Faculteit Wetenschappen
Vakgroep Subatomaire en Stralingsfysica
Academiejaar 2002-2003

A Study of Fragmentation Processes in the HERMES experiment using a Ring Imaging Čerenkov Detector

Fragmentatie Processen in het HERMES experiment bestudeerd met een Ring Imaging Čerenkov Detector

Promotor: Prof. Dr. D. Ryckbosch

Proefschrift ingediend tot het verkrijgen van de graad van
Doctor in de Wetenschappen, groep Natuurkunde
door
Brecht Hommez

Any unexplained phenomenon passes through three stages before the reality of it is accepted. During the first stage it is considered laughable. During the second stage, it is adamantly opposed. Finally, during the third stage, it is accepted as self-evident.

Arthur Schopenhauer

Contents

Chapter 1

Introduction

1.1 Elementary particles

The search for the elementary, indivisible building blocks of matter is one of the most fundamental problems in physics. Early in the previous century, physicists discovered that the atom is not an elementary indivisible unit; each atom consists of a nucleus and electrons orbiting around it. Then they discovered that the nucleus is not an elementary unit either; each nucleus consists of protons and neutrons packed tightly together. In the 1930's, physicists began to build accelerating machines, producing beams of energetic protons or electrons that could serve as projectiles to split the nucleus. In the 1950's they built much larger and more powerful accelerating machines with which they tried to split the proton and the neutron. But the result of these attempts was chaos: when bombarded by very energetic projectiles, the proton and the neutron do not split into simple subnucleonic pieces. Instead, the violent collisions generate a variety of new, exotic particles by the conversion of kinetic energy into mass. For want of a better name, these new particles were called 'elementary particles'. Only in the last decades, physicists have found convincing evidence that protons, neutrons and other elementary particles are made of very small, compact subunits. These subnucleonic building blocks are called quarks.

Quarks were introduced by M. Gell-Mann and independently by G. Zweig to interpret the information available on the properties of the elementary particles. Based on their quantum numbers and their behaviour in reactions, the particles could be classified into groups or families of similar particles. To explain these similarities, Gell-Mann and Zweig proposed a theory in which all the particles are constructed out of three fundamental building blocks: up, down and strange quarks, or simply quarks with "flavour" u, d and s. They all have spin 1/2 and fractional electric charges of 2/3, $-1/3$ and $-1/3$ respectively.

The quark model aimed at describing the mass, charge and spin of the particles. For instance, the proton has a mass[1] of about 1 GeV, a fractional charge +1 and spin equal to 1/2. According to the model, a proton with its spin, for instance, in the upward direction is made of two quarks with flavour up and charge 2/3 plus one quark with flavour down

[1]Throughout this text the natural units will be used in which $\hbar = c = 1$.

and charge $-1/3$. Two of the quarks have spin 1/2 in the upward direction and one has its spin 1/2 in the down direction. Each of the three quarks carries about one third of the mass of the proton. Looking at the "static" properties of particles like their mass, charge and spin, was not enough to unravel the details of their structure. To glance at the inside of particles, physicists use nowadays deep-inelastic scattering (DIS) experiments.

1.2 Deep-inelastic scattering

In scattering experiments a focused beam of particles is dispersed by the interaction with a target. The way this dispersion takes place yields information on the structure of the target. For instance, the existence of the nucleus was suggested by Rutherford as an explanation of the scattering experiments of Geiger and Marsden, who used thin foils of gold as targets and bombarded these with a beam of alpha particles. To Rutherfords amazement, some of the alpha particles were deflected by such a large angle that they came out backward. Rutherford recognised that the large deflection must be caused by a close encounter between the alpha particle and a very small but very massive kernel inside the gold atom: the nucleus. The knowledge of the structure of the nucleus increased considerably when it was realised that atomic nuclei consist of nucleons: protons and neutrons.

While the nucleus was discovered by elastic scattering, the building blocks of the nucleon could only be explored by *inelastic scattering*. To discover these building blocks, physicists tried to break protons into pieces by bombarding them with projectiles that are accelerated to very high energies. Basically, particle accelerators such as the one at SLAC (Stanford Linear Accelerator Center) are exploited as microscopes of extremely high resolution. The experiments at SLAC scattered electrons off protons. The interaction proceeds via the exchange of a virtual photon between the electron and the proton.

A measure of the resolution of the experiment is given by the four-momentum squared of the virtual photon probe, Q^2. The SLAC experiments reached a maximum[2] Q^2 of 7.4 GeV^2, corresponding to a resolution much smaller than the proton size, hence the name *deep-inelastic scattering*.

The results of the SLAC experiments in 1969 showed that the scattering data (i.e. the cross section) did not exhibit a (strong) dependence on Q^2, indicating that the resolution of the scattering probe did not change the outcome of the experiment. The only reason could be that the electrons scattered off *point-like* entities in the proton. This Q^2-independence, called *scaling*, was predicted by J. Bjorken and could be explained by R. Feynman, who introduced the parton model: the proton was pictured to be a collection of almost free point-like constituents off which electrons scatter incoherently. The constituents of the proton were initially called partons, but it soon became clear that they had a lot in common with the quarks of Gell-Mann and Zweig, up to then only a mathematical concept.

Feynman's partons have spin 1/2, fractional electric charge and flavour, but they have a very small mass compared to Gell-Mann's quarks, a few MeV against about 300 MeV.

[2] In order to accelerate a small rocket of 1000 kg to the same energy, $100 \times 100 \times 100$ km^3 fuel would be needed, roughly 10 000 times the world 's reserves of crude oil.

But this is not the only difference between the two models. Further experiments showed that the protons could not be composed solely of quarks. To explain the measurements, theoretical physicists postulated a "sea" of virtual quarks and anti-quarks. Virtual particles are particles that don't last long. The quarks whose flavour determines the nature of the particle, for example the up, up, and down quarks of the proton, are called "valence quarks." The valence quarks float in the sea of virtual quarks that are constantly popping in and out of the vacuum. A new theory called Quantum Chromo Dynamics, or QCD, added neutral particles called gluons to the protons constituents. According to QCD, gluons are the particles that bind quarks together. These gluons can interact with each other, and can produce virtual quark anti-quark pairs. The picture of a proton as a sea of gluons and virtual quarks with three valence quarks floating in it (the so-called quark parton model) reminded theoretical physicist James D. Bjorken of Thomson's model of the atom. He referred to it as "the plum-pudding proton".

Deep-inelastic scattering experiments are performed in some of the world's largest experimental facilities for high energy physics, such as CERN (Switzerland), SLAC (USA) and DESY (Germany). They usually employ beams of electrons, positrons, muons or, more rarely, neutrinos. They scatter off different kinds of fixed targets or off a beam of protons, and they operate at different energies and kinematic coverage, opening an exciting world of complex scattering phenomena.

1.3 The spin of the proton

Since the first SLAC experiments, the DIS process with charged lepton beams has been widely used to probe the structure of the nucleon. With the technological development of polarised beams and targets the spin structure of the nucleon became accessible. Starting from the late 1970's, a variety of experiments has been carried out to investigate the nucleon spin structure. The SLAC experiments E80 and E130 confirmed the general ideas of scaling and QPM interpretation of the nucleon structure. In addition, the Ellis-Jaffe sum rule prediction for the proton spin structure was verified, though within large experimental uncertainties. What the Ellis-Jaffe sum rule actually predicts is the so-called spin dependent structure function g_1 of the proton which according to Feynman's quark parton model is a measure of the sum of the quark spins in the proton. Ellis and Jaffe assumed that the strange sea quarks were unpolarised and made a theoretical calculation for g_1. In 1987, a high precision measurement of the proton spin structure function was performed by the European Muon Collaboration (EMC) at CERN with a polarised muon beam. Their result significantly deviated from the prediction of the Ellis-Jaffe sum rule. As a consequence, if no polarisation of the strange quark sea is assumed, the EMC data imply that only about 12% of the nucleon spin is carried by the quarks. This surprising result lead to the proposal of a new series of experiments which could test the correctness of the EMC measurement with increased accuracy. Also this second generation of experiments (Spin Muon Collaboration at CERN and E142 and E143 at SLAC) confirmed the deviation from the naive quark parton model predictions found in the previous polarised scattering experiments and the "spin puzzle" was born. In order to have a complete picture of the nucleon spin, also the spin contributions from the gluons and the sea quarks in the nucleon must be taken into account:

$$\frac{1}{2} = \frac{1}{2}\Delta\Sigma + \Delta G + L_z \tag{1.1}$$

where $\Delta\Sigma$ and ΔG are the net spin contributions from the quarks and the gluons, respectively, and L_z is the contribution from the orbital angular momentum of the partons. Up to now, the spin puzzle has not been solved and especially the contributions ΔG and L_z are unkown.

1.4 Semi-inclusive DIS at HERMES

The HERMES experiment at DESY was designed to disentangle the contributions from the different quark flavours to the nucleon's spin in *semi-inclusive* deep-inelastic scattering reactions. In such reactions, *hadrons* are detected in coincidence with the scattered lepton. These hadrons are created by bombarding a target (e.g. a proton) with projectiles (e.g. electrons) with high energy. If the target breaks apart during the collision, new particles (hadrons) can be created. These hadrons are particularly interesting since the flavour of the quark probed in the scattering process can be deduced from the charge and the type of the observed hadron. This method is called *flavour tagging* and allows a direct separation of the spin contributions by the individual quark flavours. An important issue in this method is the ability to identify the type of the produced hadron and therefore the HERMES spectrometer was equipped with a RICH (Ring Imaging Čerenkov detector) in 1998.

The HERMES (HERa MEasurement of nucleon Spin) experiment is located at the Deutsches Elektronen SYnchrotron (DESY) laboratory in Hamburg, Germany. In DESY's Hadronen Elektronen Ring Anlage (HERA) a 27.5 GeV naturally polarised positron beam circulates. The HERMES target contains nuclei in the positron path, allowing deep-inelastic interactions. Since both the positron beam and the target nuclei are polarised, HERMES is able to study the spin stucture of the nucleon. A large acceptance spectrometer behind the target detects the forward going positron and hadronic fragments from the beam-target interactions. HERMES has been taking data since 1995, and this thesis reports on an analysis of polarised and unpolarised data taken in 2000.

1.5 Outline

This thesis is divided in 6 chapters. Chapter 2 reviews the theoretical framework of unpolarised and polarised deep-inelastic scattering and its interpretation in the quark parton model. A number of predictions for quantities describing the nucleon structure and recent experimental results in this field are reviewed. Chapter 3 describes the HERMES experiment. In chapter 4 the specific detector responsibility of the author, the RICH detector is detailed. Chapter 5 explains the inclusive analysis of the HERMES 2000 data and the extraction of the F_2 structure function ratio of deuterium and hydrogen. In Chapter 6 a number of semi-inclusive studies will be presented: the ratio of the down and up valence quarks in the proton will be studied, the contribution from the strange quarks to the proton spin will be investigated and transverse momentum spectra of different hadron species will be presented.

Chapter 2

Deep-inelastic lepton nucleon scattering

2.1 Introduction

A scattering process is called *inelastic* if the number and/or types of the particles in the final state are different from those in the initial state. In contrast to an elastic process, inelastic scattering can be related to the break-up of at least one of the particles present in the initial state and can therefore be used to investigate the substructure of the participating probes.

In inelastic lepton nucleon scattering, an incoming lepton $l = e^{\pm}, \mu^{\pm}, \nu_{e,\mu}, \bar{\nu}_{e,\mu}$ interacts with a target nucleon N in such a way that the nucleon breaks up and forms a final hadronic state:

$$l + N \to l' + X \tag{2.1}$$

where X denotes any set of hadrons allowed by conservation laws. As explained in chapter 1 the substructure resolved by the incoming lepton in this case is nothing else than the quark content of the nucleon, hence the name *deep*-inelastic scattering (DIS).

The DIS process is called *inclusive* if only the outgoing lepton l' is detected, while in the *semi-inclusive* case also at least one of the final state hadrons X is recorded in coincidence. If all particles in the final state and their kinematics are known, the measurement is called *exclusive*.

In the semi-inclusive case hadrons originating from the quark that is struck by the incoming lepton carry information on the flavour of this struck quark. These so-called *current fragments* must be separated from the hadrons originating from the target remnants, the so-called *target fragments*. The separation between the current and target fragmentation regions can be accomplished by imposing requirements on the kinematics of the detected hadrons, as will be explained later in this chapter.

The DIS process is mediated by the exchange of a virtual boson ($\gamma^*, W^{\pm}, Z^0$) between the incoming lepton l and one of the partons inside the target nucleon. In the case of e^- (or e^+) scattering and for energy transfers (ν) from the lepton to the nucleon that are small compared to the masses of the carriers of the weak force ($W^{\pm}, Z^0$), electromagnetic

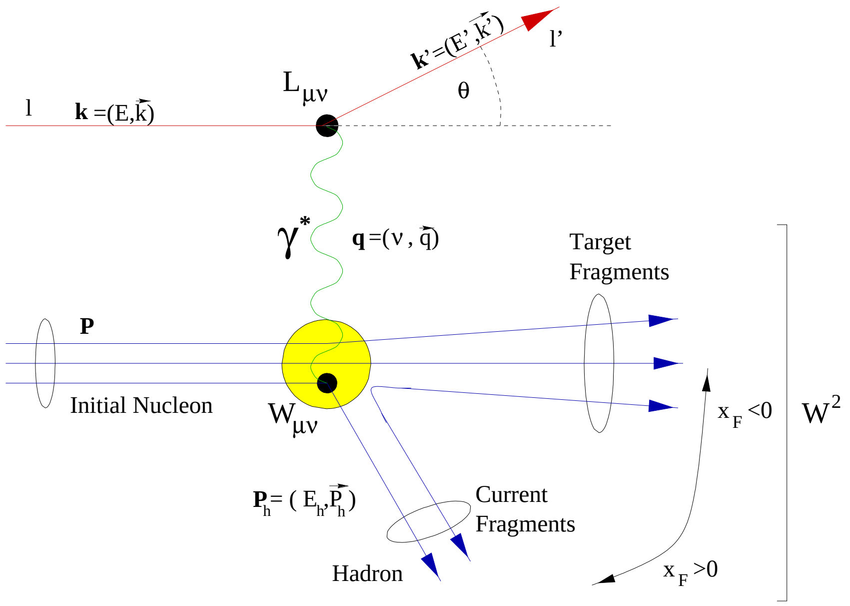

Figure 2.1: Schematic diagram of deep-inelastic lepton nucleon scattering in the one-photon exchange approximation. The kinematical quantities shown in this figure are defined in section 2.2.

processes are dominant. The lowest order electromagnetic contribution to deep-inelastic lepton nucleon scattering (the one-photon exchange graph) is schematically depicted in fig. 2.1.

Because of the particular circumstances at HERMES where electrons or positrons with an energy of 27.5 GeV interact with a stationary target, the following sections will focus on charged leptons and electromagnetic processes within the one-photon approximation. Multi-photon exchange diagrams are suppressed by a factor of $\alpha_{em} = 1/137$ and may partially be taken into account via so-called radiative corrections.

In this chapter the formalism of inclusive and semi-inclusive deep-inelastic lepton nucleon scattering will be introduced, along with the interpretation in the quark parton model. For the semi-inclusive case a discussion on hadron formation (also called *fragmentation*) will be included.

Since HERMES is a double polarised experiment (where both the spin of the leptons in the beam and the spin of target are aligned (anti-)parallel to the momentum of the incident lepton), the description of DIS will be extended to the polarised case.

2.2 Kinematics

In fig. 2.1 the semi-inclusive DIS process is shown. Note that in a fixed target experiment as HERMES, neglecting the thermal motion of the target atoms and the Fermi motion of the nucleons within the atoms, the laboratory frame coincides with the nucleon rest frame. An incoming lepton with four-momentum $\mathbf{k} = (E, \vec{k})$ is scattered off a target nucleon with four-momentum $\mathbf{P} = (M, \vec{0})$, where M denotes the restmass of the target nucleon. After the scattering process, the lepton has a four-momentum $\mathbf{k}' = (E', \vec{k'})$ and the scattering angle between the incoming and scattered lepton momenta is θ. The four-momentum of an observed hadron h in the final state is $\mathbf{P_h} = (E_h, \vec{P_h})$

The scattering process occurs through the exchange of a virtual photon that transfers four-momentum from the incoming lepton to the target. The four-momentum of the exchanged virtual photon is the loss of four-momentum of the incoming lepton:

$$\mathbf{q} = \mathbf{k} - \mathbf{k}' \tag{2.2}$$

The negative square of the four-momentum of the virtual photon or its squared invariant mass is given by

$$Q^2 = -\mathbf{q}^2 = -(\mathbf{k} - \mathbf{k}')^2 \overset{lab}{\approx} 4\,E\,E' \sin^2(\frac{\theta}{2}) \tag{2.3}$$

which has always positive values. In this equation the lepton restmass is neglected, as it is much smaller than the incident and scattered momentum. The Bjorken scaling variable x is defined as

$$x = \frac{Q^2}{2\,\mathbf{P}\cdot\mathbf{q}} = \frac{Q^2}{2\,M\,\nu} \tag{2.4}$$

with

$$\nu = \frac{\mathbf{P}\cdot\mathbf{q}}{M} \overset{lab}{=} E - E' \tag{2.5}$$

so that ν represents (in the laboratory frame) the energy that is transferred by the virtual photon from the incoming lepton to the target nucleon. The Bjorken scaling variable y is given by

$$y = \frac{\mathbf{P}\cdot\mathbf{q}}{\mathbf{P}\cdot\mathbf{k}} \overset{lab}{=} \frac{\nu}{E} \tag{2.6}$$

which in the laboratory system denotes the fractional energy transfer of the virtual photon with respect to the beam energy.

Defining $\mathbf{P_h^{tot}} = \sum_h \mathbf{P_h}$ as the sum of the four-momenta of the outgoing hadron fragments X, the squared invariant mass of the final hadronic state is given by

$$W^2 = (\mathbf{P_h^{tot}})^2 = (\mathbf{q} + \mathbf{P})^2 \overset{lab}{=} M^2 + 2\,M\,\nu - Q^2 = M^2 + Q^2\frac{1-x}{x} \tag{2.7}$$

The variable W^2 is a measure of the invariant mass that is available to create the final hadronic state in the scattering process.

All the above defined variables are inclusive quantities, and they are not independent. Assuming that the incoming energy E is known, only two variables are necessary to determine the kinematics of the scattering process: the scattering angle θ and the energy of the scattered lepton E'. With this knowledge Q^2, x, y and W^2 can be calculated. The kinematics of an inclusive scattering event are fully determined by any combination of two of these variables and mostly the combination of x and Q^2 is chosen.

The interpretation of x is easiest in the so-called Breit reference frame in which the target nucleon has a very large momentum. In such a frame the momenta of the partons will be almost collinear with the nucleon momentum, so that to a good approximation the target can be viewed as a stream of quarks, each carrying a fraction of the nucleon momentum. Furthermore, for $Q^2 \gg M^2$ it can be shown that the fraction of the momentum of the target nucleon carried by the struck quark is precisely x. Apart from that, the variable x is also a measure for the inelasticity of the scattering process. In an elastic scattering process the target nucleon remains intact, and consequently $W^2 = M^2$, which implies $2M\nu = Q^2$ or $x = 1$. For inelastic processes, $W^2 > M^2$ and $0 < x < 1$.

The squared invariant mass of the virtual photon, Q^2 , is a measure of the spatial resolution in the scattering process. In the Breit frame, it can be shown that the wavelength of the virtual photon is proportional to $1/\sqrt{Q^2}$. Hence the resolving power of the photon probe increases with higher Q^2 which allows an investigation of the substructure of the nucleon on the parton scale.

In semi-inclusive scattering, the variables x and Q^2 alone are not sufficient to describe the event, but additional kinematic variables are required for a detected hadron. The fraction of the energy of the virtual photon which is transferred to a detected hadron is defined as:

$$z = \frac{\mathbf{P} \cdot \mathbf{P_h}}{\mathbf{P} \cdot \mathbf{q}} \stackrel{lab}{=} \frac{E_h}{\nu} \tag{2.8}$$

The Feynman scaling variable x_F is defined in the center of mass system of the virtual photon and the nucleon, and scales the longitudinal component of the hadron's momentum to its maximum possible value:

$$x_F = \frac{P_h^L}{|\vec{q}|} \tag{2.9}$$

where P_h^L is the longitudinal momentum of a hadron h in the center of mass system of the virtual photon and the nucleon:

$$P_h^L = \vec{P}_h \cdot \frac{\vec{q}}{|\vec{q}|} \tag{2.10}$$

The transverse hadron momentum in the same frame can be defined[1] as:

$$p_t = P_h^T = \sqrt{\vec{P}_h^{\,2} - (P_h^L)^2} \tag{2.11}$$

The kinematically allowed ranges for the above quantities are $0 \leq z \leq 1$ and $-1 \leq x_F \leq 1$. The Feynman variable x_F separates hadrons going backwards in the $\gamma^* N$ reference frame ($x_F < 0$) from hadrons going forwards ($x_F > 0$) along the direction of the virtual photon. Selecting hadrons with high enough z assures that they go fast enough in the laboratory frame ($z > z_0$, eg. $z_0 = 0.2$). For the HERMES kinematical region, x_F and z are strongly correlated. These variables can be used to select the current fragmentation region, as hadrons propagating along the direction of the virtual photon most probably originate from the struck quark.

In addition to the two inclusive variables, three extra variables are needed to describe semi-inclusive DIS. For instance the combination (Φ^γ, z, p_t) can be chosen (fig. 2.2): the angle Φ^γ defines the relative orientation of the lepton scattering plane and the hadron

[1] The more common notation p_t will be used instead of P_h^T if it is clear from the context which type of hadron is meant.

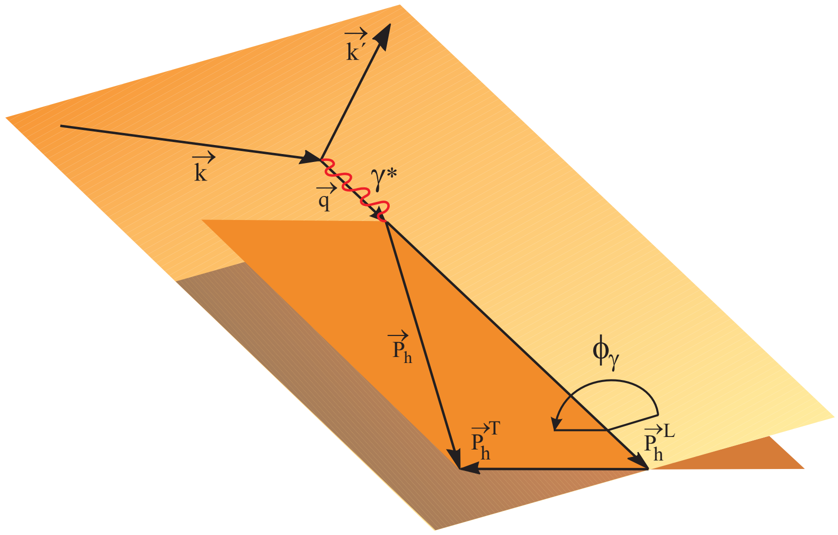

Figure 2.2: Definition of the hadron transverse momentum P_h^T and the azimuthal angle Φ^γ. The lepton scattering plane is defined by the lepton momenta $\vec{k}$ and $\vec{k}'$, the hadron production plane is spanned by the vectors $\vec{q}$ and $\vec{P}_h$. The azimuthal angle Φ^γ is the angle between the lepton scattering plane and the hadron production plane. The hadron transverse momentum $P_h^T \equiv p_t$ is defined as the component of the hadron momentum P_h perpendicular to the direction of the virtual photon.

production plane, the variables z and P_h^T are respectively a measure of the longitudinal and transverse hadron momentum in the hadron production plane and define the hadron momentum components in the hadron production plane.

2.3 Inclusive deep-inelastic scattering

2.3.1 Cross section

The fundamental quantity for scattering studies is the differential cross section as a function of the measured variables. This cross section is proportional to the probability of scattering or equivalently to the particle detection rate measured in the experiment. The lowest order Quantum Electrodynamics (QED) inclusive DIS cross section can be calculated by evaluating the Feynman graph for the one-photon exchange [1]:

$$\frac{\mathrm{d}^2\sigma}{\mathrm{d}x\,\mathrm{d}Q^2} \propto L_{\mu\nu}W^{\mu\nu} \tag{2.12}$$

where $L_{\mu\nu}$ is the leptonic tensor and $W^{\mu\nu}$ the hadronic tensor (cfr. fig. 2.1). The leptonic tensor describes the coupling of the incoming point-like lepton with the exchanged virtual photon. It is well-known from QED and can be written down using the Feynman rules. At DIS energies the photon is able to resolve the target structure and so the target nucleon cannot be approximated as pointlike. For this reason the hadronic tensor cannot be explicitly written down from first principles. $W^{\mu\nu}$ can be calculated in general form, however, and after the use of symmetry arguments and conservation laws, a useful form of $W^{\mu\nu}$ is arrived at in terms of dimensionless structure functions. As $W^{\mu\nu}$ can be written in terms of these structure functions, so can the DIS (unpolarised) cross section, after contraction of $L_{\mu\nu}W^{\mu\nu}$:

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

where $\alpha \approx 1/137$ is the electromagnetic coupling constant and $F_{1,2}(x,Q^2)$ are the dimensionless structure functions of the nucleon that have a simple interpretation in the quark parton model.

2.3.2 The quark parton model and QCD

The quark parton model (QPM) is a combination of the static quark model introduced by Gell-Mann and Zweig to explain the hadron mass spectrum and the parton model by Bjorken and Feynman, introduced in conjunction with the electron scattering experiments at SLAC in the 1960's. The QPM is formulated in the earlier mentioned Breit reference frame, where the variable x represents the fraction of the nucleon momentum carried by the struck quark. The quarks are for short times quasi-free and deep-inelastic scattering off a nucleon can be treated as elastic scattering off a quark. The advantage of the QPM lies in the fact that it helps to visualise the mechanics of DIS processes and that it provides the possibility of quick approximate calculations. In the basic 'naive' quark parton model the interactions between the quarks are neglected, but it is possible to include these by means of Quantum Chromo Dynamics corrections, which leads to the QCD corrected QPM, as will be explained later in this section.

Before explaining the basic features of the QPM we mention that the DIS lepton nucleon cross section can be calculated in the framework of the naive QPM as the incoherent sum of elastic scattering off quarks in the nucleon. Comparison of this QPM cross section to eq. (2.13) leads to the following interpretation of the nucleon structure functions F_1 and F_2:

$$F_1(x,Q^2) = \frac{1}{2}\sum_f e_f^2 q_f(x,Q^2) \tag{2.14}$$

$$F_2(x,Q^2) = \sum_f e_f^2\, x q_f(x,Q^2) \tag{2.15}$$

where e_f is the charge of the quark with flavour f, and the *parton density function* (PDF) $q_f(x,Q^2)$ denotes[2] the number density of a quark f with momentum fraction x. First the structure function F_2, i.e. the charge-weighted sum of the parton momentum distributions will be discussed and later in this section the interpretation of F_1 will be given in terms of the so-called Callan-Gross relation

$$2xF_1(x,Q^2) = F_2(x,Q^2) \tag{2.16}$$

[2]Often $q_f(x)$ is referred to as the probability to find a quark of flavour f with momentum fraction x. Although intuitively clearer, this interpretation is somewhat misleading since the integral $\int u_v(x)dx$ is not 1, as the proton contains two valence u quarks.

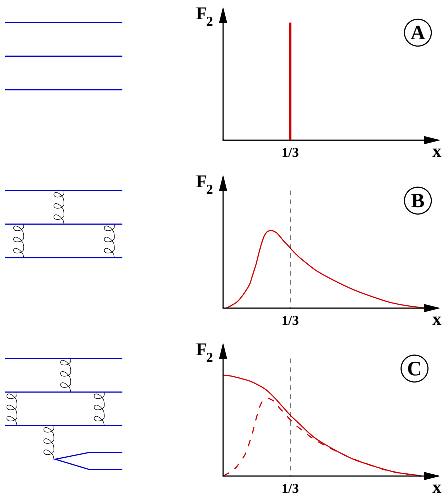

Figure 2.3: Different contributions to the parton momentum distribution F_2. If the nucleon only contained the 3 valence quarks each of them would carry one third of the nucleon momentum (a). The exchange of gluons shifts and smears the momentum distributions of the valence quarks and hides parts of the nucleonic momentum from the virtual photon probe (b). The gluons may split up into quark anti-quark pairs which give an additional sea contribution at small x (c).

In the naive QPM, the proton is pictured as an association of three *non-interacting* quarks (2 *up* quarks and one *down* quark), hence each carries 1/3 of the proton momentum. In this model with only so-called *valence* quarks the structure function F_2 would be a δ-function at $x = 1/3$ (fig. 2.3a). Although the naive QPM can explain the quantum numbers of the hadrons, it needs modification since for some hadrons several apparently identical quarks are found in the same quantum state, violating Pauli's exclusion principle. To avoid this violation, QCD introduced a new degree of freedom for the quarks: *colour*. There are three colours (labelled red, green and blue) and the carrier of the colour force is the gluon, a spin 1 object. The exchange of a gluon between two coloured quarks is analogous to the exchange of a photon between two charged particles. However, the gluons themselves have colour whereas the photons do not have an electric charge, making the description of the *strong* force (QCD) more complicated then QED. The introduction of the interactions between the three valence quarks via gluons smears the momentum distribution F_2 into a broad peak with a maximum around a value lower than $x = 1/3$, as depicted in fig. 2.3b. Moreover, like a naked electron in QED is surrounded by a cloud of virtual photons, electrons and positrons, a quark may emit a virtual gluon which splits up into a quark anti-quark pair. Quarks produced in this process are called sea quarks. A valence quark together with its cloud of gluons and sea quarks is called a *constituent* quark of the hadron. The introduction of sea quarks leads to an additional contribution to F_2 at small x (fig. 2.3c). In the QCD corrected QPM model the summation in the structure functions (2.14) and (2.15) runs over the three flavours (d, u, s) of the valence quarks and sea quarks and anti-quarks, but not over gluons, since the virtual photon cannot couple electromagnetically to the charge-neutral gluons.

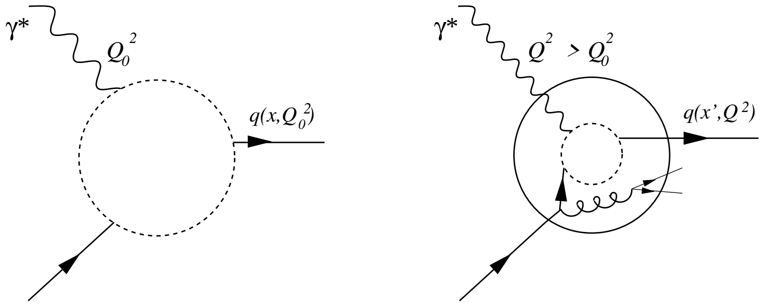

Figure 2.4: Schematic representation of the Q^2-dependence of the quark distributions due to the increasing resolving power of the virtual photon. In the left diagram the virtual photon with momentum Q_0^2 probes the quark distribution $q_f(x, Q_0^2)$. For $Q^2 > Q_0^2$ (right diagram) the photon probes 'softer' quarks due to the emission of a gluon.

The proton and deuterium structure function $F_2(x, Q^2)$ has been measured by several fixed target experiments: NMC (muon scattering [2]), SLAC (electron scattering [3]) and BCDMS (muon scattering [4]), all confirming the functional x-dependence of fig. 2.3c, hence strengthening the belief in the QCD corrected QPM. A striking feature of the experimental data was their dependence on Q^2 for fixed values of x. As explained earlier the Q^2-scale defines the 'resolution' of the scattering process. In the naive QPM one would expect that the scattering occurs on point-like quarks that are much smaller than the distance scale of the process and hence no scale-dependence is expected. The scaling violation of F_2 (i.e. the Q^2 dependence) should not be interpreted as an indication of the quarks not being point-like, but confirms the picture of the QCD corrected QPM, as schematically represented in fig. 2.4 and explained as follows: for a certain value of $Q^2 = Q_0^2$, the probed quark distribution is $q_f(x, Q_0^2)$. Increasing the resolution of the virtual photon by using $Q^2 > Q_0^2$, one resolves that the quark momentum has changed due to the emission of a gluon. Thus, in the DIS process the quark with momentum fraction x' is now probed, which originated from a parent quark with momentum fraction $x > x'$. In the same way the virtual photon can couple to a quark which was generated by a gluon splitting into a quark anti-quark pair. As a consequence of these processes the probability to find a parton with large x decreases with increasing Q^2 since the momentum fraction can be shared among several more partons. QCD allows the calculation of these effects by the Dokshitzer-Gribov-Lipatov-Altarelli-Parisi (DGLAP) equations [5, 6, 7] which can be used to evolve quark distribution functions and structure functions measured at a certain Q^2 to a fixed Q_0^2. The DGLAP equations do not only take the gluon-emission and absorption by quarks into account but also the coupling of gluons to each other (unlike photons in QED, gluons may interact with each other through their colour charge). An important parameter in the DGLAP equations is the *coupling strength* α_s, which is (in first order perturbative QCD) given[3] by

$$\alpha_s = \frac{4\pi}{9} \frac{1}{\ln (Q^2/\Lambda^2)} \tag{2.17}$$

[3]In general α_s depends on the number of active quark flavours, the eq. (2.17) has been written down for the three flavours (u,d,s) active at HERMES energies.

where Λ^2 is the QCD scale parameter which gives the limit where perturbative QCD can be applied and is about 300 MeV. The value of the strong coupling "constant" α_s depends on Q^2, corresponding to a dependence on the spatial separation. For small distances (or photon wavelengths) and corresponding high values of Q^2, α_s vanishes logarithmically as if the quarks are no longer bound inside a nucleon ("asymptotic freedom"). The DIS process is then an incoherent scattering from the individual partons as free constituents. This key assumption in the QPM is hence reproduced by QCD in the so-called Bjorken limit $Q^2 \to \infty$. All experiments are performed at finite Q^2 and perturbative QCD corrections expanded in $\alpha_s(Q^2)$ are necessary to adjust parton model interpretations and to reproduce the Q^2-dependence of the structure functions and quark and gluon distributions.

The Callan-Gross relation (2.16) is valid in the naive quark parton model and holds only if the partons are spin 1/2 particles. For spin 0 particles, the Callan-Gross relation would reduce to $F_1(x,Q^2) = 0$. Experimental data [2, 3] rule decisively in favour of the former relation, confirming that quarks have a half integer spin. The experimental data measured in fact the ratio R that is defined as:

$$R(x,Q^2) = \frac{F_2(x,Q^2)}{2xF_1(x,Q^2)}(1+\gamma^2) - 1 \tag{2.18}$$

and found deviations from zero. In this equation, R is the ratio of the absorption cross section of longitudinally (σ_L) and transversely (σ_T) polarised virtual photons by the nucleon:

$$R(x,Q^2) = \frac{\sigma_L}{\sigma_T} \tag{2.19}$$

and γ is a kinematical factor that is defined as $\gamma = 4M^2x^2/Q^2$ and is $\ll 1$, since $4M^2x^2$ is small compared to Q^2 for the kinematical ranges that were investigated. It should be noted that the terms 'transverse' and 'longitudinal' here do not refer to the spin of the photon, but to the direction of the electric and magnetic field vectors with respect to the direction of motion of the virtual photon[4]. The naive QPM predicts σ_L to be zero; only interactions with transversely polarised photons are allowed. The experimental results on R contradict this and allow also interactions with longitudinal polarised photons. Hence we can not simplify the inclusive DIS cross section (2.13) by the Callan-Gross relation, but R needs to be introduced, yielding the following expression for the inclusive cross section:

$$\frac{\mathrm{d}^2\sigma}{\mathrm{d}x\,\mathrm{d}Q^2} = \frac{4\pi\alpha^2}{xQ^4}F_2(x,Q^2)\left[\frac{y^2+Q^2/E^2}{2(1+R(x,Q^2))} + \left(1 - y - \frac{M\,x\,y}{2E}\right)\right] \tag{2.20}$$

The deviations from the naive QPM where masses are neglected as well as transverse momenta of partons inside the nucleons can be proven to be described by the ratio $R(x,Q^2)$.

2.3.3 Polarised inclusive DIS

The information on the spin distribution inside the nucleon is contained in the antisymmetric part of the hadronic tensor $W^{\mu\nu}$. In the unpolarised case the rigorous calculation of the cross section averages over the spin state of the incoming leptons, forcing the

[4] A real photon can only have helicity $\lambda = \pm 1$ and electric and magnetic field vectors that are transverse to its direction of motion. A virtual photon can apart from being transversely ($\lambda = \pm 1$) polarised, also have a longitudinally ($\lambda = 0$) polarised component. See [8] for a complete description.

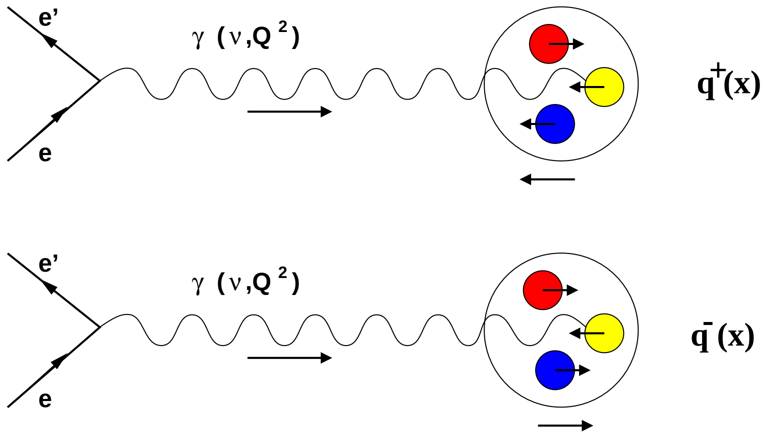

Figure 2.5: Diagram of polarised DIS. The virtual photon is always absorbed by a quark with opposite spin direction. The top panel shows the process that is sensitive to quarks polarised parallel to the nucleon's spin ($q_f^+(x)$), while the bottom panel is for quarks with their spin direction anti-parallel to the nucleon spin ($q_f^-(x)$).

anti-symmetric parts of $W^{\mu\nu}$ to vanish while calculating the tensor product $L_{\mu\nu}W^{\mu\nu}$. In the polarised case, when looking at the cross section difference of leptons with opposite spin state, the symmetric parts cancel (i.e. the unpolarised, spin-independent structure functions drop out) and the asymmetric parts become evident, giving access to the nucleon spin. Experimentally either the beam or target orientation is flipped; for the following discussion the latter case will be considered. HERMES is a double polarised experiment, where both the incoming lepton beam and the target are longitudinally polarised, i.e. the spins are aligned along the momentum of the incident lepton. The difference between the cross sections in the cases where the spins of the incoming lepton and the target nucleon are parallel ($\uparrow\Uparrow$) and anti-parallel ($\uparrow\Downarrow$), is given by

$$\frac{\mathrm{d}^2\sigma^{\uparrow\Downarrow} - \mathrm{d}^2\sigma^{\uparrow\Uparrow}}{\mathrm{d}x\,\mathrm{d}Q^2} = \frac{8\pi\alpha^2}{Q^4}\frac{y}{E}\left[(E + E'cos\theta)g_1(x,Q^2) - \frac{Q^2}{\nu}g_2(x,Q^2)\right] \tag{2.21}$$

where $g_{1,2}(x,Q^2)$ are the dimensionless polarised structure functions of the nucleon. The orientation of the lepton spin is indicated with a single arrow and the spin of the target with a double arrow. Just as the unpolarised structure functions can be interpreted in the QPM, so can the polarised ones. The unpolarised PDFs can be rewritten and the polarised PDFs are introduced as

$$q_f(x,Q^2) = q_f^+(x,Q^2) + q_f^-(x,Q^2) \tag{2.22}$$

$$\Delta q_f(x,Q^2) = q_f^+(x,Q^2) - q_f^-(x,Q^2) \tag{2.23}$$

where $q^{+(-)}(x,Q^2)$ denotes the number density of a quark with flavour f, momentum fraction x and spin parallel (anti-parallel) to the nucleon spin. Extending the analogy with (2.14) and (2.15) the polarised structure functions $g_{1,2}$ can be written as

$$g_1(x,Q^2) = \frac{1}{2}\sum_f e_f^2 \Delta q_f(x,Q^2) \tag{2.24}$$

$$g_2(x,Q^2) = 0. \tag{2.25}$$

where the summation runs again over the flavours of the valence and sea quarks. This allows a transparent interpretation of the polarised structure function $g_1(x)$ in the QPM, as illustrated in fig. 2.5: in the absorption process of a virtual photon by a quark, the photon can only couple to quarks whose spin is aligned opposite to the spin of the photon because helicity must be conserved and because a quark can only have spin 1/2 and not spin 3/2. In both panels of fig. 2.5, the incoming polarised lepton at the left emits a virtual photon with spin projection "pointing" to the right in this picture. In the top panel of fig. 2.5, the target polarisation points to the left so that we measure the cross section for photon and target polarisations anti-parallel. This cross section is labelled[5] $\sigma_{1/2}$ and is sensitive to the distribution of quarks with their spins in the same direction as the spin of the nucleon, or $q_f^+(x)$. In the bottom panel of fig. 2.5 the target spin direction has been reversed so that we measure the cross section for photon and target polarisations parallel ($\sigma_{3/2}$). However, the elementary process is not changed. The photon is absorbed by a quark with spin projection pointing to the left and this process is sensitive to $q_f^-(x)$. Since g_1(x) is related to the difference of $q_f^+(x)$ and $q_f^-(x)$ it can be deduced by measuring the cross section difference $\sigma_{1/2} - \sigma_{3/2}$, as will be explained later in this section.

