.002	0.002
0.525	0.114	0.003	0.001	0.002	0.004
0.575	0.101	0.003	0.002	0.002	0.003
0.625	0.086	0.003	0.003	0.002	0.003
0.675	0.067	0.004	0.005	0.002	0.001
0.725	0.062	0.004	0.006	0.002	0.000
0.775	0.029	0.004	0.006	0.003	0.007
0.825	0.037	0.004	0.006	0.003	0.001
0.875	0.026	0.002	0.003	0.002	0.005

Table B.11: Born level K^+ multiplicity, no VMD correcion applied. The systematic uncertainties respectively stand for the contribution from the RICH, propagated through the smearing and acceptance correction ($\delta\mu_{RICH}$), the systematic contribution due to the limited statistics in the Monte Carlo used for this smearing correction ($\delta\mu_{MC}$) and the assigned uncertainty due to the fragmentation model in the Monte Carlo ($\delta\mu_{FF}$).

K^+ **multiplicity versus** z ***(diffr. background subtracted)***					
$\langle z \rangle$	μ^{K^+}	$d\mu^{K^+}$	$\delta\mu^{K^+}_{EXP}$	$\delta\mu^{K^+}_{MC}$	$\delta\mu^{K^+}_{FF}$
0.175	0.576	0.009	0.043	0.004	0.038
0.225	0.366	0.006	0.027	0.003	0.014
0.275	0.272	0.004	0.016	0.002	0.015
0.325	0.235	0.004	0.010	0.002	0.006
0.375	0.188	0.003	0.006	0.002	0.000
0.425	0.161	0.003	0.004	0.002	0.001
0.475	0.136	0.003	0.003	0.002	0.002
0.525	0.111	0.003	0.003	0.002	0.004
0.575	0.097	0.003	0.003	0.002	0.003
0.625	0.084	0.003	0.004	0.002	0.003
0.675	0.067	0.004	0.005	0.002	0.001
0.725	0.061	0.004	0.006	0.002	0.000
0.775	0.028	0.004	0.006	0.003	0.007
0.825	0.037	0.004	0.006	0.003	0.000
0.875	0.025	0.002	0.003	0.001	0.004

Table B.12: Born level K^+ multiplicity, with diffractive background subtracted. The uncertainties have the same meaning as in the non-vmd corrected case, with the execption that $\delta\mu_{EXP}$ stands for the propagated total uncertainty on the experimental multiplicities, so due to the RICH **as well as** the diffractive background correction.

K^- **multiplicity versus** z ***(no diffr. correction)***					
$\langle z \rangle$	μ^{K^-}	$d\mu^{K^-}$	$\delta\mu^{K^-}_{RICH}$	$\delta\mu^{K^-}_{MC}$	$\delta\mu^{K^-}_{FF}$
0.175	0.322	0.007	0.039	0.002	0.003
0.225	0.204	0.004	0.023	0.002	0.002
0.275	0.143	0.003	0.013	0.001	0.003
0.325	0.106	0.002	0.007	0.001	0.004
0.375	0.0791	0.0019	0.0044	0.0014	0.0018
0.425	0.0640	0.0018	0.0023	0.0013	0.0002
0.475	0.0451	0.0016	0.0014	0.0011	0.0000
0.525	0.0390	0.0016	0.0014	0.0011	0.0003
0.575	0.0277	0.0014	0.0022	0.0009	0.0003
0.625	0.0192	0.0015	0.0040	0.0009	0.0007
0.675	0.0142	0.0015	0.0053	0.0008	0.0005
0.725	0.0081	0.0015	0.0064	0.0007	0.0000
0.775	0.0047	0.0016	0.0083	0.0007	0.0005
0.825	0.0031	0.0013	0.0076	0.0006	0.0002
0.875	0.0007	0.0006	0.0035	0.0003	0.0007

Table B.13: Born level K^- multiplicity, no VMD correcion applied. The systematic uncertainties respectively stand for the contribution from the RICH, propagated through the smearing and acceptance correction ($\delta\mu_{RICH}$), the systematic contribution due to the limited statistics in the Monte Carlo used for this smearing correction ($\delta\mu_{MC}$) and the assigned uncertainty due to the fragmentation model in the Monte Carlo ($\delta\mu_{FF}$).

K^- **multiplicity versus** z ***(diffr. background subtracted)***					
$\langle z \rangle$	μ^{K^-}	$d\mu^{K^-}$	$\delta\mu^{K^-}_{EXP}$	$\delta\mu^{K^-}_{MC}$	$\delta\mu^{K^-}_{FF}$
0.175	0.318	0.007	0.039	0.002	0.003
0.225	0.201	0.004	0.023	0.002	0.002
0.275	0.140	0.003	0.013	0.001	0.003
0.325	0.104	0.002	0.007	0.001	0.004
0.375	0.0764	0.0019	0.0045	0.0013	0.0017
0.425	0.0610	0.0018	0.0024	0.0012	0.0001
0.475	0.0423	0.0016	0.0016	0.0011	0.0000
0.525	0.0370	0.0016	0.0015	0.0010	0.0003
0.575	0.0257	0.0014	0.0023	0.0009	0.0003
0.625	0.0185	0.0015	0.0040	0.0009	0.0006
0.675	0.0138	0.0015	0.0053	0.0008	0.0005
0.725	0.0078	0.0015	0.0064	0.0007	0.0000
0.775	0.0043	0.0016	0.0083	0.0007	0.0004
0.825	0.0030	0.0013	0.0076	0.0006	0.0002
0.875	0.0008	0.0006	0.0035	0.0003	0.0007

Table B.14: Born level K^- multiplicity, with diffractive background subtracted. The uncertainties have the same meaning as in the non-vmd corrected case, with the execption that $\delta\mu_{EXP}$ stands for the propagated total uncertainty on the experimental multiplicities, so due to the RICH **as well as** the diffractive background correction.