The structure function $g_2(x)$ can be interpreted in terms of binding effects in the nucleon and it is related to transverse degrees of freedom of the quarks in the nucleon [1]. In the QPM that considers non-interacting partons moving collinearly within the nucleon, the structure function $g_2(x)$ has no simple interpretation and vanishes. Experiments [10] have shown that g_2 is indeed very small, so that it usually can be neglected.

Information about spin degrees of freedom can be obtained from measurements of cross section asymmetries, instead of cross section differences, that are more difficult to access from the experimental point of view. The cross section asymmetry for a longitudinally polarised lepton beam and nucleon target is defined as:

$$A_{\|} = \frac{\sigma^{\uparrow\Downarrow} - \sigma^{\uparrow\Uparrow}}{\sigma^{\uparrow\Downarrow} + \sigma^{\uparrow\Uparrow}} \tag{2.26}$$

Provided the time intervals between the flipping of the target spin are short enough, efficiency and acceptance effects which are not correlated to the relative orientation of the beam and target spins, cancel out by measuring cross section asymmetries instead of cross section differences. Thus, asymmetry measurements are less susceptible to systematic effects than measurements of differences of absolute cross sections. The asymmetry $A_{\|}$ can be measured by simply counting[6] the number of events when the beam and target polarisations are in the same ($N^{\uparrow\Uparrow}$) and opposite ($N^{\uparrow\Downarrow}$) direction

$$A_{\text{meas}} = \frac{N^{\uparrow\Downarrow} - N^{\uparrow\Uparrow}}{N^{\uparrow\Downarrow} + N^{\uparrow\Uparrow}} \tag{2.27}$$

Multiplication of this measured asymmetry by a factor that accounts for experimental inefficiencies (limited polarisations of beam and target and the fraction of nucleons in the target that can be polarised) yields the true asymmetry $A_{\|}$.

In lowest order, the fundamental process in deep-inelastic scattering is the interaction of a virtual photon γ^* with the target nucleon N and the physically significant asymmetry

[5] The subscript 1/2 refers to the net spin of the virtual photon and target, in order to avoid confusion with the notation $\sigma^{\uparrow\Downarrow}$ where the arrows refer to the spin of the lepton and target.

[6] For simplicity, the luminosity for both spin-states has been assumed to be equal.

for photon absorption on the nucleon is given by

$$A_1 = \frac{\sigma_{1/2} - \sigma_{3/2}}{\sigma_{1/2} + \sigma_{3/2}} \tag{2.28}$$

It can be shown [9] that A_1 and $A_{\parallel}$ are connected by a kinematical factor

$$A_{\parallel} \approx DA_1 \tag{2.29}$$

where the depolarisation factor $D(y, R(x, Q^2))$ accounts for the fact that the lepton does not transfer its complete polarisation to the virtual photon. Finally, the photon asymmetry is related to the structure function g_1 [9] as

$$A_1 \approx \frac{g_1}{F_1} \tag{2.30}$$

Using eq. (2.18) we can therefore derive g_1 from the longitudinal spin asymmetry $A_{\parallel}$ experimentally determined, the known unpolarised structure functions F_2 and R and the depolarisation factor D (which can be calculated from the measured kinematics):

$$g_1(x, Q^2) = \frac{A_{\parallel}(x, Q^2)}{D(y, R(x, Q^2))} \frac{F_2(x, Q^2)(1 + \gamma^2)}{2x(1 + R(x, Q^2))} \tag{2.31}$$

The structure function g_1 has been measured for proton, deuterium and ^{3}He targets by several experiments (EMC, SMC, HERMES, SLAC E143, SLAC E154), and the interpretation of these measurements can be interpreted in terms of sum rules which will be described in the next section.

2.3.4 Sum rules

Up to now most of the theoretical models fail to make precise predictions for the x-dependence of the structure functions. The main reason for this is the breakdown of perturbative QCD calculations at low energy scales due to the increasing α_s according to eq. (2.17). However, the same models often can give information about the value of n-th moment of a certain structure function $f(x)$ that is defined as $\int_0^1 x^{n-1} f(x) dx$, which then can be compared to the corresponding experimental results. Unfortunately, the distributions measured experimentally cannot cover the full x-integration range and have to be extrapolated into some unmeasured region. The models used for this extrapolation introduce an additional uncertainty on the moments quoted by the experimentalists. In the following a brief overview of the most important sum rules is given.

2.3.4.1 Gottfried sum rule

The parton density functions are usually split into valence and sea contributions

$$q(x) = q_v(x) + q_s(x) \tag{2.32}$$

Because of the symmetric production mechanism of the sea, the relation

$$\int_0^1 q_s(x) dx = \int_0^1 \bar{q}_s(x) dx \tag{2.33}$$

holds for all flavours. As the PDFs give the number densities of quarks inside the nucleus, integration of the valence quark PDFs for the proton gives

$$\int_0^1 u_v(x)dx = 2 \qquad \int_0^1 d_v(x)dx = 1 \tag{2.34}$$

The assumption of isospin symmetry between the proton and the neutron leads to the following relations (the subscript indicates the nucleon type):

$$u(x) \equiv u_p(x) = d_n(x) \tag{2.35}$$

$$d(x) \equiv d_p(x) = u_n(x) \tag{2.36}$$

$$s(x) \equiv s_p(x) = s_n(x) \tag{2.37}$$

Using the above relations, one can easily calculate the 0^{th} moment of the structure function difference $F_2^p(x) - F_2^n(x)$, known as the Gottfried sum rule [11]:

$$S_G = \int_0^1 \frac{1}{x}\left(F_2^p(x) - F_2^n(x)\right)dx = \frac{1}{3} + \frac{2}{3}\int_0^1 [u_s(x) - d_s(x)]dx \tag{2.38}$$

A flavour symmetric light quark sea, $u_s(x) = d_s(x)$, leads to $S_G = 1/3$. A measurement of the Gottfried Sum by NMC [2] resulted in $S_G = 0.2281 \pm 0.0065$. If isospin symmetry holds, a global light flavour asymmetry $\int_0^1 [u_s(x) - d_s(x)]dx \approx -0.16$ would account for the NMC result. The light flavour asymmetry has been confirmed by HERMES [12] and was found to be x-dependent, showing an excess of d_s over u_s-quarks in the proton.

2.3.4.2 Bjorken sum rule

A similar sum rule for the difference of the spin structure functions $g_1(x)$ of the proton and neutron is their first moment, which can be derived from fundamental QCD principles to be:

$$\int_0^1 \left(g_1^p(x) - g_1^n(x)\right)dx = \frac{1}{6}\left|\frac{g_A}{g_V}\right| + \text{QCD corrections} \tag{2.39}$$

The Bjorken sum rule [13] is expressed in terms of basic quantities, g_A/g_V is the ratio of axial- and vector coupling in the weak neutron β-decay. At a scale of Q^2= 5 GeV^2 one obtains a value for the Bjorken sum rule of 0.182 ± 0.003. From the combined results of EMC, SMC, and SLAC E143 data, an experimental value of 0.170 ± 0.012 [14] was obtained, thus confirming the theoretical prediction within the experimental and theoretical uncertainties. This provides a stringent verification of perturbative QCD.

2.3.4.3 Ellis-Jaffe sum rule

Another spin-related sum rule is the Ellis-Jaffe sum rule which predicts the first moment of $g_1(x)$ for the proton and neutron separately [15]. Additional assumptions have to be made, namely SU(3) flavour symmetry and the assumption that the strange sea is unpolarised:

$$\int_0^1 g_1^p(x)dx = 0.164 \pm 0.005 \qquad \int_0^1 g_1^n(x)dx = -0.018 \pm 0.004 \tag{2.40}$$

In 1987 EMC published a value of $0.126\pm 0.010\pm 0.015$ for the Ellis-Jaffe proton sum rule. This measurement was later confirmed by measurements at SLAC. The most plausible explanation for the violation of the Ellis-Jaffe sum rule and the experimental confirmation of the Bjorken sum rule is that the main assumption made in the Ellis-Jaffe sum rule (i.e. an unpolarised strange sea) is wrong. This assumption was not made for the Bjorken sum rule, since the strange contribution vanishes in the difference between proton and neutron.

The violation of the Ellis-Jaffe sum rule was a first hint that the simple picture that the nucleon spin is the quantum mechanical sum of only the three valence quark spins is not realised in nature. In general, from conservation of angular momentum, all the possible contributions to the nucleon spin are

$$\frac{1}{2} = \frac{1}{2}(\Delta u + \Delta\bar{u} + \Delta d + \Delta\bar{d} + \Delta s + \Delta\bar{s}) + \Delta G + L_z \tag{2.41}$$

where L_z is the orbital angular momentum of partons in the nucleon and ΔG is the spin contribution from the gluons. The inclusive measurement of g_1 for one nucleon is only sensitive to the charge weighted sum of all quark flavours and not to the individual quark distributions. The semi-inclusive reaction, however, can measure all quark polarisations in a more direct manner without the necessity for the same theoretical assumptions as used for the analysis of the inclusive channel. The semi-inclusive scattering will be discussed in the next sections.

2.4 Semi-inclusive DIS

2.4.1 Introduction

In the previous sections it has been shown that inclusive deep-inelastic scattering has an easy interpretation in the framework of the QCD corrected QPM: there are only four structure functions: F_1 and F_2 (connected by R) in the unpolarised case and g_1 and g_2 in the polarised case. Both F_2 and R have been measured, yielding a rather good understanding of the unpolarised DIS process. In polarised inclusive scattering the only relevant structure function in first order is g_1 which has been measured as well. However, the inclusive structure functions do not constrain the contributions from all individual quark flavours to the momentum and spin of the nucleon. This information is contained in the flavour of the quark that is struck in the DIS process and can be accessed by detecting at least one final state hadron in coincidence with the scattered lepton. This process is semi-inclusive DIS.

As depicted in fig. 2.1 the semi-inclusive DIS process is described in the QPM as a single quark which is struck by the virtual photon and ejected from the nucleon. Due to the *confinement* requirement of QCD, which demands that only colour neutral states exist as free particles, both the struck quark and the target remnants have to form a set of colour neutral final state hadrons. This hadronisation process involves long-range interactions between the struck quark and the target remnant as they move apart. Since long-range interactions are equivalent to low values of Q^2, they involve large values of the coupling constant $\alpha_s(Q^2)$. Hence the perturbative expansions of QCD in powers of α_s are not longer justified and QCD cannot describe the hadronisation. Instead the hadronisation process is parametrised by fragmentation functions $D_f^h(z, Q^2)$ which give the probability density that a struck quark of flavour f probed at a particular scale Q^2, produces a final state hadron h with fractional energy z. Since the total energy of the produced hadrons needs to conserve the energy of the virtual photon, the fragmentation functions are normalised:

$$\sum_h \int_0^1 z \cdot D_f^h(z, Q^2) dz = 1 \tag{2.42}$$

Fragmentation functions are studied by selecting the current fragmentation region, because the correlation between the struck quark and the detected hadron needs to be enhanced. In the next sections, the properties of the rather phenomenological fragmentation functions will be described and a physically motivated fragmentation model will be discussed. Fragmentation functions will be frequently used in the physics analysis described later in this thesis.

2.4.2 Properties of fragmentation functions

The *factorisation* theorem states that the quark fragmentation into hadrons is independent of the photon-quark scattering process. Factorisation naturally implies that the fragmentation process is assumed to be universal, i.e. independent of the original environment of the initial quark, which starts the hadronisation process. This *environmental independence* was experimentally confirmed by comparing fragmentation functions determined in DIS with hadronisation processes from $q\bar{q}$ pair production in e^+e^-collisions [16, 17]. The same data support the independence of x, while earlier data showed [18, 19] that fragmentation functions vary with Q^2. Although QCD is unable to describe the fragmentation itself, the Q^2-evolution of the fragmentation functions can be calculated by perturbative QCD and a prediction can be made for the scaling violations of the fragmentation functions [20]. The underlying mechanism for scaling violations in QCD is gluon radiation from the scattered quark. Some evidence for a slight x-dependence has been found in the older EMC data [21, 22] and more recent HERMES data [23]. However, since the variables x, Q^2 and W^2 are correlated ($W^2 = M^2 + Q^2[1/x - 1]$), the x-dependence for fixed W^2 can also be interpreted as a variation with Q^2 for fixed W^2, so that it is experimentally difficult to distinguish between scaling violations (Q^2-dependence of D_f^h) and a breakdown of factorisation (x-dependence of D_f^h).

At HERMES energies, the fragmentation of six quarks and anti-quarks (u, d, s, $\bar{u}$, $\bar{d}$, $\bar{s}$) dominantly produces six charged hadrons: π^+, π^-, K^+, K^- , p and $\bar{p}$. The strange quarks are suppressed compared to up and down quarks because of the extra mass which must be generated to form a $s\bar{s}$-pair, while the creation of heavier charm, bottom and top quark pairs can be neglected in the fragmentation process. Thirtysix independent fragmentation functions then describe semi-inclusive particle production. However, symmetry assumptions are usually employed to reduce the number of independent fragmentation functions. In the case of charged pions, using charge conjugation invariance ($D_u^{\pi^+} = D_{\bar{u}}^{\pi^-}$) and isospin symmetry ($D_u^{\pi^+} = D_d^{\pi^-}$), it is possible to define the following functions:

$$D^+ \equiv D_u^{\pi^+} = D_{\bar{d}}^{\pi^+} = D_d^{\pi^-} = D_{\bar{u}}^{\pi^-} \tag{2.43}$$

$$D^- \equiv D_d^{\pi^+} = D_{\bar{u}}^{\pi^+} = D_u^{\pi^-} = D_{\bar{d}}^{\pi^-} \tag{2.44}$$
$$D^s \equiv D_s^{\pi^+} = D_{\bar{s}}^{\pi^+} = D_s^{\pi^-} = D_{\bar{s}}^{\pi^-} \tag{2.45}$$

The functions $D^+(z, Q^2)$, $D^-(z, Q^2)$ and $D^s(z, Q^2)$ are called *favoured*, *unfavoured*, and *strange* (pion) fragmentation functions, respectively. They obey the hierarchical relationship $D^+ > D^- > D^s$. The favoured fragmentation function relates to pions, which have the initial quark in their ground state wavefunction. Such fragmentation processes are more probable than the unfavoured or strange quark cases. The D^+ and D^- fragmentation functions have been measured in semi-inclusive DIS with pions in the forward region by EMC [24] and HERMES [23, 25].

Finally, the concept of fragmentation functions allows to relate the (unpolarised) semi-inclusive cross section to the expression for the inclusive cross section, given in eq. (2.20), by

$$\frac{\mathrm{d}^3\bar{\sigma}^h}{\mathrm{d}x\,\mathrm{d}Q^2\,\mathrm{d}z} = \frac{\sum_f e_f^2 q_f(x, Q^2) D_f^h(z, Q^2)}{\sum_f e_f^2 q_f(x, Q^2)} \cdot \frac{\mathrm{d}^2\sigma}{\mathrm{d}x\,\mathrm{d}Q^2} \tag{2.46}$$

This expression explicitly displays the factorisation property, as the cross section is splitted into two terms which depend separately on x and z. The notation $\bar{\sigma}$ indicates that we have integrated the two other hadron variables required to describe semi-inclusive DIS.

2.4.3 The Lund fragmentation model

Since it is not possible to calculate the fragmentation functions in perturbative QCD directly from first principles, several phenomenological models have been developed to describe the fragmentation process. Historically the need for explanations of the observed hadronic final states in e^+e^- annihilation triggered the development of various models, generally realised by Monte Carlo simulations. The most important approaches are the independent fragmentation model of Field and Feynman [26], the cluster fragmentation [27], and the Lund string model [28]. The underlying model for the Monte Carlo studies in this work was the Lund model and only this model will be described here. The Lund string model is nowadays the most successful model to match the experimental fragmentation data. It is implemented into the JETSET Monte Carlo generator [29] and adjustable parameters have been tuned [25, 30, 31] to reproduce the HERMES measured particle multiplicities and distributions of kinematical spectra.

The starting point for the Lund model is the QCD prediction of confinement, which implies that free coloured quarks cannot exist. In the case of DIS, the Lund model does not treat the struck quark independently but considers it as connected to the nucleon remnant via the colour field. While the struck quark and the target remnant move apart from each other, the colour field is stretched in between them. This field can be imagined as a cylindrical flux tube or simply as a *string* between the quark and the anti-quark. The string may be approximated by a long distance (r) linear potential $V(r)$ with a strength κr, where κ is the string constant with a value of about 1 GeV/fm, i.e. the amount of field energy stored per unit length. As the string stretches, the potential energy builds up and when the stored energy exceeds the mass of a $q\bar{q}$-pair, such a pair is produced and the string is divided into two substrings. One of the substrings is still connected to the struck quark, the other to the target remnant. For each string the process of stretching and breaking continues until the on-shell mass of a hadron is reached.

In a quantum mechanical sense the quark and the anti-quark are produced in one point and tunnel out of their production point with some finite longitudinal and transverse momentum gained from the energy of the field. The probability for a $q\bar{q}$ pair with quark mass m and quark transverse momentum $p_\perp$ is proportional to:

$$\exp\left(-\frac{\pi m_\perp^2}{\kappa}\right) \tag{2.47}$$

where $m_\perp^2 = p_\perp^2 + m^2$ is the transverse mass of the quarks. The string itself has no transverse momentum, so the $p_\perp$ of each quark produced in a pair is balanced by its partner. The Lund model assumes $p_\perp$ to be distributed Gaussian and the width of this Gaussian is a parameter that has to be determined from experimental data. The resulting hadrons will have a transverse momentum $p_t \equiv P_h^T$ made up by the sum of the $p_\perp$ of their quarks.

The decision which final state hadron type is chosen for a certain string depends on the relative occurrence of quark flavours participitating in the $q\bar{q}$ production. This is governed by the quark masses and eq. (2.47), giving $u\bar{u}$: $d\bar{d}$: $s\bar{s} = 1$: 1: 0.3. This mechanism can be controlled in the LUND model by changing the probability for the production of strange quarks compared to non-strange quarks by adjusting the so-called strangeness suppression factor.

The way in which the (longitudinal) energy fraction z is assigned to the hadrons is defined by the so-called 'Lund symmetric fragmentation function'

$$f(z) = z^{-1}(1-z)^a \exp\left(-\frac{bm_\perp^2}{z}\right) \tag{2.48}$$

where a and b are free parameters that in principle depend on the quark flavour.

2.4.4 Transverse momentum of hadrons

An important parameter of hadrons is their transverse momentum, that was defined by eq. (2.11) and illustrated in fig. 2.2. In general there are four different ways in which hadrons can acquire transverse momentum p_t in deep-inelastic scattering:

- p_t can be accumulated in the fragmentation process, as explained in the previous section.

- The intrinsic transverse momentum of the struck quark (k_T) can also contribute. The concept of intrinsic k_T was originally introduced to take corrections to the Breit frame approximation into account, and a value of about 0.3 GeV could be understood in terms of the Heisenberg uncertainty relation for quarks confined in the proton.

- In the Lund model without gluons, the colour force field is treated like a string spanned between the colour charges and stretched from the quark to the target

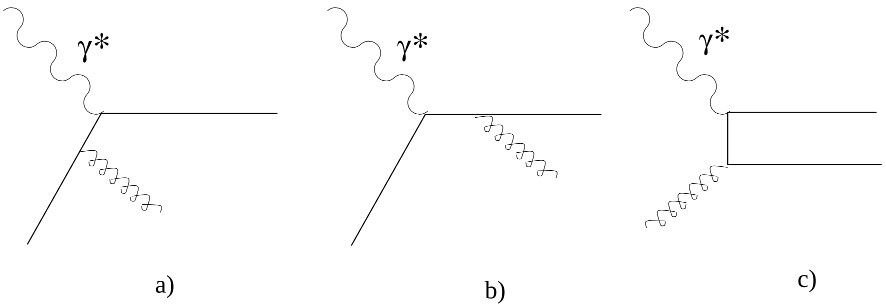

Figure 2.6: Feynman diagrams of first order QCD corrections to deep-inelastic scattering: a) and b) represent initial and final state gluon Bremsstrahlung, c) shows the photon-gluon fusion process where the gluon splits in a virtual quark anti-quark pair, one of which absorbs the virtual photon.

remnant, as explained in the previous section. Gluons are introduced as energy and momentum carrying kinks on the string, that is now stretched from the quark via the gluon to the target remnant. These kinks provide a source of transverse momentum with respect to the virtual photon. These processes generate in general hadrons with a large p_t.

To first order, perturbative QCD corrections introduce two processes of this kind, gluon Bremsstrahlung and photon-gluon fusion, as illustrated in fig. 2.6. Events arising from these processes appear as "three jet events": the remainder of the nucleon that fragments, and two jets that arise from the quark and the gluon in the case of gluon Bremsstrahlung or the quark and anti-quark for the photon-gluon fusion. These latter two jets have a transverse momentum with respect to the virtual photon. Gluon radiation and photon-gluon fusion are called *hard* QCD processes since they involve the fragmentation of an extra parton.

- Radiated gluons (fig. 2.6 a and b) which are too *soft* to give rise to separate jets of particles can still have effects on the p_t of hadrons. For a soft gluon the energy in the field between e.g. that gluon and the struck quark is too small to create quark anti-quark pairs, hence the string does not break, resulting in a 2-jet configuration just like for the leading order process that does not contain gluon radiation. However, the bend in the colour field due to the (soft) gluons gives rise to some extra p_t of the hadrons. But in general this effect is small.

Experimentally disentangling these contributions is extremely difficult, if not impossible. For example, intrinsic k_T is only present in DIS and not in e^-e^+ annihilation. Therefore a comparison between the average $< p_t >$ of hadrons in these two processes could allow to determine k_T. However, two effects complicate this comparison [32]. First, the magnitude of the QCD effects in annihilation is not the same as in lepton-nucleon scattering. Second, the methods used to determine the jet axis are different[7]. Hence no firm conclusions about

[7]In DIS, hadron transverse momenta are measured with respect to a predetermined direction: the momentum of the photon, while in e^-e^+ annihilation one looks event per event for an axis (*thrust axis*) to which the sum of the transverse momenta of the secondary particles is minimum.

the primordial k_T can be drawn by comparing the two sets of data. A more successful method to study k_T is the so-called *prompt photon production* process that was studied e.g. by the ZEUS experiment [33] using the positron-proton collider HERA. When positrons collide with protons accelerated in the opposite direction, most of the interactions take place by the positron first radiating a quasi-real photon (i.e. $Q^2 \approx 0$) which then interacts with the proton. This photon can hit a quark in the proton and bounce off, while the quark rebounds in the opposite direction. If the quark were initially stationary, the prompt photon and the jet from quark fragmentation would emerge exactly at opposite angles. The initial quark momentum shows itself in that the prompt photon is found to be not exactly aligned with the jet direction, but pushed to one side. In ZEUS, the distribution of this sideways momentum of the photon has been measured and related to a value for $k_T = 1.69$ GeV. The fixed target Tevatron experiments [34, 35] found a k_T around 1 GeV, and the Tevatron collider data [36] indicate that k_T is around 4 GeV. These extremely large values (compared to the few hundred MeV expected from the uncertainty relation) are not yet fully understood, but recent QCD theoretical calculations [37] relate the k_T-broadening to the fact that intrinsic k_T and soft gluon radiation can not be disentangled and yield a combined 'effective' k_T.

To conclude we can state that a complete understanding of how hadrons build up their p_t is not yet available. Experimental techniques cannot fully distinguish between the different mechanisms and only the aid of Monte Carlo simulations can shed more light on this topic: by varying different parameters steering different effects one can compare the results from the simulation with the experimental data, and gain more insight in the magnitude of the different contributions. In the case of HERMES, the Lund model is the workhorse.

2.4.5 Polarised quark distributions

The power of the Lund model to describe the fragmentation process has been exploited maximally by HERMES to determine the contributions of the individual quark flavours to the nucleon spin. The determination of Δq_f relies on the relations between experimentally measured semi-inclusive hadron asymmetries and so-called purities (based on the Lund model), both of which will be defined below.

The measured semi-inclusive asymmetry A^h_{meas} for a certain hadron h can be defined in analogy with eq. (2.27) as

$$A^h_{\text{meas}} = \frac{N_h^{\uparrow\Downarrow} - N_h^{\uparrow\Uparrow}}{N_h^{\uparrow\Downarrow} + N_h^{\uparrow\Uparrow}} \tag{2.49}$$

where $N_h^{\uparrow\Uparrow}$ ($N_h^{\uparrow\Downarrow}$) is the observed number of semi-inclusive hadrons when the incoming lepton beam and target are in the same (opposite) spin-polarisation state. As for the inclusive case, the cross section asymmetry $A^h_{\|}$ can be obtained from A^h_{meas} by accounting for experimental inefficiencies, and in analogy with eq. (2.29) the physically relevant photon-nucleon asymmetry A^h_1 can then be obtained by taking the depolarisation factor D into account. This latter asymmetry can now be related to the quark polarisations Δq_f if it is assumed that the fragmentation process is spin-independent, i.e. the probability to produce a hadron h from a struck quark q_f is independent of the relative orientation of the quark spin and the nucleon spin. With this assumption it is possible to define generalised structure functions F_1^h and g_1^h relevant to the semi-inclusive scattering. In the QPM, according to (2.14) and (2.24), these may be written as

$$F_1^h(x,Q^2,z) = \frac{1}{2}\sum_f e_f^2 q_f(x,Q^2) D_f^h(z,Q^2) \tag{2.50}$$

$$g_1^h(x,Q^2,z) = \frac{1}{2}\sum_f e_f^2 \Delta q_f(x,Q^2) D_f^h(z,Q^2) \tag{2.51}$$

Assuming $g_2^h = 0$, the z-integrated semi-inclusive spin asymmetries are related to the polarised and unpolarised quark distributions and fragmentation functions in analogy with the inclusive case (2.30) as follows

$$A_1^h(x,Q^2) = \frac{\int g_1^h(x,Q^2,z)\mathrm{d}z}{\int F_1^h(x,Q^2,z)\mathrm{d}z} = \frac{\sum_f e_f^2 \Delta q_f(x,Q^2)\int D_f^h(z,Q^2)\mathrm{d}z}{\sum_f e_f^2 q_f(x,Q^2)\int D_f^h(z,Q^2)\mathrm{d}z} \tag{2.52}$$

This equation can be rewritten as follows:

$$A_1^h = \sum_f P_f^h \frac{\Delta q_f}{q_f} \tag{2.53}$$

where we have factored out the quark polarisations ($\Delta q/q$) and the *purities*, which can be expressed as:

$$P_f^h(x,Q^2) = \frac{e_f^2 q_f(x,Q^2)\int D_f^h(z,Q^2)\mathrm{d}z}{\sum_{f'} e_{f'}^2 q_{f'}(x,Q^2)\int D_{f'}^h(z,Q^2)\mathrm{d}z} \tag{2.54}$$

P_f^h represents the conditional probability that an observed hadron of type h was produced by a struck quark of flavour f with momentum fraction x. In a sense it is the inverse of a fragmentation function.

The HERMES experiment has measured inclusive asymmetries on hydrogen and deuteron targets and semi-inclusive asymmetries for pions and kaons on the same targets, which can be collected in a vector $\vec{A}$:

$$\vec{A} = (A_{1,p}, A_{1,d}, A_{1,p}^{\pi^\pm}, A_{1,d}^{\pi^\pm} A_{1,d}^{K^\pm}) \tag{2.55}$$

If the vector of quark polarisations

$$\vec{Q} = (\frac{\Delta u}{u}, \frac{\Delta d}{d}, \frac{\Delta \bar{u}}{\bar{u}}, \frac{\Delta \bar{d}}{\bar{d}}, \frac{\Delta s + \Delta \bar{s}}{s + \bar{s}}) \tag{2.56}$$

is introduced, eq. (2.53) can be written in matrix form

$$\vec{A} = \mathsf{P} \cdot \vec{Q} \tag{2.57}$$

where the concept of a purity is generalised to inclusive scattering by setting the fragmentation functions to 1 in eq. (2.54).

The purity matrix P can be determined by a Monte Carlo simulation based on the Lund model: each element P_f^h was determined as $N_f^h / \sum_{f'} N_{f'}^h$ where N_f^h is the number of

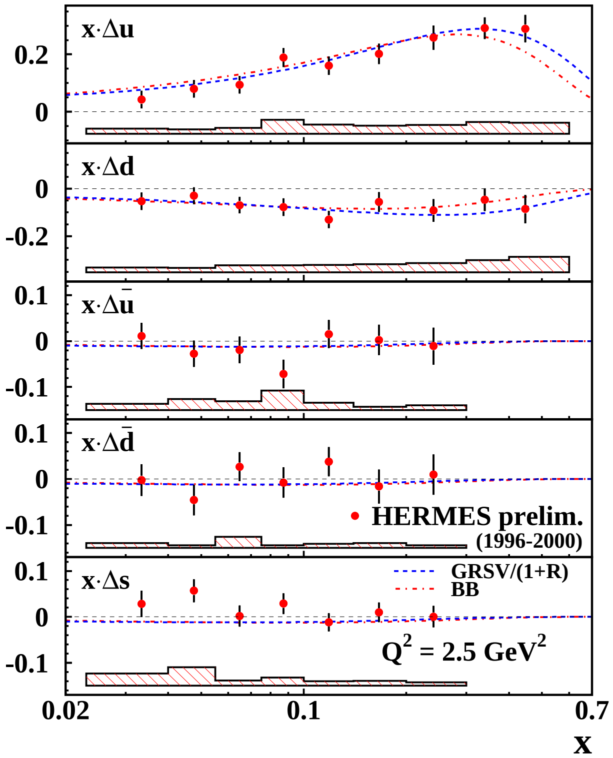

Figure 2.7: The x-weighted polarised quark densities as determined by HERMES. The plots show the result of a five parameter fit to the data assuming a symmetric strange sea polarisation ($\Delta s/s = \Delta\bar{s}/\bar{s}$). The data have been evolved to a common Q^2=2.5 GeV. Two NLO QCD fits to the world inclusive data are shown by the dashed [39] and the dashed-dotted [40] line.

hadrons of type h, in the geometric experimental acceptance, that were produced[8] from quarks of flavour f. The overconstrained system of equations (2.57) with 8 asymmetries was solved by a so-called *five parameter fit*, yielding the polarised quark distributions as shown in fig. 2.7. The polarisation of the up quarks is positive with the largest polarisation at high x where the valence quarks dominate. The down quark polarisation is negative over the measured x-range. The polarisation of the light sea flavours $\bar{u}$ and $\bar{d}$ is consistent with zero. The result for the strange quarks, however, favours a slightly positive polarisation at small values of x.

[8]The determination of purities requires the unpolarised quark density functions to be known in the Monte Carlo simulation. Several parametrisations of q_f exist, based on measured cross sections in a variety of processes (DIS and related hard scattering processes) and theoretical considerations. The existing parametrisations were fit to the world data on F_2. Since the formalism yielding eq. (2.53) started from the structure function F_1, an extra kinematical factor $C_R = [1 + R(x, Q^2)]/[1 + \gamma^2]$ (based on eq. (2.18)) was introduced [38] in the right hand side of eq. (2.53).

2.5 The scope of this work

The intention of this thesis is to study fragmentation processes in the framework of semi-inclusive DIS. The experimental tool to examine the DIS process will be described in the next chapter: a high energy positron beam that circulates in the HERA storage ring and collides with a gaseous target, producing scattered leptons and hadron fragments that are both detected in the HERMES spectrometer. Up to 1998 this spectrometer could only discriminate between scattered leptons and produced hadrons without the ability to identify different hadron types. The identification of pions, kaons and protons is possible since 1998 due to the installation of a so-called RICH detector. In chapter 4 a detailed description of the algorithms that were developed to identify hadrons with the HERMES RICH will be given. An overview of the implementation of the RICH simulation program which describes the experimental data well, will also be presented.

After this description of the 'detector-related' results of this thesis, the next chapters will place emphasis on physics. In Chapter 5 the structure function ratio $F_2^d(x)/F_2^p(x)$ will be extracted from experimental data. Although this ratio is rather well known from earlier world data, this study is nevertheless interesting since it will confirm the agreement of HERMES data with existing parametrisations of this ratio. Moreover, the determination of the F_2-ratio involves major corrections to the experimental data, hence this (inclusive) analysis provides a stringent check for the consistency of these corrections and the validity of the analysis method. In Chapter 6 the inclusive extraction method will be extended to the semi-inclusive case. Here the use of the RICH detector will play a major role and also the developed RICH simulation program will enter in the picture. Three different semi-inclusive topics will be studied: the ratio of *down* and *up* valence quarks as a function of Bjorken x ($d_v(x)/u_v(x)$), the polarisation of the strange sea quarks Δs and the transverse momentum distributions of pions, kaons and protons.

The valence quark distributions $d_v(x)$ and $u_v(x)$ have been parametrised by different theoretical groups, based on existing world data. Since the parametrisations differ substantially, the extraction of their ratio (section 6.2) can constrain the theoretical fits and help to determine the errors on the parametrisations, which are currently not available.

The strange quark polarisation $\Delta s(x)$ has already been determined by HERMES using the five-parameter fit that relies on the purities from the Lund model (section 2.4.5). In section 6.3, an attempt will be made to determine Δs without these purities, starting from simple basic assumptions such as symmetry relations between fragmentation functions (cfr. eq. (2.45)). It will be shown that this approach does not work and that the failure can be related to the ab initio assumptions.

A measurement of the semi-inclusive transverse momentum spectra for positively and negatively charged pions, kaons and protons will be presented in section 6.4. The shape of these spectra is closely related to adjustable parameters in the Lund model, hence a measurement of these distributions is particularly useful for improving the Lund description at HERMES energies. Moreover, in earlier efforts [31] to tune these parameters only combined experimental HERMES data for *all* hadrons were used, because at that time HERMES could not yet discriminate between pions, kaons and protons. Using the RICH detector it will be possible to disentangle the transverse momentum spectra with respect to the different hadron types. This will be important for future Lund tunings at HERMES.

Chapter 3

The HERMES experiment

The HERMES detector is located in the east section of the HERA collider at DESY in Hamburg, Germany. The experiment was proposed in 1990 and approved in 1993, the installation was carried out in the 1994/1995 HERA winter shutdown and data taking started in spring 1995.

The HERA storage ring facility consists of two separate accelerators. The proton ring was built for an energy of 820 GeV before and of 920 GeV after the accelerator upgrade in 1998. In the lepton ring polarised positrons (or electrons) of 27.5 GeV circulate in the opposite direction. Fig. 3.1 shows HERA with its four experiments. While the HERA lepton beam can be polarised, the current machine configuration does not allow a polarisation of the proton beam. As the primary goals of the HERMES experiment (the measurement of double spin asymmetries in inclusive and semi-inclusive deep-inelastic scattering) require both leptons and nucleons to be polarised, HERMES operates in a fixed target mode and only the 27.5 GeV polarised lepton beam is used, while the unpolarised proton beam passes through the apparatus without interaction.

The experiments H1 and ZEUS operate in colliding mode to study unpolarised deep-inelastic scattering in a kinematical region that is not accessible for fixed target experiments. HERA-B is a fixed target experiment and uses only the proton beam. This experiment was designed to use particles in the halo of the proton beam which are scattered off target wires at high luminosities. The intended goal was to study CP violation in B-meson systems. HERMES uses the lepton beam and an internal gas target.

As the data analysed in this work originate from the 2000 data taking period of HERMES, special emphasis in the following sections will be given to the experimental setup during that year. The reason why only the 2000 data set has been used, is that the RICH detector was only operational since the fall of 1998 and the 2000 data set has much higher statistics and better data quality as compared to the ones from 1998 and 1999. In the next sections the HERA lepton ring and the HERMES spectrometer will be described in more detail. Although the fact that HERMES is in principle a double polarised experiment (beam and target are polarised) is not relevant for most of the physics described in this work, some words will be spent on the polarisation to illustrate the unique nature of HERMES.

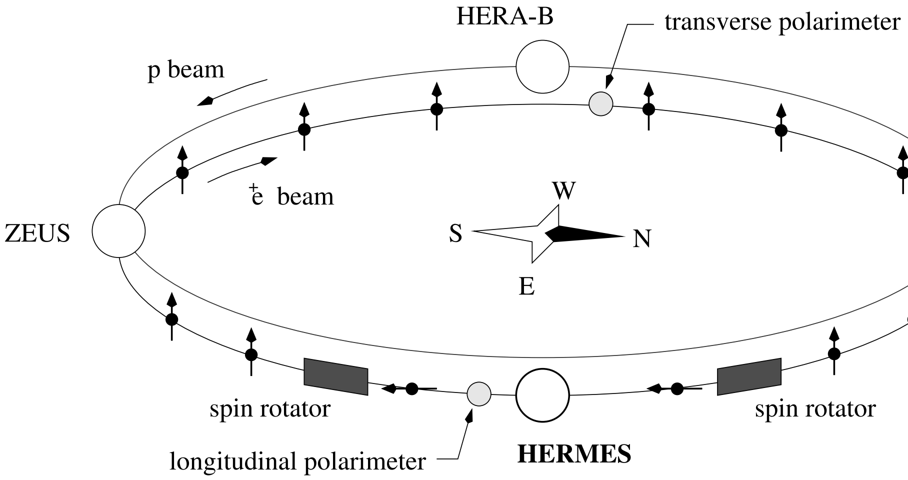

Figure 3.1: Schematic layout of the HERA accelerator. The positions of the four experiments making use of the HERA ring are shown, together with the positions of the 2 beam polarimeters and the spin rotators on either side of the HERMES experimental site.

3.1 The polarised lepton beam

HERA can store electrons as well as positrons. Since the lepton beam consisted of positrons in the year 2000, the word "positron" will be used to refer to the lepton beam in the following sections. The HERA ring has a circumference of 6.3 km and the positron beam is initially unpolarised after injection. However, positrons that circulate in the beam pipe become polarised transverse to their direction of motion. This is called the Sokolov-Ternov effect [41]: being deflected by the bending magnets in the accelerator, positrons emit synchrotron radiation. For most of the time the synchrotron radiation does not flip the spin of the positron. However, for positrons moving in a plane perpendicular to a uniform magnetic field, there is a small amplitude to flip the positron spin. This spin flip amplitude has a small associated asymmetry between the cases where the positron spin is oriented parallel or anti-parallel to the field. The spin flip to the parallel orientation proceeds more rapidly than the spin flip to the anti-parallel case [42]. As time goes by, a difference in the two spin configurations builds up, resulting in a net polarisation of the beam in the direction parallel to the magnetic field (i.e. perpendicular to the direction of motion of the positrons). This transverse polarisation of the beam rises exponentially with time

$$P(t) = P_{max}(1 - e^{-t/\tau}) \tag{3.1}$$

Here P_{max} stands for the theoretical polarisation limit. The characteristic polarisation rise time is denoted by τ. Both parameters depend on the beam energy and the accelerator radius. For an accelerator with energy 27.5 GeV and bending radius of 707 m with homogeneous magnetic field these values are $P_{max} = 92.4\%$, $\tau = 37$ min. In a real machine depolarising effects exist, like small misalignments of the magnets in the ring that cause non-uniformity of the magnetic fields, interactions between the lepton and proton beams and so-called spin diffusion (i.e. shifts in energy and position of the positron while it radiates a synchrotron photon, resulting in a spin tilt with respect to the magnetic field).

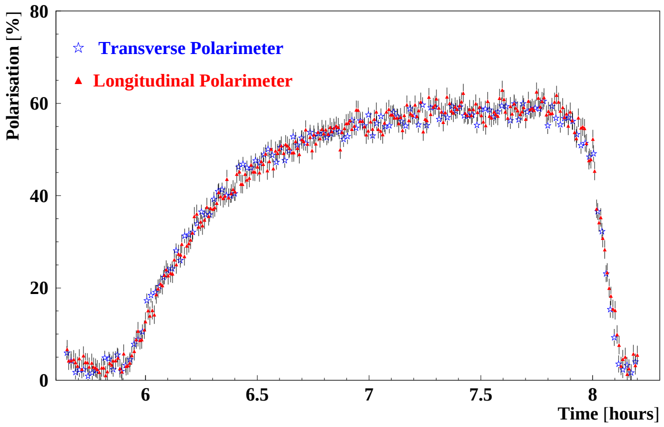

Figure 3.2: The rise time curves of the HERA lepton beam polarisation as measured by both the transverse and longitudinal Compton polarimeter.

These effects reduce the self-polarisation by the Sokolov-Ternov effect and hence decrease the maximum reachable polarisation at HERA to a value of about 70% with a rise time of about 20 min. The average beam polarisation achieved at HERA is about 55%.

The polarisation obtained by the Sokolov-Ternov effect is transverse, but most of the polarised physics processes studied at HERMES require a polarisation parallel to the direction of the positron momentum. To obtain a longitudinal polarisation of the beam two spin rotators are installed enclosing the HERMES interaction region, as shown in fig. 3.1. They consist of horizontal and vertical dipole magnets [43] which by means of successive rotations cause the direction of the spin to be longitudinal without affecting the position and slope of the beam. The spin rotator upstream of the experiment rotates the transverse polarisation of the beam into a longitudinal one. The spin rotator downstream of the experiment rotates the longitudinal polarisation back into a transverse one, ensuring that the positrons remain polarised throughout the storage ring. The helicity of the longitudinally polarised beam can be made positive or negative. Reversal of the beam helicity requires a mechanical shift of the rotator magnets and can only be performed during a short shutdown of the accelerator.

Since the polarisation of the beam is strongly time dependent, it has to be measured with good accuracy. Therefore, two independent polarimeters are installed which are also shown in fig. 3.1. Since no loss of polarisation is expected due to the spin rotators, the degree of transverse polarisation in the ring is equal to that of the longitudinal polarisation in the region of the HERMES experiment. Both polarimeters use photon positron (electron) scattering (Compton scattering). The transverse polarimeter (TPOL) [44] uses the spin dependence of the Compton cross section: the scattering of circularly polarised laser light off transversely polarised positrons shows an azimuthal dependence with respect to the positron spin direction. This yields a spatial up-down asymmetry of the backscattered photons in a position sensitive calorimeter. The transverse beam polarisation is determined from the magnitude of this asymmetry.

The longitudinal beam polarisation is measured with the longitudinal polarimeter (LPOL)

[45], positioned close to the HERMES experimental area. It was designed to obtain an independent, more precise measurement of the beam polarisation. Like the transverse polarimeter the longitudinal polarimeter makes use of asymmetries in the Compton scattering of a circularly polarised laser beam off the polarised positron beam. The longitudinal positron polarisation produces an asymmetry in the energy spectrum of the backscattered Compton photons. Energy weighted spin asymmetries can be formed and used to measure the beam polarisation. The exponential rise of the polarisation to its asymptotic value can be seen in fig. 3.2 where a comparison between measurements of the transverse and longitudinal polarimeter is given.

3.2 The internal gas target

The HERMES internal gas storage cell [46] utilises atomic gas particles for the target interactions. The target cell is an integral part of the positron beam pipe and is embedded in the target vacuum chamber. A schematic view of the HERMES target region is shown in fig. 3.3. The target cell is 40 cm long and has an elliptical profile which has the same shape as the beam profile. The wall material of the cell (aluminium) and exit window of the chamber (stainless steel) are as thin as possible to allow scattered particles to escape the target region with minimum disturbance. The ends of the target cell are open to allow the positron beam to pass through the target cell, interacting solely with the target gas. The target gas is injected through a tube at the center of the cell. The gas target thickness is approximately 100 times greater than compared to a free gas jet crossing the beam [47] since the interaction probability with the beam is large as the particles bounce off the walls numerous times before exiting the cell at either end. The density of the gas is increased by cooling the target cell to 100 K, which reduces the velocity of the target gas particles and increases the occupancy time in the cell. The target gas atoms which have left the cell are pumped away by a powerful pump system around the target area to ensure a marginal reduction of the life time of the stored positron beam due to the target gas in the beam path.

Two collimators protect the cell from synchrotron radiation, which may damage the target cell and increase the detector background. The first collimator (C1) is located 2 m upstream of the target and is movable to allow injection and tuning of the beam. The C1 consists in fact of two collimators, one split horizontally, the other vertically, which are closed to a position near the beam (but not close enough to affect the beam quality) during data taking and reduce the synchrotron radiation accompanying the positron beam. The second collimator (C2) is situated just upstream of the target (see fig. 3.3). It is designed to stop any radiation reflected from the C1 collimator.

Behind and in front of the storage cell, so-called wake-field suppressors are installed which provide a smooth transition between the cross section of the HERA beam tube and the storage cell in the target chamber. The bunched positron beam in HERA induces mirror currents in the walls of the beam pipe which cause the emission of strong radio frequency (rf) fields (*wake fields* [48]) at discontinuities in the impedance of the pipe. These rf fields can deposit a sizeable amount of energy in the target area, which both heats up the storage cell and the feedback with the positron beam itself destabilises the beam orbit. Thanks to the wake-field suppressors only a marginal load of rf fields is generated in the target area.

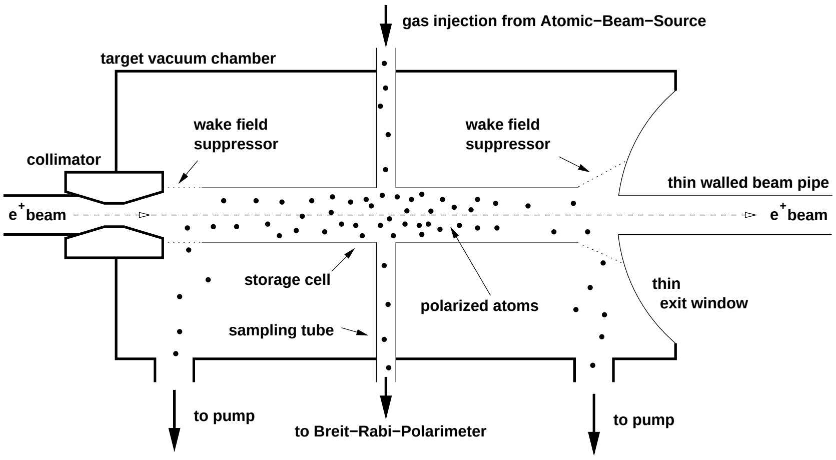

Figure 3.3: Schematic view of the target chamber and the storage cell. The C2 collimator is shown on the upstream side of the target. Wake field suppressors consisting of small metal tubes are used to join the target cell to the beam pipe and reduce heating from the wake fields.

The HERMES target uses polarised as well as unpolarised gases. Helium-3, hydrogen and deuterium have been used as polarised targets. Unpolarised targets of these same gases as well as nitrogen and krypton have been used. For the physics analysis in this thesis, only data collected during 2000 with the deuterium and hydrogen target were used. The deuterium data include both the polarised and unpolarised deuterium target gases. The target polarisation from the polarised deuterium target has been averaged (summed) over the polarisation values. Since the polarised target switches its polarisation states every 1-2 minutes, summing over both polarisation states effectively results in an unpolarised target. The majority of the data taken with the unpolarised deuterium and hydrogen were so-called high density data: at the end of a HERA fill, when the positron beam current is too low to be useful for the collider experiments, HERMES is allowed to increase its target density, reducing the beam life time to a few hours. This enables HERMES to collect a significant luminosity on unpolarised targets in a short time.

The unpolarised hydrogen and deuterium gas was fed from bottles of pure gas which provides a constant density in the target cell on the order of 10^{15} nucleons/cm^2 ($\geq 10^{16}$ nucleons/cm^2 in the case of 'end of fill' high density data). The higher density of the unpolarised target compared to the polarised target (10^{14} nucleons/cm^2) allows calibration measurements of the target and spectrometer to be made in a relatively short time, as well as providing data for unpolarised analyses.

Polarised deuterium is provided by the Atomic Beam Source (ABS) [49]. In the ABS dissociator, a strong radio frequency field splits deuterium molecules into atomic deuterium. Sextupole magnets polarise the electrons in deuterium via Stern-Gerlach separation. High frequency magnetic fields induce hyperfine transitions which exchange electron and nuclear polarisation, yielding polarised deuterons that are injected in the storage cell. A magnetic field was maintained in the target to provide a quantisation axis for the polarised gas. The initial polarisation of the atomic deuterium could be diluted from recombination of

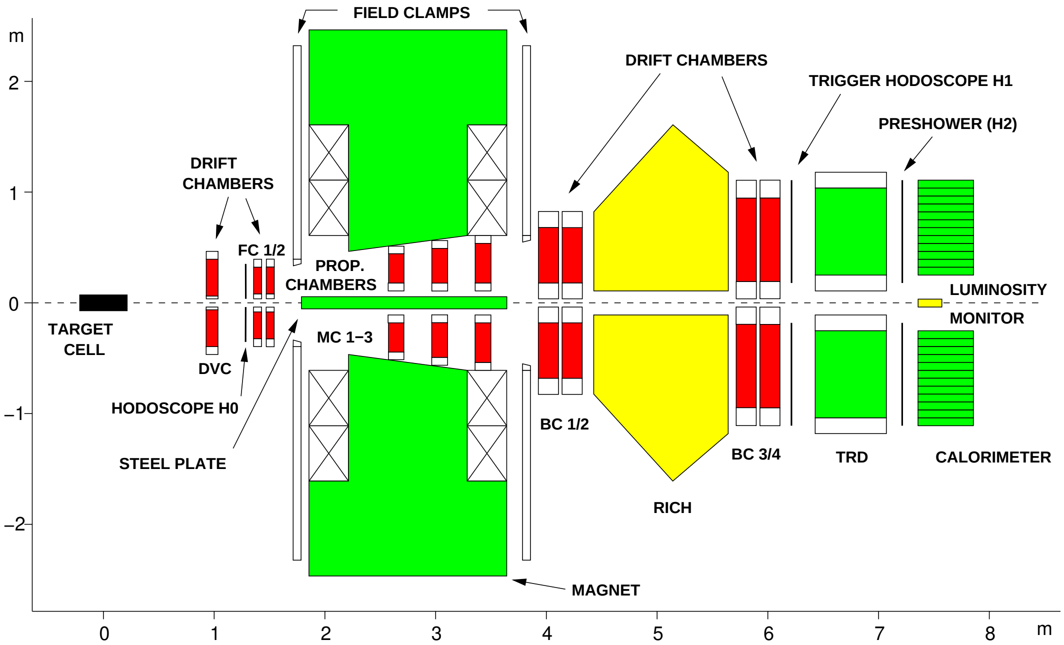

Figure 3.4: Schematic side view of the HERMES spectrometer. The positron beam enters from the left. The spectrometer is split into two halves, one above the beam line and one below. See the text for details on the detectors.

atomic into molecular deuterium, cell wall collisions and spin exchange collisions. A small fraction of the injected gas atoms is extracted through the sampling tube and fed into a target gas analyzer (TGA) [50] to measure the fraction of the deuterium in atomic form. A Breit-Rabi Polarimeter (BRP) [51] measures the polarisation of the atomic deuterium. Combining the measurement of the TGA and BRP allows a calculation of the deuterium target polarisation:

$$P_T = \alpha_0[\alpha_r + (1 - \alpha_r)\beta]P_T^{atom} \tag{3.2}$$

where α_0 is the initial fraction of deuterium atoms leaving the ABS, and is large, typically 0.99. α_r is the fraction of atoms that do not recombine to molecules in the target cell, as measured by the TGA. β is the nuclear polarisation of the recombined molecules relative to the nuclear polarisation of the atomic deuterium measured by the BRP, P_T^{atom}. The factor β is generally unknown, and a significant part of the target systematic uncertainty is attributed to it. The deuterium target polarisation in 2000 was 84.5 $\pm$ 4 %.

3.3 The HERMES spectrometer

The HERMES spectrometer is described in detail in [52]. It is a forward spectrometer with a large acceptance that can detect the scattered positron as well as hadrons in coincidence. This allows inclusive and semi-inclusive measurements of the DIS process, which are the backbone of this thesis.

A diagram of the spectrometer is shown in fig. 3.4. Briefly, the HERMES spectrometer consists of multiple tracking chambers before and after a magnet, as well as several particle identification detectors. The luminosity monitor is installed at the rear of the spectrometer.

The spectrometer is split into two halves (top/bottom) due to the need for a septum plate in the midplane of the magnet to eliminate deflection of the lepton and proton beams, which go through the spectrometer. Each spectrometer half is equipped with the same sequence of detectors. This configuration allows to perform two independent measurements to control the systematic uncertainties. The origin of the HERMES coordinate system is defined at the center of the target. The positive z-axis is oriented along the positron beam direction, the y-axis is pointing vertically upwards, and the x-axis is pointing towards the outside of the ring, making a right-handed orthogonal axis system. The geometrical acceptance of the spectrometer is $|\theta_x| < 170$ mrad in the horizontal direction and $40 < |\theta_y| < 140$ mrad in the vertical direction.

3.3.1 Magnet and tracking detectors

The dipole magnet provides a magnetic field in the vertical (y) direction, so that particles are bent in the horizontal direction. The momentum of the particles can be determined from the bending in the magnet. The field has a deflection power of $\int Bdl = 1.3$ Tm and to minimise the stray magnetic field outside the physical magnet, large iron plates are used as field clamps on the upstream and downstream magnet faces.

The tracking system performs several tasks:

- Determine the event vertex[1] to ensure the event came from the gas, not from the walls of the target cell or from the collimators upstream of the target.
- Measure the scattering angles of all particles for the determination of kinematical variables.
- Measure the particle momentum from the deflection of the track in the spectrometer magnet.
- Identify hits in the particle identification detectors associated with each track.

The tracking system of the HERMES spectrometer may be divided into the front region (between the target and the magnet), the magnet region and the back region. In the front region, the responses of the Drift Vertex Chambers (DVC) and Front Chambers (FC) allow to measure the scattering angles of charged tracks leaving the target and to identify if a track originates in the target gas or not. The spectrometer magnet is located behind these front chambers and bends the charged tracks in the horizontal plane. Inside the magnet are placed Magnet Chambers (MC) which help to constrain this bending directly. Behind the magnet the Back Chambers (BC) measure the deflected trajectories. This rear tracking not only determines the track momenta but it also localises each trajectory

[1]The event vertex is calculated as the point from which two or more tracks in the event have the smallest distance. In the case of only one track, the positron beam axis is taken as the second track.

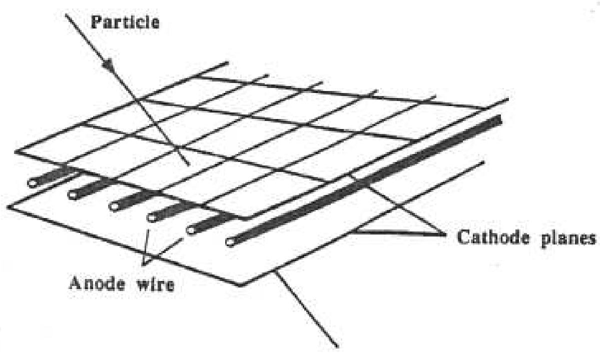

Figure 3.5: The principle of the multiwire proportional chamber. A charged particle ionises the gas between the cathode planes and the created electrons and ions move towards the anode and the cathodes, respectively.

inside the particle identification detectors. The location of all tracking detectors is shown in fig. 3.4.

All tracking chambers are drift chambers, except the MC which are multiwire proportional chambers (MWPC). MWPC were invented by Charpak (Nobel prize in physics 1992). Charpak's invention consists of using an earlier development, the proportional counter, in an unconventional way. The classical proportional counter, like the Geiger Müller tube, consists of a thin anode wire in the middle of a tube with a diameter of about a centimetre. Between the wire and the wall of the tube a high voltage of a few kilovolts is applied. A charged particle passing through the gas-filled tube will ionise the gas, and the electrons that are released as part of the electron-ion pairs will drift to the anode. Near the wire the electric field is very strong and results in a rapid acceleration of the electrons. They then have enough energy to ionise the gas and more electrons are liberated, which in their turn are accelerated and so on. This results in an avalanche of electrons and positive ions which gives rise to an electric signal on the wire. The position of the charged particle that started the ionisation in the gas can, however, only be determined with a precision of about a centimetre, the size of the tube.

The break-through occurred with Charpak's idea [53] that the resolution is greatly improved if many anode wires are arranged in a plane between two common cathode planes, as shown in fig. 3.5.

MWPC with high resolution are expensive because signals from a very large number of wires have to be read out. This cost can be reduced in a drift chamber that uses the fact that electrons take time to drift from their point of production to the anode. Thus the time delay between the passage of a charged particle through the chamber and the creation of a pulse at the anode is related to the distance between the particle track and the anode wire. In practice a reference time has to be defined, which is provided by the trigger that will be discussed in section 3.4. Drift chambers operate worse in high magnetic fields than MWPC [55], therefore the MC were chosen to be MWPC. In case of the HERMES drift chambers, a different setup compared to the MC has been used: instead of a plane of anode wires between the cathode foils (as depicted in fig. 3.5), the middle plane consist of alternating anode (sense) wires and cathode (potential) wires. This configuration creates an electric field that optimises the drift time towards the anode wires.

All tracking chambers consist of several planes with three different wire orientations, adding up to a total number of 51 planes. The wires in one plane (X) are oriented along the vertical axis shown in fig. 3.4, the wires in the two other planes are tilted by $+30^0$ (U plane) and -30^0 (V plane) with respect to the vertical axis. This arrangement allows the reconstruction of both horizontal and vertical tracking information without wires oriented in the horizontal plane. The lateral extension of the large BC (that have to cover the enlarged acceptance area behind the magnet) precluded the use of horizontally strung wires. The identical arrangement of wire planes in each of the tracking detectors simplifies the reconstruction code significantly [54]. The transverse measurement in the drift chambers has an inherent ambiguity as the particle could have crossed by either side of a sense wire in the drift chambers. To eliminate this left-right ambiguity another set of planes with offset wires is used. The offset planes have the wires offset by one-half the cell width. These staggered planes are labeled with (X',U',V'). For example the DVC consist of 6 planes in the configuration (XX',UU',VV').

More information about the HERMES tracking chambers can be found in [55, 56, 57]. The tracking algorithm that combines the information from hits in the tracking chambers to geometrical trajectories is called HRC (Hermes Reconstruction Code) and is described in detail in [54].

The combination of the tracking chambers and the magnet allows a momentum resolution $\Delta p/p$ of less than 3% and an angular resolution of less than 1.5 mrad over the entire kinematic range of the HERMES experiment.

3.3.2 Particle identification

Measuring semi-inclusive deep-inelastic scattering processes requires the identification of the scattered lepton and coincident charged hadrons. The HERMES particle identification (PID) system employs four subsystems: a lead-glass calorimeter, two plastic scintillator hodoscopes, one of which is preceded by a lead curtain and acts as a preshower detector, a transition radiation detector, and a Ring Imaging Čerenkov detector (RICH). Each of these detectors responds differently to positrons and hadrons, allowing particles to be distinguished as one or the other. The PID detectors are located in the back region of the spectrometer, so an extrapolation of the trajectories as determined by the back region tracking system is used to assign PID detector responses to tracks. All PID detectors are shown in fig. 3.4. The RICH detector will be extensively discussed in chapter 4. Here we just mention that it has two Čerenkov radiators, gas and aerogel (since this information will be needed later in this section). The goal of the RICH is mainly to identify the produced hadrons as pions, kaons or protons.

3.3.2.1 The transition radiation detector

Transition radiation is produced when a relativistic charged particle crosses a boundary between mediums with different dielectric constants [59]. The transition radiation (TR) for ultra-relativistic particles lies in the X-ray region (several keV). The probability for

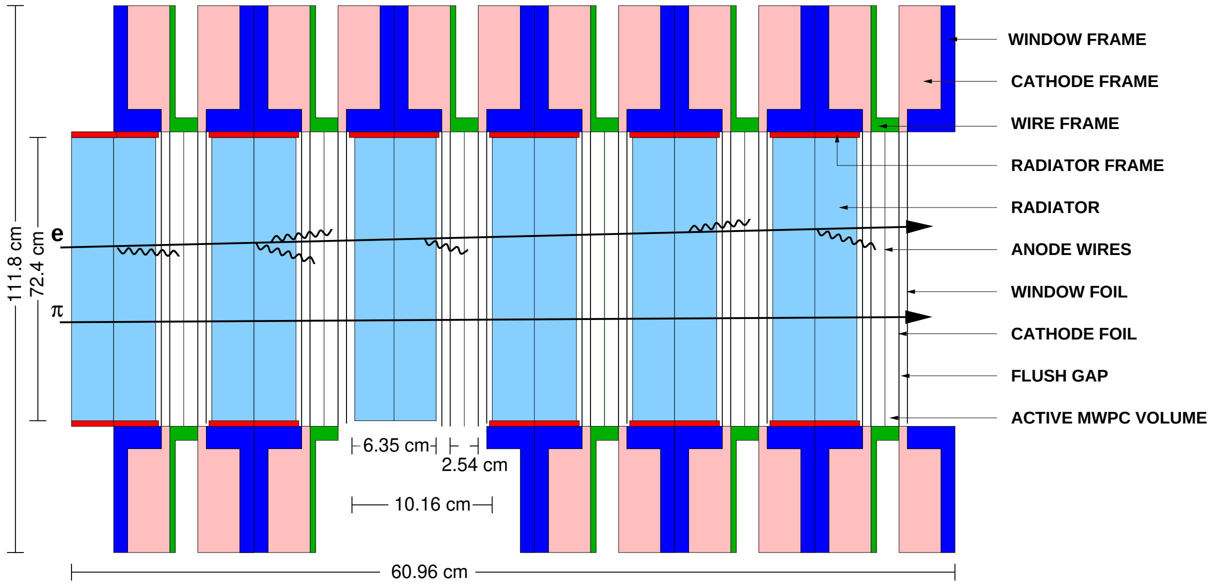

Figure 3.6: Schematic picture of the six TRD modules.

emission of TR photons is proportional to the Lorenz factor $\gamma = 1/\sqrt{1-\beta^2}$. For example, a 5 GeV positron has a γ of 10^4 and may radiate X-rays; a 5 GeV pion has a γ of only 36, so that the probability that the positron emits a TR photon will be 300 times larger than for the pion. Hence leptons and hadrons can efficiently be separated if the TR is detected.

The Transition Radiation Detector (TRD) consists of six identical modules per detector half and is shown in fig. 3.6. Each module is built of a 6.35 cm thick radiator and a 2.54 cm thick X-ray detector. The radiator is made of loosely packed layers of polyethylene/polypropylene fibers. A charged particle passing through the radiator crosses a random matrix of dielectric transistions and produces transition radiation. Because transition radiation is a relatively small effect, many dielectric surfaces are required for a practical detector. The layers of fibers achieve an equivalent of about 250 surfaces. The X-ray detector is a conventional multiwire proportional chamber, containing a plane of 256 vertical anode wires between two cathode foil sheets (cfr. fig. 3.5). The working principle of this MWPC is somewhat different compared to the explanation given in section 3.3.1 since here X-rays need to be detected. The X-ray TR photons are detected by photoelectric absorption. A photon with energy E_γ can be absorbed only in the j shell of an atom, if the binding energy E_j of the shell is smaller than the photon energy. Absorption of the photon results in the emission of a photoelectron with energy $E_e = E_\gamma - E_j$. These photoelectrons are detected in the MWPC as explained in section 3.3.1. Since the photon flux that is absorbed in the gas of the MWPC increases according to Z^4 (Z being the atomic number of the gas), an efficient absorption requires a high atomic number. A gas mixture with mainly Xe is the best absorber for TR radiation in the HERMES case [58]. Apart from photoelectrons from the transition radiation, the MWPC will also record energy that is deposited due to gas ionisation of the chamber gas when a charged particle transverses it. However, the MWPC signal resulting from TR (positrons) is much larger

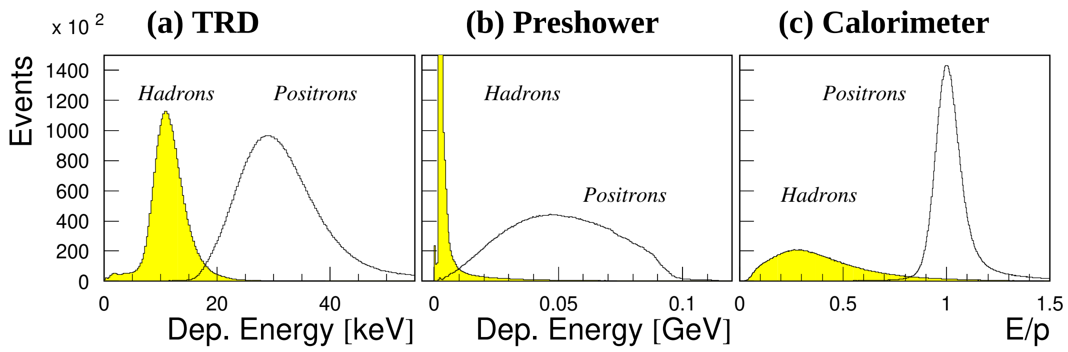

Figure 3.7: Number of counts for positrons (empty histograms) and hadrons (shaded histograms) in the three particle identification detectors: transition radiation detector, hodoscope H2 and the calorimeter.

than signals from ionisation (positrons and hadrons), allowing for a separation between positrons and hadrons. The mean energy deposited in the TRD modules is shown in fig. 3.7(a) for hadrons and positrons, already indicating that leptons and hadrons can be separated rather well. More information about the TRD can be found in [58].

3.3.2.2 The calorimeter

The calorimeter [60] in the rear of the spectrometer (fig. 3.4) is a key component of both the trigger (see section 3.4) and the particle identification system. As shown in fig. 3.8, each half of the calorimeter is a 42 × 10 array of 9 × 9 × 50 cm blocks. These blocks are made of F101 lead-glass, selected for its resistance to radiation damage [61]. Each block is coupled to a photomultiplier tube. As radiation near the beamline can damage the glass during injection of HERA, both halves of the calorimeter are mounted on a movable frame and can be moved away from the beam until a stable beam is available.

The basic principle that the calorimeter uses for particle identification is the topological difference between electromagnetic and hadronic showers. *Electromagnetic showers* are created by the interactions of a high energy electron, positron or photon with matter. For energies above 100 MeV the dominating processes are Bremsstrahlung for electrons/positrons and pair production for photons. These processes successively produce a cascade of particles until finally the low energy secondary particles lose their energy primarily by ionisation. The parameter that sets the typical scale for the development of electromagnetic showers is the radiation length X_0, defined as the mean distance over which a high energy positron loses all but e^{-1} of its energy by Bremsstrahlung. The composition of a *hadronic shower* is much more complex compared to an electromagnetic shower. Typically a hadron, for example a pion, will first lose some energy by ionisation before a hadronic interaction takes place. This hadronic interaction will most likely be either the fission or spallation of a nucleus, producing pions and nuclear fragments which may in turn decay again. Neutral pions will decay into two photons which will start electromagnetic cascades within the hadronic cascade. This results in a mixture of pions, protons, neutrons, muons, nuclear fragments, neutrinos and electromagnetic show-

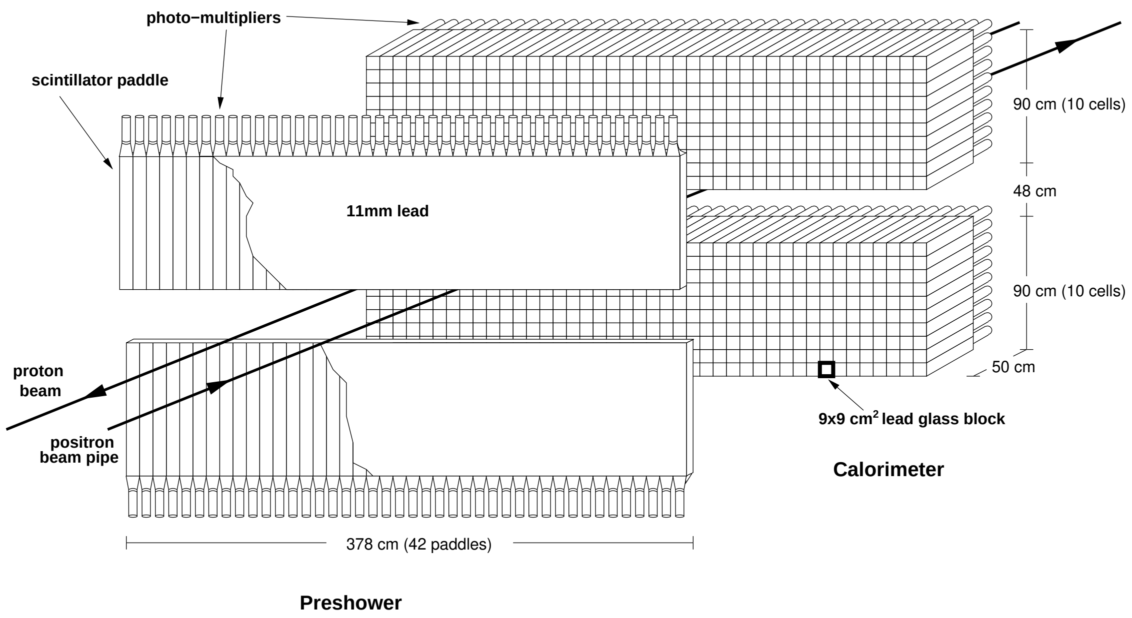

Figure 3.8: Schematic view of the preshower counter H2 and the calorimeter. The paddles of H2 are aligned with the corresponding columns of the calorimeter and are preceded by a Pb sheet.

ers. Other than in electromagnetic cascades not all the energy is deposited in a detectable form. However, the development of hadronic showers can be parametrised for practical purposes. To this end the nuclear interaction length λ is introduced, which plays a similar role as X_0 for the electromagnetic case. It is defined as the mean free path between inelastic interactions.

The maximum transverse cell size of the lead-glass blocks is determined by the requirement that the cell occupancy is less than 1 for most triggers as it is impossible to separate tracks showering in the same cell. The depth of the lead-glass corresponds to 18 radiation lengths, so a positron striking the blocks will deposit nearly all of its energy inside. Since $\lambda = 137.7$ g/cm^2 is much larger than $X_0 = 10.73$ g/cm^2, less of a hadron's energy is lost in the lead-glass. If the information on longitudinal and transverse shower development is combined, a simplified picture of electromagnetic and hadronic showers like fig. 3.9 can be drawn [62]. An electromagnetic shower starts practically instantaneously. It is almost contained within one calorimeter block and completely contained within an array of 3×3 blocks. Only a very small fraction of the shower energy leaks out at the back of the blocks. A hadronic shower starts typically only in the second half of the calorimeter and spreads over the boundaries of a 3×3 block array. A large part of the hadronic shower energy leaks out of the calorimeter. Because of this topological difference the ratio of the detected shower energy $E = E_{cal}$ in a 3×3 block array to the momentum p of the incident particle, is a valid particle identification parameter. This ratio E/p is ≈ 1 for positrons but $E/p << 1$ for hadrons, as illustrated in fig. 3.7(c).

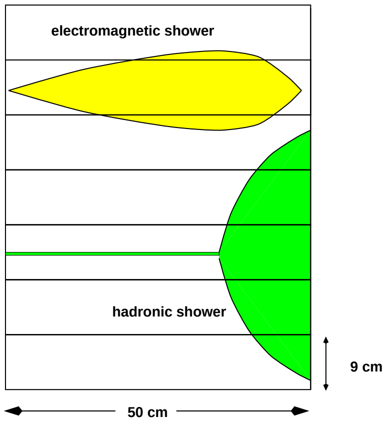

Figure 3.9: Schematic view of the topology of electromagnetic and hadronic showers in the HERMES calorimeter.

3.3.2.3 The hodoscopes

There are three hodoscopes (H0, H1, and H2) in the spectrometer. The hodoscopes are scintillation counters: charged particles can excite the atomic electrons of the counter medium; a small fraction of the excitation energy emerges during de-excitation as visible light that is directed to the face of a photomultiplier.

H0 and H1 only provide trigger signals (the trigger will be discussed in section 3.4), while H2 serves as a preshower detector for particle identification. A diagram of the preshower detector is shown in fig. 3.8. The hodoscope H2 consists of 42 vertical scintillator paddles that overlap slightly to avoid holes in the acceptance. The signals are read out by photomultiplier tubes located at the end of each paddle. An 11 mm thick lead plate (2 radiation lengths) is placed in front of H2. Positrons passing through the lead initiate an electromagnetic shower that deposits its energy in the scintillators, while hadrons do not initiate showers in the lead curtain since the lead corresponds to only 0.02 nuclear interaction length λ. Hadrons thus only deposit energy by ionisation. The preshower response to positrons and hadrons is plotted in figure 3.7(b), illustrating how both particle species can be separated.

The preshower detector H2 is also used in the trigger together with the calorimeter and the hodoscope H0 and H1. The latter is constructed in the same way as the preshower, but there is no lead in front of it. The hodoscope H0 is installed directly in front of the front drift chambers (fig. 3.4). It consists of a single sheet of plastic scintillator and the scintillation light is detected with two phototubes on the side edges. The reason why H0 and H1 are installed in the HERMES spectrometer will become clear while discussing the trigger in section 3.4.

3.3.2.4 Particle identification algorithm

In every physics analysis of HERMES data it is necessary to separate hadrons and leptons in the data sample. This is accomplished by a probabilistic algorithm that utilises the responses of the four particle identification detectors, that are different for the two particle types.

From the deflection of the particle in the spectrometer magnet, it is possible to calculate its momentum p. The polar angle θ can be reconstructed with the different tracking detectors. In each PID detector the particle will leave some energy E. The task of the particle identification algorithm is then to find the probability $P(i|E,p)$, given E and p, that the particle is a lepton ($i = l$) or a hadron ($i = h$). By introducing the different particle fluxes ϕ_i and making use of Bayes' theorem, one can show mathematically [58] that this probability for a certain detector '*det*' is given by

$$P_{det}(i|E,p) = \frac{\phi_i(p,\theta)P_{det}(E|i,p)}{\sum_{j=l,h}\phi_j(p,\theta)P_{det}(E|j,p)} \tag{3.3}$$

where $P_{det}(E|l(h),p)$ is the probability that a lepton (hadron) with momentum p will deposit an energy E in the detector. These probability distributions, commonly called parent distributions, could be measured at test-beam facilities with appropriate lepton and hadron beams. At HERMES, however, a different approach is used. The parent distributions are extracted from real data by imposing stringent cuts on the PID detectors other than the one considered (cfr. fig. 3.7). For example, to generate lepton parent distributions for the TRD, cuts are imposed on the responses of the calorimeter and the preshower detector. This procedure assumes that the responses of the detectors for a given particle type are uncorrelated. Generating clean particle samples has the advantage that the parent distributions take into account aging effects and varying detector conditions during data taking periods.

Once obtained the probabilities $P_{det}(l|E,p)$ and $P_{det}(h|E,p)$ are combined into a logarithmic ratio $PID^{'}_{det}$

$$PID^{'}_{det} = log_{10}\frac{P_{det}(l|E,p)}{P_{det}(h|E,p)} \tag{3.4}$$

In terms of the parent distributions and the flux factors this quantity becomes

$$PID^{'}_{det} = log_{10}\frac{\phi_l(p,\theta)P_{det}(E|l,p)}{\phi_h(p,\theta)P_{det}(E|h,p)} = PID_{det} - log_{10}\Phi \tag{3.5}$$

where

$$PID_{det} \equiv log_{10}\frac{P_{det}(E|l,p)}{P_{det}(E|h,p)} \qquad \text{and} \qquad \Phi \equiv \frac{\phi_h(p,\theta)}{\phi_l(p,\theta)} \tag{3.6}$$

The quantity PID_{det} is defined for the calorimeter, the preshower detector, the RICH detector and the TRD. At HERMES the following combinations of PID values are commonly defined

$$PID3 \equiv PID^{calorimeter} + PID^{preshower} + PID^{RICH} \tag{3.7}$$

where PID^{RICH} is the sum over the PID values for the gas and the aerogel radiator.

$$PID5 \equiv PID_{TRD} = \sum_{i=1}^{6} PID_i^{TRD} \tag{3.8}$$

The sum in the last equation runs over the six TRD modules. In practice the flux ratio Φ is often neglected (i.e. uniform fluxes for leptons and hadrons are assumed) and only the

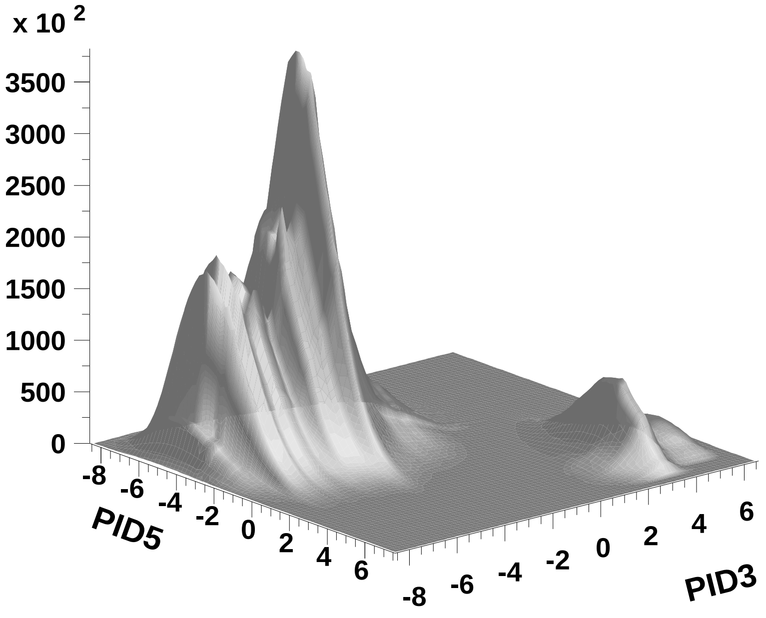

Figure 3.10: Correlation of PID3 and PID5. A clean separation of leptons and hadrons is visible.

quantity $PID3 + PID5$ is used for particle identification. This approximation is valid as long as the flux ratio is fairly constant and/or sufficiently hard cuts on $PID3+PID5$ are used. In fig. 3.10 the correlation of PID5 and PID3 is shown; the positrons (small bump) are seen to be clearly separable from the hadrons (large peak).