B.4 Q^2 scaling correction factors

Q^2 **scaling correction vs** z**,** $Q_0^2 = 2.5$ **GeV**^{2}						
$\langle z \rangle$	$C_{Q^2}^{\pi^+}$	$dC_{Q^2}^{\pi^+}$	$\delta C_{Q^2}^{\pi^+}$	$C_{Q^2}^{\pi^-}$	$dC_{Q^2}^{\pi^-}$	$\delta C_{Q^2}^{\pi^-}$
0.175	1.000212	0.000012	0.000505	1.000700	0.000016	0.001449
0.225	0.999977	0.000008	0.000072	0.999833	0.000007	0.001030
0.275	1.00001	0.00003	0.00001	0.99906	0.00003	0.00086
0.325	0.99996	0.00005	0.00001	0.99827	0.00006	0.00076
0.375	0.99974	0.00009	0.00003	0.99687	0.00010	0.00065
0.425	0.99908	0.00013	0.00003	0.99551	0.00015	0.00062
0.475	0.99852	0.00017	0.00004	0.9936	0.0002	0.0006
0.525	0.9977	0.0002	0.0001	0.9913	0.0003	0.0008
0.575	0.9970	0.0003	0.0001	0.9897	0.0004	0.0008
0.625	0.9958	0.0003	0.0002	0.9877	0.0005	0.0010
0.675	0.9951	0.0005	0.0003	0.9847	0.0006	0.0011
0.725	0.9954	0.0006	0.0003	0.9837	0.0007	0.0011
0.775	0.9959	0.0008	0.0003	0.9797	0.0010	0.0013
0.825	0.9933	0.0010	0.0005	0.9827	0.0012	0.0010
0.875	0.9988	0.0013	0.0001	0.9935	0.0016	0.0003
$\langle z \rangle$	$C_{Q^2}^{K^+}$	$dC_{Q^2}^{K^+}$	$\delta C_{Q^2}^{K^+}$	$C_{Q^2}^{K^-}$	$dC_{Q^2}^{K^-}$	$\delta C_{Q^2}^{K^-}$
0.175	1.0028	0.0002	0.0019	1.0038	0.0003	0.0024
0.225	1.0006	0.0002	0.0004	1.0036	0.0003	0.0020
0.275	1.00000	0.00017	0.00000	1.0030	0.0003	0.0019
0.325	0.99976	0.00013	0.00021	1.0034	0.0002	0.0027
0.375	0.99994	0.00003	0.00023	1.00179	0.00011	0.00288
0.425	1.00002	0.00008	0.00002	1.00001	0.00002	0.00296
0.475	0.99979	0.00019	0.00010	0.99810	0.00016	0.00303
0.525	0.9989	0.0004	0.0002	0.9940	0.0004	0.0030
0.575	0.9982	0.0005	0.0002	0.9898	0.0006	0.0031
0.625	0.9966	0.0007	0.0003	0.9844	0.0009	0.0029
0.675	0.9956	0.0009	0.0004	0.9793	0.0012	0.0026
0.725	0.9977	0.0011	0.0002	0.9743	0.0015	0.0030
0.775	0.9985	0.0015	0.0001	0.973	0.002	0.003
0.825	0.9983	0.0019	0.0001	0.972	0.003	0.003
0.875	1.006	0.003	0.0006	0.964	0.004	0.006

Table B.15: Q^2 scaling correction factors versus z for $\pi^\pm$ and $K^\pm$ with $Q_0^2 = 2.5$ GeV^2. The uncertainties quoted are statistical ($dC_{Q^2}^h$) as well as systematical ($\delta C_{Q^2}^h$).

Q^2 **scaling correction vs** z**,** $Q_0^2 = 25$ **GeV**^{2}						
$\langle z \rangle$	$C_{Q^2}^{\pi^+}$	$dC_{Q^2}^{\pi^+}$	$\delta C_{Q^2}^{\pi^+}$	$C_{Q^2}^{\pi^-}$	$dC_{Q^2}^{\pi^-}$	$\delta C_{Q^2}^{\pi^-}$
0.175	0.949601	0.000012	0.062191	0.953786	0.000015	0.064943
0.225	0.887923	0.000007	0.035673	0.891196	0.000006	0.036526
0.275	0.84461	0.00003	0.02284	0.84658	0.00003	0.02151
0.325	0.81368	0.00004	0.01779	0.81442	0.00005	0.01393
0.375	0.77244	0.00007	0.01587	0.77168	0.00008	0.00758
0.425	0.74148	0.00010	0.01669	0.73990	0.00011	0.00502
0.475	0.71816	0.00012	0.01802	0.71584	0.00015	0.00393
0.525	0.68478	0.00016	0.01994	0.68226	0.00019	0.00278
0.575	0.66265	0.00019	0.02107	0.6604	0.0002	0.0024
0.625	0.6340	0.0002	0.0222	0.6328	0.0003	0.0029
0.675	0.5988	0.0003	0.0233	0.5981	0.0004	0.0039
0.725	0.5739	0.0003	0.0237	0.5741	0.0004	0.0043
0.775	0.5329	0.0004	0.0238	0.5334	0.0005	0.0048
0.825	0.5008	0.0005	0.0234	0.5063	0.0006	0.0049
0.875	0.4461	0.0006	0.0217	0.4580	0.0008	0.0048
$\langle z \rangle$	$C_{Q^2}^{K^+}$	$dC_{Q^2}^{K^+}$	$\delta C_{Q^2}^{K^+}$	$C_{Q^2}^{K^-}$	$dC_{Q^2}^{K^-}$	$\delta C_{Q^2}^{K^-}$
0.175	1.1596	0.0003	0.1144	1.1718	0.0003	0.1210
0.225	1.0585	0.0002	0.0576	1.0780	0.0003	0.0667
0.275	0.98814	0.00017	0.03101	1.0109	0.0003	0.0449
0.325	0.93714	0.00012	0.01833	0.9620	0.0002	0.0385
0.375	0.86668	0.00003	0.01064	0.89151	0.00009	0.03490
0.425	0.81325	0.00007	0.01079	0.83664	0.00002	0.03274
0.475	0.77334	0.00014	0.01206	0.79530	0.00012	0.03054
0.525	0.7180	0.0003	0.0146	0.7375	0.0003	0.0267
0.575	0.6842	0.0003	0.0163	0.7016	0.0004	0.0245
0.625	0.6444	0.0004	0.0183	0.6611	0.0006	0.0224
0.675	0.6005	0.0006	0.0202	0.6198	0.0008	0.0223
0.725	0.5742	0.0007	0.0210	0.5949	0.0009	0.0240
0.775	0.5333	0.0008	0.0220	0.5659	0.0012	0.0286
0.825	0.5046	0.0010	0.0224	0.5477	0.0014	0.0327
0.875	0.4601	0.0014	0.0232	0.515	0.002	0.039

Table B.16: Q^2 scaling correction factors versus z for $\pi^\pm$ and $K^\pm$ with $Q_0^2 = 25$ GeV^2. The uncertainties quoted are statistical ($dC_{Q^2}^h$) as well as systematical ($\delta C_{Q^2}^h$).