A more detailed description of the HERMES PID can be found in [63].

3.3.2.5 The gain monitoring system

To monitor the gain and ageing effects, every detector component equipped with photomultipliers (hodoscopes, calorimeter and luminosity monitor) is connected to a light source via glass fibres. A dye laser at 500 nm is used to send pulses of various intensities through the fibers to all photomultipliers of the aforementioned detectors and additionally to a reference photodiode. As the photodiode has a stable gain, the comparison of the PMT signal to that in the photodiode can be used to monitor relative gain changes over time and to find dead channels. More information on the Gain Monitoring System (GMS) can be found in [64].

3.4 Triggers

The HERMES experiment is equipped with a variety of triggers that are used to extract information that is useful for either physics studies or to check the (mal)functioning of the detector. In order to keep the overall trigger rate low enough, most non-physics triggers are prescaled, meaning that only every second, fourth, eighth... one is recorded.

This prohibits an overload of the data acquisition due to a too high overall trigger rate. Further information on the HERMES trigger system can be found in [65]. The most important trigger at HERMES is Trigger 21, defining a potential deep-inelastic scattering event. Trigger 21 is not prescaled in order to collect the largest DIS sample possible. To fire the DIS Trigger 21, the following four conditions have to occur in one detector half in coincidence:

- Hodoscope H0 and H1 give signals. Photons and other neutral particles are suppressed by this condition since they do not leave signals in the scintillators. Photons are potentially dangerous, because they can create large signals in the preshower H2 and the calorimeter, which might be interpreted as DIS signals.

  In addition, the condition on H0 reduces the chance of triggers from particles originating from the proton beam (that travels in the other direction compared to the lepton beam): protons that leave the storage ring and hit the proton beam pipe can produce hadronic showers of high energy and create signals in the calorimeter and preshower. H0 is placed in front of the magnet and therefore has enough separation from H1 and H2 to determine whether a particle is going forward or backward in the spectrometer. This suppresses the proton beam background.

- The signal in the preshower is required to be above the minimum ionising level, in order to avoid hadronic background.

- The sum of two adjacent calorimeter *columns* results in more than E_{thr} deposited energy.
  A minimum ionising particle (hadron) would deposit 0.8 GeV of energy in the calorimeter; to ensure that a positron shower has occurred and to reduce the amount of unusable data written to data storage media, a threshold $E_{thr} > 0.8$ GeV was set on the calorimeter column sums. Since the majority of a scattered positron's energy will be deposited within a 2×2 array of calorimeter blocks, the energy sum of two complete columns of the calorimeter provides a fast way to check if there was indeed a scattered positron.

  For running with unpolarised target gas at high densities, the calorimeter threshold E_{thr} was set to 3.5 GeV, equivalent to an event selection cut on $y = (E - E')/E < 0.87$ where E is the beam energy and E' is the energy of the scattered track. The threshold ensures that a triggered positron has enough energy to require only minimal radiative corrections that are the largest for $y > 0.87$ [131], and reduces the low energy background events at the same time. For the polarised running the threshold was set to 1.4 GeV to extend the kinematic coverage of the recorded DIS events to higher values of y ($y < 0.91$).

While Trigger 21 is required in the offline analysis for a potential deep-inelastic scattering event, trigger 16 has to be vetoed. This trigger 16 is defined as the trigger from the Gain Monitoring System. The artificial trigger signals from the GMS do not correspond to deep-inelastic scattering events and have to be rejected for physics purposes.

3.5 The luminosity monitor

The luminosity monitor [66] serves to measure the luminosity L, defined as the ratio of a certain measured scattering rate and the cross section of this process:

$$L = \frac{R}{\sigma} \tag{3.9}$$

The determination of the luminosity[2] by this equation is possible by measurement of a scattering rate, if the corresponding cross section is known. Therefore, a reference scattering reaction is used. This reaction is the beam particle scattering off target shell electrons. In case of a positron beam the elastic scattering $e^+e^- \rightarrow e^+e^-$ is called Bhabha scattering. In addition the beam positron and the shell electron can annihilate $e^+e^- \rightarrow \gamma\gamma$ (pair production). If the beam particle is an electron, only elastic scattering $e^-e^- \rightarrow e^-e^-$ (Møller scattering) is possible. The cross sections of these reactions are exactly calculable [67] and the luminosity can be determined by measuring the rates of these reference processes

$$L = \frac{R_{ref}^{measured}}{\sigma_{ref}} \tag{3.10}$$

The total cross section σ_{ref} is the integral of the differential cross section over the acceptance Acc of the luminosity monitor:

$$\sigma_{ref} = \int_{Acc} \epsilon \frac{d\sigma_{ref}}{d\Omega} d\Omega \tag{3.11}$$

taking into account the detection efficiency ϵ of the luminosity monitor. The lumi constant is defined as $C_{lumi} = 1/\sigma_{ref}$. Its value has been determined by detailed Monte Carlo simulations [68]. The two final state particles of the scattering reaction, leptons or photons, are detected by two electromagnetic calorimeters, which are mounted on both sides close to the positron beam pipe 7.2 m downstream of the storage cell (see fig. 3.4). Each calorimeter is composed of 12 crystals. Each crystal is read out by an individual photomultiplier tube. To select Bhabha, Møller and photon pair events, the trigger for the luminosity measurement requires two coincident signals in both calorimeters corresponding to an energy deposition above 4.5 GeV. Background events typically deposited a significant amount of energy in only one of the two calorimeter blocks. Radiation that occurs while injecting the beam can cause severe damage to the calorimeters. Therefore they are moved away from the beam pipe during beam injection.

A complication occurs for the polarised measurements due to the spin dependence of the Bhabha cross section. Even though the polarised target is operated in a mode to deliver only nuclear polarisation and no polarisation for the shell electrons, small inefficiencies in the high frequency transitions for the hyperfine states as well as a transfer of the nuclear spin to the electron shell in atomic spin exchange collisions lead to a remaining electron polarisation of 1-2% in the target cell [69]. To avoid a bias in the luminosity measurement due to the Bhabha asymmetry, it has therefore been decided to fit the luminosity over

[2] In principle the luminosity can also be determined using the target areal density ρ_T and the beam current I: $L = \frac{\rho_T I}{e}$. Because of the unprecise indirect (pressure) measurement of the density in the center of the target, the use of a reference scattering reaction is preferred.

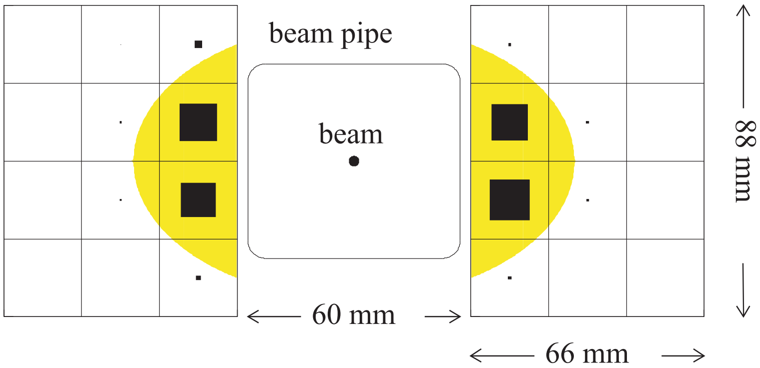

Figure 3.11: Hit distribution in the calorimeters of the luminosity monitor. The size of the boxes is proportional to the number of hits per channel. The shaded area shows the beam pipe acceptance

time intervals much longer than the typical spin flip time of the target. As the luminosity is proportional to the beam current, and since the target density is normally kept constant during polarised data taking, the time dependence of the luminosity should be given by the time dependence of the beam current, which follows an exponential function. Typically a fit extends over a whole fill of positrons in the HERA ring, which on average lasts several hours. Only in case of instabilities in the beam current because of partial beam losses or in the target density, a fill is split into up to three separate fit intervals.

Chapter 4

The HERMES Ring Imaging Čerenkov Detector

4.1 Introduction

One of the major physics goals of the HERMES experiment is determining the different quark flavour contributions to the spin of the nucleon. Several semi-inclusive processes, when one or more hadrons besides the scattered lepton are detected in the final state, can provide a direct measurement of these contributions. Different types of light hadrons (π, K and p) are involved in these processes and their identification is crucial if one wants to disentangle the spin structure of the nucleon.
In late 1996 the HERMES collaboration had successfully operated its spectrometer for almost two years. The TRD, calorimeter and a pair of scintillator hodoscopes provided excellent lepton/hadron separation. The only available hadron identification at that time was by means of a threshold Čerenkov detector which could only separate pions from heavier hadrons in a limited momentum range (roughly 4 to 14 GeV). To extend the momentum range and to provide full hadron identification, the threshold Čerenkov detector was replaced by a Ring Imaging Čerenkov (RICH) detector which was installed in 1998. The HERMES RICH detector is the first RICH detector in a major experiment that uses aerogel as a Čerenkov radiator to identify pions, kaons and protons.
In this chapter we will summarise the theory of Čerenkov counters and describe the design of the HERMES RICH detector. The algorithms that are used for particle identification[1] (PID) will be detailed as well as the development of the RICH simulation program which reproduces the experimental data well.

4.2 Čerenkov radiation

The speed of light in an optical medium with refractive index n is $c_n = c/n$. A charged particle moving with velocity v in this material will emit Čerenkov radiation [70] whenever v exceeds c_n (this radiation is an electromagnetic analogy of pressure waves formed by

[1]In this chapter, the term PID will refer to pion, kaon and proton identification, not to lepton/hadron separation as in the previous chapter.

breaking the speed of sound). This Čerenkov radiation is cone-shaped and has a polar emission angle θ relative to the particle direction given by

$$\cos\theta = \frac{1}{\beta n} \tag{4.1}$$

with $\beta = v/c$. The production of Čerenkov photons by particles of unit charge is governed by the Frank-Tamm relation [71]

$$\frac{d^2N}{dxd\lambda} = \frac{2\pi\alpha}{\lambda^2}(1 - \frac{1}{\beta^2 n^2(\lambda)}) \tag{4.2}$$

where N is the number of photons produced, x is the path length through the radiator, λ is the photon wavelength and $n(\lambda)$ is the refractive index of the radiator.

Combining relation (4.1) with

$$\beta = \frac{p}{\sqrt{m_0^2c^2 + p^2}} \tag{4.3}$$

(where p is the momentum[2] of the particle and m_0 its restmass) yields

$$\cos\theta = \frac{1}{n}.\sqrt{1 + \frac{m_0^2c^2}{p^2}} \tag{4.4}$$

This equation can be used in two ways to identify particles: the first method is applied in *threshold* Čerenkov detectors, the second method in *Ring Imaging* Čerenkov detectors. The HERMES RICH will combine both methods at the same time, as will be explained later in this chapter. In a threshold counter one employs the existence of a minumum (threshold) momentum p_{th} for the production of Čerenkov light:

$$p_{th} = \frac{m_0 c}{\sqrt{n^2 - 1}} \tag{4.5}$$

The measurement of the particle momentum provides a simple particle identification method: pions, kaons and protons have a different m_0 and thus a different p_{th}. Whether or not a charged particle radiates Čerenkov light at a given momentum yields a clue about its mass and identity. For example, the former HERMES threshold Čerenkov detector ran with a mixture of 70% nitrogen and 30% perfluorobutane (C_4F_{10}) with refractive index $n = 1.000629$, which gives momentum thresholds of 3.9, 13.9 and 26.4 GeV for pions, kaons and protons respectively. Hadrons with momentum p in the window $3.9 < p < 13.9$ GeV could thus be identified as pions. This was the only information on the hadron identity which could be obtained with the former HERMES threshold Čerekov counter.

In a Ring Imaging Čerenkov detector PID is realised by measuring the momentum p of the particle *and* the Čerenkov angle θ. The relation (4.4) allows to calculate the restmass m_0 which identifies the particle type. The HERMES RICH detector will be described in the following sections.

[2]To avoid confusion with the P matrix introduced later in this chapter, the momentum of electrons **and** hadrons will be denoted with p in this chapter.

4.3 Two Čerenkov radiators for the HERMES RICH

The momentum range over which the HERMES RICH detector should provide clear particle identification is defined by the hadron momenta spectra in the HERMES acceptance. Simulations [72] showed that about 95% of all the hadrons (π, K and p) are found in the range of 2.0 to 15.0 GeV. The coverage of this range was the first requirement for the construction of the HERMES RICH. It is impossible to cover this range with a single radiator: gases under atmospheric pressure are useful only for energies above 9 GeV since the kaon threshold for Čerenkov radiation is typically > 9 GeV. The use of a high pressure gas system was technically not feasible in HERMES. Clear liquid radiators can only be employed for hadron identification below 4 GeV. The only solution for HERMES is to make use of a dual radiator detector: one radiator covers the high-momentum range, the other one covers the low-momentum range. The high-momentum region (above 9.4 GeV) relies on a gas (C_4F_{10}) radiator with an index of refraction $n_{C_4F_{10}} = 1.00137$. The low-momentum end can only be spanned by a material with a refraction index of about $n = 1.03$. With the recent development of new clear aerogel (a silicon-based sponge-like structure in which more than 90% of the volume is air) this difficult low-momentum range can be covered. New production techniques [73] have yielded aerogels that are both hydrophobic and much more transparant than the earlier (hydrophilic) aerogels. The Čerenkov angles produced by the combination of this aerogel and the heavy C_4F_{10} gas for electrons, pions, kaons and protons are plotted in fig. 4.1 as a function of particle momentum. The corresponding threshold momenta are listed in table 4.1. All pion momenta within the spectrometer acceptance are above the pion threshold momentum for aerogel of 0.6 GeV, 90% of the kaon and 78% of the proton momenta are above the kaon threshold of 2.0 GeV.

An important parameter for hadron separation with RICH detectors is p_{max}, the maximum separation momentum [70]. This is defined as the maximum momentum for which the average photon emission angle of two particle types (with masses m_1 and m_2) is separated by a certain number of standard deviations n_σ:

$$p_{max} = \sqrt{\frac{m_2^2 - m_1^2}{2k_f n_\sigma}} \tag{4.6}$$

where k_f is the RICH detector constant that can be calculated using the design parameters of the RICH [72, 75], yielding the values for p_{max} listed in table 4.2.

Fig. 4.2 illustrates the overlap between the momentum regions for both radiators. The lightly shaded region indicates where the particle can be identified based on whether or not Čerenkov light is present at all. In this region the detector acts like a threshold Čerenkov counter. In the darkly shaded region the identification is based on the average reconstructed angles. The plot considers each radiator separately but the PID algorithms will combine the information from the aerogel and the gas. The momentum region for which the identification of pions, kaons and protons is possible, is limited by the kaon threshold momentum for aerogel at 2.0 GeV and by the p_{max} for π/K separation in C_4F_{10} at 15.0 GeV.

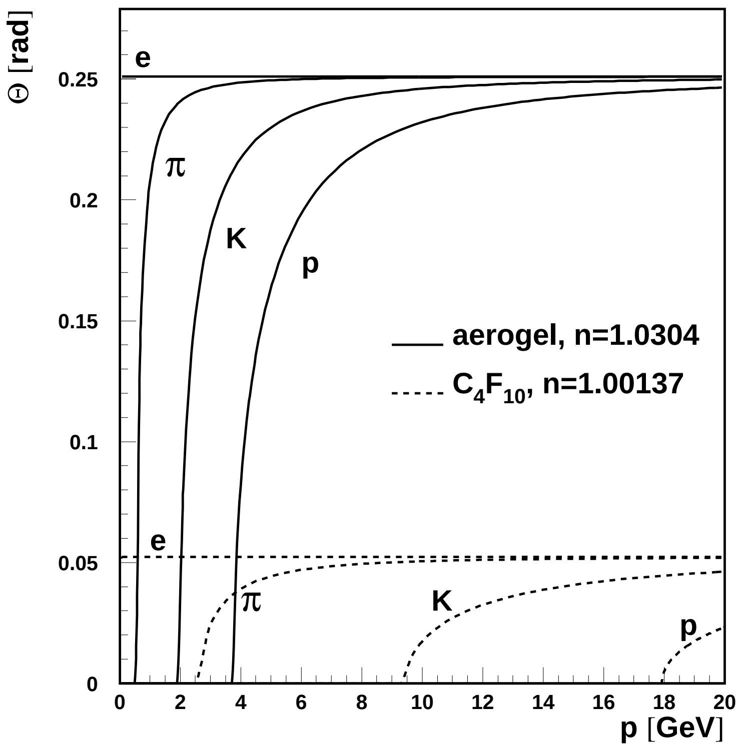

Figure 4.1: The Čerenkov angle θ as a function of the particle momentum for electrons, pions, kaons and protons for the aerogel and gas radiator.

	aerogel	C_4F_{10}
n	1.0304	1.00137
π	0.6 GeV	2.7 GeV
K	2.0 GeV	9.4 GeV
p	3.8 GeV	17.9 GeV

Table 4.1: Čerenkov light thresholds for pions, kaons and protons.

	aerogel	C_4F_{10}
$p_{max}(\pi/K)$	6.7 GeV	15.0 GeV
$p_{max}(K/p)$	11.2 GeV	25.3 GeV

Table 4.2: Maximum separation momenta p_{max}. $n_\sigma = 4.652$ was chosen, corresponding to a misidentification of the particle of 1% [75].

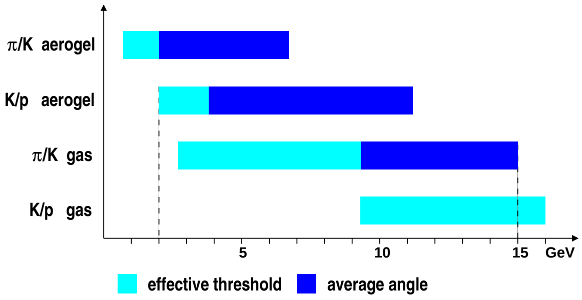

Figure 4.2: Momentum ranges for hadron separation in aerogel and C_4F_{10}. Between the dashed lines the hadrons can be separated.

4.4 Detector design

The RICH detector consist of two symmetric halves, above and below the HERA beam pipes (see fig. 3.4). A schematic view of the top RICH is shown in fig. 4.3 and fig. 4.4. The Čerenkov photons produced in the aerogel and the gas radiators are reflected by a spherical mirror array onto a photon detector plane. For particles with momenta above the gas threshold this results in two concentric 'rings', a smaller gas ring and a larger aerogel ring. An example for electrons is shown in fig. 4.5. These rings are simply the projection of the Čerenkov light cone on the photon detector. This picture already illustrates the principal idea behind RICH PID: by observation of the rings on the photon detector plane, one can calculate the Čerenkov angle θ. Together with the measured momentum p of the particle this allows (see fig. 4.1) to determine the particle type. A more detailed explanation of the particle identification will be given in section 4.5, but first the different components that can be distinguished in fig. 4.3 and fig. 4.4 will be described briefly (more details can be found in [74, 75]). The knowledge of the RICH geometry and some technical details is relevant as a considerable amount of this thesis work was devoted to the development of the RICH Monte Carlo simulation program that will be described in section 4.8.

4.4.1 The aerogel radiator

The aerogel was manufactured by Matshushita Electric Works (Osaka, Japan) in the form of tiles of average dimensions (x, y, t) $11.4 \times 11.4 \times 1.13$ cm^3. These tiles are not perfectly quadratic, they have one flat xy surface and one concave xy surface: due to capillary forces acting on the the upper xy surface of the mold containing the tile during its manufacturing process, the tiles have a top xy surface with a meniscus shape and a flat bottom xy surface. This results (fig. 4.6) in the tiles being thicker at their edges, with differences in thickness t up to 2 mm between the center of the tile and the edge. The tile edges have some influence on the experimental reconstruction of the Čerenkov angle, as will be illustrated later in this chapter. The refractive index was measured for all tiles [76], which gave an average of $n = 1.0304$ with a spread of $3.6\ 10^{-4}$.

The use of these tiles for the HERMES RICH was influenced by many considerations,

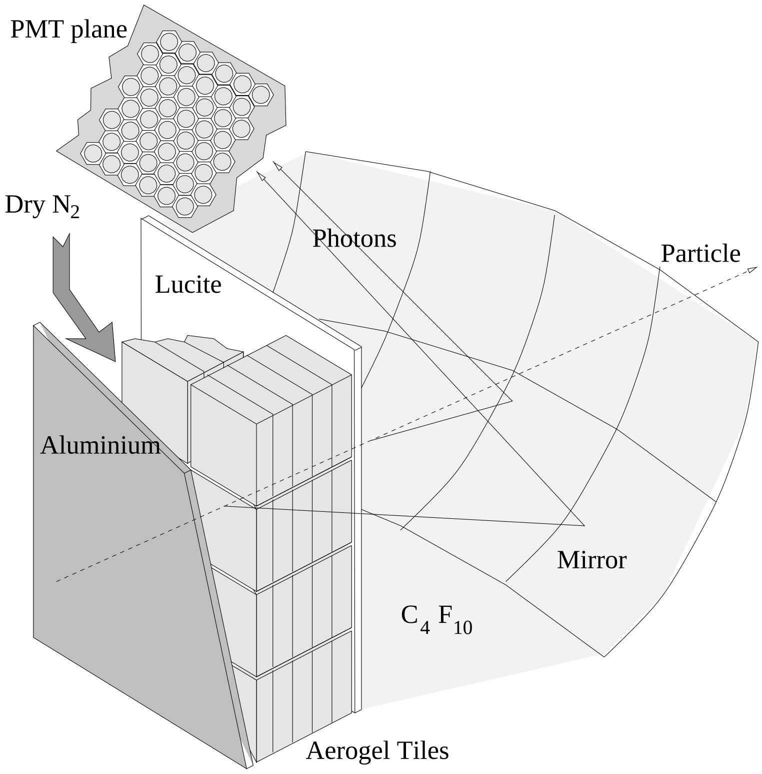

Figure 4.3: Basic geometry and radiator configuration of the HERMES RICH (not to scale).

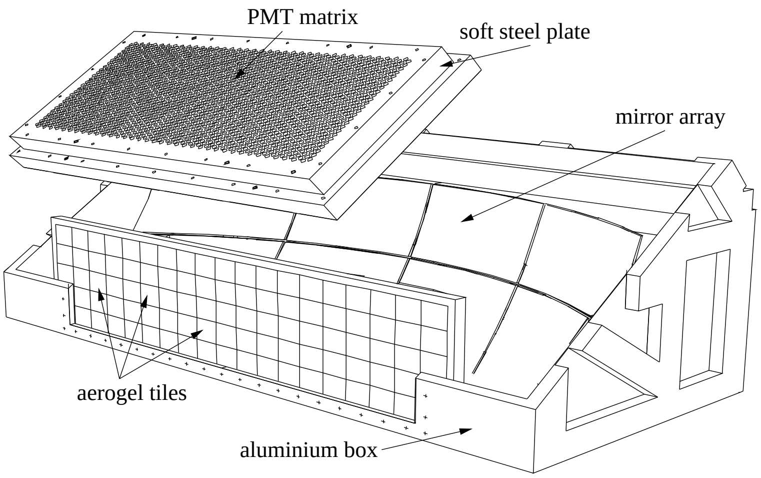

Figure 4.4: A cutaway schematic sideview of the top RICH counter.

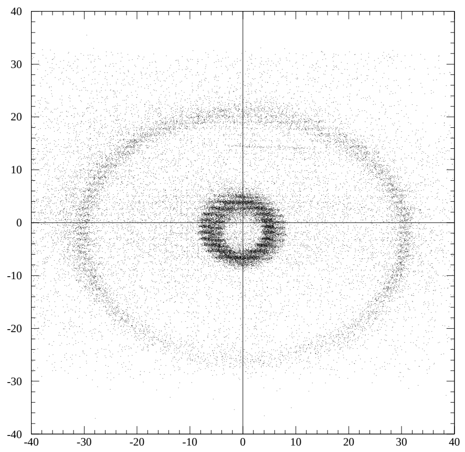

Figure 4.5: Gas and aerogel rings from electron tracks as observed in the RICH. The rings from individual tracks were all transposed such that the (reflected) track center lies at the origin. The offset of the center of the rings is due to the imperfect implementation of the detector geometry which will be discussed in section 4.5.1.2. Dimensions are in cm.

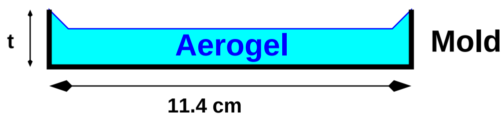

Figure 4.6: The shape of an aerogel tile, due to the manufacturing process. The mean thickness $< t >$ is 1.13 cm. The mold is removed after drying the tile in an autoclave.

some related to the optical characteristics of aerogel and others more practical in nature. The main optical property that determines the behaviour of aerogel as a Čerenkov radiator is the transparency T. The transmission of light through aerogel is not perfect due to Rayleigh scattering. This occurs in a medium when it contains inhomogeneities that are smaller than the wavelengths of the light passing through that medium; these inhomogeneities then deflect rays of light. The transmission T of light with wavelength λ through an aerogel tile of known thickness t is usually parametrised by the Hunt formula [77]:

$$T = A \cdot e^{\frac{-C \cdot t}{\lambda^4}} \tag{4.7}$$

The factor A is a wavelength independent transmission factor, C is the intensity of the Rayleigh scattering. This formula thus assumes that the absorption of light $(1 - A)$ in aerogel is wavelength independent, while the (Rayleigh) scattering cross section varies with λ^{-4}. The value of the clarity coefficient C is proportional to the Čerenkov radiation which is scattered per unit of sample length. The values of A and C have been measured [78] using a sample of 200 tiles for the wavelength range of 200-900 nm. The measurements could be described well by $C.t = 0.0094\ \mu m^4$ and $A = 0.964$.

The individual aerogel tiles are stacked in 5 horizontal rows and 17 vertical columns. Each of these 85 stacks is 5 layers of tiles thick, as can be seen in fig. 4.3. The configuration with 5 rows and 17 columns allows to cover the full HERMES acceptance. The aerogel wall thickness (i.e. the stack thickness) of 5 layers of tiles ($= 5 \times t \approx 5$ cm) was chosen as the optimal point for maximising the unscattered light yield relative to the scattered background photons [78, 79], using formula (4.7). The aerogel tiles are held in an aluminium frame. Black plastic spacers of appropriate thicknesses between the aluminium frame and the tiles prevent them from shifting while the radiator is moved. The 85 individual tile stacks are separated from one another by opaque black tedlar sheets that absorb photons crossing stack boundaries. The aim of this tedlar sheets is to avoid internal reflection of the photons off the tile sides and to reduce the background due to refraction of photons at the tile boundaries. The tedlar sheets also absorb part of the Rayleigh scattered photons since Rayleigh scattering has a nearly isotropic angle distribution. There is no danger that the tedlar sheets will absorb a large portion of the non-scattered photons, since the Čerenkov cone is small compared to the (x, y) dimensions of the aerogel tiles.
To achieve the best ring resolution, tiles with similar refractive indices are placed together in the same stack, so that particles passing through the radiator emit photons with very similar Čerenkov angles.

The aluminium frame that holds the tile stacks has a 1 mm aluminium entrance window and a 3.2 mm UVT lucite exit window. The choise of the entrance window was only driven by the concern to minimise the radiation length in the HERMES spectrometer, while the lucite exit window material was chosen primarily for its transmission properties. As can be seen from formula (4.7) the distribution of Rayleigh scattered light is concentrated at small wavelengths due to the λ^{-4} dependence of the Rayleigh cross section. The aim of the exit window is to transmit as many unscattered photons as possible while suppressing the scattered ones. Since lucite has a cut-off wavelength of 290 nm, it cuts away about half of the Rayleigh scattered photons. So, the useful photons emitted in the aerogel are largely limited to visible wavelengths.

To prevent possible degradation of the aerogel by the C_4F_{10} gas, the tile container is tightly sealed from the gas radiator. In addition dry N_2 gas is flushed continuously through the aerogel container at a slow rate.

4.4.2 The gas radiator

The body of the detector is constructed of aluminium and has a volume of about 4000 litre. A gas control system recirculates the C_4F_{10} gas through this volume, keeping the gas at a slight overpressure with respect to the atmosphere. The inner walls of this box are painted black to reduce the reflections of photons at the walls. It is not necessary to consider attenuation in the C_4F_{10} gas. One of the reasons that these fluorocarbons are good radiators is that the light transmission is very high and the dispersion very small. Čerenkov radiation emitted in the gas is mainly situated in the visible and UV part of the wavelength spectrum.

4.4.3 The photon detector

The photon detector array consist of a soft steel 'honeycomb' matrix that holds 1934 hexagonally closed packed photomultipliers. The sensitive area of each photomultiplier tube (PMT) is a circle with a diameter of 15 mm [80]. To guard the sensitive area (the bialkali photocathode) against degradation from the C_4F_{10} gas, a quartz window is placed in front of the sensitive area to provide a seal from the C_4F_{10} gas. To increase the amount of light collected on the photon detector, a conical funnel is inserted in front of the PMT. The opening diameter of the funnels is 23.27 mm; these funnels increase the geometric coverage from 38% to 81%. The photomultipliers are read out digitally (0 or 1), recording only the identities of the photomultipliers that have fired.

4.4.4 The mirror array

The mirror array consists of eight individual mirror segments. Due to the position of the RICH in the spectrometer, it was important to keep the mirrors as thin as possible in terms of radiation thickness. Hence, the mirror segments were fabricated from a graphite fiber composite, coated with an epoxy film yielding an optically smooth surface, and finally aluminised. A carbon fiber frame with adjustable three point mounts holds the mirror segments to form a mirror array with the total dimensions of 252.4×79.4 cm^2.

4.4.5 Photon detection efficiency

Čerenkov photons are emitted in the gas and the aerogel. Not all produced photons will be detected on the photon detector, since the efficiency for transmission and detection is limited. The efficiencies for transmission and detection of aerogel and gas photons is shown in fig. 4.7. The included contributions are the mirror reflectivity, the PMT cone efficiency, the effect of the air gap between PMT and quartz window (loss), the light transmission through lucite and the quantum efficiency of the PMT bialkali cathode. The sharp edge in the cone efficiency results from the sharp cut-off in reflectivity at 200 nm. The shaded histogram is the total efficiency to detect an emitted photon.

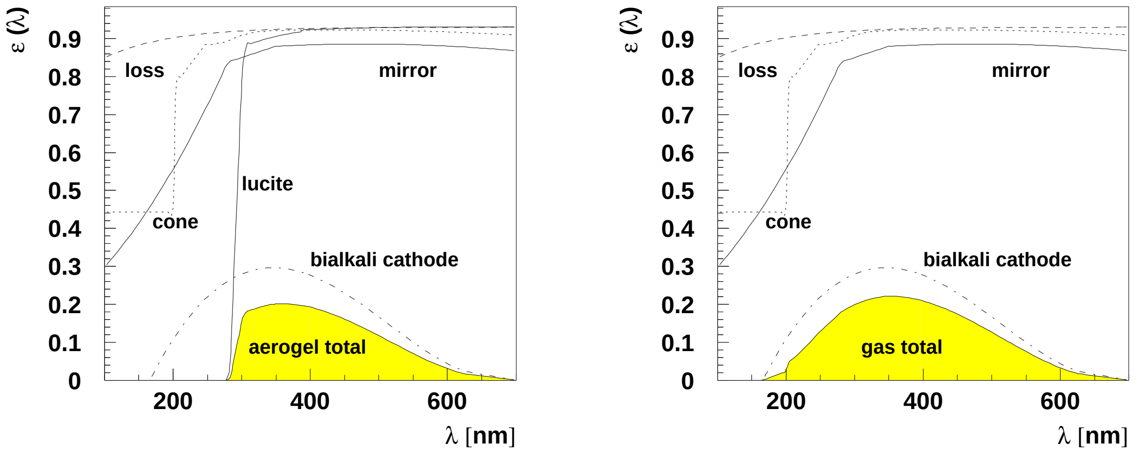

Figure 4.7: Efficiencies for transmission and detection of aerogel (left) and gas photons (right) as a function of the photon wavelength λ.

4.5 Particle identification with the RICH

The aim of the particle identification is to make use of a direct measurement in the RICH detector, i.e an observation of a number of fired PMTs in a certain event, to retrieve information about the type of the particle(s) in the event concerned. This is a far from simple task and two different algorithms, *Indirect Ray Tracing* (IRT) and *Direct Ray Tracing* (DRT), were developed by the RICH group to fulfil this task. Before these algorithms could be used to identify pions, kaons and protons, it was necessary to perform an "alignment" of the RICH detector. The short time scale of the RICH installation (13 months after approval) only allowed a simple alignment procedure during installation and when the RICH detector was finally inserted in the HERMES experiment, it could no longer be reached to perform more detailed geometrical measurements of the complete device and in particular the mirror. Already the first test measurements in the RICH, immediately after the installation in 1998, revealed that the limited knowledge of the alignment would compromise the PID. The final alignment of the RICH detector had to be determined from the experimental data itself, and the IRT algorithm was perfectly suited for this task. After this alignment procedure the RICH was ready to act as a "hadron separator" and the IRT and DRT methods could be used in all their glory. The efficiency of both methods for particle identification efficiency was analysed, using experimental data only.

The description in the next sections will be more or less chronologically, following the steps that were taken to proceed from the first disappointing test results to a fully operational RICH that has excellent hadron identification properties. First we will describe the IRT algorithm and show how one can reconstruct the Čerenkov angle with IRT. The knowledge of this angle made it possible to execute the alignment procedure. The alignment procedure will be described and it will be shown how the design resolution of the RICH was reached. After this achievement, the particle identification properties of IRT could be investigated and compared to experimental results. Since the particle identification with IRT was not completely satisfactory, it was decided to investigate if DRT performs better. Studies on the behaviour of DRT revealed some weaknesses and the RPS (Rich Pid Scheduler) was developed to decide which of the two PID methods gives the best outcome. Finally, results of this RPS method were compared to experimental data and the PID efficiency could be studied.

4.5.1 Indirect Ray Tracing

4.5.1.1 Principle of IRT

Both PID routines have the same input, consisting of two tables that are filled on an event by event basis. The first table is the so-called *track table* that contains the information registered by the other HERMES detectors for all tracks in a certain event. These other detectors are on the one hand the tracking detectors (FC, DVC, MC, BC) and on the other hand the PID detectors (calorimeter, TRD, hodoscopes) which provide an excellent hadron/lepton separation. The momentum of each track is determined by its bending in the spectrometer magnet. The second table is the *dataRICH table* that contains all the numbers of fired PMTs in the event. A RICH "mapping" file translates the integer identification number of a certain PMT to coordinates in the PMT plane.

Since the PMT plane is not spherical it does not correspond to the real focal plane of the spherical mirror. The detected rings are thus not perfect circles, but rather elliptical in shape, with a degree of distortion that depends on the direction of the track within the acceptance. Instead of attempting to fit these rings directly, a great simplification is achieved by reconstructing the Čerenkov angle θ. An algorithm was developed to eliminate the influence of the imaging system, *inverse or indirect ray tracing*. The problem of inverse ray tracing can be stated as: given a track and a hit in the RICH photon detector plane, we want to know under what Čerenkov angle θ (with respect to this track) the photon causing this hit was emitted. Using the terminology of fig. 4.8 the geometrical problem can be formulated as:

Given the emission point E (*the photon vertex*), point D (*the PMT hit*) and point C (*the center of the spherical mirror*); find the point S on the spherical mirror where the photon was reflected.

The mathematical solution of this problem can be found in [81, 75]. This solution allows the inverse ray tracing routine to reconstruct the Čerenkov angle θ, of course under the assumption that the photon vertex is known. Since it is not possible to know where the Čerenkov photon was emitted along the particle track, we assume (for each track in the event) that the hit could be coming from the aerogel or the gas, and the emission vertex is estimated accordingly to be in the middle of the particle track through the aerogel or gas radiator. For each radiator hypothesis (gas and aerogel) the emission angle θ is then reconstructed. This results in a spectrum of reconstructed angles as shown in fig. 4.9, which is clearly dependent on the radiator hypothesis. Due to the very different index of refraction, it is rarely a problem to distinguish which radiator hypothesis is correct. A likelihood analysis will discard the wrong hypothesis as will be explained later on.

An important difference between the treatment of reconstructed aerogel and gas angles lies in the refraction of aerogel photons at the aerogel-gas boundaries. This is approximately corrected for in the reconstruction program [82] under the assumption of a flat aerogel-gas boundary and coplanar refraction.

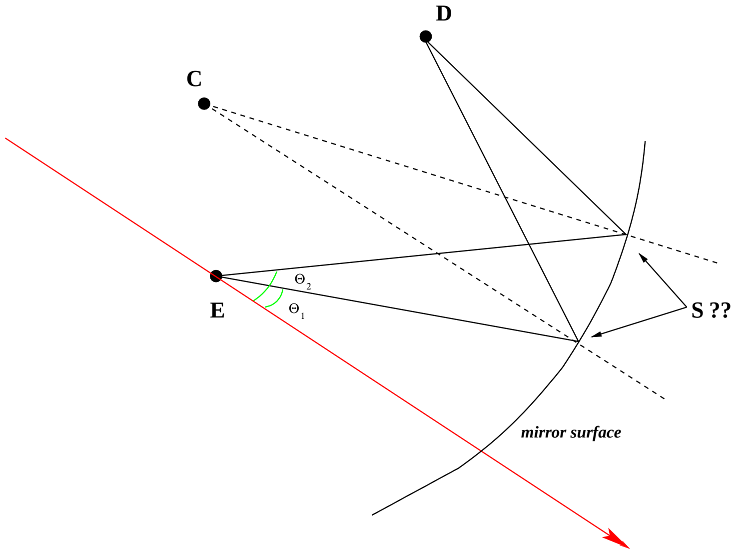

Figure 4.8: The problem of inverse ray tracing. The (red) arrowed line is the track of the charged particle, both indicated values of θ are two possible values of the Čerenkov angle, both indicated values of S are two possible reflection points of the photon on the mirror. The inverse ray tracing allows to find the (unique) position S and hence the Čerenkov angle θ.

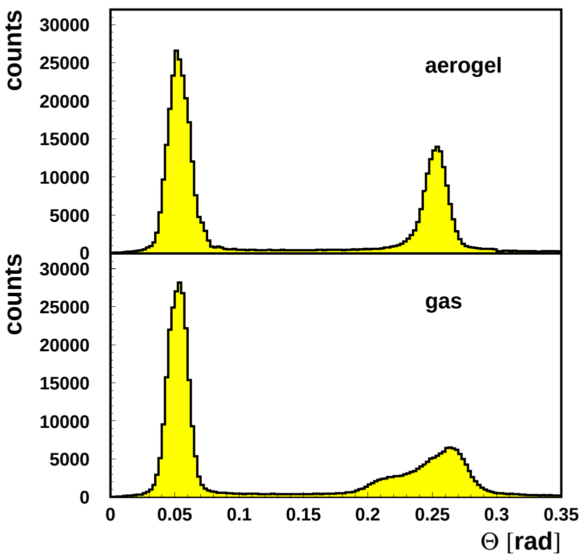

Figure 4.9: Reconstructed Čerenkov angle spectra for the assumption that the photon was emitted in the middle of the aerogel (top) or in middle of the gas (bottom). The data are from single track, low background electron events with $p > 5$ GeV.

Before explaining how this IRT algorithm can be used for hadron identification, we will first focus on the use of IRT for the alignment of the RICH. The reconstructed Čerenkov angle, an intermediate result of the complete IRT algorithm, is the perfect tool for alignment.

4.5.1.2 Alignment of the RICH detector with IRT

The standard deviation of the distribution of the reconstructed photon angles is called the *single photon resolution* σ_θ. This quantity plays an important role in the alignment procedure. Several contributions determine σ_θ, estimates are listed in table 4.3. The contribution $\Delta\theta_{em}$ results from the uncertainty in the emission vertex along the track in the radiator as discussed in the previous section. The detector granularity determines $\Delta\theta_{pix}$ which follows from the size of the "pixels" (the size of the individual photomultipliers). These two contributions affect both the aerogel and the gas angles. The chromatic aberration $\Delta\theta_{chr}$ derives from the variation of the index of refraction with respect to the wavelength and is only important for the resolution in the aerogel. These quantities do not add quadratically since they are strongly correlated and are only listed separately to indicate their relative importance. Their combined effect is calculated in a Monte Carlo simulation program [72] and is denoted as $\Delta\theta_{MC}$.

In a first phase, when IRT was just implemented, it was applied on a data sample of events with a single electron track. With the other PID detectors in HERMES, it is relatively easy to define such a data sample since the efficiency of hadron/lepton separation is above 97%. From fig. 4.1 it is obvious that the Čerenkov angle θ for electrons does not depend on the momentum, so electrons are an ideal testcase for the behaviour of the RICH. If one then used IRT to make a plot similar to fig. 4.9, it became clear that the distribution of reconstructed photon angles was not centered around the theoretical value expected from fig. 4.1 and that the width of the peak was about 3 times larger than the design value $\sigma_\theta \approx 7$ mrad. Another problem that occurred was the incorrect centering of the aerogel and gas rings, as can be seen in fig. 4.5.

Since the resolution is a crucial parameter for hadron identification, both problems asked for an explanation and had to be solved. The problem of the rings not being centered was caused by the fact that the geometry of the RICH detector was not properly implemented in the IRT algorithm. The RICH was constructed and installed within 13 months after approval. After installation there was no time left for precise position measurements of the different building blocks and only some rough information was available. Design studies showed that the position of some components (the position of the photon detector plane and the complete aluminium body of the counter) was not really crucial for the

	aerogel	C_4F_{10}
$\Delta\theta_{em}$	1.8 mrad	2.2 mrad
$\Delta\theta_{pix}$	5.6 mrad	5.2 mrad
$\Delta\theta_{chr}$	2.5 mrad	-
$\Delta\theta_{MC}$	7.1 mrad	7.2 mrad

Table 4.3: Individual contributions to the single photon resolution in aerogel and C_4F_{10}.

performance of the detector, but the geometrical values in the reconstruction algorithm should of course correspond to the real values, otherwise the reconstructed angles won 't agree with the real ones. Since it was technically not feasible to survey the RICH again (and in more detail) after installation, the alignment of the detector was determined from experimental data in an iterative procedure.

First of all the position of the PMT plane was optimised. The orientation of this plane is crucial for the RICH “mapping” file that translates the (integer) number of a certain PMT to the location in space of the corresponding hit. Obviously, the coordinates of the hits strongly determine the indirect ray tracing. To optimise the position of this plane, a 3-dimensional scan of the assumed position was performed and it was studied which setting gave the best centering of the gas and aerogel rings.

A second step in the alignment was the variation of the position of the aluminium counter body in three dimensions. This position is also important for the IRT as it determines the coordinates of the particle tracks in the RICH, a track being the reference axis for the measurement of the reconstructed angles. On the other hand, the position of the counter body also fixes the position of the aerogel and the gas volumes and the IRT assumes that Čerenkov photons are emitted in the middle of the radiator, as already mentioned. Hence the position of the counter body also affects the indirect ray tracing results.

Both optimisations depend on each other and therefore both scans were done iteratively and a common optimisation was found. Variations of 30 cm in all directions were studied and a best position for the photon detector plane and the body was found. A precision of only 5 cm was necessary. These optimisations solved the problem of the rings not being centered and reduced the single photon resolution to ≈ 2 times the design value. The remaining difference could only be due to the mirror.

The mirror array consists of eight individual mirror segments (two adjacent rows of 4 mirror pieces, see fig. 4.4). The radius of curvature of the mirror array is 2.20 m. The short time scale of the RICH installation only allowed for a simple alignment procedure when the mirrors were installed. A point light source at the nominal radius of curvature of the array was imaged on a screen. The mirror segments were then adjusted to minimise the size of the reflected image. To neutralise the unexpected flexibility of the mirrors, their mounting was fine-tuned by adjusting additional pressure points (at the edges of the mirror pieces) where necessary. A nominal 90% of the reflected light was contained in a spot corresponding in size to the pixel size of 23 mm. The actual mirror alignment was determined from the experimental data itself. It consisted of three steps: global alignment, individual alignment and local mirror alignment.

For global mirror alignment the mirror array was treated as a single unit. A 3-dimensional scan of the assumed position of the mirror center was performed and the influence on the reconstructed aerogel and gas rings was studied again using a set of high energy single electron tracks. The performance was characterised using 24 parameters, including the average angles, their spread, the single photon resolution, the yields and the azimuthal uniformity of the gas rings. A careful study resulted in a small range of acceptable values for the position of the mirror center and a shallow optimum was selected. Global alignment reduced the single photon resolution from 14 mrad to 10 mrad. The difference with the design value of 7 mrad could only be due to the individual mirror pieces. The results and the procedure of the global alignment are summarised in a RICH internal report [83].

The junction of the 8 mirror pieces relative to one another is not necessarily perfect. In addition, the construction of the mirror allows for local variations of the radius of curvature of the surface, especially around the stress points introduced to compensate for the flexibility of the individual mirror segments (positioned at the edges of the mirrors). These variations in the mirror radius introduce local variations in the reconstructed angle. This translates into a larger spread on the reconstructed angle, and (if not corrected) a different average angle for some parts of the mirror. In the next two steps of the alignment procedure, the individual mirrors were thus aligned relative to one another (*individual mirror alignment*) and then each mirror was further subdivided and the pieces were individually adjusted (*local alignment*). The complete procedure is explained in [75] and [84].

After completion of the alignment, the single photon resolution was 7.6 mrad for the aerogel and 7.5 mrad for the C_4F_{10} gas, i.e. sufficiently close to the design value of 7 mrad. Once this aim was achieved, hadron identification with the RICH became possible as will be explained in the next section.

4.5.1.3 Particle identification with IRT

For each event the IRT method reconstructs one Čerenkov angle θ for each PMT hit/track combination per radiator, i.e. one gas angle and one aerogel angle per PMT hit for each track. The step from these reconstructed angles to the identification of hadrons is made through a likelihood analysis [84, 85]. In this analysis the average reconstructed angles and the threshold behaviour for the two radiators have to be combined in a coherent fashion. Finally there should also be a possibility to assess the quality and reliability of the particle identification.

If a track is defined as a hadron track by the other PID detectors in HERMES, there are three possible particle hypotheses: the hadron can be a pion, kaon or proton. The method to assign a particle type to a certain track depends on the momentum of the track. First we will describe the 'pure' RICH behaviour of the Čerenkov counter and add the threshold aspects in a second step.

As a starting point we consider only one radiator. It is rather simple to set up a likelihood algorithm if the momentum of the track is above the Čerenkov threshold for all particle type hypotheses. In this case a certain number of Čerenkov photons is expected for each hypothesis, so it makes sense to reconstruct the Čerenkov angles for each PMT hit. From the reconstructed angles an average angle $\langle\theta\rangle$ could in principle be determined. The angular range over which the average is formed must be limited to minimise the impact of background hits and hits from other tracks and from the other radiator. Consequently, 'windows' are imposed on the spectrum of reconstructed angles that are centered around the theoretically expected angles for the three particle type hypotheses. Hence, we do not calculate a single average $\langle\theta\rangle$, but an average for each particle hypothesis $\langle\theta\rangle^\pi$, $\langle\theta\rangle^K$, $\langle\theta\rangle^p$ around respectively θ^π_{theo}, θ^K_{theo} and θ^p_{theo}. The width of the windows around θ^π_{theo}, θ^K_{theo} and θ^p_{theo} is a parameter that can be optimised. In the present case this width (σ_w) is set to a fixed value of $\pm$ 16 mrad, which corresponds to about 4 times the single photon resolution σ_θ.

For a large number of tracks of a given particle type and momentum the resulting average

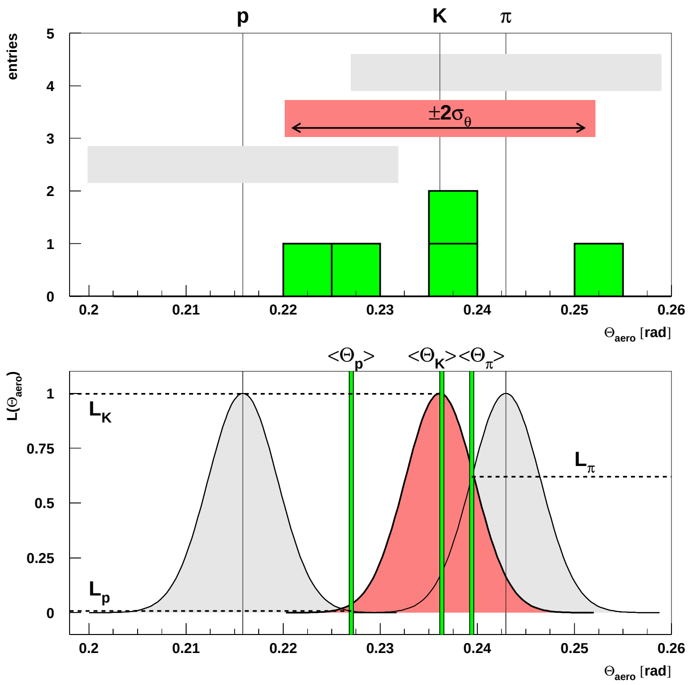

Figure 4.10: Principle of the IRT likelihood algorithm. The upper figure shows the five reconstructed aerogel angles for an event with a 8.2 GeV kaon as squares, together with the theoretically expected angles and the respective windows. The lower figure shows the resulting average angles, the Gaussian distributions around the theoretical angles and the resulting likelihood values for the three hypotheses pion, kaon and proton.

angles also follow a Gaussian distribution around the theoretical angle, thus a Gaussian likelihood can be formed for each particle hypothesis i ($i = \pi, K, p$):

$$L_i \equiv L(\langle\theta\rangle^i) = exp\left[-\frac{(\theta^i_{theo} - \langle\theta\rangle^i)^2}{2\sigma^2_{\langle\theta\rangle}}\right] \tag{4.8}$$

where $\sigma_{\langle\theta\rangle}$ is the average angle resolution which is calculated from the single photon resolution σ_θ as

$$\sigma_{\langle\theta\rangle} = \frac{\sigma_\theta}{\sqrt{\langle N_{PMT}\rangle - 1}} \tag{4.9}$$

The average expected number of PMT hits $\langle N_{PMT}\rangle$ is in principle a function of particle type, momentum and radiator. Due to the acceptance effects resulting from the limited sizes of the aerogel tiles, the mirrors and the photon detector it also depends on the parameters of the individual track. Hence, the best possible value for $\langle N_{PMT}\rangle$ would require a detailed simulation of the track. As will be shown further on, the IRT method works fairly good with the assumption that $\langle N_{PMT}\rangle$ is a constant, set to fixed values of 6 for aerogel and 11 for the gas. Fig. 4.10 [84] illustrates the principle of the likelihood calculation.

For particles with momenta above the threshold momentum p^i_{th} with a number of hits N^i_{PMT} in the window belonging to the particle hypothesis i, the likelihood L_i is calculated

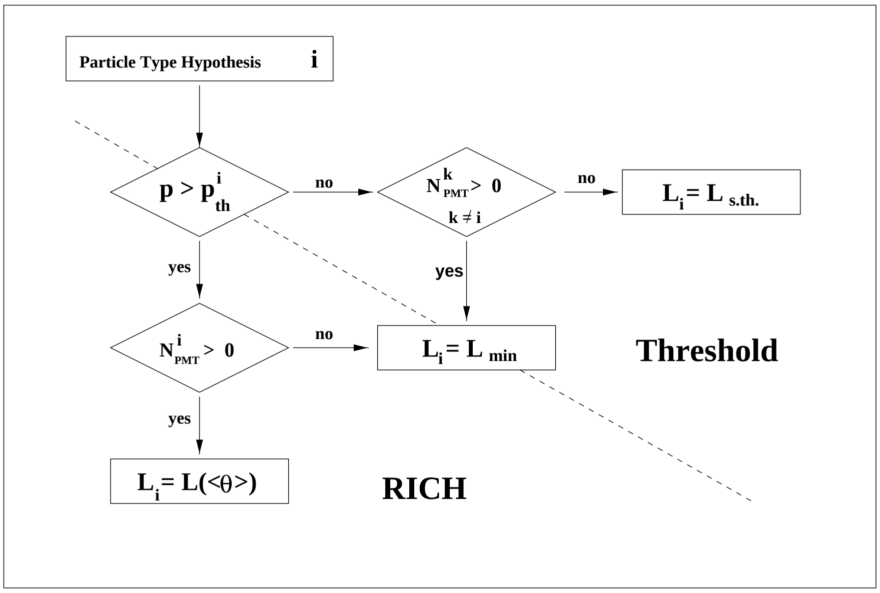

Figure 4.11: Flowchart of the IRT likelihood algorithm (for one radiator) including threshold treatment. The dashed line separates the RICH and the threshold related parts of the algorithm.

according to eq. (4.8). However, there are two other cases that must be considered: the case that the particle momentum is below the threshold, were no hits are expected and consequently no window is imposed, and the case that the particle momentum is above threshold, but no hits are detected. For a coherent analysis a likelihood value must be assigned to both cases. This likelihood value should connect smoothly to the values calculated according to eq. (4.8).

In the first case (when $p < p^i_{th}$) a 'sub-threshold likelihood' $L_{s.th.}$ is assigned. This sub-threshold value should be somewhat lower than the typically expected likelihoods for measured angles. However, if hits are detected in one of the windows that belong to another particle type hypothesis k ($N^k_{PMT} > 0$), this sub-threshold value is replaced by the minimal likelihood L_{min}. This ensures that the likelihood for the hypothesis k will be larger than the likelihood for the sub-threshold hypothesis i. The minimum likelihood value is defined as the likelihood for an average angle at the edge of the window:

$$L_{min} = L(\theta_{th} + \sigma_w) \tag{4.10}$$

This represents the smallest value that can occur for actually detected hits. This setting of the minimum likelihood guarantees that it is lower than any 'real' likelihood, that it connects smoothly to it and that there is a well defined lower limit.

In the second case, where the momentum is above threshold but no hits are detected, the likelihood is set to the same minimum value L_{min}. This reflects the fact that both cases should carry about the same weight and should be equally un-likely. The flowchart for this algorithm is shown in fig. 4.11.

Two additional features have been added to the scheme presented in fig. 4.11 to improve the hadron identification. First of all the background has to be taken into account. Any number of observed hits that is not higher than the expected number of background hits (B_{PMT}) is indistinguishable from background. To reflect this, the average angle is only formed (and the corresponding likelihood calculated) if the number of fired PMTs is larger than the expected background. So, instead of comparing N^i_{PMT} with 0, as shown in fig. 4.11, we compare N^i_{PMT} with B_{PMT}. Secondly, right at the Čerenkov threshold a sizeable fraction of all particles does not generate any PMT hits. This is simply a matter of statistics. The threshold momentum p_{th} is therefore replaced by an effective threshold momentum $p^{'}_{th} = p_{th} + \Delta p$ where Δp was determined by a Monte Carlo simulation and is 0.6 for the gas and 0.0 for the aerogel [84].

Finally the likelihoods for the two radiators, aerogel and gas, are combined to an overall likelihood by multiplication

$$L_i^{tot} = L_i^{aerogel} \cdot L_i^{gas} \tag{4.11}$$

The total IRT likelihood L_i^{tot} is then used to define the most likely hadron type and allows to calculate a 'quality parameter' rQp that quantifies how well the particle was identified. The most likely hadron type is the one with the highest likelihood. The quality parameter rQp is defined as

$$rQp = log_{10}(\frac{L_1^{tot}}{L_2^{tot}}) \tag{4.12}$$

where L_1^{tot} and L_2^{tot} are the likelihoods of the most likely and second most likely hadron type hypothesis. In the case that $L_1^{tot} = L_2^{tot}$ the particle is considered 'unidentified'.

The efficiency of the IRT method to identify pions, kaons and protons will be studied in detail in sections 4.7 and 4.8. These studies will reveal that the IRT method has problems to identify kaons in the low momentum region. Hence, it is worthwhile to consider another algorithm for PID with the RICH: the direct raytracing.

4.5.2 Direct Ray Tracing

The DRT method starts with the track parameters and, similar to a Monte Carlo simulation, simulates hits in the photon detector for each particle hypothesis. From these photon hits, the probability to fire a certain PMT is calculated and the combined hit probabilities for the complete PMT plane are called a 'hit probability distribution'. This hit probability distribution is calculated for each of the three different particle type hypotheses, and is then compared to the actual hit pattern in the RICH. The best matching hypothesis determines the actual particle type. There is no intermediate result from DRT (like the reconstructed angles for IRT) but the hit probabilities are directly used for the particle identification.

First of all there has to be an algorithm to generate the Čerenkov photons and to determine where they hit the photon detector plane. For each track t (with momentum p) and each particle hypothesis h (π, K, p) and each radiator r (gas/aerogel) 360 photons are generated, uniformly distributed in their azimuthal angle around the particle track. Also the photon vertices are uniformly distributed along the particle trajectory in the radiator. Since the momentum p is known also the Čerenkov angle θ (see fig. 4.1) is known for each particle hypothesis, so the photons are generated with this angle (relative to the track).

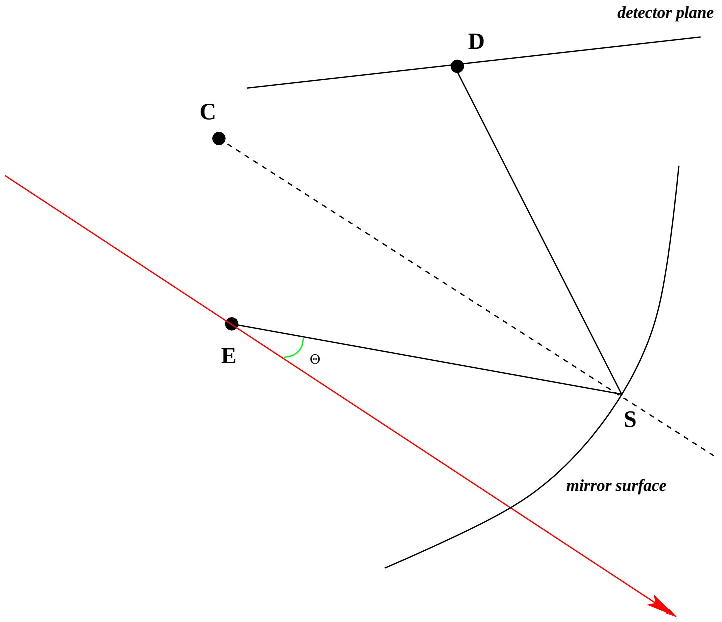

Figure 4.12: The problem of direct ray tracing. The (red) arrowed line is the track of the charged particle. S is the reflection point of the photon on the mirror. The direct ray tracing allows to find the (unique) position S and hence the point D where the photon hits the detector plane.

After generating the photons they have to be propagated through the RICH to find their impact point on the photon detector after reflection on the mirror. This is the aim of direct ray tracing. Using the terminology of fig. 4.12, the geometrical problem can be formulated as:

Given the emission point E of a photon (*the photon vertex*), its Čerenkov angle θ and point C (*the center of the sperical mirror*), determine the point D where the photon hits the detector plane.

A mathematical solution of this problem can be found in [81, 86]. This technique allows to find the impact point on the photon detector for all the generated photons. Photons that miss the mirror are ignored. The number of photons incident on the i-th PMT for a particle hypothesis h, a track t and a radiator r is stored in $N_{PH}^{(h,t,r)}(i)$. The Monte Carlo probability $P^{(h,t,r)}(i)$ of the i-th PMT to be struck by the photons of the (h,t,r) ring follows as

$$P^{(h,t,r)}(i) = \frac{N_{PH}^{(h,t,r)}(i)}{N_{PH}^{(h,t,r)}} \tag{4.13}$$

where $N_{PH}^{(h,t,r)} = \sum_i N_{PH}^{(h,t,r)}(i)$. This probability is not yet suitable for the particle identification: it is radiator specific, it takes only the geometrical tracking of the photons into account and it does not contain any information about specific detector related effects (e.g. background, the probability that a PMT will effectively fire if hit by a photon, the experimentally observed number of hits in a ring, ...) or physical processes (e.g. absorption, scattering of photons). This probability has to be translated into another 'more realistic' probability $P_{PMT}^{(h,t)}(i)$ that expresses the probability that a PMT i will fire under the hypothesis h for track t (and does not depend anymore on the radiator hypothesis). This happens in the following way:

The mean number of photoelectrons (i.e. photons which give a response in a PMT) in the i-th PMT can be written as:

$$S_{PE}^{(h,t,r)}(i) = n^{(h,t,r)} P^{(h,t,r)}(i) \tag{4.14}$$

where $n^{(h,t,r)}$ is the total number of photoelectrons expected for the (h,t,r) ring. This number can be estimated from the Frank-Tamm relation (4.2) according to which the number of produced Čerenkov photons is proportional to

$$1 - \frac{1}{\beta^2 n^2} \tag{4.15}$$

where we neglect the wavelength dependence of the refractive index since the RICH can not yield any information about it. Using this relation, the total number of photoelectrons expected for the (h,t,r) ring can be approximated as

$$n^{(h,t,r)} = N_0^r + N_{relat}^r \left(\frac{1 - \frac{1}{\beta_{(h,t)}^2 n_r^2}}{1 - \frac{1}{n_r^2}} \right) \tag{4.16}$$

The value of $\beta_{(h,t)}$ can be calculated using eq. (4.3) for a track t with a given momentum and hypothesis h. N_{relat}^r is the average expected number of photoelectrons for a fully relativistic particle for radiator r, and N_0^r is the expected number of photoelectrons if there should not be a ring (i.e. the noise for sub-threshold behaviour). The values of N_{relat}^r and N_0^r were determined from experimental data where only so-called 'ideal tracks' were used to count the number of PMTs in an experimentally observed ring. These are tracks that are not influenced by the limited acceptance of the photon detector plane and the mirror. It is important to avoid these acceptance effects in the determination of N_{relat}^r and N_0^r, since the direct ray tracing itself accounts for them. These effects are thus already implicitely included in the probability $P^{(h,t,r)}(i)$ and they should not be double counted. Studies of experimental data yielded values for $N_{relat}^{gas} = 12$, $N_{relat}^{aerogel} = 8$ and $N_0^{gas} = N_0^{aerogel} = 0.5$

Summing both radiators for a given hypothesis, the total number of photoelectrons in the i-th PMT becomes:

$$N_{PE}^{(h,t)}(i) = \sum_r S_{PE}^{(h,t,r)}(i) \tag{4.17}$$

Assuming a Poisson distribution of the photoelectrons $N_{PE}^{(h,t)}(i)$, the probability that the i-th PMT does *not* respond can be evaluated as

$$\bar{P}_{PMT}^{(h,t)}(i) = exp(-N_{PE}^{(h,t)}(i)) \tag{4.18}$$

The probability for the PMT to fire becomes accordingly

$$P_{PMT}^{(h,t)}(i) = 1 - \bar{P}_{PMT}^{(h,t)}(i) \tag{4.19}$$

This last probability can be used for particle identification by calculating the likelihood for each particle hypothesis h for the track t:

$$L^{(h,t)} = \sum_i log_{10}[P^{(h,t)}_{PMT}(i)C_{PMT}(i) + \bar{P}^{(h,t)}_{PMT}(i)(1 - C_{PMT}(i))] \tag{4.20}$$

where $C_{PMT}(i)$ is 1 or 0 depending on whether the PMT i did or did not fire for the actual observed hitpattern in the concerned event. The hypothesis which maximises $L^{(h,t)}$ will be assumed to be true for track t.

There are no distinct likelihoods for the gas and the aerogel, as was the case for IRT, and there is no detailed treatment of sub-threshold behaviour or the minimum number of hits. Just as for IRT, a quality parameter rQp is defined according to eq. (4.12).

The DRT algorithm is also equipped with an aerogel/gas boundary refraction for aerogel photons. This treatment is different compared to IRT where an aerogel refraction correction was a posteriori applied to the reconstructed angles, while for DRT we simulate the refraction for the generated photons and compare afterwards to the actual observed hitpattern. Another feature of the DRT algorithm is called ‘blobbing’. The aim of blobbing is to simulate the scattering effects that occur in the RICH detector (e.g scattering of photons at the aerogel tile edges, Rayleigh scattering, scattering due to imperfect mirror alignment, ...). These scattering effects broaden the ring widths and blobbing was developed to take this into account [87]. For each PMT that was hit by a photon generated by DRT, each adjacent pixel is assigned the hit of a fractional photon. This effectively enlarges the ring with an appropiate fall-off towards the outer edges of the ring, as shown in fig. 4.13. Blobbing compensates for imperfections in the ‘experimental’ RICH and improves the hadron identification of the DRT method.

The disadvantage of the DRT method is the large number of Monte Carlo calculations which have to be performed for each track, particle hypothesis and radiator, which makes this algorithm several times slower than the IRT based method. This also explains why a track based DRT method was implemented and not an event based method, as was originally aimed for [81]. An event based method would use combined particle hypotheses H for each event instead of a single particle hypothesis h for each track. A combined particle type hypothesis can be written as

$$H \equiv \{(h_k) \in (\pi, K, p)), k = 1..N_{track}\} \tag{4.21}$$

For n unidentified tracks and 3 particle hypotheses, the number of combined hypotheses H (and thus photon patterns to generate) would increase from $3.n$ to 3^n, which would make the DRT method slower. Moreover, the format of the HERMES data does not easily allow for an event based PID method.

The algorithms for IRT and DRT described above already existed before this Ph.D. work began and a first rough implementation was available. The main effort in this work was in understanding the algorithms, debugging and improving them and estimating their efficiencies, together with using the IRT for the alignment procedure. The final code for hadron identification with the HERMES RICH was a common effort of the complete RICH group and is documented in [88]. The programming language is C.

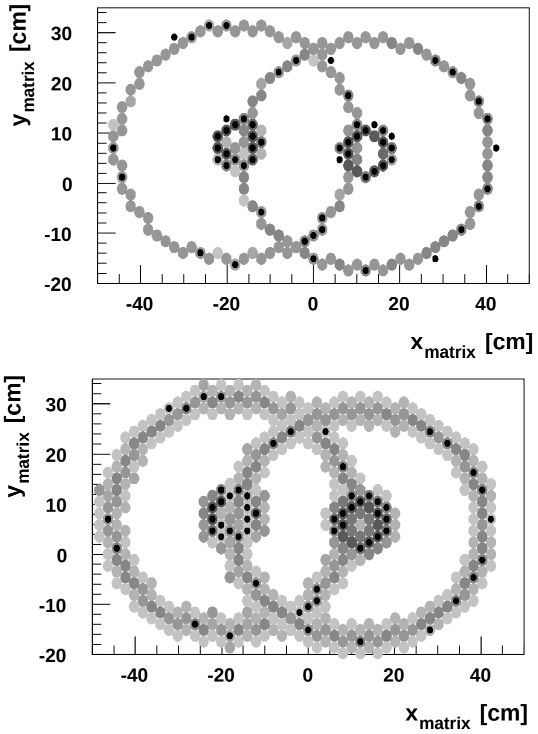

Figure 4.13: The upper panel shows a typical event without blobbing, the lower panel shows the effect of blobbing for the same event. The black points are the actual PMT hits in this event. The grey scale indicates the DRT probability of a PMT to be hit by a photon.

4.5.3 RICH PID Scheduler

Both PID methods have different strengths and weaknesses. DRT is more time consuming than IRT. In general DRT seems to work better for lower momenta and IRT works better for higher momenta, as will be shown in section 4.7. In addition it has become clear that their performance can be heavily dependent on the topology of the event. As a result the RICH PID Scheduler (RPS) was developed [87]. It decides which of the two reconstruction methods gives the more reliable result for a given track. While it is based on two track-level methods it has an event-level component since it takes into account the topology of the event (number of tracks per detector half, overlap of the resulting rings). The method represents a decision network based on a lookup table that is trained with Monte Carlo data.

The flowchart for the RPS is shown in fig. 4.14. For each track in a certain event three properties of the track are compared to a predefined binning (the momentum of the track, the multiplicity of the event and the overlap parameter; see below). A lookup table will decide which PID algorithm is the best for the bin to which the track belongs. The choice of the binning is based on the following arguments:

- The performance of DRT and IRT is momentum dependent. The momentum of the track thus plays a role to decide which method to take. The momentum is binned in 1 GeV bins, ranging from 0.0 to 30.0 GeV.
- The track multiplicity of the event is binned in 4 bins, according to the case with 1, 2, 3 and > 3 tracks in the concerned event. These bins contain respectively 83.9, 14.6, 1.4 and 0.1% of all the events, so it is not necessary to define a finer binning for higher track multiplicities.
- The "overlap parameter" of the track expresses the degree to which the concerned Čerenkov ring for a track overlaps with another ring. This parameter can of course not be defined for a single track event. For the events with 3 and > 3 tracks, statistics are too low to allow the unfolding of the multiplicity bin with an overlap bin. Hence this overlap parameter (x_{olp}) is only defined for events with 2 tracks:

$$x_{olp} = 1 - \frac{|\Delta x|}{r_1 + r_2 + r_{PMT}} \tag{4.22}$$

  where $|\Delta x|$ is the distance between the centers of both rings, r_1 and r_2 are the radii of the rings and r_{PMT} is the radius of the photomultiplier. Since the rings are in general not circles but asymmetric distorted ellipses, their center and their radius has to be approximated. The center of the ring can be approximated by the 'virtual track hit point', i.e. the point where the track would hit the photon detector plane if it would have been reflected by the mirror. This virtual track hit point can easily be calculated with a DRT method. The radius of a ring is then measured along the line that connects both virtual track hit points [89]. Rings that are just touching each other have a x_{olp} of 0, for ring with the same virtual track hit point, x_{olp} is 1. As the maximum aerogel ring size corresponds to about 15 PMT diameters, the range in x_{olp} from 0 to 1 was divided in 15 bins in order to have approximately one bin for one PMT overlap. Rings that do not overlap have a negative overlap

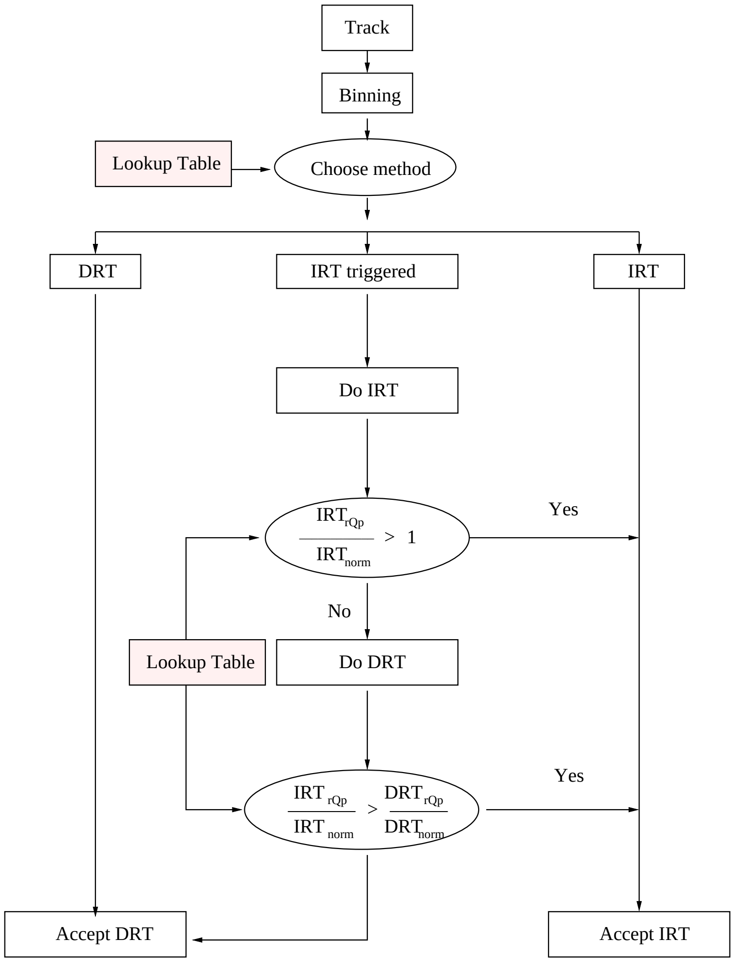

Figure 4.14: The decision flowchart for the RICH PID Scheduler.

parameter and are all stored in the same bin. If it is not possible to calculate the overlap parameter (e.g. in the case of near-threshold rings) a separate bin for x_{olp} is taken. Remark that cases where both a gas and an aerogel ring are present for a certain track do not cause ambiguities: aerogel rings are larger and hence the first to overlap, so the aerogel ring is considered in these cases. If only an aerogel ring exists, there is no problem at all and the case where only a gas ring is present does not occur (see fig. 4.1).

The resulting binning in momentum, track multiplicity and overlap parameter is summarised in table 4.4. For each track its location in the binning table is determined and a lookup table (generated with a Monte Carlo simulation) decides which PID method performs the best for this track. The choice between IRT and DRT is not always obvious: in some bins it is hard to predict which method will perform the best. In that case the 'IRT triggered' method is used. IRT triggered will always first try the IRT method (since this is the fastest method) and check if the resulting quality parameter is higher than a certain constant IRT_{norm} (which is also bin dependent and is stored in a second lookup table). If this is the case the result from IRT is accepted; if not, the DRT method will

Multiplicity bin	number of momentum bins	number of bins in x_{olp}
1	31	1
2	31	17
3	31	1
> 3	31	1

Table 4.4: Binning for the RICH PID Scheduler.

be tried. The normalised quality parameter from DRT is compared to the one from IRT and if

$$\frac{DRT_{rQp}}{DRT_{norm}} > \frac{IRT_{rQp}}{IRT_{norm}} \tag{4.23}$$

the result from DRT is accepted. If not, IRT will be accepted. The RPS only provides the identified particle type, not an additional quality parameter as it can not be defined anymore coherently.

As can be seen from the flowchart in fig. 4.14 the RPS method is based on two lookup tables: the first one decides which of the three PID methods (IRT, DRT or IRT triggered) performs the best for a certain bin (*method lookup table*), the second lookup table is only needed for IRT triggered and stores the values for IRT_{norm} and DRT_{norm} (*IRT triggered lookup table*). Both lookup tables were generated by a training procedure, using Monte Carlo data. The parameter maximised in the training process is the so-called 'method quality' F:

$$F = \frac{\sum_{(i=\pi,K,p)} C_i^P P_i^i + \sum_{(i=\pi,K,p)} C_i^Q Q_i^i}{\sum_{(i=\pi,K,p;\ j=P,Q)} C_i^j} \tag{4.24}$$

where e.g P_K^K is the *kaon identification efficiency*, i.e. the probability that a real kaon is identified by the RICH as a kaon, and e.g Q_K^K is the *kaon purity*, i.e. the probability that a particle that was identified by the RICH as a kaon is truly a kaon. The reason why the 'method quality' F takes both the identification efficiencies and the purities into account, can be illustrated with a simple example: consider a sample with 100 hadrons in total with truly 70 pions, 20 kaons and 10 protons. If the RICH would identify all these 100 hadrons as kaons, this would yield a perfect kaon efficiency ($P_K^K = 1.0$) but the kaon purity would be very bad ($Q_K^K = 0.2$). On the other hand, if the RICH identifies 1 of the true kaons as a kaon and the other 99 hadrons as pions, the kaon purity is perfect ($Q_K^K = 1.0$), but the kaon efficiency would be poor ($P_K^K = 0.05$). Only a combination of efficiencies and purities can thus be used as a valid parameter to judge the quality of a PID method. The 'method quality' F is a weighted average of the efficiencies and purities. The chosen values for the method quality coefficients $C_i^{P(Q)}$ are listed in table 4.5. This setting was chosen under the assumption that purity is more important than efficiency, and that kaons are more important than pions or protons. However, it was shown that a 'flat' setting of $C_i^{P(Q)} = 1$ for all coefficients produces virtually identical results. Only more extreme settings select for a particular behaviour [87]. Both the method lookup table and the IRT triggered lookup table were generated by using a brute force method: for each bin from table 4.4 the three PID methods (DRT, IRT and IRT triggered) were executed and the one with the maximum F was chosen for the method lookup table. In case IRT triggered was preferred, the lookup table for IRT_{norm} and DRT_{norm} was generated by using a binning of these normalisation parameters and searching for the ones that maximised F.

C_π^P	C_K^P	C_p^P	C_π^Q	C_K^Q	C_p^Q
1	2	1	2	4	2

Table 4.5: Standard setting for the method quality coefficients.

Results from the RPS will be presented in section 4.7 and it will be shown that especially in the low momentum region improvements in the particle identification can be noticed, while only in roughly 20% of the cases DRT is used. Therefore the computing time does not increase too significantly.

4.6 Tracking with the RICH detector?

About 5% of the observed rings in the HERMES RICH cannot be associated with a track in the front and/or back region of the HERMES spectrometer. So, the question arises if one can use the RICH as a tracking device to reconstruct the track of the particle, starting only from its hitpattern seen in the RICH. A small stand-alone Monte Carlo code based on the Hough transform method was written to answer this question. The Hough transform is a powerful method for detecting edges and is frequently used in different types of pattern recognition problems (e.g. detection of landmines). Some high-energy experiments (e.g. [116]) applied this technique successfully for Ring Imaging Čerenkov detectors. Since Hough transforms are mathematically rather complex and not easy to understand intuitively no further information will be given here. At first sight Hough transforms were not only promising for track independent ring fitting but also for dealing with the problems of overlap and noise in the HERMES RICH. More information and some results of the stand-alone Monte Carlo code that investigates the results of Hough transforms can be found in the internal report [117]. The conclusion from this report is that track independent ring fitting is not possible for the HERMES RICH, due to the rather large pixel size of the photomultipliers and the small number of hits that are observed. Nevertheless, some ideas from [117] were used in the alignment procedure described in section 4.5.1.2.

4.7 The PID efficiency of the RICH

Any hadron that is identified by the RICH detector as either a pion, kaon or proton, is only identified with a certain probability and there is always a chance for misidentification. Efficiencies and misidentifications are collected in the so-called P matrix [90]. This P matrix is not only important to judge the PID performance of the RICH but it will also play a crucial role in the *unfolding of hadron momentum distributions*, as will be described in section 4.9. A detailed knowledge of the P matrix is thus required.

4.7.1 The P matrix: hadron identification efficiencies

Each hadron has a 'true' particle type which is in principle only known in a Monte Carlo simulation, but not in experimental data. Every hadron is also assigned an 'identified' particle type based on the result of the RICH PID algorithm. Depending on the choice of identification algorithm and cuts it may be impossible to identify some particles. Their particle type is 'unidentified', labelled as X.