Q^2 **scaling correction vs** x**, for pions** $Q_0^2 = 2.5$ **GeV**^{2}						
$\langle z \rangle$	$C_{Q^2}^{\pi^+}$	$dC_{Q^2}^{\pi^+}$	$\delta C_{Q^2}^{\pi^+}$	$C_{Q^2}^{\pi^-}$	$dC_{Q^2}^{\pi^-}$	$\delta C_{Q^2}^{\pi^-}$
0	0.94049	0.00012	0.01998	0.94097	0.00011	0.01975
1	0.95110	0.00010	0.01501	0.95131	0.00010	0.01461
2	0.96031	0.00009	0.01174	0.96089	0.00009	0.01116
3	0.97148	0.00014	0.00794	0.97306	0.00014	0.00748
4	0.99353	0.00006	0.00248	0.99465	0.00006	0.00281
5	1.02046	0.00005	0.00513	1.01986	0.00005	0.00573
6	1.05782	0.00003	0.01181	1.05615	0.00003	0.01167
7	1.09646	0.00002	0.01737	1.09450	0.00002	0.01633
8	1.13476	0.00002	0.02192	1.13177	0.00002	0.01975
9	0.90266	0.00019	0.01010	0.90142	0.00019	0.00505
10	0.91999	0.00017	0.00859	0.91838	0.00018	0.00400
11	0.93454	0.00016	0.00756	0.93310	0.00016	0.00354
12	0.94864	0.00019	0.00605	0.94926	0.00019	0.00268
13	0.98386	0.00012	0.00106	0.98521	0.00012	0.00176
14	1.03096	0.00008	0.00212	1.02995	0.00008	0.00228
15	1.08922	0.00005	0.00737	1.08759	0.00005	0.00388
16	1.14936	0.00003	0.01274	1.14824	0.00003	0.00566
17	1.20818	0.00002	0.01806	1.20644	0.00002	0.00737
18	0.8643	0.0003	0.0107	0.8615	0.0003	0.0052
19	0.8872	0.0002	0.0098	0.8839	0.0002	0.0044
20	0.9045	0.0002	0.0090	0.9025	0.0002	0.0042
21	0.9196	0.0002	0.0084	0.9191	0.0002	0.0042
22	0.9682	0.0002	0.0021	0.96738	0.00018	0.00197
23	1.03785	0.00012	0.00245	1.03662	0.00012	0.00198
24	1.12053	0.00008	0.00990	1.12067	0.00008	0.00293
25	1.20765	0.00005	0.01795	1.20588	0.00005	0.00350
26	1.29253	0.00003	0.02609	1.29260	0.00003	0.00395
27	0.8172	0.0004	0.0139	0.8145	0.0004	0.0094
28	0.8449	0.0003	0.0128	0.8434	0.0003	0.0080
29	0.8649	0.0003	0.0119	0.8627	0.0003	0.0075
30	0.8857	0.0003	0.0111	0.8849	0.0003	0.0071
31	0.9436	0.0003	0.0051	0.9442	0.0003	0.0036
32	1.04117	0.00017	0.00343	1.04267	0.00017	0.00266
33	1.15804	0.00011	0.01445	1.15878	0.00011	0.00606
34	1.28324	0.00007	0.02658	1.28124	0.00007	0.00804
35	1.40962	0.00005	0.03943	1.40459	0.00005	0.00878

Table B.17: Q^2 scaling correction factors versus x for $\pi^{\pm}$ with $Q_0^2 = 2.5$ GeV^2. The uncertainties quoted are statistical ($dC_{Q^2}^h$) as well as systematical ($\delta C_{Q^2}^h$). The binning indices represent the binning used in figure 6.8.

Q^2 **scaling correction vs** x**, for kaons** $Q_0^2 = 2.5$ **GeV**^{2}						
$\langle z \rangle$	$C_{Q^2}^{K^+}$	$dC_{Q^2}^{K^+}$	$\delta C_{Q^2}^{K^+}$	$C_{Q^2}^{K^-}$	$dC_{Q^2}^{K^-}$	$\delta C_{Q^2}^{K^-}$
0	1.005017	0.000017	0.029138	1.011184	0.000016	0.033564
1	1.004430	0.000005	0.020255	1.010596	0.000005	0.024160
2	1.004945	0.000005	0.014243	1.011124	0.000005	0.017365
3	1.00900	0.00004	0.01138	1.01568	0.00004	0.01526
4	1.00520	0.00006	0.00400	1.00595	0.00009	0.00419
5	0.991050	0.000002	0.009487	0.984952	0.000008	0.013403
6	0.993885	0.000005	0.015812	0.983343	0.000012	0.021911
7	1.001175	0.000005	0.019980	0.987086	0.000003	0.028250
8	1.011474	0.000005	0.022756	0.994540	0.000003	0.033065
9	0.94778	0.00012	0.00599	0.95582	0.00012	0.01490
10	0.95849	0.00011	0.00392	0.96592	0.00011	0.01243
11	0.96845	0.00010	0.00328	0.97620	0.00010	0.01084
12	0.98193	0.00013	0.00541	0.98805	0.00015	0.01172
13	0.99858	0.00003	0.00377	1.001989	0.000007	0.005672
14	1.01211	0.00004	0.00629	1.00575	0.00003	0.01184
15	1.04370	0.00003	0.00915	1.03166	0.00002	0.02053
16	1.07794	0.00002	0.01108	1.061314	0.000017	0.027447
17	1.11300	0.00002	0.01264	1.090939	0.000012	0.033049
18	0.88847	0.00014	0.01012	0.8979	0.0002	0.0164
19	0.90894	0.00017	0.00866	0.91703	0.00019	0.01358
20	0.92331	0.00019	0.00754	0.93017	0.00018	0.01167
21	0.94063	0.00019	0.00615	0.9481	0.0002	0.0094
22	0.97989	0.00014	0.00337	0.98592	0.00011	0.00660
23	1.03134	0.00010	0.00404	1.02484	0.00008	0.00850
24	1.10004	0.00006	0.00901	1.08728	0.00006	0.01812
25	1.17162	0.00004	0.01419	1.15062	0.00004	0.02636
26	1.24279	0.00003	0.01907	1.21637	0.00002	0.03416
27	0.8212	0.0004	0.0199	0.8400	0.0003	0.0246
28	0.8508	0.0003	0.0160	0.8653	0.0003	0.0189
29	0.8688	0.0003	0.0141	0.8829	0.0003	0.0153
30	0.8860	0.0003	0.0122	0.8972	0.0003	0.0125
31	0.9409	0.0003	0.0057	0.9522	0.0002	0.0058
32	1.03937	0.00017	0.00350	1.03381	0.00015	0.00436
33	1.15443	0.00011	0.01396	1.14553	0.00010	0.01540
34	1.27333	0.00007	0.02440	1.25090	0.00006	0.02538
35	1.40030	0.00005	0.03528	1.36564	0.00004	0.03605

Table B.18: Q^2 scaling correction factors versus x for $K^{\pm}$ with $Q_0^2 = 2.5$ GeV^2. The uncertainties quoted are statistical ($dC_{Q^2}^h$) as well as systematical ($\delta C_{Q^2}^h$). The binning indices represent the binning used in figure 6.8.

Q^2 **scaling correction vs** x**, for pions** $Q_0^2 = 25$ **GeV**^{2}						
$\langle z \rangle$	$C_{Q^2}^{\pi^+}$	$dC_{Q^2}^{\pi^+}$	$\delta C_{Q^2}^{\pi^+}$	$C_{Q^2}^{\pi^-}$	$dC_{Q^2}^{\pi^-}$	$\delta C_{Q^2}^{\pi^-}$
0	0.78314	0.00009	0.03624	0.78587	0.00009	0.03330
1	0.79197	0.00008	0.03228	0.79450	0.00009	0.02906
2	0.79964	0.00008	0.02969	0.80250	0.00008	0.02621
3	0.80819	0.00011	0.02655	0.81189	0.00012	0.02333
4	0.82654	0.00005	0.02195	0.82990	0.00005	0.01959
5	0.84895	0.00004	0.01696	0.85093	0.00004	0.01318
6	0.88002	0.00003	0.01198	0.88121	0.00003	0.00862
7	0.91217	0.00002	0.00802	0.91321	0.00002	0.00530
8	0.94403	0.00002	0.00488	0.94431	0.00002	0.00304
9	0.68898	0.00015	0.02219	0.68922	0.00016	0.00927
10	0.70166	0.00013	0.02131	0.70162	0.00014	0.00834
11	0.71276	0.00012	0.02073	0.71287	0.00013	0.00784
12	0.72351	0.00015	0.01982	0.72522	0.00016	0.00740
13	0.75037	0.00009	0.01645	0.75268	0.00009	0.00729
14	0.78629	0.00006	0.01519	0.78686	0.00006	0.00545
15	0.83072	0.00004	0.01209	0.83089	0.00004	0.00435
16	0.87660	0.00003	0.00898	0.87723	0.00003	0.00332
17	0.92145	0.00002	0.00588	0.92169	0.00002	0.00234
18	0.60119	0.00019	0.02393	0.60065	0.00019	0.00449
19	0.61612	0.00017	0.02384	0.61529	0.00017	0.00414
20	0.62814	0.00016	0.02367	0.62822	0.00016	0.00393
21	0.63767	0.00015	0.02360	0.63886	0.00015	0.00386
22	0.67084	0.00013	0.02042	0.67188	0.00013	0.00306
23	0.71963	0.00008	0.01908	0.72050	0.00008	0.00378
24	0.77696	0.00005	0.01574	0.77893	0.00005	0.00398
25	0.83673	0.00004	0.01210	0.83753	0.00003	0.00378
26	0.89555	0.00002	0.00825	0.89775	0.00002	0.00299
27	0.49995	0.00019	0.02665	0.50248	0.00019	0.00775
28	0.51690	0.00019	0.02676	0.52029	0.00019	0.00716
29	0.52862	0.00018	0.02682	0.53170	0.00019	0.00698
30	0.54088	0.00018	0.02691	0.54495	0.00018	0.00689
31	0.57518	0.00016	0.02483	0.58044	0.00015	0.00474
32	0.63460	0.00011	0.02259	0.64092	0.00011	0.00309
33	0.70583	0.00007	0.01901	0.71229	0.00007	0.00299
34	0.78214	0.00005	0.01490	0.78757	0.00004	0.00324
35	0.85840	0.00003	0.01036	0.86264	0.00003	0.00294

Table B.19: Q^2 scaling correction factors versus x for $\pi^{\pm}$ with $Q_0^2 = 25$ GeV^2. The uncertainties quoted are statistical ($dC_{Q^2}^h$) as well as systematical ($\delta C_{Q^2}^h$). The binning indices represent the binning used in figure 6.8.

Q^2 **scaling correction vs** x**, for kaons** $Q_0^2 = 25$ **GeV**^{2}						
$\langle z \rangle$	$C_{Q^2}^{K^+}$	$dC_{Q^2}^{K^+}$	$\delta C_{Q^2}^{K^+}$	$C_{Q^2}^{K^-}$	$dC_{Q^2}^{K^-}$	$\delta C_{Q^2}^{K^-}$
0	0.973498	0.000017	0.051566	1.000147	0.000016	0.068851
1	0.972929	0.000005	0.043860	0.999566	0.000005	0.061599
2	0.973427	0.000005	0.03883	1.000088	0.000005	0.056533
3	0.97612	0.00004	0.03639	1.00337	0.00004	0.05552
4	0.97245	0.00005	0.02870	0.99376	0.00009	0.04383
5	0.958755	0.000002	0.016077	0.973014	0.000008	0.025301
6	0.961498	0.000005	0.009528	0.971425	0.000002	0.016523
7	0.968549	0.000005	0.005496	0.975122	0.000003	0.010257
8	0.978513	0.000004	0.003008	0.982486	0.000003	0.005613
9	0.80601	0.00010	0.01014	0.83505	0.00011	0.04181
10	0.81415	0.00009	0.01097	0.84287	0.00009	0.04108
11	0.82260	0.00008	0.01183	0.85184	0.00008	0.040550
12	0.83405	0.00011	0.01435	0.86218	0.00013	0.04196
13	0.84819	0.00002	0.01333	0.874344	0.000006	0.037065
14	0.85969	0.00004	0.00590	0.87763	0.00003	0.02164
15	0.88652	0.00003	0.00417	0.90023	0.00002	0.01477
16	0.91561	0.00002	0.00296	0.926114	0.000015	0.009555
17	0.94538	0.00002	0.00178	0.951962	0.000011	0.005441
18	0.65172	0.00018	0.01953	0.67958	0.00018	0.03565
19	0.66515	0.00015	0.01887	0.69242	0.00015	0.03382
20	0.67567	0.00014	0.01828	0.70233	0.00014	0.03268
21	0.68673	0.00017	0.01761	0.71424	0.00018	0.03134
22	0.71457	0.00010	0.01546	0.74186	0.00008	0.03020
23	0.75296	0.00007	0.01263	0.77203	0.00007	0.02000
24	0.80312	0.00005	0.00977	0.81907	0.00004	0.01436
25	0.85438	0.00003	0.00702	0.86579	0.00003	0.00975
26	0.90628	0.00002	0.00441	0.915268	0.000018	0.005541
27	0.50507	0.00019	0.02762	0.5402	0.0002	0.0333
28	0.52323	0.00019	0.02618	0.55655	0.00018	0.03013
29	0.53375	0.00018	0.02552	0.56731	0.00017	0.02806
30	0.54369	0.00017	0.02487	0.57594	0.00017	0.02643
31	0.57608	0.00015	0.02210	0.61014	0.00015	0.02376
32	0.63635	0.00010	0.01869	0.66237	0.00010	0.01921
33	0.70680	0.00007	0.01475	0.73395	0.00006	0.01479
34	0.77959	0.00005	0.01091	0.80147	0.00004	0.01077
35	0.85640	0.00003	0.00701	0.87416	0.00003	0.00642

Table B.20: Q^2 scaling correction factors versus x for $K^{\pm}$ with $Q_0^2 = 25$ GeV^2. The uncertainties quoted are statistical ($dC_{Q^2}^h$) as well as systematical ($\delta C_{Q^2}^h$). The binning indices represent the binning used in figure 6.8.

Appendix C

Data tables for $\langle \cos\phi \rangle_{UU}$ and $\langle \cos 2\phi \rangle_{UU}$ extraction

$\langle z \rangle$	$\langle \cos\phi \rangle_{UU}$	$\sigma_{\langle \cos\phi \rangle_{UU}}$	$\langle \cos 2\phi \rangle_{UU}$	$\sigma_{\langle \cos 2\phi \rangle_{UU}}$
moments for π^+				
0.25	-0.0165	0.0019	0.005	0.002
0.35	-0.013	0.002	0.018	0.002
0.45	-0.043	0.003	0.020	0.003
0.55	-0.108	0.005	0.008	0.004
0.65	-0.180	0.005	0.009	0.005
0.75	-0.219	0.006	-0.027	0.006
moments for π^-				
0.25	-0.017	0.002	0.010	0.002
0.35	-0.008	0.003	0.032	0.003
0.45	-0.040	0.004	0.040	0.004
0.55	-0.093	0.005	0.055	0.005
0.65	-0.160	0.006	0.059	0.006
0.75	-0.176	0.008	0.025	0.008

Table C.1: Extracted $\langle \cos\phi \rangle_{UU}$ and $\langle \cos 2\phi \rangle_{UU}$ moments versus z using the subtractive method for π^+ and π^-. In this extraction we integrated over all other kinematical variables. Note that on both data and Monte Carlo the hadron momentum was restricted to $2 < p < 25$ GeV.