The most fundamental observables are the numbers of particles N_t^i of each particle type t that cross the detector and that are identified as particle type i.

$$N_t^i \quad \begin{array}{l} \leftarrow \textit{identified particle type} \\ \leftarrow \textit{true particle type} \end{array} \tag{4.25}$$

A useful set of basic parameters for the description of the performance of the RICH are the *identification probabilities* P_t^i. They give the probability that a particle with true type t is identified as type i:

$$P_t^i = \frac{N_t^i}{\sum_j N_t^j} \tag{4.26}$$

Together they form the matrix P, which in the case of three particles (π, K, p) has the form

$$\mathsf{P} = \mathsf{P}(\pi, K, p) = \begin{bmatrix} P_\pi^\pi & P_K^\pi & P_p^\pi \\ P_\pi^K & P_K^K & P_p^K \\ P_\pi^p & P_K^p & P_p^p \\ P_\pi^X & P_K^X & P_p^X \end{bmatrix} \tag{4.27}$$

The last row of this matrix concerns the unidentified particles. The diagonal elements P_i^i of the 3×3 submatrix of the P matrix without the last row are called *identification efficiencies*

$$\epsilon_i = P_i^i \tag{4.28}$$

The remaining elements of the P matrix are called *misidentifications*. Due to the definition of the probabilities, the columns of this matrix are normalised to 1

$$\sum_i P_t^i = 1 \tag{4.29}$$

The P matrix elements are sensitive to many effects (detector properties, the particle identification algorithm, topology of the events, momentum of the particles, position of the track in the detector,...). Binning the P matrix in all parameters defining its behaviour is not feasible, and it was decided to bin only the 2 most important parameters (for IRT and RPS): the momentum of the track and the track multiplicity of the event (with a distinction between 1, 2 and more than 2 tracks per detector half).

In principle there are two methods to determine the P matrix, i.e. a Monte Carlo simulation and the use of decaying particles. In a Monte Carlo simulation both the identified particle type and the true particle type are known, which makes it possible to calculate P matrix elements. An example for the proton efficiency $P_p^p = \epsilon_p$ resulting from an (obsolete) Monte Carlo production is shown in fig. 4.15 [87]. It is clear that RPS is always better than IRT or DRT and in some momentum bins even better than both of them. The reason why the Monte Carlo used for this plot is obsolete and how one can determine identification efficiencies from experimental data only will be discussed in the next section.

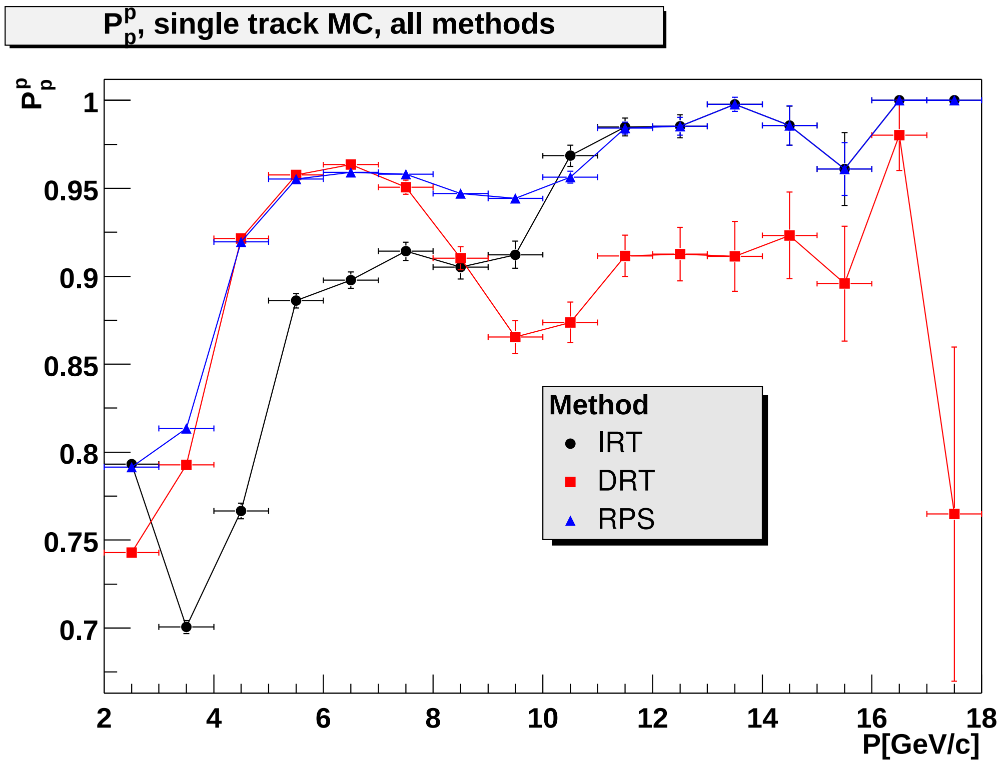

Figure 4.15: Proton efficiency from a (obsolete) Monte Carlo for IRT, DRT and RPS (the figure was taken from [87]). Only events with one track in the detector half were used to calculate the efficiency.

4.7.2 The P matrix and decaying particles

The RICH is the only detector in the HERMES experiment that can distinguish between different hadron types. Accordingly there is only one possibility to know the true particle type by using experimental data only, that is by making use of the decay products of different hadrons. As an illustration, the decay of the Φ meson in two kaons (Φ(1020 MeV) $\rightarrow K^+K^-$) can be used (fig. 4.16). If the invariant mass for two hadrons

$$M_{h_1h_2} = \sqrt{p_1^2 + m_1^2 + p_2^2 + m_2^2} \tag{4.30}$$

is calculated by using the kaon mass for both hadrons and if a peak in the invariant mass spectrum M_{KK} around the Φ mass (1020 MeV) is observed, one knows that every pair of hadrons that gives an invariant mass underneath the peak, is a pair of two kaons. Taking 2 random hadrons to calculate the invariant mass M_{KK} does not give a clear peak, hence one of the hadrons is required to be identified by the RICH as a kaon. For the second hadron 4 different histograms are filled, depending on the particle type that the RICH PID assigned to the second hadron (π, K, p, X). The second hadron is required to have a charge opposite to that of the first identified kaon.

In the case of a 100% efficient RICH PID the invariant mass in the histograms where the second hadron is not identified as a kaon should be purely background with a continuous distribution. The peaks observed in the histograms (K,π), (K,p) and (K,X) indicate

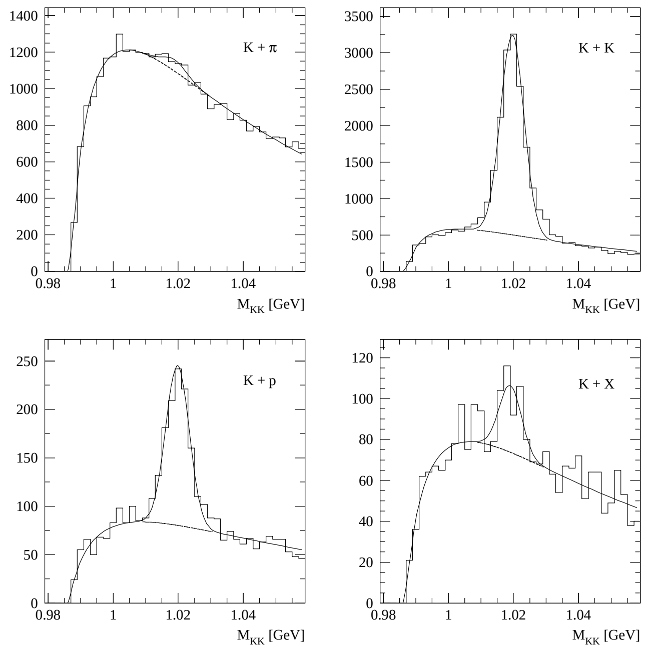

Figure 4.16: Invariant mass distributions of two hadrons where for both hadrons the kaon mass has been used to calculate the invariant mass. One of the hadrons is identified by the RICH as a kaon, for the other hadron in the pair the hadron type assigned by the RICH is indicated by the label under the horizontal axis on each panel. This second hadron has a momentum between 3 and 4 GeV in this example. The curves represent fits to the Φ peak and the background.

that the RICH is not perfectly efficient. The peaks in these latter 3 M_{KK} histograms are formed by Φ decays where the second hadron is truly a kaon but misidentified by the RICH as respectively a π, p and X particle. The efficiency and the misidentifications collected in the kaon column P_K^i of the P matrix can now be calculated as:

$$P_K^K = N(KK)/N(KH) \tag{4.31}$$
$$P_K^\pi = N(K\pi)\,/N(KH) \tag{4.32}$$
$$P_K^p = N(Kp)\,/N(KH) \tag{4.33}$$
$$P_K^X = N(KX)/N(KH) \tag{4.34}$$

where $N(KH) = N(KK)+N(K\pi)+N(Kp)+N(KX)$ and where e.g. $N(K\pi)$ represents the number of events in the peak of the invariant mass distribution $(K\pi)$ after background subtraction. This procedure is repeated for each momentum bin for the second hadron (the momentum is binned in 1 GeV bins, from 2 to 15 GeV) and for each topology case for the P matrix (in fact, decaying Φ mesons only allow to study the cases with 2 and more than 2 tracks per detector half). As an example the kaon column of the P matrix for the topology with 2 tracks per detector half is shown in fig. 4.17 for IRT, DRT and RPS. Although the kaon efficiency ϵ_K for the RPS method is in general higher than the one for DRT and IRT, the RPS seems to fail in the momentum bins between 2 and 3 GeV and between 10 and 11 GeV. The reason is twofold: as explained in section 4.5.3 the RPS

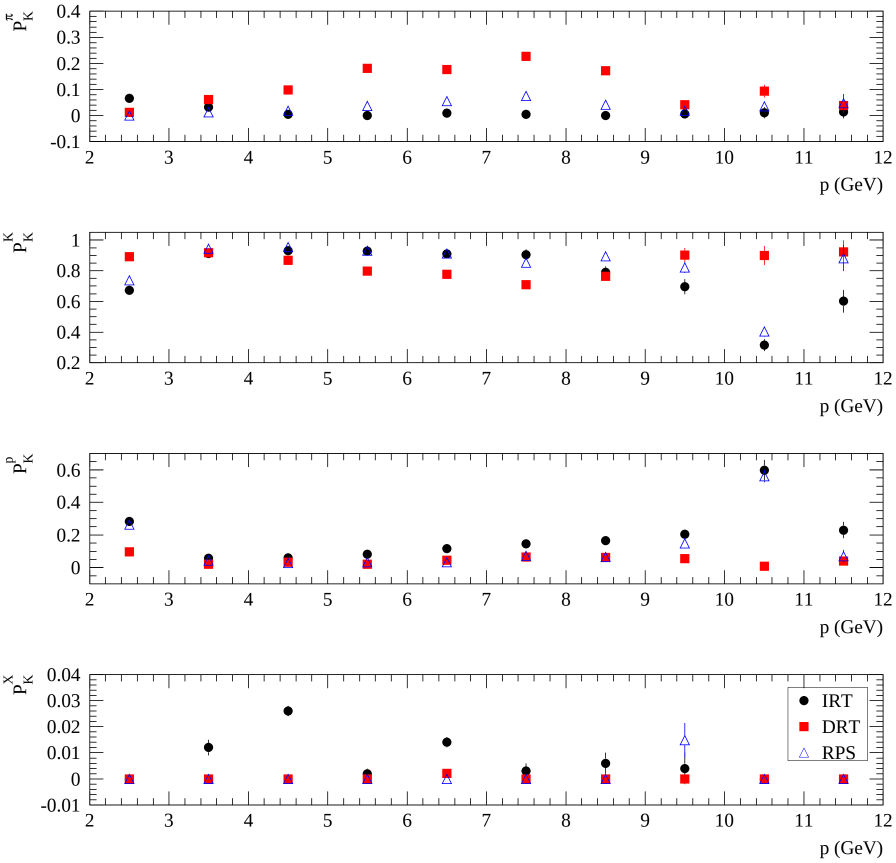

Figure 4.17: Momentum dependence of the kaon efficiencies and misidentifications from decaying Φ mesons, using experimental data only, for the topology case with 2 tracks per detector half. Each panel has a different vertical scale.

uses two lookup tables which were trained by Monte Carlo data. At the time that the RPS was programmed only a preliminary version of the Monte Carlo (MC) program was available which did not simulate the real experimental RICH well. This version of the MC was used to generate the lookup tables for RPS and therefore it is understandable that the RPS has some problems with experimental data. In this Ph.D. work the Monte Carlo program has been completely redesigned (as will be described in the next section) and this new MC will be used to 'train' the RPS again and to generate new lookup tables. Nevertheless, if one uses the RPS lookup tables from the 'bad' Monte Carlo to analyse data from the same 'bad' MC, the RPS in principle works as was illustrated in fig. 4.15. The Argonne research group is currently generating the new RPS lookup tables with the improved version of the Monte Carlo program. The second reason why the RPS does not outperform IRT and DRT in all momentum bins with respect to the kaon efficiency, can be found in the parameter that was maximised in the RPS training process; the method quality F (eq. (4.24)). It was shown in [87] that losing a few percent of kaon effiency can improve the kaon purity by an amount larger than a factor of two. Therefore, if RPS chooses the PID method which yields the lowest kaon efficiency, it sometimes means that

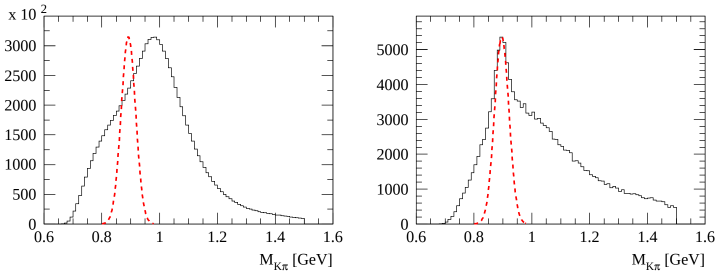

Figure 4.18: Experimental πK invariant mass spectrum. Left panel: first particle identified by the RICH as π^+, second particle identified as negative hadron. Right panel: first particle identified as π^+, second particle identified as K^-. The dashed (red) line represents the mass and the width of the K^* as expected from the Particle Data Book [92].

the kaon purity is superior for this PID method. Also the pion and proton identification efficiencies and purities are included in F and the RPS attempts to find a compromise for the three hadron types.

Apart from Φ decays, other decaying particles can be used to study the P matrix elements

$$\rho^0 \rightarrow \pi^+ + \pi^- \tag{4.35}$$
$$K_s \rightarrow \pi^+ + \pi^- \tag{4.36}$$
$$\Lambda \rightarrow p + \pi^- \tag{4.37}$$

These decays have been studied in [91] for the IRT method and for the decaying Φ mesons an excellent agreement with the IRT points in fig. 4.17 has been found. Apart from the Φ, another good candidate to study the kaon elements from the P matrix would be the K^*(892 MeV) decay into a kaon and a pion. Although the RICH is capable of identifying the K^* meson, as can be seen in fig. 4.18, it turned out that its decay is not useful to obtain information about the P_K^i elements. If one wants to study these matrix elements, the 4 invariant mass distributions where the first hadron is identified as a pion ((π,π); (π,K); (π,p); (π,X)) have to yield a clear peak around the K^* mass. The problem here is the abundant number of ρ^0(770 MeV) mesons that dilute the peak in the (π,π) case, making it impossible to identify the K^* peak.

As mentioned, there are two methods to determine the P matrix: a Monte Carlo simulation and the use of decaying particles. The advantage of the decaying particles is that no other information than the one recorded by the RICH in the experiment itself is used and there is no need to rely on a simulation or other assumptions. The disadvantages are twofold: first of all it is impossible to determine the complete P matrix with decaying particles (e.g. kaons from Φ decays tend to end up in the same detector half and very little statistics is available for the topology case with 1 track per detector half, which makes it impossible to determine the kaon column of the P matrix for this topology case). Secondly, the P matrix derived from the decaying particle analysis is not usable for HERMES physics analyses that mostly deal with the DIS process; the kinematics of DIS are different compared to the ones for decaying particles. For example, the opening angle between both kaons from the Φ decay has a restricted phase space and both kaons will have a specific topology in the RICH, while two kaons resulting from DIS will have another topology (both rings will usually overlap in a different way compared to the Φ decay process). As a consequence the P matrix from the decaying particles can only be used for the decay process and cannot be applied to DIS processes. No method exists to determine a P matrix for DIS experimentally. Another strategy has thus to be applied which is based on a Monte Carlo simulation: if the *experimental* P matrix from the decaying particles agrees with the one from a *Monte Carlo simulation* of decaying particles, confidence in the simulation is gained. In this case one can trust that the simulation is also capable of generating the P matrix for DIS. Therefore a reliable Monte Carlo simulation of the RICH had to be developed as will be described in the following sections. Results from the simulation with respect to the P matrix will be presented in section 4.8.5. The exact knowledge of the DIS P matrix is extremely important for the unfolding of hadron momentum distributions, as will be described in section 4.9.

4.8 Monte Carlo description of the RICH

The complete HERMES Monte Carlo (MC) is a suite of programs with the goal of simulating the complete HERMES experiment. Two main parts of the MC can be considered: the first part concerns all the physics processes and the event generation, the second part describes the response of the HERMES spectrometer to these events. This last part also involves the RICH detector and a main part of this Ph.D. work consisted in developing a reliable model for the RICH in the MC simulation. In this section we will explain how the geometrical description of the RICH detector was included in the Monte Carlo program, which effects were implemented to simulate the response of the RICH (and argumentations why these effects are required) and finally a comparison between experimental data and Monte Carlo results will be presented.

4.8.1 GMC, HMC and HRC

The most important parts of the software chain for the HERMES Monte Carlo are depicted in fig. 4.19. The generation of physics events is done by GMC (Generator Monte

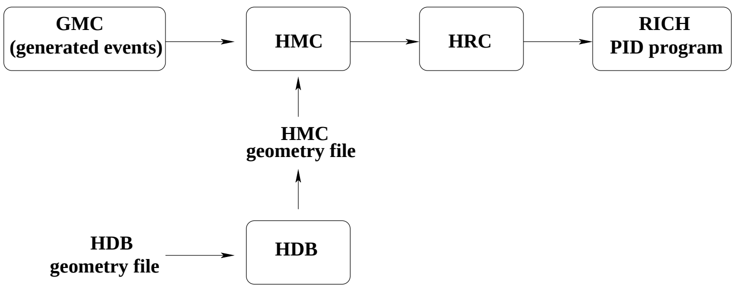

Figure 4.19: The flowchart for Monte Carlo data.

Carlo). Several versions of GMC exist. In this work the standard version (*GMC_dis*) was used in which the LEPTO event generator [93] first models the deep-inelastic scattering process in the electron nucleon interaction, and JETSET [29] (a package based on the LUND string model) then simulates the hadronisation process. The LUND model contains several parameters which are not (yet) perfectly tuned for an optimum description of the hadron multiplicity spectra measured at HERMES [31]. Since the only goal of the following studies was to simulate the response of the RICH detector, the physics input is of secondary importance and the fact that JETSET is not exactly fine-tuned to HERMES conditions is not relevant. The final result of GMC for each event is a kinematical description of all particles present in the final state. The output of GMC is handed to HMC (HERMES Monte Carlo) that simulates which particles are actually seen by the detector, tracks the particles through the spectrometer and mimics the responses of the individual detectors in the spectrometer. HMC is based on the GEANT package [94], a very comprehensive detector simulation package developed at CERN and frequently used in the field of high-energy physics. GEANT provides not only tools for the tracking of particles through matter and detectors but also simulates secondary interactions like Bremsstrahlung, multiple scattering, bending of the tracks in a magnetic field,... . GEANT also allows a graphical representation of the detector setup and of the particle trajectories. The final result of the HMC simulation of an event is a set of signals that the detectors would register if real particles from that event passed through them and is at an equivalent level (and format) to that of the experimental data. The output of HMC is dealt with in exactly the same manner as the experimental data: HRC (Hermes Reconstruction Code) reconstructs the charged particle trajectories from the information recorded by the tracking chambers and determines the momentum of each track. Finally, the RICH PID program uses the information on the reconstructed tracks to identify the hadrons in each reconstructed event.

4.8.2 The inclusion of the RICH in HDB

The main input for HMC is the HMC geometry file, that describes every detail of each detector in HERMES: the placement of its component pieces, the materials out of which it is constructed, etc. Also the target and the beam pipe are described in the HMC geometry file. This geometry file is managed by the HDB (HERMES DataBase) program. HDB reads in one input file, the “HDB geometry file”, and after performing some calculations writes out another file, the “HMC geometry file”. Both of these files, are ASCII files in ADAMO [95] format. ADAMO is a FORTRAN based database structure and provides a way of defining tabular data structures and manipulating them. The purpose of the HDB input file is to provide the user with an easy-to-use format for the creation and modificaton of geometric entities. HDB checks if no geometrical ambiguities are present in this file, executes some coordinate transformations when necessary and produces an output file which can be used by HMC. The production of the “HMC geometry file” has to be done only once and HDB is a stand-alone application which is independent of HMC.

Part of this work consisted of including the RICH detector in the HDB geometry file starting from the geometry file that contained the old threshold Čerenkov detector. The possibility to switch back to the old threshold Čerenkov counter had to be assured and thus the source code related to it was kept and this detector remained in the geometry

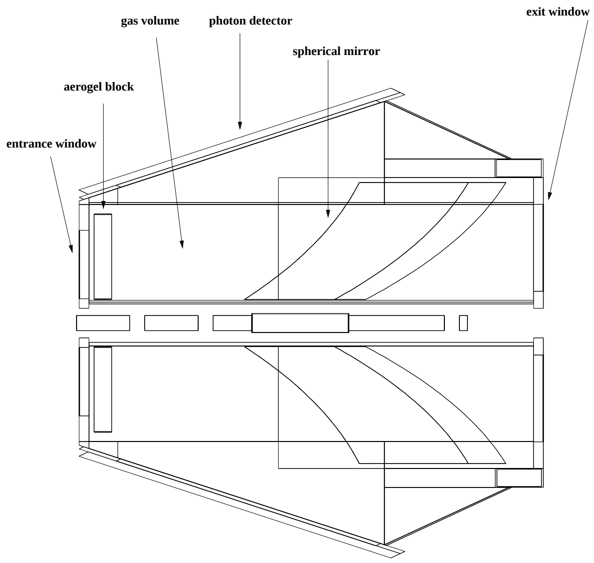

Figure 4.20: Side view of the (top and bottom) RICH in the geometry file of HMC. Top and bottom RICH are separated by the beam pipe. The spherical mirror is composed of different pieces and only a projection is visible here.

files. Switching back to the old Čerenkov is made possible by giving a timestamp when running HDB and letting HDB decide (by the timestamp) if there is either a RICH or a threshold Čerenkov counter.

The complete geometry can only be composed out of boxes, cones, (elliptical) tubes, spheres and wedges, since only these volumes are supported by HDB. The GEANT visualisation of the RICH is shown in fig. 4.20. The photon detector is implemented as one big sensitive plane without granularity, i.e. the individual photomultipliers are not modelled in the geometry file; a software routine will handle the digitisation, as will be explained later on. The aerogel consists of one massive wall, not of separate tiles. In a first version of the Monte Carlo an aerogel tile structure (with all 425 individual tiles) was implemented [76, 96, 97], resulting in a very large geometry file which decreased the speed of the Monte Carlo significantly without improving the results of the simulation. Although the aerogel is implemented as a massive wall, the tile structure is somehow accounted for, since the geometry file contains the opaque black sheets of tedlar between the aerogel stacks. The advantage of this setup is that it can be accomplished with a limited number of geometrical volumes, keeping the size of the geometry file manageable. An example of a track that radiates photons near an aerogel tile edge can be seen in fig. 4.21.
The lucite window behind the aerogel wall (which suppresses the Rayleigh scattered photons, see fig.4.7) was not put into the geometry, since GEANT tracks a photon that is trapped by internal reflection in the lucite forever (i.e. until the maximum number of steps in GEANT is reached), wasting CPU-power. The wavelength cut-off below 290 nm will be modelled in the software as will be explained later in this section.

Further details about the inclusion of the RICH in the HDB geometry file were summarised in the report [98].

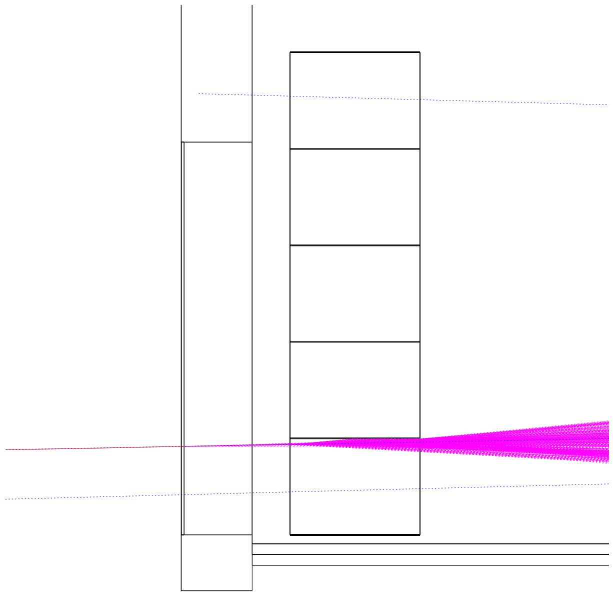

Figure 4.21: The absorption of photons by the black sheets of tedlar between the aerogel stacks. The figure is a side view of a detail around the aerogel region in the RICH, showing the entrance window and the aerogel wall. A charged particle track (shown in red) enters the aerogel wall and radiates Čerenkov photons. Photons hitting the black sheet of tedlar are absorbed by it.

4.8.3 Structure of the source code and effects included

After the geometry of the RICH was implemented in the HDB geometry file, all physics processes related to Čerenkov photons had to be programmed in the HMC source code. The tracking of photons and related effects (reflection, absorption, ...) can be handled by default GEANT routines, other effects (described below) had to be programmed and tested separately. The programming language of HMC is FORTRAN 77. The source code is completely documented [99] and flowcharts were made for all the individual routines. All the effects needed for a reliable description of the RICH were added step by step and intermediate checks were performed after each step, but here only the final version of the source code will be motivated (based on experimental data or physics arguments) and described briefly. Finally, some effects that were not included in the Monte Carlo description will also be discussed along with the reasons to omit them. For simplicity the structure of the source code will be divided into three parts: initialisation, tracking and digitisation.

4.8.3.1 Initialisation

The initialisation routine first decides (based on the time information in the geometry file) if the old threshold Čerenkov detector or the RICH has to be simulated by HMC. In case the RICH is active the initialisation routine performs the following tasks:

- First some geometrical data is extracted from the geometry file (the position of the center of the aerogel wall, the size of the aerogel wall, the position of the photon matrix, ...) since this information is needed by the digitisation routine, described in section 4.8.3.3. Also a geometry file is written out, which contains information about the RICH needed by the PID program (e.g. the mirror center and radius are required for the IRT reconstruction of the gas and aerogel angles).

- The gas and the aerogel are present in the geometry file as "dead materials" and have to be set up as Čerenkov radiators by a GEANT subroutine. This GEANT subroutine requires some parameters; the refractive index of the radiator and its absorption length are the two most important ones. Both parameters are typically wavelength dependent and the input to GEANT is based on theoretical models.

  Aerogel consists of clusters of silica particles arranged in a highly porous, three-dimensional netlike structure. The optical dispersion in aerogel can be modelled approximately [100] by the dispersion relation of fused silica and air:

$$n_{aerogel}(E_\gamma).V(total) = n_{SiO_2}(E_\gamma).V(SiO_2) + n_{air}(E_\gamma).[V(total) - V(SiO_2)] \quad (4.38)$$

  where E_γ is the photon energy and V is the volume. Assuming that $n_{air} = 1$, this equation can be written as

$$n_{aerogel}(E_\gamma) = n_{SiO_2}(E_\gamma).(1 - B) + B \quad (4.39)$$

  where $B = 1 - V(SiO_2)/V(total)$. The parameter B was determined from experimental data as follows: by taking high energy electrons ($\beta = 1$) and measuring the average reconstructed aerogel angle, formula (4.3) yielded a value of 1.03052 for $n_{aerogel}$. The wavelength for this index of refraction was determined from a simple stand-alone Monte Carlo program that simulated all the effects shown in fig. 4.7. The wavelength distribution of the photons that survived the efficiencies from that figure was determined and the mean photon energy value was $\langle E_\gamma \rangle = 2.925$ eV. The value for the index of refraction of fused silica at that photon energy is [101, 102] 1.4677, yielding a value for $B = 0.93474$ (this value also learns that aerogel consist for 93.5% of air). This determines in principle the dispersion relation (4.39). However the values for the dispersion relation of fused silica [101, 102] are only available in a tabular form, no functional photon energy dependence is available. The tabulated values could be used to calculate the values of $n_{aerogel}$ which were fitted by a function, yielding [103]

$$n_{aerogel}(E_\gamma) = 1.02821 + 0.789308 * 10^{-3} * exp(0.368298 * E_\gamma) \quad (4.40)$$

  The refractive index of the C_4F_{10} gas can be described by [104]

$$n_{C_4F_{10}}(E_\gamma) = 1.0013178 + (1.723 * 10^{-7})(E_\gamma) + (5.495 * 10^{-6})(E_\gamma^2) \quad (4.41)$$

  where E_γ is expressed in eV. The constant terms of these dispersion relations will strongly influence the mean value of the reconstructed aerogel and gas angle. As will be explained in section 4.8.4 these terms were tuned in order to reach agreement with the experimental observed values for the angles. For further reference the constant term for the gas will be called A_{gas} and the constant term for the aerogel will be called A_{aer}.

It is not necessary to consider wavelength dependent absorption in the C_4F_{10} gas. One of the reasons that these fluorocarbons are good radiators is that the light transmission is very high and the dispersion very small. The effect of attenuation in C_4F_{10} is undetectable in the HERMES RICH since the absorption edge is at about 150 nm and by 220 nm the attenuation is not measurable. For the aerogel the absorption length as a funtion of wavelength is given by the Hunt formula (4.7).

The gas and aerogel refractive indices and the aerogel absorption length must be given to the GEANT subroutine in a number of bins where each bin represents a certain value of the photon energy. The upper and lower limits of the range for the binning in photon energy can immediately be derived from fig. 4.7: for the aerogel the photon energy is 1.77 eV $< E_\gamma <$ 4.43 eV (or 280 nm $< \lambda_\gamma <$ 700 nm). For the gas the photon energy is 1.77 eV $< E_\gamma <$ 7.3 eV (or 170 nm $< \lambda_\gamma <$ 700 nm). Beyond these limits no photons are detected so it is useless to generate them. The number of bins in photon energy was determined by a small stand-alone program [99]. The *minimum* number of bins (the HMC CPU-time increases significantly with the number of bins) that reproduced the functional shapes of the dispersion relations and the aerogel absorption length was determined, resulting in 3 bins in photon energy for the gas and 10 bins for the aerogel.

- Optical properties have to be assigned to some other volumes in the geometry file, which is accomplished by a GEANT routine. The mirror is made 100% reflecting, the photon detector plane is set up as a perfectly efficient detector, the black sheets of tedlar in the aerogel wall are made fully absorbing, etc.

4.8.3.2 Tracking

After the initialisation phase is executed, HMC can process the events that were generated by GMC. Each generated particle is tracked through the whole HMC simulation of the HERMES spectrometer. If a charged particle enters the aerogel or the gas, GEANT starts generating Čerenkov photons according to eq. (4.1) and (4.2). Once generated these Čerenkov photons are tracked one by one through the RICH until the photon either "disappears" or is detected on the photon detector plane. GEANT tracks each photon in *steps*. The length of each step is calculated by GEANT and basically corresponds to 'how far the photon travels before something happens to it'. This "something" can be either an in-flight absorption (as already described above for the aerogel) or a "boundary action", when the photon crosses a boundary between two different volumes. When a photon arrives at the boundary of the dielectric medium in which it travels, its behaviour depends on the nature of the two materials which join at that boundary and three cases are distinguished in GEANT:

1. Case dielectric $\rightarrow$ dielectric: the photon can be transmitted (refracted ray) or reflected. If the photon can only be reflected, total internal reflection takes place.
   This transition dielectric $\rightarrow$ dielectric occurs e.g. when an aerogel photon is refracted at the aerogel/gas boundary.

2. Case dielectric $\rightarrow$ metal: the photon can be absorbed by the metal or totally reflected back into the dielectric. If the photon is absorbed it can be detected according to the detection efficiency of the metal.
   This case occurs e.g. when a photon is reflected on the mirror or if it is detected on the photon detector plane.

3. Case dielectric $\rightarrow$ black material: a black material is a tracking medium for which the user has not defined any optical property. In this case the photon is immediately absorbed undetected.
   An example is the photon absorption by the aluminium frame of the RICH for which no optical properties are defined.

GEANT accounts for the roughness of a surface (the quality of the surface finish between two media) by choosing an angle α which is defined as the angle between two normals: that to the effective reflection plane and that to the nominal one (i.e. the boundary between the two media). The effective normal is built by randomly choosing a vector of length $(1 - REFL)$ and adding it to the nominal normal. The limits of α are thus:

$$\sin\alpha = \pm(1 - REFL) \tag{4.42}$$

where $REFL$ may vary from 0 to 1. It is given as input by the user. A value of 1 means perfect smoothness while 0 stands for maximum roughness. In between the surface is modelled by a bell-shaped distribution in α.

For the simulation of the RICH the roughness of a surface was only used to describe the mirror surface. The value of $REFL$ will be called MR (*mirror roughness*) in this case and it was derived from a tuning procedure, as will be described in section 4.8.4.

4.8.3.3 Digitisation

The term digitisation refers to the Monte Carlo simulation of the response of the RICH photon detector plane. The digitisation includes the simulation of the reponse of the actual instrument, including efficiencies, instrument limitations, etc. The ultimate purpose of the digitisation is to fill the so-called dataRICH table that contains all the (integer) identification numbers of fired PMTs in the event. The digitisation routine does not use any GEANT subroutines but is completely home-written. It performs the following tasks:

- A comparison of Monte Carlo results and experimental data (section 4.8.5) learns that the number of produced photons from GEANT is systematically too high, and hence the number of fired PMTs will be too high in the Monte Carlo. Therefore an artificial transparency factor is applied to the photons. This factor was tuned to reproduce the experimental yields, as will be described in section 4.8.4. Before anything else concerning digitisation is done, these (wavelength independent) transparency factors are applied, separately for gas and aerogel photons. For further reference the transparency for the aerogel will be denoted as T_{aero} and the transparency for the gas as T_{gas}.

- As already mentioned, the photon detector is implemented in the geometry file as one big sensitive plane, without granularity; the individual photomultiplier tubes are not modelled in the geometry file. A software routine translates a hit (i.e. a photon that hits the Monte Carlo sensitive plane) to a fired PMT. This routine needs to know the geometrical center of the real photon detector plane (which is obtained from the geometry file in the initialisation routine), the size of the photomultipliers (taking into account the funnel glued in front of each PMT), and the honeycomb closed packing of the PMTs. For each photon it is checked whether it hits a certain PMT.

- For photons that have hit a PMT, all the survival probabilities described in section 4.4.5 and depicted in fig. 4.7 are applied. Parametrisations based on theoretical models have been used to describe the functional shapes of fig. 4.7 and can be found in [99]. For the photon collection efficiency of the PMT cones a more complicated model has been used, taking into account the angle of incidence of the photon on the funnel and the number of reflections that a photon undergoes in the funnel [105, 106].
- For photons that have hit a PMT and that survived all the photon detection probabilities, the PMT number is written to the dataRICH table. If several photons hit the same PMT, the PMT number is listed only once in that table since the PMTs in the 'experimental' RICH only record their information digitally (0 or 1).

4.8.3.4 Effects omitted in the Monte Carlo

Some effects are at present not simulated in the Monte Carlo because they are either too complicated to incorporate or not needed. The RICH is from the optical point of view a complicated device and modelling all optical effects is not feasible. Although some processes are not implemented, it will be shown in section 4.8.5 that the agreement between Monte Carlo and experimental data is satisfactory. The most important omitted effects are described below and it is pointed out why they were not included in the Monte Carlo simulation.

- As described in section 4.8.3.1 the Rayleigh scattering is realised as a wavelength dependent transmission rather than a real scattering process. Only the photons which are not lost by scattering are kept in the simulation and the scattered ones are simply considered as "lost", whereas in reality they cause a change of the Čerenkov angle: the scattering angle δ of a photon (with respect to the direction of the incoming photon) is distributed according to a $(1+cos^2\delta)$-law, thus the Rayleigh scattering is nearly isotropic. GEANT is not able to model the Rayleigh scattering for aerogel adequately and therefore is was decided to neglect these scattered photons entirely, as their influence on the particle identification is small: for the IRT algorithm the only Čerenkov angles which are considered to calculate the average Čerenkov angle $\langle\theta\rangle$ from eq. (4.8) are the ones within the $\pm\sigma_w$ angle window. Rayleigh scattered photons are mostly scattered outside this window and hence do not make any difference for the particle identification with the IRT algorithm. For the reconstruction of Monte Carlo data, the number of background PMTs (B_{PMT}) introduced in section 4.5.1.3 is set to zero in the IRT program. For the DRT method the Rayleigh scattered photons will not cause a hitpattern that is simular to a hitpattern from an actual hadron and hence the likelihood from eq. (4.20) will not be influenced. Of course, Rayleigh scattered photons cause a background which will not be reproduced by the Monte Carlo.
- The mirror consist in reality of eight different pieces which are not perfectly aligned but in the Monte Carlo it is implemented as one massive piece. The alignment procedure, described in section 4.5.1.2, also revealed geometrical differences between the top and bottom detector, e.g. the radius of curvature of both mirrors and the relative position of the RICH in the complete spectrometer. These imperfections are not implemented in the geometry file, since the IRT algorithm is equipped with an alignment correction routine and the "blobbing" technique for DRT eliminates the influence of improper alignment.

Differences in mirror reflectivity between the eight mirror pieces are not considered in the Monte Carlo. Attempts have been made to determine the relative piecewise mirror reflectivities [107, 108] from experimental data, using electron gas rings and pion aerogel rings. These studies revealed that the obtained values for the mirror reflectivities are dependent on how the mirror segments are sampled and that differences in reflectivity within one segment are of the same order of magnitude as between two mirror pieces. It is therefore questionable if a single numerical value per mirror segment is sufficient for a realistic model of the data. Hence it was decided that in a first approximation the model for the global mirror efficiency from fig. 4.7 is sufficient. Since the Monte Carlo and experimental data show a satisfactory agreement without a more sophisticated model for the piecewise mirror reflectivity, it was decided to neglect this effect.

- If one plots the experimental dependence of the aerogel Čerenkov angle for e.g. pions versus their momentum, a figure simular to fig. 4.1 is obtained. A fit to this figure, using the theoretically expected formula (4.1) with the refractive index as a fit parameter, allows to determine the aerogel refractive index, yielding a value of 1.03070 $\pm$ 0.00002 and 1.03040 $\pm$ 0.00002 for top and bottom detector [109] respectively. These differences between the aerogel refractive indices for top and bottom were not programmed in the Monte Carlo where top and bottom half are identical detectors. The RICH particle identification program compensates for this in the case of experimental data: for the IRT algorithm the theoretical angles θ^i_{theo} $(i = \pi, K, p)$ from eq. (4.8) were calculated with the appropriate index of refraction, depending on the track being in the top or bottom detector. The correction for the refraction at the aerogel/gas boundary is also calculated with the appropriate index of refraction. In general small variations in the aerogel index of refraction have a negligible influence on the PID performance of the IRT algorithm, as was shown by a detailed study [110]. For the DRT method the hitpatterns C_{PMT} for each track are generated with the proper index of refraction, as well as the simulation of the refraction at the aerogel/gas boundary. The above describes the reconstruction of experimental data. In the case of MC data the PID program utilises identical settings for the top and bottom RICH.

- As already mentioned, the aerogel wall is implemented without any tile structure, apart from the inclusion of the opaque sheets of black tedlar between the aerogel stacks in the geometry file. As will be illustrated in section 4.8.5 these sheets only absorb photons and thus reduce the number of fired PMTs for tracks passing close to the sheets but do not change the Čerenkov angle. Nevertheless, the tile structure is clearly visible in experimental data and especially the edges of the aerogel tiles seem to affect the behaviour of the aerogel Čerenkov angle. The influence of the tile edges on the reconstructed and average Čerenkov angles was studied [111] for electrons (using IRT) where two methods were applied. The first method examines the global properties of all tiles together while the second method focuses on single edges. The former method has to its advantage that a small data sample will yield enough statistics, although the edge effects may be diluted if the position in space of the aerogel wall is not exactly known or if various tile stacks have different dimensions. These two uncertainties can cause errors in the calculation of the tile edge positions. The second method only looks at one specific tile edge. The choice of a tile edge in the middle of the acceptance maximises the statistics.