$\langle z\rangle$	$\langle\cos\phi\rangle_{UU}$	$\sigma_{\langle\cos\phi\rangle_{UU}}$	$\langle\cos 2\phi\rangle_{UU}$	$\sigma_{\langle\cos 2\phi\rangle_{UU}}$
moments for K^+				
0.25	0.087	0.005	0.010	0.005
0.35	0.069	0.005	0.002	0.005
0.45	0.014	0.006	-0.020	0.006
0.55	-0.060	0.007	-0.037	0.008
0.65	-0.129	0.009	-0.052	0.009
0.75	-0.130	0.011	-0.060	0.012
moments for K^-				
0.25	0.056	0.006	0.010	0.006
0.35	0.045	0.008	-0.005	0.008
0.45	0.056	0.010	-0.011	0.010
0.55	-0.005	0.013	0.013	0.014
0.65	-0.108	0.018	0.013	0.019
0.75	-0.15	0.03	-0.097	0.03

Table C.2: Extracted $\langle\cos\phi\rangle_{UU}$ and $\langle\cos 2\phi\rangle_{UU}$ moments versus z using the subtractive method for K^+ and K^-. In this extraction we integrated over all other kinematical variables. Note that on both data and Monte Carlo the hadron momentum was restricted to $2 < p < 25$ GeV.

$\langle p_T\rangle$	$\langle\cos\phi\rangle_{UU}$	$\sigma_{\langle\cos\phi\rangle_{UU}}$	$\langle\cos 2\phi\rangle_{UU}$	$\sigma_{\langle\cos 2\phi\rangle_{UU}}$
moments for π^+				
0.10	-0.022	0.002	-0.001	0.002
0.30	-0.0457	0.0015	0.0023	0.0015
0.50	-0.073	0.002	0.009	0.002
0.70	-0.096	0.003	-0.001	0.003
0.90	-0.126	0.006	-0.022	0.006
1.10	-0.145	0.012	-0.064	0.012
moments for π^-				
0.10	-0.010	0.002	0.008	0.003
0.30	-0.0295	0.0018	0.0201	0.0018
0.50	-0.063	0.002	0.030	0.002
0.70	-0.080	0.004	0.034	0.004
0.90	-0.085	0.008	0.030	0.007
1.10	-0.134	0.015	0.000	0.015

Table C.3: Extracted $\langle\cos\phi\rangle_{UU}$ and $\langle\cos 2\phi\rangle_{UU}$ moments versus p_T using the subtractive method for π^+ and π^-. In this extraction we integrated over all other kinematical variables. Note that on both data and Monte Carlo the hadron momentum was restricted to $2 < p < 25$ GeV.

$\langle p_T \rangle$	$\langle \cos\phi \rangle_{UU}$	$\sigma_{\langle \cos\phi \rangle_{UU}}$	$\langle \cos 2\phi \rangle_{UU}$	$\sigma_{\langle \cos 2\phi \rangle_{UU}}$
moments for K^+				
0.10	-0.029	0.005	-0.002	0.005
0.30	0.040	0.004	0.007	0.004
0.50	0.104	0.004	0.014	0.005
0.70	0.042	0.007	0.004	0.007
0.90	-0.038	0.013	-0.079	0.013
1.10	-0.07	0.02	-0.10	0.02
moments for K^-				
0.10	-0.017	0.007	-0.008	0.007
0.30	0.063	0.005	0.017	0.005
0.50	0.138	0.007	0.059	0.007
0.70	0.040	0.011	0.023	0.011
0.90	-0.066	0.019	-0.055	0.019
1.10	-0.11	0.04	-0.07	0.04

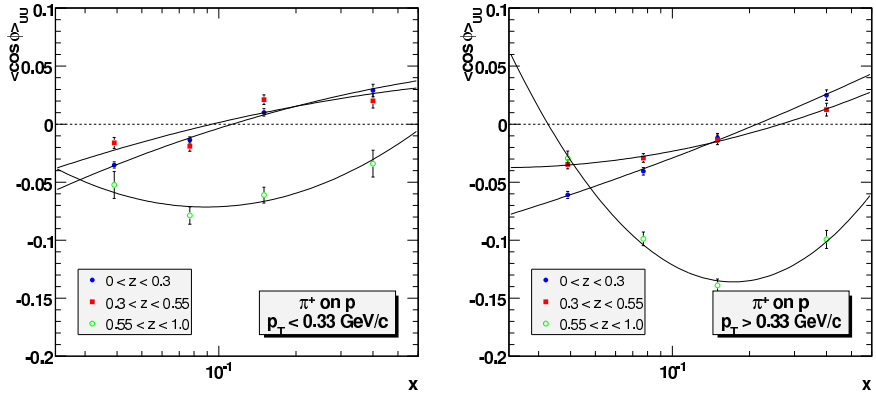
Table C.4: Extracted $\langle \cos\phi \rangle_{UU}$ and $\langle \cos 2\phi \rangle_{UU}$ moments versus p_T using the subtractive method for K^+ and K^-. In this extraction we integrated over all other kinematical variables. Note that on both data and Monte Carlo the hadron momentum was restricted to $2 < p < 25$ GeV.

$\langle x \rangle$	$\langle \cos\phi \rangle_{UU}$	$\sigma_{\langle \cos\phi \rangle_{UU}}$	$\langle \cos 2\phi \rangle_{UU}$	$\sigma_{\langle \cos 2\phi \rangle_{UU}}$
moments for π^+				
0.032	-0.068	0.003	0.015	0.003
0.048	-0.057	0.002	0.005	0.003
0.066	-0.041	0.002	-0.001	0.002
0.088	-0.043	0.002	-0.002	0.003
0.150	-0.0237	0.0018	-0.0066	0.0019
0.400	0.011	0.003	-0.004	0.003
moments for π^-				
0.032	-0.064	0.003	0.021	0.003
0.048	-0.045	0.003	0.022	0.003
0.066	-0.033	0.003	0.017	0.003
0.088	-0.025	0.003	0.016	0.003
0.150	-0.004	0.002	0.014	0.002
0.400	0.045	0.004	0.032	0.004

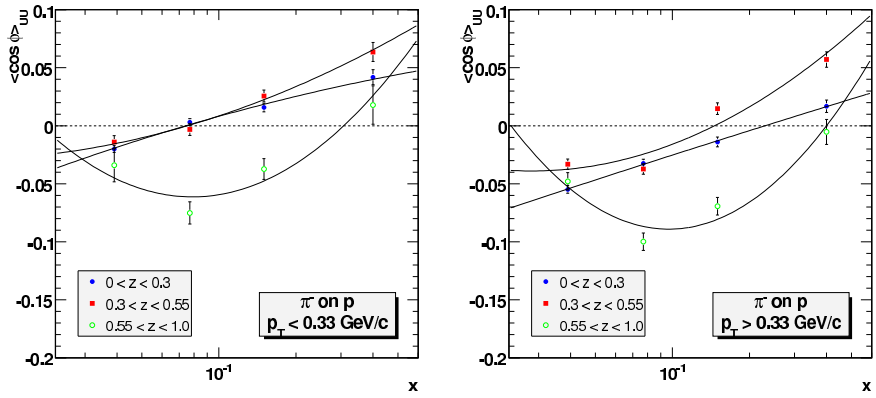
Table C.5: Extracted $\langle \cos\phi \rangle_{UU}$ and $\langle \cos 2\phi \rangle_{UU}$ moments versus x using the subtractive method for π^+ and π^-. In this extraction we integrated over all other kinematical variables. Note that on both data and Monte Carlo the hadron momentum was restricted to $2 < p < 25$ GeV.

$\langle x\rangle$	$\langle\cos\phi\rangle_{UU}$	$\sigma_{\langle\cos\phi\rangle_{UU}}$	$\langle\cos 2\phi\rangle_{UU}$	$\sigma_{\langle\cos 2\phi\rangle_{UU}}$
moments for K^+				
0.032	0.078	0.006	-0.021	0.006
0.048	0.092	0.006	-0.010	0.006
0.066	0.069	0.005	0.006	0.006
0.088	0.046	0.006	0.014	0.006
0.150	0.017	0.004	0.003	0.004
0.400	0.045	0.006	0.016	0.007
moments for K^-				
0.032	0.066	0.008	0.006	0.008
0.048	0.069	0.008	0.013	0.008
0.066	0.063	0.007	0.023	0.008
0.088	0.050	0.008	0.016	0.009
0.150	0.031	0.006	0.010	0.007
0.400	0.063	0.010	0.010	0.012

Table C.6: Extracted $\langle\cos\phi\rangle_{UU}$ and $\langle\cos 2\phi\rangle_{UU}$ moments versus x using the subtractive method for K^+ and K^-. In this extraction we integrated over all other kinematical variables. Note that on both data and Monte Carlo the hadron momentum was restricted to $2 < p < 25$ GeV.

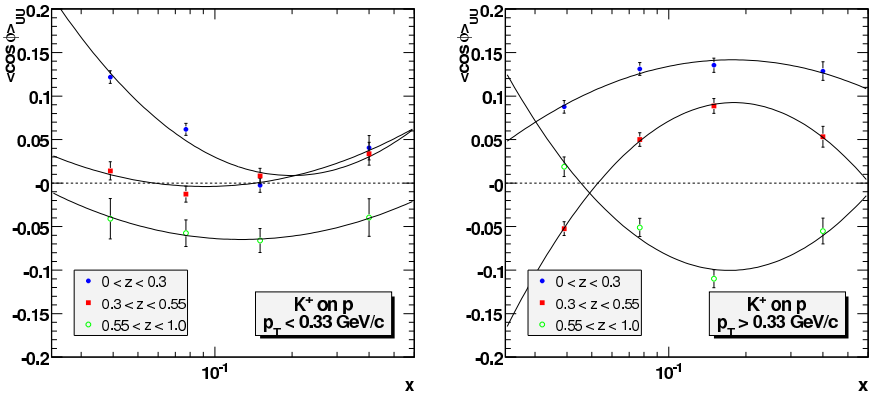


(a) Positive Pions

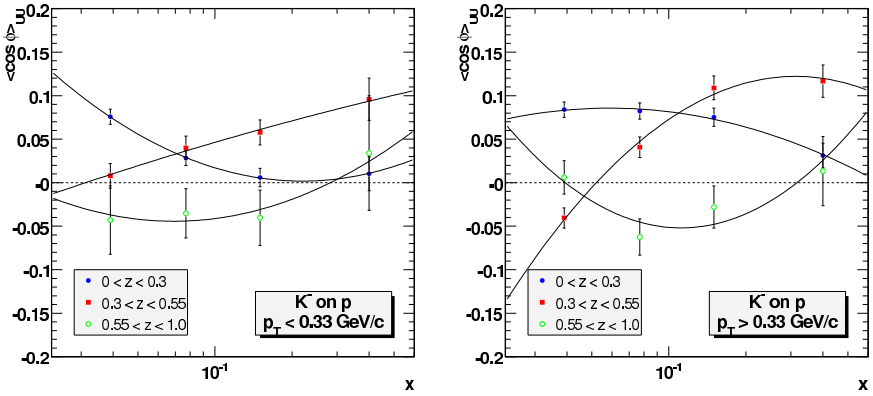


(b) Negative Pions

Figure C.1: Extracted $\cos\phi$ moments for pions in multidimensional binning.



(a) Positive Kaons

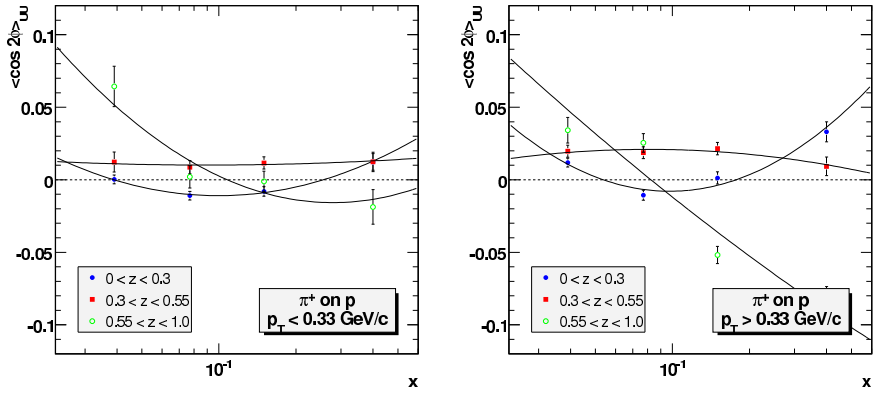


(b) Negaive Kaons

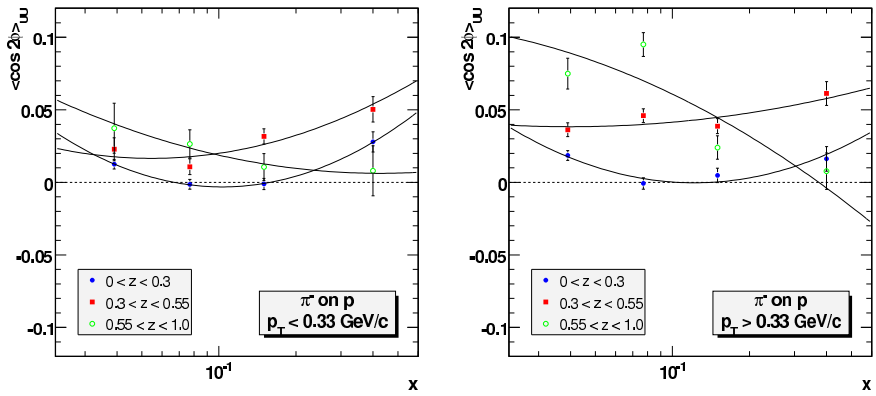
Figure C.2: Extracted $\cos\phi$ moments for kaons in multidimensional binning.