  A result from the first method to study the overall tile edge effect for the bottom RICH

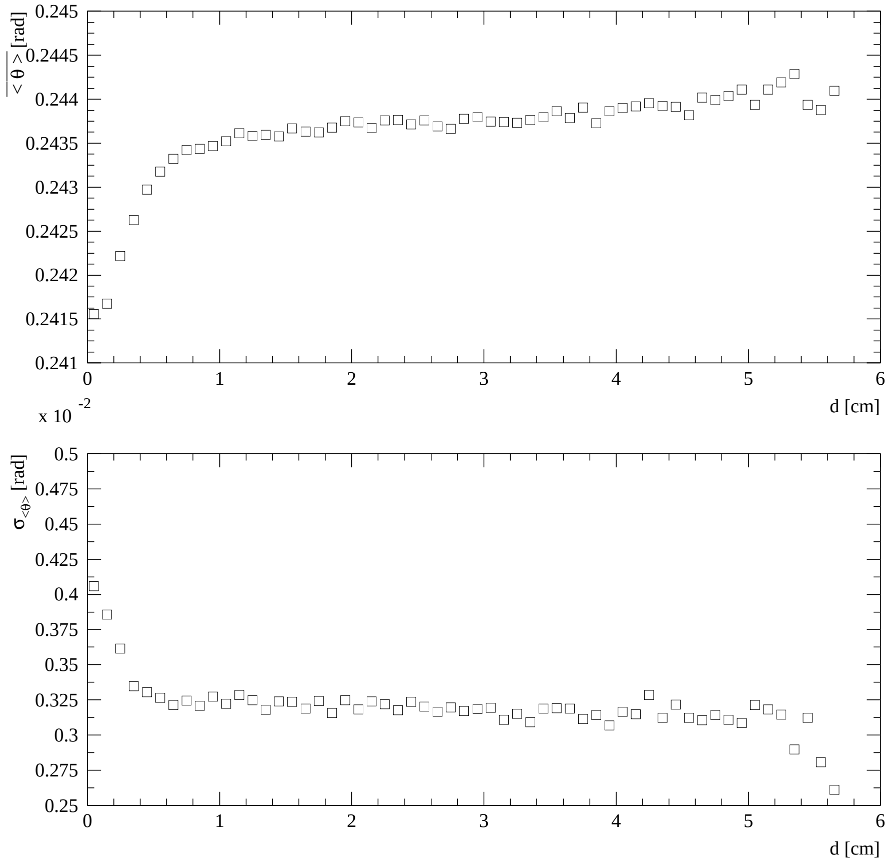

Figure 4.22: Mean value and width of the Gaussian distribution of $\langle\theta\rangle$ versus d (the distance from the electron track to the nearest tile edge). The zero on the vertical scale is suppressed.

is shown in fig. 4.22. On the horizontal axis, the quantity d is plotted which is the distance from an electron track to the *nearest* tile edge. Since the tiles have vertical dimensions of 11.4 by 11.4 cm the maximum value for d is 5.7 cm. The mean aerogel angle $\langle\theta\rangle$ for all the electrons in the data sample was then plotted for each bin in d and this distribution was subsequently fitted with a Gaussian and a constant background term. The mean value ($\overline{\langle\theta\rangle}$) and the width ($\sigma_{\langle\theta\rangle}$) of this Gaussian distribution are shown on the vertical axis. It is clear that near the edges of the aerogel tiles the mean value of the Čerenkov angle shows a fall-off, while the spread of its distribution increases. From fig. 4.22 it is hard to pin down the origin of this phenomenon. A possible reason could be the shape of the aerogel tiles as shown in fig. 4.6: photons passing through the edges of the aerogel tiles see a wedge-like structure and undergo reflection or refraction. To investigate the behaviour of the aerogel angle near the edges three studies were performed: the wedge-like structure was implemented (in a simplified model) in the geometry file, a small stand-alone Fortran program was written to simulate the refraction and reflection of photons at the tile edges and an "effective edge function" was included in the HMC source code to mimic the behaviour of the photons near the edges. None of these studies could explain or reproduce the behaviour observed in fig. 4.22, although a clear deviation in the Čerenkov angle was noticed but without the appropriate shape. Another additional reason for this conduct had to be sought and a good candidate is the thickness variation near the tile edges: thickness variations imply variations in the density of the aerogel (ρ) and these density fluctuations are related to a change in the index of refraction:

$$\frac{dn}{n-1} = \frac{d\rho}{\rho} \tag{4.43}$$

Variations in refractive index due to local density fluctuations and variations in path length due to tile surface features can deflect the trajectories of photons away from the Čerenkov cone and the deflection is approximately [79]

$$d\vec{\theta} = \vec{z} \times \nabla((n(x,y)-1) \cdot t(x,y)) \tag{4.44}$$

where x and y are the lateral dimensions along the tile, $\vec{z}$ is the normal to the surface and $t(x,y)$ is the tile thickness. The light deflection near the tile edges was measured for a single aerogel tile using a laser setup [79], yielding a deflection corresponding to the functional shape of fig. 4.22 and with a deflection of ≈ 7 mrad at 0.2 cm from the tile edge. This value is rather large compared to fig. 4.22 but the laser measurement is not really representative for what is happening with Čerenkov light passing through the aerogel installed in the RICH:

- The laser measurement was performed only for perpendicular incidence on the aerogel tile, while in the RICH the distribution of the angle of incidence is not known.
- The stacks of aerogel tiles in the RICH consist of 5 tiles each with different geometrical properties and different density fluctuations. Moreover, the sides of the aerogel stacks are covered with black tedlar sheets. The laser measurement was performed for a single naked tile.
- The laser measurement employed a single wavelength (633 nm HeNe laser), while the emission of Čerenkov photons in the RICH is governed by the Frank-Tamm relation (4.2) for aerogel. The dispersion is also wavelength dependent.

The laser measurement thus only gives a rough indication about the tile edge effects occurring in the RICH and therefore it was decided to perform more detailed studies, starting from experimental data only and considering all stacks of aerogel tiles separately instead of looking at all tiles together.

First of all, a method was developed to identify stacks of tiles, using experimental RICH data only. In fig. 4.23 one single tile is shown for the bottom detector. This plot was made (using electrons) in the following way: Let *xpos* and *ypos* respectively be the x and y position of the electron track at the z-middle of the aerogel wall (where (x,y,z) are now the coordinates in the HERMES reference frame), then a histogram *ypos* versus *xpos* can be created with the ranges as seen on the plot. The histogram is filled for each ($xpos, ypos$)-pair belonging to an electron track in the data sample with a weight equal to the number of PMTs for the electron track (i.e. the number of PMTs in the $\pm\sigma_w$ window around the theoretical aerogel electron angle as explained in section 4.5.1.3). As can be seen from the colourscale on the plot, the tedlar sheets reduce the number of

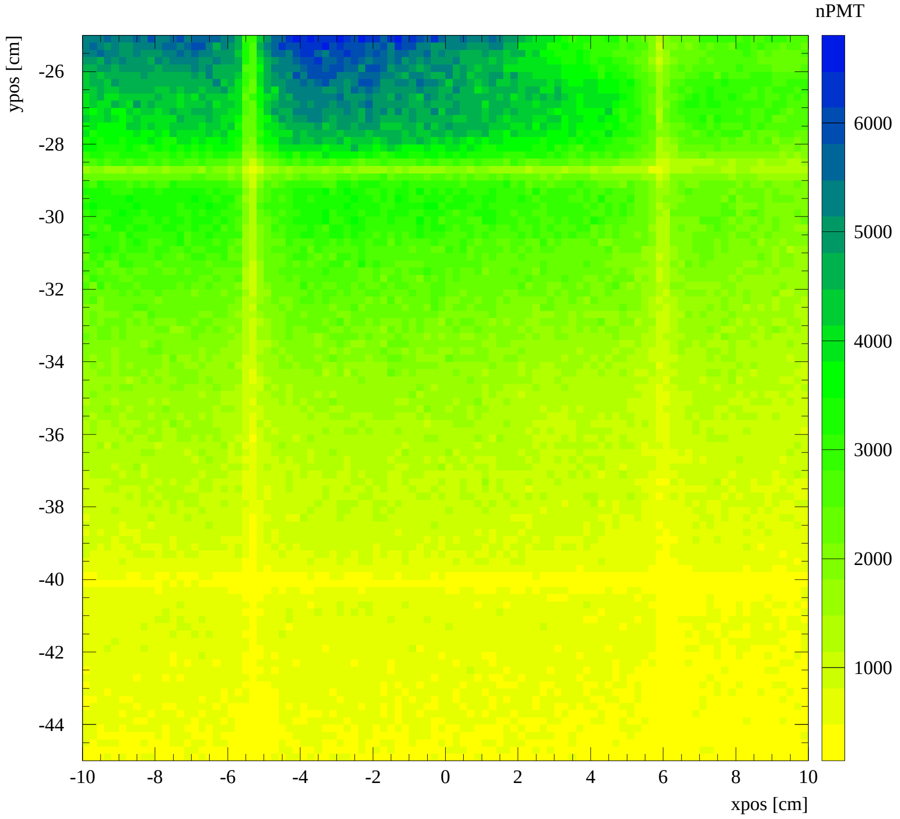

Figure 4.23: One single tile in the bottom detector. The colourscale indicates the number of fired PMTs. The gradient in occupancy is due to the geometrical acceptance for electron tracks.

PMTs and make it possible to determine the location of the stack borders. Each border can be studied and an example will be given for the upper horizontal border of fig. 4.23. Now a new quantity n' is introduced:

$$n' = \frac{1}{\beta.cos(\theta)} = \frac{1}{\sqrt{(1/(1+(m_e/p)^2))}.cos(\theta)} \tag{4.45}$$

In theory n' should be a constant equal to the index of refraction n_{aero}. But the Čerenkov angle θ is influenced by all the properties of the tiles, thus n' is rather an *effective* index of refraction, reflecting both optical and geometrical properties of the tile. The position dependence of n' can be studied by making use of the already obtained information about the position of the tile borders. This is illustrated in fig. 4.24 for the upper horizontal border of fig. 4.23. The drop of n' near the edge of the tile is clearly visible.

This procedure was repeated for all the tile stacks in the aerogel wall [111, 112] and for each stack the mean effective refractive index was fitted with a function

$$< n' > (r) = A + B \cdot \frac{t(l=0)}{t(l)} \tag{4.46}$$

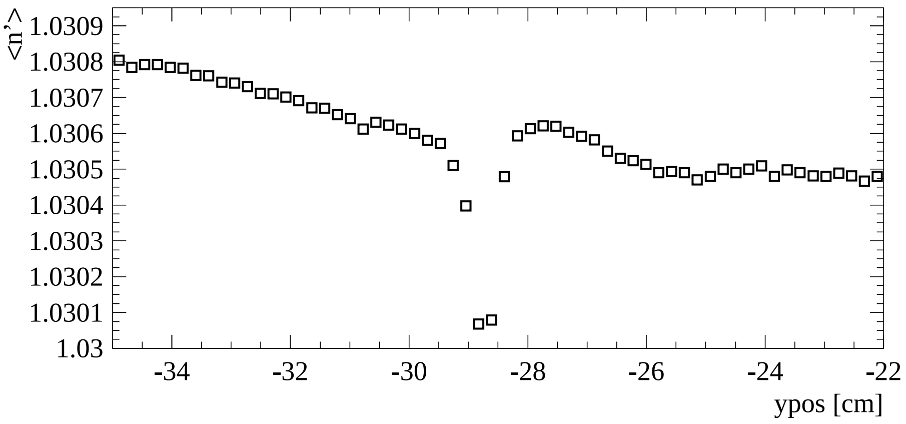

Figure 4.24: Mean value of n' versus *ypos* for the upper horizontal border between two tile stacks as illustrated in fig. 4.23. This border has been isolated with the restriction -4.0 cm $< xpos <$ 4.0 cm.

where l is the distance between the particle track and the center of the tile and where $t(l)$ describes the geometrical shape of the tile, according to fig. 4.6. Both *xpos* and *ypos* projections were considered. To examine all stacks, pions were used instead of electrons since they provide a better coverage of the acceptance. Once the parametrisation for all tile stacks was calculated, in principle two things could be done with it: after implementing it into the Monte Carlo, one could check if the behaviour in fig. 4.22 is reproduced or not. The drawback of this solution would be the overhead in the number of calculations for each particle track. A more elegant solution is the incorporation of the parametrisation into the RICH PID program that reconstructs the *experimental* data. Using this parametrisation and the knowledge of each particle track position, the theoretical Čerenkov angles θ^{π}_{theo}, θ^{K}_{theo} and θ^{p}_{theo} for the IRT method can be calculated with the appropriate index of refraction as well as the refraction correction for the IRT method. A detailed study of the PID efficiency [112] learned that the parametrisation has a negligible effect on the performance of the IRT algorithm, mainly since the $\pm\sigma_w$ window around the theoretical expected angle is broad enough to include also Čerenkov angles that undergo deflection at the tile edges. Hence it was decided that the parametrisation of n' was not needed, neither in the Monte Carlo, nor in the RICH PID program for experimental data. Nevertheless, this study was important because there was some concern that the tile edges could bias the reconstruction of the Čerenkov angle and have an impact on the PID.

To conclude this discussion about the tile structure of the aerogel, it can be stated that the inclusion of the opaque black spacers between the aerogel stacks in the geometry file is sufficient to mimic the tile structure in the Monte Carlo: these spacers reduce the number of firing PMTs for tracks that pass nearby tile edges and this reduction has a sizeable influence on the particle identification. Edge effects as a consequence of geometrical properties and density fluctuations of the tile have an impact on the aerogel Čerenkov angle but play a negligible role for the PID and hence can be omitted in the Monte Carlo description.

4.8.4 Tuning the Monte Carlo

After updating the RICH geometry in the HERMES Monte Carlo and after introducing several processes into the source code to improve the description of the RICH detector (and studying which effects could be omitted) the final task still to be done, was tuning the different parameters that describe various effects. These effects are: the dispersion relations for the gas and the aerogel, the mirror roughness and the transparency for gas and aerogel. The adjustable parameters related to these effects are:

- The constant terms A_{aer} and A_{gas} of the aerogel and the gas dispersion relations (cfr. eq. (4.40) and eq. (4.41)).
- The value for the mirror roughness MR (cfr. eq. (4.42)).
- The artificial wavelength independent transparency factors for the aerogel and the gas (T_{aer} and T_{gas}), as described in section 4.8.3.3.

The final goal of the RICH Monte Carlo description is to reach an accurate agreement with the experimental data. Therefore, the experimental quantities used in the tuning procedure (and how well they can be determined) must be defined first. The starting point for tuning the Monte Carlo will be electrons: Monte Carlo results will be compared with experimental data for

- high energetic electrons with $p > 5.0$ GeV (in order to get rid of background),
- which are the only track in the detector half (in order to obtain a clean sample)
- and which have rings that are completely on the mirror and on the PMT matrix (to avoid acceptance effects).
- For the aerogel (both for experimental data and Monte Carlo), only tracks which are sufficiently far away from the aerogel tile edges will be considered, to bypass the (small) tile edge effects in the tuning procedure.

For these electrons the *reconstructed* angles and the *average* angles will be studied for the IRT method only. The reconstructed angle is obtained if the IRT reconstructs one Čerenkov angle for each PMT hit/track combination per radiator. For the reconstructed angles, a "big window" around the theoretically expected value for the Čerenkov angle will be used:

$$0.0 < \text{reconstructed gas angle} < 0.1 \tag{4.47}$$

$$0.2 < \text{reconstructed aerogel angle} < 0.3 \tag{4.48}$$

The average angle is the average of all the reconstructed angles in a $\pm\sigma_w$ window around the theoretical electron angle as explained in section 4.5.1.3.
For simplicity, we will refer to the distribution of the reconstructed angle in the big window as *angles in the big window* and we will call the distributions of the average angle the *angles in the small window* (both for aerogel and gas).

For the number of PMTs ($nPMT$), the same windows will be considered. The experimental distributions for the angles and $nPMT$ are presented as the shaded histograms in fig. 4.28; their mean value and spreads will be the basic quantities used in the tuning of the aforementioned free parameters.

The necessity of the small window is obvious from the discussion in section 4.5.1.3, the big window is here considered in addition to take into account the overlap between rings which is important for kaons from Φ decays.

4.8.4.1 Tuning the dispersion relations for gas and aerogel

Tuning the constant terms of the dispersion relations (A_{gas} and A_{aer}) is rather simple: the only quantity that should determine these values is the mean value of the histogram that contains the "angles in the small window". In this subsection, we will only consider this quantity and for brevity refer to it as *mean angle*.

The values of A_{gas} and A_{aer} were varied in the Monte Carlo within the following ranges:

$$1.00126 < A_{gas} < 1.00134 \tag{4.49}$$

$$1.0275 < A_{aer} < 1.0285 \tag{4.50}$$

and it was examined how the mean angle changed correspondingly. For this scan the mirror roughness was set to 1.0 (i.e. there is no mirror roughness) and the transparencies for the gas and aerogel were set to 1.0 (i.e. there is no artificial attenuation). The result of the scan can be seen in fig. 4.25. On the horizontal axis the constant term of the dispersion relation is shown (but for clarity only the significant digits of A_{gas} and A_{aer} are shown: the value plotted on the horizontal axis for the aerogel is $(A_{aer} - 1.0) \times 10^3$ and for the gas it is $(A_{gas} - 1.001) \times 10^5$). On the vertical axis, the *mean angle* is plotted. The (red) horizontal line is the experimental value of the *mean angle* (the error could be neglected), while the (green) vertical lines are the theoretical values for A_{gas} or A_{aer} that were mentioned in eq. (4.40) and eq. (4.41).

The points on the plot were fitted with a first order polynomial and the intersection with the horizontal line was calculated, yielding the following values for the constant terms in the dispersion relations:

$$A_{aer} = 1.02804 \tag{4.51}$$

$$A_{gas} = 1.001287 \tag{4.52}$$

To conclude we mention that both for aerogel and gas:

- Within the range of the scan for A_{aer} and A_{gas} no differences for the number of PMTs was seen (both for the mean value and the width of the $nPMT$ distribution), neither in the big nor the small window.

- Also the spread on the angles (both for the big and small window) does not depend on the constants of the dispersion relation.

- The mean value of the angle in the big window shows the same behaviour as a function of A_{aer} and A_{gas} as the mean value of the angle in the small window.

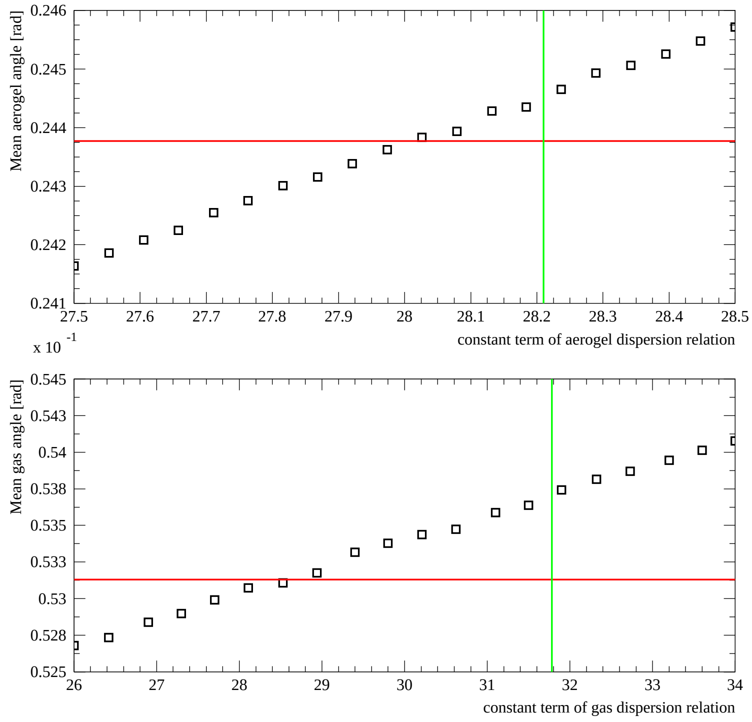

Figure 4.25: Tuning the constant terms of the aerogel (top panel) and gas (bottom panel) dispersion relations. See text for a description.

4.8.4.2 Tuning the mirror roughness and the number of PMTs

Both the mirror roughness (MR) and the artificial transparencies for the gas and the aerogel (T_{gas} and T_{aer}) influence the number of PMTs and the spread on the angles as will be shown in the following figures. It must be stressed that optimising all distributions of the angles and nPMT in the big and small windows simultaneously is far from simple. Therefore, the aim of the tuning effort will be first translated to the point of view of hadron identification.

Since the RICH was mainly built to identify kaons the Monte Carlo needs to reproduce the kaon efficiencies (i.e. the probability that a real kaon is identified by the RICH as a kaon) as well as possible. The majority of the kaons in HERMES have momenta below 10 GeV and are thus mainly identified by the aerogel, as can be seen in fig. 4.1. Therefore the aerogel is considered as the most important radiator. The tuning of the Monte Carlo will be done in such a way that first the aerogel reproduces the experimental data and thereafter we will do our best to have the gas angles and $nPMT$ in agreement with the experimental results.

First the parameters MR, T_{gas} and T_{aer} were varied within the following ranges (more extreme values could be excluded a priori from fast checks):

$$0.980 < MR < 0.995 \tag{4.53}$$

$$0.3 < T_{aer} < 1.0 \tag{4.54}$$

$$0.2 < T_{gas} < 0.8 \tag{4.55}$$

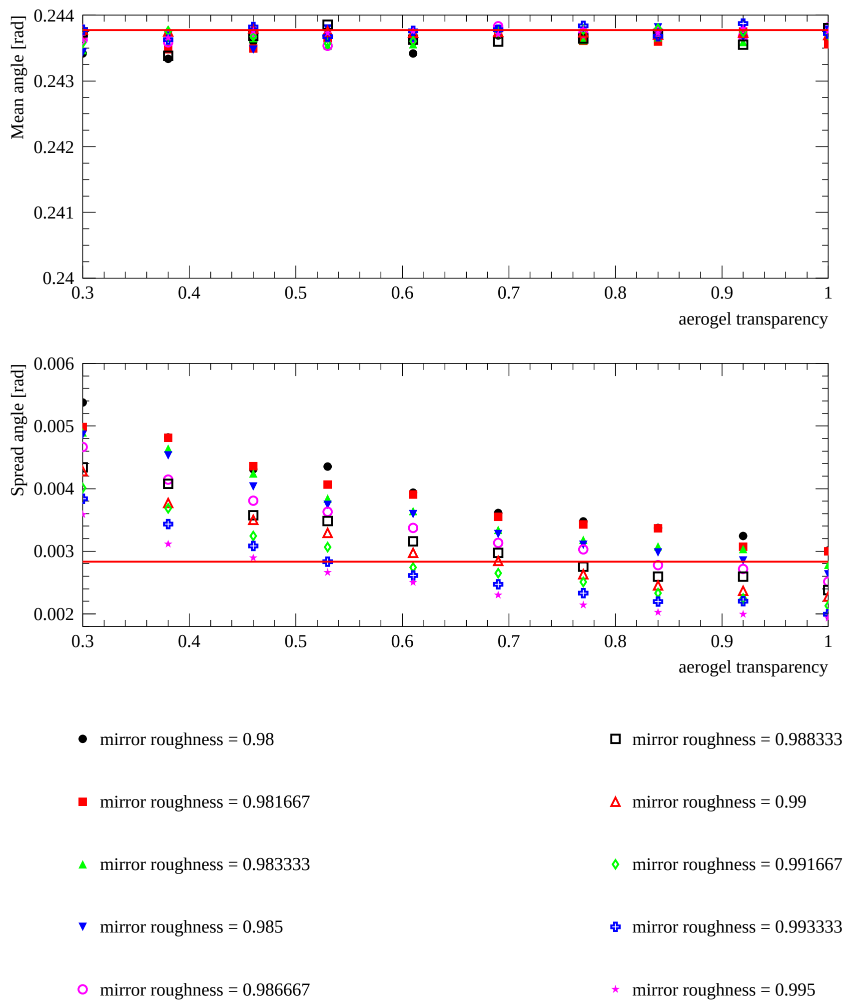

Figure 4.26: The mean value and the spread of the distribution of the aerogel angle in the small window: behaviour as a function of MR and T_{aer}. The horizontal lines represent the experimental values.

The results of that scan for gas and aerogel angles and $nPMT$ (both in the small and big window) were studied in detail and the comparison with experimental results was the basis for obtaining the final values for MR, T_{gas} and T_{aer}. Presenting all the plots and describing the entire procedure for the tuning is not feasible within this thesis; all the relevant information was summarised in a RICH internal note [113]. Here only the principle of the tuning method will be described and one example will be illustrated: the aerogel angle in the small window. The average value of the aerogel angle in the small window is already determined from the dispersion relation and does not vary significantly as a function of transparency or mirror roughness as can be seen in fig. 4.26. So only the spread on the average aerogel angle is affected by T_{aer} and MR. fig. 4.26 shows that different combinations (T_{aer},MR) are able to reproduce the experimental spread on the aerogel angle (the (red) horizontal line on the plot). First all possible combinations (T_{aer},MR) that match with the experimental data were determined in the following way: for each fixed value of the MR (cfr. legenda of fig. 4.26), a quadratic function was fitted

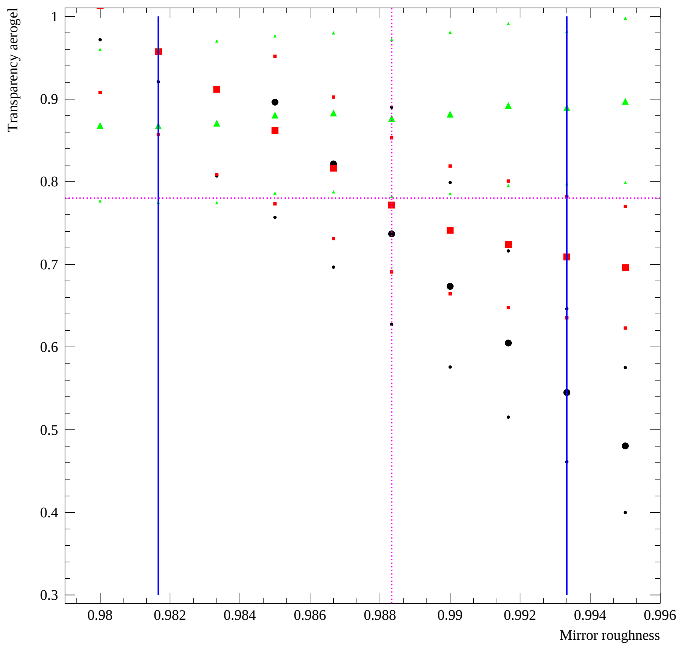

Figure 4.27: Tuning the MR and T_{aer}. Black dots: criterium 1 (spread on aerogel angle in small window). Red squares: criterium 2 (mean aerogel nPMT in small window). Green triangles: criterium 3 (mean aerogel nPMT in big window). For criterium 1, 2 and 3 the big symbols correspond with the experimental value, the small symbols represent the error bars if a ± 10 % error on the experimental value is allowed. Full blue lines: criterium 4 (spread on aerogel angle in big window). Magenta dashed lines: final choice of MR and T_{aer}. See text for more explanation.

to the Monte Carlo points and the intersection of this funtion with the horizontal line was calculated. These *matching points* (T_{aer},MR) are represented in fig. 4.27 as the big (black) dots. To have a feeling for the error on the big black dots, this procedure was repeated twice, where a ± 10% error on the measured value was allowed, resulting in the small (black) dots in fig. 4.27. If some points are not on the plot it means that no intersection could be found. The described procedure about the spread on the aerogel angle in the small window will be called *criterium 1*. *Criterium 2* for the aerogel is the mean number of PMTs in the small window. The same method as for *criterium 1* was repeated and the result (and the ±10 % error) is represented as the big (small) (red) squares in fig. 4.27. *Criterium 3* is the mean number of aerogel PMTs in the big window and *criterium 4* is the spread on the reconstructed aerogel angle in the big window. This last criterium is treated somewhat differently than the previous three, since the spread of the experimental distribution of the aerogel angle in the big window is strongly dependent on the background. Therefore it was decided to be less stringent for criterium 4 and the only requirement is that the MR lies somewhere between the two vertical (blue) full lines in fig. 4.27 [113].

For the C_4F_{10} gas exactly the same procedure was repeated. As the two radiators cannot be tuned independently (both T_{aer} and T_{gas} are strongly correlated with MR), an optimum was determined for the four gas and four aerogel criteria together, where the priority was given to the aerogel and where the big window was considered less important than the small window which is more crucial for hadron PID.

The final values of the tune (represented by the dashed (magenta) lines in fig. 4.27 are:

- mirror roughness $MR = 0.9883$
- gas transparency $T_{gas} = 0.45$
- aerogel transparency $T_{aer} = 0.78$

The final value for the mirror roughness is not really physical: experimental measurements of the mirror roughness yielded a value of approximately 0.994-0.995, which corresponds to a 2-3 cm aberration of the rings seen in the RICH. The value resulting from the tuning procedure corresponds to much larger aberrations, so the final value used for MR is somehow compensating for other effects in the 'real' RICH. The rather low values for the transparencies are due to the fact that GEANT generates in general too many photons. This problem is well known and occurred also for simulations in other experiments but is up to now not solved by the authors of GEANT.

Including the RICH geometry in the Monte Carlo, programming all effects to describe the physics processes that occur in the RICH (from the generation of Čerenkov light to firing PMTs), studying and deciding which effects could be omitted and finally tuning the remaining 'free' parameters, took a considerable amount of time of this Ph.D. work. Each minor change in either the geometry file or the source code meant that a new Monte Carlo production had to be run in order to visualise the impact of that change. The Monte Carlo productions were done on a PC cluster in Gent, consisting of 18 CPUs (Pentium III, 650 Mhz), running the Linux operating system. A single production consisted of the complete chain depicted in fig. 4.19 (the CPU time for GMC is negligible) and it was always executed in parallel on the complete PC cluster, taking about 24 hours to accumulate enough statistics. Several Linux shell scripts were written in order to distribute the processes over the CPUs of the cluster, to monitor the available disk space and to redirect the output to a single PC, for an online analysis of the running production and to compare the results with experimental data. More than 40 MC productions were done, which consumed in total 100 Gigabyte of disk space. After this the comparison between Monte Carlo and experimental results (which will be described in the next section) was good enough to include all the changes in the official version of HMC.

4.8.5 Comparison Monte Carlo and experimental results

4.8.5.1 Angles and number of PMTs for electrons

With all the tuned parameters a new Monte Carlo production was done and the results are presented in fig. 4.28. In these plots the experimental results for electrons are compared with the Monte Carlo outcome. The cuts that were applied to these electrons were already described in section 4.8.4. For the aerogel we emphasize that only electrons which are sufficiently far away from the tile edges were selected. The agreement in the small window

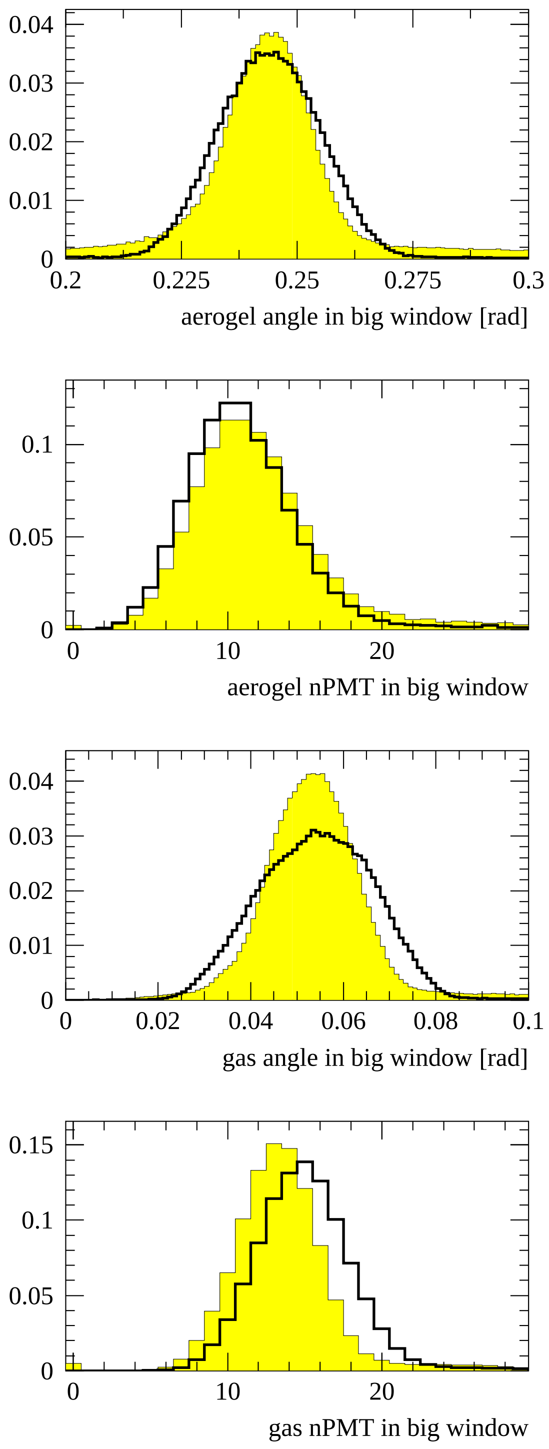

Figure 4.28: Comparison for aerogel and gas (angles and $nPMT$) between experimental data (yellow shaded histograms) and Monte Carlo simulation (black open histograms) for electrons. The comparison is shown for the 'big' and 'small' window.

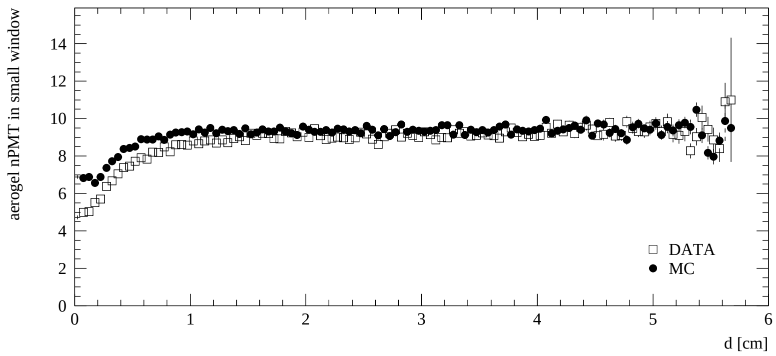

Figure 4.29: Comparison for aerogel $nPMT$ in the small window (for electrons) between experimental data (squares) and Monte Carlo simulation (dots) as a function of d (where d is the distance from the electron track to the nearest tile edge).

is excellent, but the agreement in the big window is not entirely satisfactory, especially for the gas angle. The reason is understood and derives from conflicting criteria in the tuning process [113] that drive the optimal settings for the tunable parameters to different values for the small window and the big window. The differences should not be considered as worrysome as the big window has a minor impact on the hadron PID. It plays however a (small) role for overlapping kaon rings from Φ decays.

The behaviour of the aerogel $nPMT$ near the tile edges is shown in fig. 4.29. The presence of the black sheets of tedlar between the aerogel stacks in the geometry file models the fall-off of $nPMT$ near the tile edges quite well.

4.8.5.2 Identification efficiencies and misidentifications

To study how the PID efficiencies and misidentifications introduced in section 4.7.1 look like, two different Monte Carlo productions were done, both with exactly the same parameter settings, i.e. the settings resulting from the tune described above. The only difference between both Monte Carlo productions is situated on the generator level (cfr fig. 4.19): the first production was using standard GMC_dis input (simulating the deep-inelastic scattering process), the second production utilised a version of GMC that generates decaying Φ mesons. The two Monte Carlo productions were analysed with IRT ($rQp > 0.0$). The reason for these two productions was already mentioned in section 4.7.2: only the decaying particle efficiencies and misidentifications can be determined experimentally. The P matrix elements for DIS can only be produced by a Monte Carlo production. To trust these latter results from the Monte Carlo the experimental efficiencies and misidentifications from decaying particles must agree with the ones from a Monte Carlo production. This is illustrated for kaons in fig. 4.30 for the topology case with 2 tracks per detector half. The agreement between the (black) squares on the plot and the (red) dots is very good, which implies that the result represented by the (green) triangles (i.e. the P matrix for DIS) can be trusted.

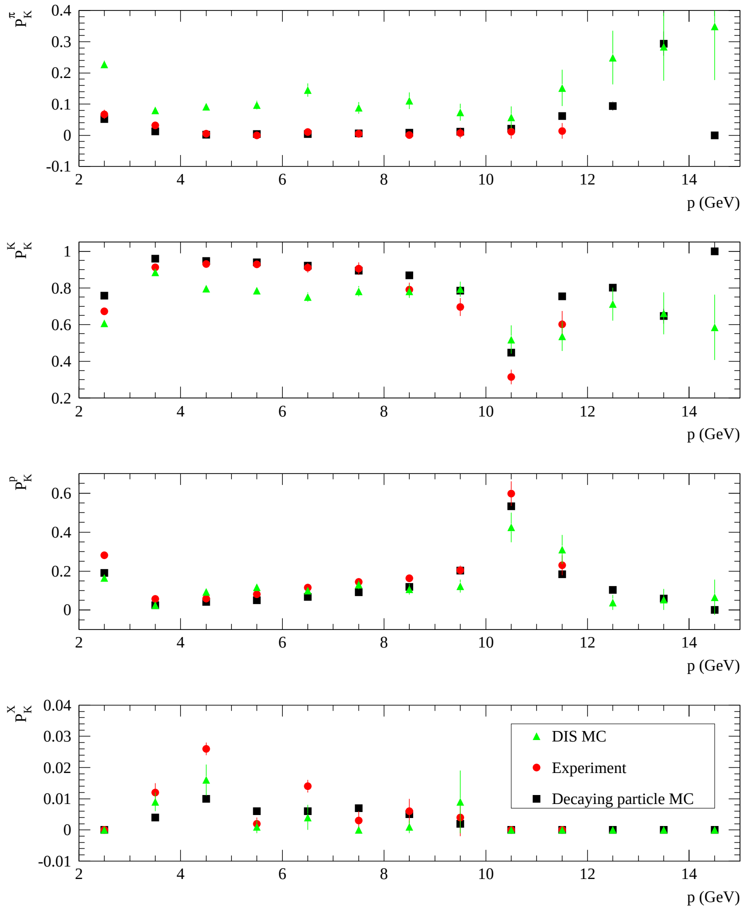

Figure 4.30: The kaon column of the P matrix for the topology case with 2 tracks per detector half: comparison between decaying particle Monte Carlo, DIS Monte Carlo and experimental results from decaying particles.

The different behaviour of the decaying particle MC and the DIS MC is due to a different topology of the events in both Monte Carlo 's: for the decaying particle MC there are always 2 kaons in the detector half, while for the DIS MC the majority of the 2 track events consists of 1 electron track and 1 kaon track. The overlap between the two rings is different in both cases due to different phase space limitations. This explanation why both Monte Carlo 's produce different efficiencies is confirmed if one considers a plot analogous to fig. 4.30, but for events with 1 track per detector half instead of 2 tracks per half [113]. For single tracks per detector half, the agreement between both Monte Carlo 's is much better since the influence of the topology of the event is negligible: there is only one ring in the complete detector half.

From the DIS MC production one can get a full P matrix which is valid for the DIS process and which will be used in the next section to 'unfold' the hadron momentum distributions for kaons, pions and protons.