P_T	$\langle z\rangle$	$\langle xrangle$	$\langle\cos\phi\rangle$ **for** π^+	$\langle\cos 2\phi\rangle$ **for** π^+	$\langle\cos\phi\rangle$ **for** π^-	$\langle\cos 2\phi\rangle$ **for** π^-
$P_T < 0.33$	0.15	0.039	-0.035 ± 0.003	0.000 ± 0.003	-0.020 ± 0.003	0.013 ± 0.003
		0.077	-0.014 ± 0.003	-0.011 ± 0.003	0.003 ± 0.003	-0.001 ± 0.003
		0.150	0.010 ± 0.003	-0.008 ± 0.003	0.016 ± 0.004	-0.001 ± 0.004
		0.400	0.030 ± 0.005	0.012 ± 0.006	0.042 ± 0.006	0.028 ± 0.007
	0.43	0.039	-0.016 ± 0.005	0.012 ± 0.007	-0.014 ± 0.006	0.023 ± 0.008
		0.077	-0.019 ± 0.004	0.009 ± 0.005	-0.003 ± 0.005	0.011 ± 0.005
		0.150	0.021 ± 0.004	0.012 ± 0.004	0.026 ± 0.005	0.032 ± 0.005
		0.400	0.020 ± 0.006	0.012 ± 0.007	0.064 ± 0.008	0.050 ± 0.009
	0.78	0.039	-0.052 ± 0.012	0.064 ± 0.014	-0.034 ± 0.014	0.037 ± 0.017
		0.077	-0.079 ± 0.008	0.002 ± 0.008	-0.075 ± 0.010	0.026 ± 0.010
		0.150	-0.061 ± 0.007	-0.001 ± 0.007	-0.037 ± 0.009	0.011 ± 0.009
		0.400	-0.034 ± 0.012	-0.019 ± 0.012	0.018 ± 0.017	0.008 ± 0.017
$P_T > 0.33$	0.15	0.039	-0.061 ± 0.003	0.012 ± 0.003	-0.055 ± 0.003	0.019 ± 0.003
		0.077	-0.041 ± 0.003	-0.011 ± 0.003	-0.032 ± 0.004	-0.000 ± 0.004
		0.150	-0.012 ± 0.004	0.001 ± 0.004	-0.014 ± 0.004	0.005 ± 0.005
		0.400	0.025 ± 0.005	0.033 ± 0.007	0.017 ± 0.005	0.016 ± 0.008
	0.43	0.039	-0.035 ± 0.004	0.020 ± 0.004	-0.033 ± 0.004	0.036 ± 0.005
		0.077	-0.029 ± 0.004	0.019 ± 0.004	-0.037 ± 0.005	0.046 ± 0.005
		0.150	-0.013 ± 0.004	0.021 ± 0.004	0.015 ± 0.005	0.039 ± 0.005
		0.400	0.013 ± 0.005	0.009 ± 0.006	0.057 ± 0.007	0.061 ± 0.008
	0.78	0.039	-0.0293 ± 0.006	0.034 ± 0.009	-0.0479 ± 0.008	0.075 ± 0.011
		0.077	-0.0987 ± 0.006	0.025 ± 0.006	-0.0998 ± 0.007	0.095 ± 0.008
		0.150	-0.1389 ± 0.006	-0.052 ± 0.006	-0.0693 ± 0.008	0.024 ± 0.008
		0.400	-0.0993 ± 0.008	-0.082 ± 0.009	-0.0053 ± 0.011	0.008 ± 0.013

Table C.7: The $\langle\cos\phi\rangle_{UU}$ and $\langle\cos 2\phi\rangle_{UU}$ moments for the pions in the multidimensional binning versus p_T, z and x.



(a) Positive Pions



(b) Negaive Kaons

Figure C.3: Extracted $\cos 2\phi$ moments for pions in multidimensional binning.

P_T	$\langle z\rangle$	$\langle x\rangle$	$\langle\cos\phi\rangle$ **for** K^+	$\langle\cos 2\phi\rangle$ **for** K^+	$\langle\cos\phi\rangle$ **for** K^-	$\langle\cos 2\phi\rangle$ **for** K^-
$P_T < 0.33$	0.15	0.039	0.122 ± 0.007	0.027 ± 0.008	0.076 ± 0.009	0.015 ± 0.009
		0.077	0.062 ± 0.007	0.013 ± 0.008	0.028 ± 0.009	0.025 ± 0.010
		0.150	-0.002 ± 0.008	0.013 ± 0.008	0.006 ± 0.011	-0.007 ± 0.011
		0.400	0.041 ± 0.014	0.028 ± 0.015	0.010 ± 0.019	-0.00 ± 0.02
	0.43	0.039	0.014 ± 0.010	-0.023 ± 0.015	0.008 ± 0.014	0.005 ± 0.02
		0.077	-0.013 ± 0.009	-0.010 ± 0.010	0.040 ± 0.014	-0.021 ± 0.014
		0.150	0.008 ± 0.009	-0.010 ± 0.009	0.058 ± 0.014	-0.010 ± 0.015
		0.400	0.034 ± 0.013	0.002 ± 0.014	0.10 ± 0.02	0.04 ± 0.03
	0.78	0.039	-0.04 ± 0.02	0.04 ± 0.03	-0.04 ± 0.04	0.10 ± 0.05
		0.077	-0.058 ± 0.015	-0.040 ± 0.016	-0.035 ± 0.03	-0.06 ± 0.03
		0.150	-0.066 ± 0.014	-0.036 ± 0.014	-0.04 ± 0.03	0.06 ± 0.03
		0.400	-0.04 ± 0.02	-0.03 ± 0.02	0.03 ± 0.07	0.03 ± 0.07
$P_T > 0.33$	0.15	0.039	0.088 ± 0.007	-0.021 ± 0.007	0.084 ± 0.009	0.034 ± 0.009
		0.077	0.131 ± 0.007	0.043 ± 0.008	0.082 ± 0.009	0.057 ± 0.010
		0.150	0.135 ± 0.008	0.114 ± 0.010	0.075 ± 0.010	0.040 ± 0.013
		0.400	0.129 ± 0.011	0.185 ± 0.015	0.031 ± 0.014	0.05 ± 0.02
	0.43	0.039	-0.052 ± 0.008	-0.065 ± 0.009	-0.041 ± 0.012	-0.022 ± 0.013
		0.077	0.050 ± 0.008	-0.011 ± 0.008	0.041 ± 0.012	-0.027 ± 0.013
		0.150	0.089 ± 0.009	0.049 ± 0.009	0.109 ± 0.014	0.060 ± 0.014
		0.400	0.053 ± 0.012	0.031 ± 0.013	0.117 ± 0.019	0.06 ± 0.02
	0.78	0.039	0.019 ± 0.011	-0.036 ± 0.016	0.006 ± 0.019	-0.02 ± 0.03
		0.077	-0.051 ± 0.010	-0.037 ± 0.012	-0.06 ± 0.02	-0.02 ± 0.02
		0.150	-0.110 ± 0.010	-0.117 ± 0.011	-0.03 ± 0.02	-0.06 ± 0.03
		0.400	-0.055 ± 0.015	-0.101 ± 0.017	0.01 ± 0.04	-0.06 ± 0.05

Table C.8: The $\langle\cos\phi\rangle_{UU}$ and $\langle\cos 2\phi\rangle_{UU}$ moments for the kaons in the multidimensional binning versus p_T, z and x.

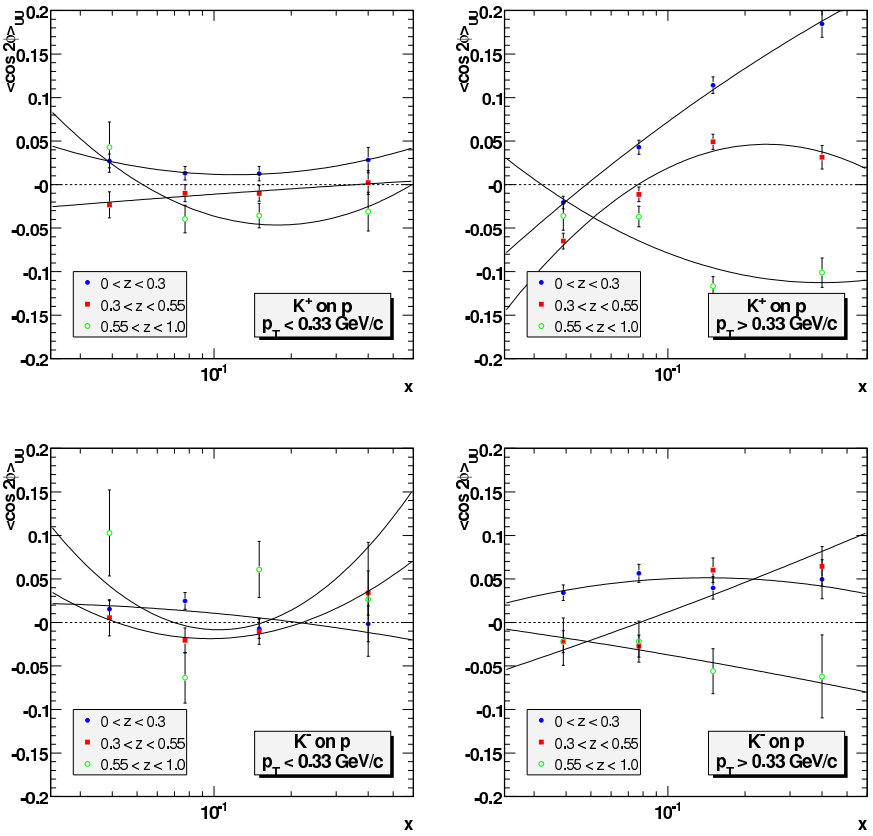


Figure C.4: Extracted $\cos 2\phi$ moments for kaons in multidimensional binning.

$\langle z \rangle$	$\Delta\mathcal{A}^{\pi^+}(\%)$	$\Delta\mathcal{A}^{\pi^-}(\%)$	$\Delta\mathcal{A}^{K^+}(\%)$	$\Delta\mathcal{A}^{K^-}(\%)$
0.12	0.31	0.12	2.00	1.62
0.15	0.31	0.29	2.33	1.78
0.18	0.15	0.26	1.51	1.50
0.22	0.09	0.09	0.98	0.64
0.25	0.36	0.16	0.03	0.14
0.28	0.54	0.39	0.42	0.15
0.32	0.38	0.48	0.66	1.04
0.35	0.43	0.62	0.66	0.99
0.38	0.51	0.68	0.70	0.89
0.42	0.58	0.72	0.76	0.99
0.45	0.50	0.68	0.81	1.18
0.48	0.40	0.64	0.68	1.36
0.52	0.08	0.35	0.44	1.16
0.55	0.54	0.26	0.26	1.07
0.58	1.05	0.06	0.35	0.73
0.62	1.70	0.43	1.62	0.55
0.65	2.45	1.16	0.60	0.44
0.68	2.54	1.09	1.56	0.59
0.72	3.06	1.27	1.84	0.23
0.75	3.02	1.26	1.22	0.31
0.78	3.49	1.56	1.65	0.02
0.82	3.03	1.73	1.00	0.36
0.85	3.15	1.34	0.51	0.75
0.88	3.83	1.40	1.06	0.59

Table C.9: Relative difference in % versus z between the acceptance correction using a reweighted Monte Carlo simulation with cosine modulations.

$\langle x \rangle$	$\Delta\mathcal{A}^{\pi^+}(\%)$	$\Delta\mathcal{A}^{\pi^-}(\%)$	$\Delta\mathcal{A}^{K^+}(\%)$	$\Delta\mathcal{A}^{K^-}(\%)$
	$0.25 < z < 0.35$			
0.030	2.11	1.83	2.61	1.14
0.048	0.91	0.43	1.21	0.99
0.065	0.13	0.29	0.38	0.00
0.088	0.23	0.58	0.56	0.24
0.120	0.32	0.31	1.66	1.75
0.170	0.03	0.61	1.76	2.94
0.250	1.01	2.58	4.23	5.58
0.350	2.72	5.96	7.46	7.85
0.500	4.77	9.41	10.43	12.08
	$0.35 < z < 0.45$			
0.030	2.85	2.79	2.30	2.52
0.048	1.55	1.23	0.59	0.00
0.065	0.45	0.10	0.09	0.71
0.088	0.27	0.66	0.26	0.27
0.120	0.31	0.40	0.18	0.86
0.170	0.00	0.79	1.15	2.55
0.250	0.72	3.16	3.32	6.58
0.350	1.89	7.31	5.30	12.30
0.500	2.98	12.20	5.88	16.97
	$0.45 < z < 0.60$			
0.030	3.83	4.04	2.21	3.44
0.048	2.91	2.46	1.66	1.11
0.065	1.39	0.48	1.19	0.34
0.088	0.19	0.91	0.01	0.19
0.120	1.39	1.13	0.81	0.43
0.170	2.06	0.62	0.48	1.05
0.250	2.57	1.42	0.15	4.11
0.350	2.20	4.80	0.49	7.96
0.500	1.51	8.16	1.13	12.52
	$0.60 < z < 0.75$			
0.030	2.93	6.18	2.36	0.30
0.048	4.38	4.73	2.20	1.85
0.065	2.82	1.14	3.13	2.17
0.088	0.23	1.61	1.15	1.16
0.120	3.71	3.47	1.85	0.51
0.170	6.62	3.28	4.78	1.83
0.250	9.38	0.79	7.46	1.96
0.350	11.25	0.90	7.75	0.28
0.500	10.71	1.86	7.06	2.00

Table C.10: Relative difference in % versus x in 4 z bins between the acceptance correction using a reweighted Monte Carlo simulation with cosine modulations.