4.9 Unfolding of hadron distributions

Any hadron that is identified by the RICH detector as either a pion, kaon or proton is only identified correctly with a certain probability. Hence most physics analyses that use the RICH detector will have to include a correction for misidentifications [114]. This section describes the basics of the unfolding of hadron momentum spectra, where 'unfolding' is defined as the extraction of spectra for the true hadron types from those for the identified hadron types.

4.9.1 The Q matrix: purities and contaminations

A useful set of basic parameters for the description of the performance of the RICH are the identification efficiencies and misidentifications, which are collected in the P matrix that was introduced in section 4.7.1. Each identification probability P_t^i gives the probability that a particle with true type t is identified as type i. The complementary set of probabilities are the true type probabilities Q_t^i that a particle that is identified as type i is truly of type t:

$$Q_t^i = \frac{N_t^i}{\sum_s N_s^i} \tag{4.56}$$

where the notations from eq. (4.25) have been used. Together they form the matrix Q, which in the case of three particles (π, K, p) has the form

$$\mathsf{Q} = \mathsf{Q}(\pi, K, p) = \begin{bmatrix} Q_\pi^\pi & Q_\pi^K & Q_\pi^p & Q_\pi^X \\ Q_K^\pi & Q_K^K & Q_K^p & Q_K^X \\ Q_p^\pi & Q_p^K & Q_p^p & Q_p^X \end{bmatrix} \tag{4.57}$$

The last column of this matrix concerns the unidentified particles. The diagonal elements Q_i^i of the 3 × 3 submatrix of the Q matrix without the last column are called *purities* and the off-diagonal elements are called *contaminations*. Due to the definition of the probabilities, the columns of this matrix are normalised to 1, but contrary to the P matrix, the summation runs over the true types t

$$\sum_t Q_t^i = 1 \tag{4.58}$$

The following may help to clarify the difference between the P and Q matrix: P_π^K is the probability that a true pion is misidentified as a kaon; Q_π^K is the probability that a particle that is identified as a kaon is in fact a pion.

The P and Q matrix can be connected if one introduces ϕ_t, the relative flux of the hadron type t

$$\phi_t = \frac{N_t}{\sum_s N_s} \qquad \text{where} \qquad N_t = \sum_i N_t^i \qquad (t = \pi, K, p) \;\; (i = \pi, K, p, X) \tag{4.59}$$

The relative flux ϕ_t gives the fraction of hadrons that are of true type t. Using the definitions of the P and Q matrix, it can easily be shown that

$$Q_t^i = \frac{P_t^i \cdot \phi_t}{\sum_s P_s^i \cdot \phi_s} \tag{4.60}$$

This last equation shows an important difference between the P and Q matrix which can best be illustrated for single track events that are not affected by the limited acceptance of the RICH mirror and PMT plane. As explained before the P matrix in this case is purely a detector property and does not depend on the composition of the particle sample. The Q matrix is different for different data samples due to its dependence on the relative particle fluxes. Hence the Q matrix reflects both the performance of the detector and the composition of the particle sample. As a result a corresponding Q matrix has to be calculated for every physics analysis. The Q matrix is needed for a physics analysis because it allows to calculate the true number of kaons, pions and protons from the measured numbers. If the numbers of identified hadrons are collected in a vector $\vec{I}$

$$\vec{I} = (I_\pi, I_K, I_p, I_X) \tag{4.61}$$

and if the true numbers of hadrons (which are of course not known a priori) are written as

$$\vec{N} = (N_\pi, N_K, N_p) \tag{4.62}$$

then $\vec{N}$ can be derived from $\vec{I}$ by the relation

$$\vec{N} = \mathsf{Q} \cdot \vec{I} \tag{4.63}$$

If the Q matrix is known, it allows to calculate the true number of hadrons $\vec{N}$. As a crosscheck a second relation between $\vec{N}$ and $\vec{I}$ can be verified afterwards:

$$\vec{I} = \mathsf{P} \cdot \vec{N} \tag{4.64}$$

The problem of unfolding can be defined as: *Given the* P *matrix and given the vector* $\vec{I}$ *(the number of identified hadrons), calculate the vector* $\vec{N}$ *(the true number of hadrons) and the* Q *matrix.*

4.9.2 Solving the unfolding problem

Before explaining how the problem of unfolding can be solved, some words need to be said about the binning: the P matrix is binned in momentum and topology case (1, 2, >2 tracks per detector half). Hence the vector $\vec{I}$ must be measured according to the same binning and in addition a distinction between positive and negative particles has to be made, since the relative particle fluxes are charge dependent. Thus $\vec{I}$, $\vec{N}$ and Q are binned in momentum, topology case and charge, since these quantities are all flux dependent. The unfolding problem has to be solved in each bin separately and this can be accomplished by an iteration method, based on eq. (4.63). From equations (4.59) and (4.60) follows that the Q matrix elements can also be expressed as

$$Q_t^i = \frac{P_t^i \cdot N_t}{\sum_s P_s^i \cdot N_s} \tag{4.65}$$

At the start of the iteration the elements of $\vec{N}$ are obviously not yet known, but the elements of $\vec{I}$ can be used to calculate the start values of the Q matrix elements:

$$Q_t^i(0) = \frac{P_t^i \cdot I_t}{\sum_s P_s^i \cdot I_s} \tag{4.66}$$

The iteration progresses then as follows:

$$\vec{N}(n+1) = \mathsf{Q}(n) \cdot \vec{I} \tag{4.67}$$

$$Q_t^i(n+1) = \frac{P_t^i \cdot N_t(n+1)}{\sum_s P_s^i \cdot N_s(n+1)} \tag{4.68}$$

A Fortran program based on the above algorithm has been implemented and an example of the results is shown in fig. 4.31 and fig. 4.32. The error bars which are visible on these figures will be discussed after explaining the behaviour of the unfolding program for this example.

The illustrated example is for the topology case with one track per detector half. Only results for positive particles will be shown, negative particles behave in a similar way. This example is not for experimental data but for Monte Carlo data, as this allows to compare the output of the unfolding program with the *exact results* one should obtain and which are only known for Monte Carlo data. This provides a method to check the accuracy of the unfolding program. The illustrated example is for the RPS method which produces very few X particles, hence these are ignored here. The input for the unfolding program (fig. 4.31) consists of the P matrix (as a function of momentum) and the measured momentum distributions (the vector $\vec{I}$). The P matrix is supposed to be charge independent, while the measured hadron fluxes are not. It is important to stress that the extraction of e.g. the true pion momentum spectrum (N^{π}) requires not only the measured pion spectrum (I^{π}) but also the spectra for kaons and protons (i.e. the complete vector $\vec{I}$). Only if these observed spectra are known, the real momentum spectra of pions, kaons and protons can be determined. The output of the unfolding program is shown in fig. 4.32, it consists of the Q matrix and the unfolded momentum distributions (the vector $\vec{N}$). Since the example concerns Monte Carlo data, both the exact vector $\vec{N}$ and the exact Q matrix are known a priori. They are also shown in fig. 4.32 and the agreement with the results produced by the unfolding program is satisfactory. The fact that the exact Q matrix elements are all on one outer side of the error bar of the unfolded results, is due to an artefact in the P matrix: the P matrix is supposed to be charge independent, but within the limited statistics of the Monte Carlo production, differences are seen between the P matrices for positive and negative charge. Due to the method of error propagation from the P matrix to the Q matrix (which will be discussed below), the distributions for positively charged hadrons favour the part of the P error that is valid for positive particles, driving the error for the iterated Q matrix towards the exact Q matrix. Nevertheless, the most important conclusion is that the Q matrix resulting from the unfolding program agrees (within its errors) with the exact Q matrix. The same conclusion is valid for the unfolded momentum distributions, shown in the lower three panels of fig. 4.32. In these panels, the unfolded distributions (the vector $\vec{N}$) produced by the unfolding program, the exact ones and the measured momentum distributions (the vector $\vec{I}$) are compared. Comparing the first two spectra shows that the unfolding program indeed works correctly, comparing the latter two illlustrates that a big difference between unfolded and measured momentum distributions can occur (e.g. kaons with momentum below 5 GeV).

P matrix elements are not known with infinite precision. They have statistical and systematic uncertainties. This results in errors on the Q matrix and the unfolded momentum distributions. The problem of propagating the errors on P to errors on Q and $\vec{N}$ is

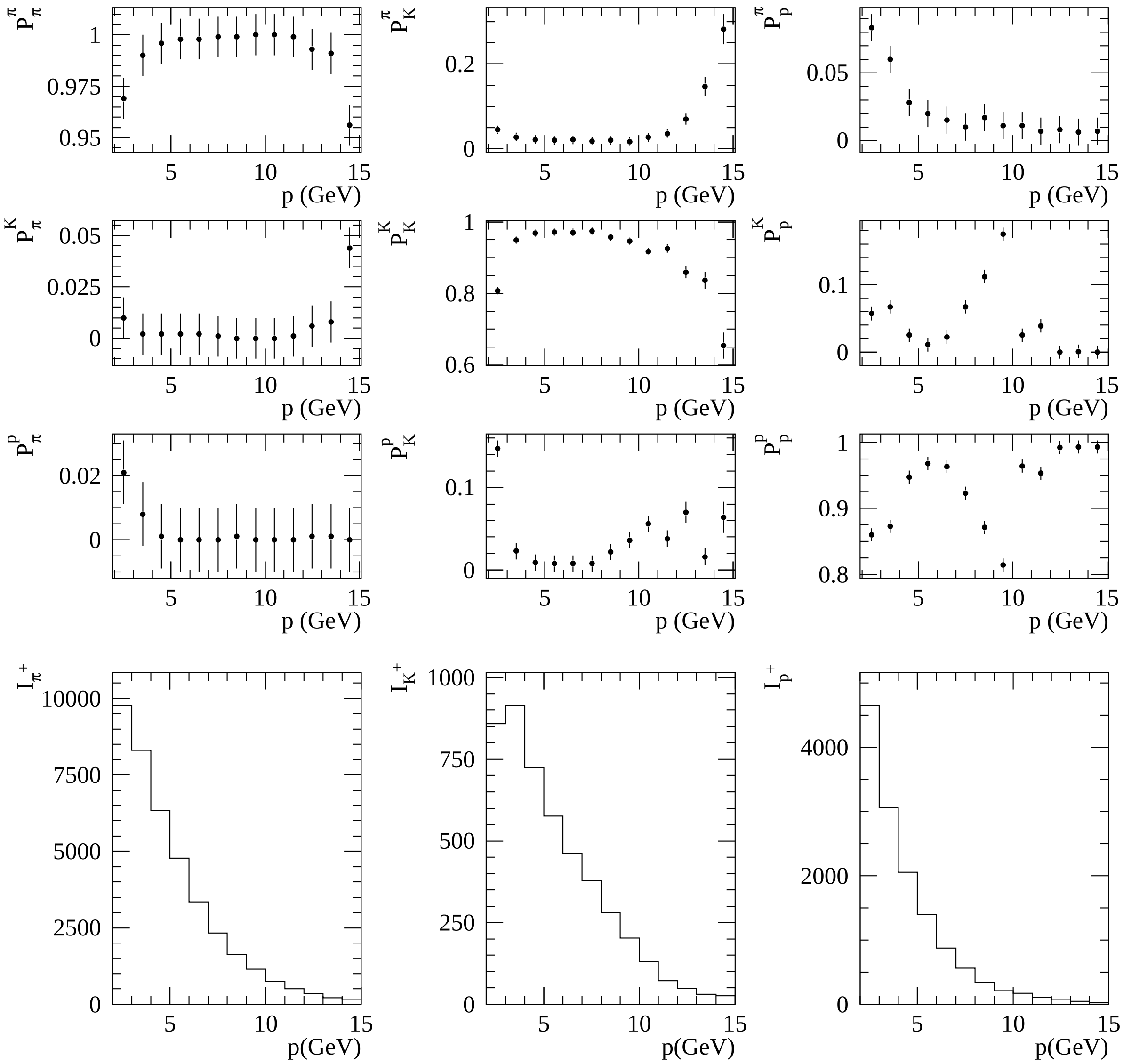

Figure 4.31: The input for the unfolding program: the P matrix (9 top panels) and the measured momentum distributions for positively charged pions, kaons and protons (bottom 3 panels). All plots are for Monte Carlo data and for the topology case with 1 track per detector half.

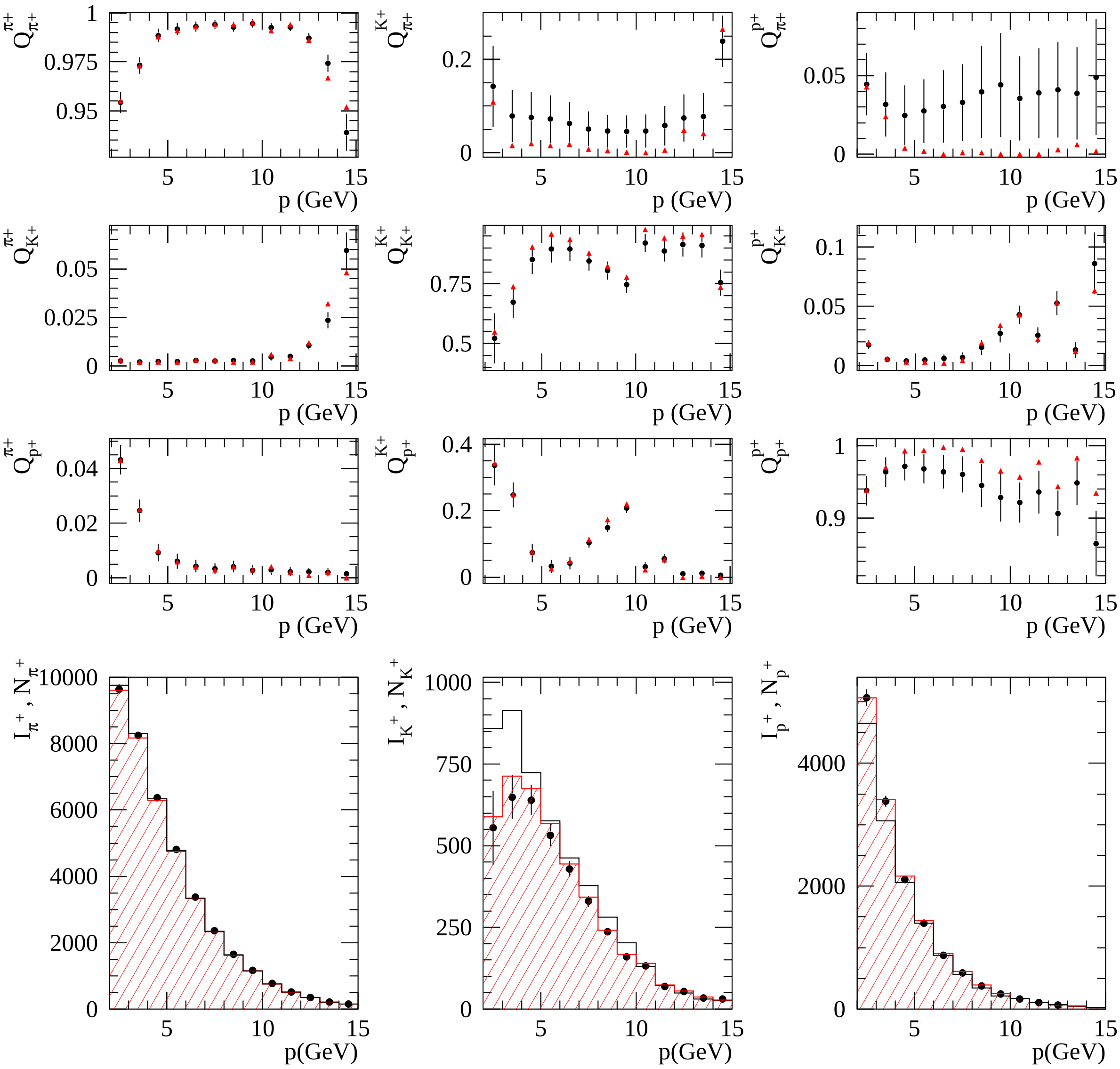

Figure 4.32: Results from the unfolding program (Monte Carlo data, for the topology case with 1 track per detector half): the Q matrix (9 top panels) and the unfolded momentum distributions (bottom 3 panels). The result from the unfolding program for the Q matrix is shown by the (black) dots, the exact Q matrix (which is known a priori for MC data) is shown by the (red) triangles. The unfolded momentum distributions (produced by the unfolding program) are shown by the (black) dots and are compared with the exact unfolded momentum distributions which are represented by the (red) hatched histograms. The (black) open histograms represent the measured momentum distributions from fig. 4.31 and are shown for reference.

addressed by a statistical method. For each P matrix element a Gaussian probability distribution is assumed, with a width that corresponds to the uncertainty of the given P matrix element. It has to be taken into account that the columns of the P matrix must stay correctly normalised to 1. Then P matrices are randomly constructed several thousand times according to these probabilities and used to unfold the hadron distributions. The number of randomly constructed P matrices must be large enough not to introduce an additional statistical uncertainty. The elements of the resulting unfolded distributions and Q matrices again follow Gaussian distributions whose widths correspond to their uncertainties.

The error propagation is necessary because small errors in the P matrix can cause large errors in the unfolded distributions and the Q matrix. The Q matrix depends on the relative hadron fluxes, as can be seen from eq. (4.68), thus the error propagation is not universal but depends on the data sample. The error propagation of course strongly depends on the P matrix errors, from which one starts. For the example above, only (binomial) statistical errors on the P matrix have been used, since both the P matrix and the measured momentum distributions come from the *same* MC production. For the unfolding of experimental data, also the systematic errors on the P matrix have to be taken into account.

Apart from the error propagation, the unfolding program also performs two other checks: as can be seen from equations (4.63) and (4.64), the vectors $\vec{I}$ and $\vec{N}$ are connected by two relationships ($\vec{N} = \mathsf{Q}\cdot\vec{I}$ and $\vec{I} = \mathsf{P}\cdot\vec{N}$). While the first relationship is always fulfilled after convergence (cfr. eq. (4.68)), this is not the case for the second relation! Therefore the unfolding program only accepts iterations (i.e. P^i_j 's) which obey both conditions to build the Gaussian distributions for the unfolded distributions and Q matrices. A drawback of introducing the error propagation, requiring the iterated $\vec{N}$ to be converged and forcing the relationship $\vec{I} = \mathsf{P}\cdot\vec{N}$ to be fulfilled, is that the unfolding program only converges if there are no X particles present in the data sample (in contradiction to a simple version of the unfolding program which is purely based on equations (4.65) - (4.68)). Since the RPS method is supposed to be used for all physics analyses in HERMES, this is not really a problem, as it produces almost no X particles and the elements $P^X_t \approx 0$ for almost all bins.

The unfolding program (without the error propagation part) has been checked by 3 other methods, which bypass the Q matrix and which are based on solving the system of equations (4.64). The first method introduces the matrix P', which consists of the first 3 rows of the P matrix. The vector $\vec{N}$ can then be obtained as $\vec{N} = \mathsf{P}'^{-1}\cdot\vec{I}'$, where $\vec{I}' = (I_\pi, I_K, I_p)$. It has to be stressed that this solution does conserve the number of particles since the relation $\sum_{i=1}^4 I_i = \sum_{j=1}^3 N_j$ will automatically be fulfilled for the solution of $\vec{N}$ that was found, due to the fact that the elements of each column of the P matrix add up to one.

The second method finds a solution for $\vec{N}$ in the Chebychev norm (minimax solution), which minimises

$$c = \max_{1\leq i\leq 4} c_i = \max_{1\leq i\leq 4} \left| I_i - \sum_{j=1}^{3} P_{ij} N_j \right| \tag{4.69}$$

The third method uses a minimisation of the square of 4 functions (Newton method):

$$f_k(N_1, N_2, N_3) = P_{k1}\cdot N_1 + P_{k2}\cdot N_2 + P_{k3}\cdot N_3 - I_k \quad (k = 1, 4) \tag{4.70}$$

and finds the vector $\vec{N}$ which minimises

$$\Phi(N_1, N_2, N_3) = \sum_{k=1}^{4} f_k(N_1, N_2, N_3)^2 \tag{4.71}$$

All three methods yield the same unfolded distributions $\vec{N}$ as the iteration method except for some bins where negative values occurred. If the unfolded distributions are known, eq. (4.56) can be used to calculate the Q matrix a posteriori, giving the same result as the iterative method in most of the cases. The computing speed was not an issue for all three methods. Since the iterative method never produces negative values for elements of $\vec{N}$ and since it yields $\vec{N}$ and Q at the same time, the iterative method is preferred. The error propagation was only programmed for the iterative method.

At present, the unfolding program is frequently used by the HERMES collaboration. The source code of the program has been documented and a brief manual has been written, available as a webpage [115].

Chapter 5

The structure function ratio $F_2^d(x)/F_2^p(x)$

After presenting the detector related results of this thesis in the previous chapter, the following two chapters will focus on the physics of inclusive and semi-inclusive deep-inelastic scattering respectively. In this chapter, the data recorded from deuterium and hydrogen targets during 2000 are used to extract the structure function ratio $F_2^d(x)/F_2^p(x)$. The determination of this inclusive quantity from experimental data is rather simple: it basically involves only the measurement of the deeply inelastic scattered positron yield on 2 different targets. Section 5.1 describes the basic formalism to derive this ratio from experimental data. Due to unanticipated difficulties in an experiment, not all of the data collected are suited for physics study. One whishes to cut the data set to eliminate instabilities that may cause a systematic bias in the result. The criteria to select the data with which the structure function ratio is formed, are outlined in section 5.2. The raw F_2-ratio that can be determined from carefully checked experimental data needs to be modified in order to reproduce the true F_2-ratio in nature. These corrections account for charge-symmetric background, detector smearing, radiative effects and detector acceptance. They will be introduced and applied in section 5.3.

5.1 Basic formalism

In the one-photon exchange approximation (fig. 2.1) the inclusive DIS cross section is related to the structure function $F_2(x, Q^2)$ and the ratio of the longitudinal and transverse cross section $R(x, Q^2)$ by eq. (2.20):

$$\frac{\mathrm{d}^2\sigma}{\mathrm{d}x\,\mathrm{d}Q^2} = \frac{4\pi\alpha^2 F_2(x, Q^2)}{Q^4 x}\left[1 - y - \frac{Q^2}{4E^2} + \frac{y^2 + Q^2/E^2}{2(1 + R(x, Q^2))}\right] \tag{5.1}$$

Hence the ratio of the deuterium and the proton cross sections at any fixed x and Q^2 can be written as

$$\frac{F_2^d}{F_2^p} = \frac{\sigma^d}{\sigma^p} \tag{5.2}$$

where we assumed that R does not depend on the target nucleus. This assumption is supported by several experimental results: at low Q^2 from SLAC [118] in the range

$0.10 < x < 0.86$ and $0.6 < Q^2 < 20$ GeV^2, from BCDMS [119] for $x > 0.07$ and $Q^2 > 8$ GeV^2 and from NMC [120] in the range $0.002 < x < 0.4$ at an average Q^2 of 0.5 GeV^2.

The F_2-ratio can be directly related to the DIS yields from deuterium and hydrogen targets: in the HERMES experiment, the number of scattered positrons is counted along with the relative luminosity measured with the luminosity monitor. According to eq. (3.9) the DIS cross section is proportional to the number of scattered positrons $N^{\Delta t}$ during a certain time interval Δt, divided by the total luminosity during that time:

$$\sigma \sim \frac{N^{\Delta t}}{L^{\Delta t}} \equiv Y \tag{5.3}$$

where Y denotes the luminosity normalised yield of DIS events. Finally, the F_2-ratio directly in terms of measured quantities becomes:

$$\frac{F_2^d}{F_2^p} = \frac{Y^d/2}{Y^p} \tag{5.4}$$

The factor 2 reflects the A/Z correction to get the nucleon density from the density of electrons of the target gas atoms: the luminosity as determined by the luminosity monitor is only valid for a hydrogen target where the density of the target protons (participating in the measured scattering reaction) is equal to the density of the target electrons (which determine the luminosity monitor reference reaction). If one has two nucleons per electron (deuterium), the factor 2 needs to be applied since the measured luminosity is only half of the luminosity of the real scattering process.

The next sections will describe how the raw yield $Y^{d/p}$ can be extracted from experimental data and which corrections must be applied to extract the desirable F_2-ratio.

5.2 Data selection

Not all the data acquired by the experiment are suitable for the extraction of physics results. The first important step in every HERMES analysis is to assure a stable detector performance and a good quality of the data sample. This is done in order to minimise systematic uncertainties caused by the experimental conditions. All subsystems of the spectrometer have to operate reliably and stable, the beam conditions must fulfil certain criteria and a high quality of target operation is required.

Additional 'physics' cuts have to be applied to select the desirable physics process, the deep-inelastic scattering in case of this F_2 analysis. Events originating from other processes and forming the 'background' for the studied one have to be removed. These particular selection criteria are chosen to minimise the background contamination in the event sample, keeping the statistics of useful events as high as possible.

The data quality selection criteria and the kinematical cuts that identify the DIS process will be described in the following two sections.

5.2.1 Data quality

The basic unit of time that characterises the detector stability is the length of a *fill* of positrons in the HERA ring. A fill corresponds to a period of typically 6-8 hours in which the experiment acquires data in an undisturbed manner. Between fills, hardware and software changes to the experimental setup may occur, introducing variations in the detector conditions from one fill to the next. The HERMES Data Acquisition (DAQ)-system separates the data collected within a fill in *runs*. When an accidental failure of a component of the detector is recorded, the corresponding runs are marked in the logbook. The rejection of these runs is the first step in the data quality selection. The smallest subdivision of the data is represented by a *burst*, corresponding to ≈ 10 seconds of data taking. Within a run, the DAQ-system generates signals that define the boundaries of each burst and enable the read-out of the slow-control data (this is the data that reflects the state of the spectrometer, target, luminosity monitor and the HERA accelerator, which is read out in time intervals from seconds to minutes). Although the detector is read out for every triggered event, each burst contains the complete information of the experimental status (e.g. beam and target polarisations and luminosity monitor rates) integrated over the 10 seconds. Events acquired within these 10 seconds consequently belong to the set of slow-control information defined in the burst.

The following paragraphs detail the criteria applied for the selection of data suitable for the analysis of the F_2-ratio. Numerical values for the criteria are summarised in table 5.1.

5.2.1.1 Data acquisition

Since the data acquisition controls the recording of the incoming data, one must ensure that this system is operating stably. An important DAQ related quantity is the dead time or equivalently the live time τ_{live}, which basically indicates the fraction of time the DAQ-system is able to accept and record new incoming data without being blocked due to the processing of the previously accepted data. This fraction can be estimated from the total number of generated triggers and triggers accepted by the DAQ-system per burst

$$\tau_{live} = \frac{T_{acc}}{T_{gen}} \qquad\qquad \tau_{dead} = 1 - \tau_{live} \tag{5.5}$$

The live time together with the burst length t_{burst} which is the length in time during which the data in that burst were collected, are restricted to certain ranges (table 5.1) since they enter the luminosity normalisation for calculating the DIS yields (eq. (5.3)):

$$L^{\Delta t} = \sum_{\text{bursts during } \Delta t} L^{\text{burst}} \cdot \tau_{dead}^{\text{burst}} \cdot t_{\text{burst}} \tag{5.6}$$

where L^{burst} is the raw luminosity rate for unpolarised bursts or the fitted rate for polarised bursts (cfr. section 3.5). To avoid DAQ problems during the beginning or termination of a run, when the DAQ has to perform several clean-up and initialisation tasks, the first and last bursts from each run were removed from the data.

Furthermore, all bursts exhibiting problems with timing or synchronisation with the slow-control data were rejected.

Nr.	Quantity	Criterium
1	Live time	$50 \leq \tau_{live} < 100$ %
2	Burst length	$9 < t_{burst} \leq 11$ s
3	Beam current	$5 < I_e < 50$ mA
4	Luminosity rate	unpol. data: $5 \leq L_{rate} \leq 2000$ Hz pol. data: $5 \leq L_{fit} \leq 80$ Hz
5	First and last burst in run	rejected
6	timing or synchronisation problems	rejected
7	HV trips in FC or BC	rejected
8	HV trips in TRD	rejected
9	PCOS initialisation in RICH	rejected
10	dead blocks in calo	rejected
11	dead blocks in H2 or lumimonitor	rejected
12	no PID (initialisation or calo. threshold)	rejected
13	bad according to logbook	rejected

Table 5.1: The data quality cuts for the 2000 data set.

5.2.1.2 Beam and luminosity

A high precision DIS experiment requires intense, stable beam currents. The positron beam current was forced to remain within the reasonable operating range of HERA.

Variations in the beam current and the target density can cause fluctuations in the rate of the luminosity monitor. The requirement for the rate to remain within a nominal operating range is a test for the stability of the HERA positron beam and the HERMES target density. The selection of this range depends on the mode of data taking. As explained in section 3.2 there are three such modes at HERMES: the most common mode is normal polarised-target data taking where the luminosity is limited by the target densities which can be obtained with the atomic beam source. Data taking with unpolarised targets can be done in two ways. Normally, unpolarised target densities are limited by the beam live time. Alternatively, data are collected in higher density running, which is done either in a dedicated run for HERMES only (the other HERA experiments do not collect data during such runs) or regularly at the end of each fill. Under these conditions the luminosity is limited by the trigger rates or the dead time of the DAQ-system. In the year 2000, all of the hydrogen data was taken with an unpolarised high density target whereas for the deuterium both polarised low density and unpolarised high density data exist.

For unpolarised data, one restricts the range for the *raw* coincidence rate in the luminosity monitor while for polarised data the *fitted* coincidence rate (cfr. section 3.5) is restricted, in a different range (due to the lower target density).

5.2.1.3 Detector and tracking conditions

Confidence in the detection and determination of particle trajectories requires the proper functioning of the tracking system in the spectrometer. The stability of the drift chambers[1] (FC and BC) is controlled by monitoring the high voltage in these chambers. If changed beam conditions or an increased proton background cause a short exceeding of the current limit imposed by the high voltage supply of a certain chamber, the high voltage of one or more planes of this chamber is switched off for a short time ('HV trip'). Until the high voltage is restored, the count rates in the concerning chamber are reduced by the HV trip. Therefore, bursts with HV trips in the FC and BC are ignored for physics purposes.

DIS studies at HERMES require the separation of leptons and hadrons through the stable operation of the PID detectors. These detectors are also the key components of the HERMES trigger. A variety of techniques is used to determine the performance of these detectors:

- For the TRD, a burst is rejected if a HV trip deteriorates the detection efficiency.

- Although the main goal of the RICH detector is the identification of the different hadron species, it is also involved in the hadron/lepton separation (cfr. eq. (3.7)). Therefore, bursts where the PCOS system (which reads the RICH PMTs out) was initialised, are rejected for this inclusive analysis.

- The GMS continuously checks the response of the photomultipliers of the calorimeter, the luminosity monitor and the preshower hodoscope H2. Bursts where the PMTs of these detectors failed, are ignored.

- Data with no available PID due to initialisation problems or an unknown calorimeter threshold are also discarded.

In total, the 2000 set contains 862 980 bursts for deuterium and 34 367 bursts for hydrogen. The application of all the criteria mentioned above yields 556 739 good bursts for deuterium and 15 543 good bursts for hydrogen. The main reason why the loss for hydrogen is higher than the one for deuterium is that synchronisation problems seem to occur more frequently for the high density data.

5.2.2 Event selection criteria

Once a collection of analysable bursts is determined, several criteria limit the data sample. From all the recorded tracks, the useful ones have to be picked out first. Then the particle identification algorithm described in section 3.3.2.4 separates the scattered positrons from the produced hadrons. Kinematic cuts on the scattered positron then define our inclusive DIS sample. These track and event criteria are motivated in the sections below and summarised in table 5.2.

[1]The DVC were not included in the tracking algorithm for the 2000 data.

Quantity	Criterium
Cuts on the track level	
reconstructed vertex inside target	$\vert z_{\mathrm{vertex}}\vert < 18$ cm
transverse vertex offset	$r_{\mathrm{vertex}} < 0.75$ cm
horizontal fiducial cut	$\vert x_{\mathrm{calo}}\vert < 175$ cm
vertical fiducial cut	$30 < \vert y_{\mathrm{calo}}\vert < 108$ cm
track momentum	$0.5 < p < 27.5$ GeV
required trigger	21
magnet tracks	rejected
PID cuts	
calo. cluster energy (DIS lepton)	$E_{\mathrm{calo}} > 3.5$ GeV
lepton identification	2 < PID3 + PID5 < 80
hadron identification	-80 < PID3 + PID5 < 0
Kinematical cuts	
select DIS region	$Q^2 > 1.0$ GeV^2
small radiative corrections	$y < 0.85$
non-resonance region	$W^2 > 10.0$ GeV^2

Table 5.2: The track and event selection criteria.

5.2.2.1 Criteria on the track level

For each track, cuts were imposed on the reconstructed vertex quantities z_{vertex} and r_{vertex}, being the position along the z axis and the perpendicular distance from the z axis of the vertex defined by the track and the beam line respectively, to select only tracks originating from the target cell.

To ensure that an electromagnetic shower created in the calorimeter is mostly contained within the calorimeter glass blocks, the center of a cluster must not be too close to the edges of the calorimeter wall. Apart from this reason, the constraint on the spatial limits of the calorimeter also acts as a *fiducial volume cut*: this cut requires the tracks to pass directly through the HERMES acceptance and removes the tracks at the edges of the spectrometer where the identification efficiency goes down.

Tracks with unphysical low or high momentum are not included in this analysis. Only tracks that belong to an event that generated a DIS trigger (trigger 21) were selected. The reconstruction method of the tracks that was used, requires a partial track in the front detector region by using FC only, and a back partial track recorded by the BC. Tracks for which only the front partial track is available, but the back partial track is missing (so-called *magnet tracks*, because the track is lost in the magnet region), are rejected.

5.2.2.2 Kinematical constraints and PID cuts

After selecting the good tracks, the DIS lepton in each event has to be found. Electrons are considered here in addition to the positrons (and kept in the data sample), as they measure charge-symmetric background in the positron sample, as will be explained later on.

First, the energy of the cluster associated with a candidate DIS lepton track is required to match the trigger condition ($E_{calo} > 3.5$ GeV). As explained in section 3.3.2.4, the relevant quantity for particle identification is the expression $PID_3 + PID_5$ which refers to the summed likelihood parameter from the various PID detectors. Using the values mentioned in table 5.2, the lepton identification efficiency is above 97% and the contamination of misidentified hadrons in the lepton sample remains below 1.0% over the entire kinematic range [121].

Once leptons have been identified in the detector, their kinematics completely specify the x, Q^2 and W^2 of the event. Kinematic cuts on the lepton delimit the phase space associated with deep-inelastic scattering. For each event, the highest momentum lepton is selected for calculation of the event properties. A cut on the negative four-momentum transfer squared Q^2 sets the scale for the deep-inelastic scattering process and allows to model the events with perturbative QCD concepts. An upper limit on the fractional lepton energy transfer y was introduced to make sure that the positron had enough energy to generate a trigger and also to limit the events into a region where the radiative corrections are not too high, as will be explained in section 5.4.2.2. Finally a minimum value for the invariant mass of the hadronic final state W^2 was required to remove the events from the baryon resonance region. Although a cut of $W^2 > 4$ GeV^2 already ensures that the nucleus is not excited into a resonant state and brakes apart incoherently, a tighter cut of $W^2 > 10$ GeV^2 was chosen in this analysis, mainly to guarantee compatibility with the d_v/u_v analysis that will be described in the next chapter. In this latter analysis, the process of interest is the *semi*-inclusive DIS where hadrons are detected in coincidence with the DIS lepton. Hadron fragmentation is analysed assuming factorisation and Monte Carlo studies [122] suggest that a cut of $W^2 > 10$ GeV^2 is the best choice for ensuring the validity of this assumption at HERMES.

After applying all the above mentioned criteria, 5.96 million good DIS events have been found for deuterium and 3.36 million for hydrogen.

5.2.3 Stability of the experimental DIS yields

To avoid possible systematic errors due to changing detector behaviour, one should make sure that the data show no instabilities after applying all the data quality requirements and the kinematical cuts mentioned above. A good probe to check this stability is the luminosity normalised yield of DIS events (Y) that was already defined in eq. (5.3). The below mentioned numerical results for Y were obtained after subtracting the charge-symmetric background from the measured DIS sample, as will be described in the next section.

As a first check, the mean yield for the deuterium target (integrated over x and Q^2) was plotted versus fill number, i.e. the yield calculated according to eq. (5.3) where Δt is the length of each fill. Fig. 5.1 shows that two periods can be isolated, separated by fill 97. At that time, the high voltage of the H0 hodoscope was changed, resulting in an overall efficiency increase of the trigger, as trigger studies [124] revealed. Therefore, data before fill 97 were discarded.

A second test for the stability of the experimental DIS yields is a comparison of the mean yields on the time scale of the burst, the run, the fill and the complete year 2000. The

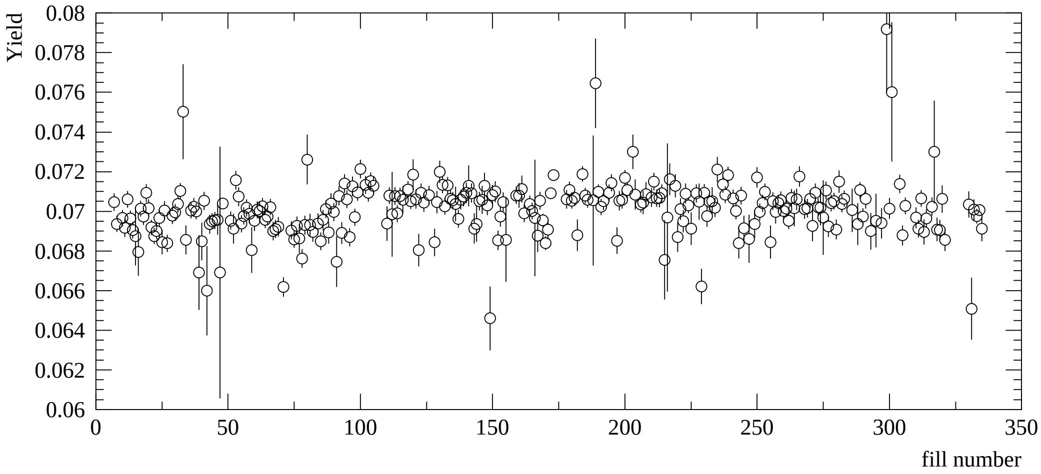

Figure 5.1: Normalised DIS yields versus fill number for the deuterium target during the year 2000.

Time scale Δt	Deuterium	Hydrogen
year 2000	0.07025 ± 0.00003	0.03924 ± 0.00002
fill level	0.0703 ± 0.0014	0.0392 ± 0.0007
run level	0.0706 ± 0.0059	0.0393 ± 0.0013
burst level	0.071 ± 0.024	0.039 ± 0.005
Time scale Δt	Unpolarised deuterium	Polarised deuterium
run level	0.0707 ± 0.0021	0.0706 ± 0.0059

Table 5.3: Mean DIS yields for different time scales.

results are listed in table 5.3. The statistical errors are somewhat smaller for hydrogen, due to the fact that all hydrogen data was taken with an (unpolarised) high density target whereas most of the deuterium data was taken with a (polarised) low density target. The last row of table 5.3 illustrates that the polarisation and thus also the target density does not affect the deuterium yields.

5.3 Determination of $F_2^d(x)/F_2^p(x)$ from experimental data

If one uses the raw data without any correction and calculates the DIS yields as a function of x, eq. (5.4) allows to extract the F_2-ratio, as shown in fig. 5.2. In this figure, the raw F_2-ratio from the 2000 data set (the numerical values are listed in table 5.4) is compared to the results of an earlier analysis (based on the 1996 data set) by M. Vincter [125]. The agreement between the two analyses is satisfactory, taking into account that slightly different kinematical cuts were used in [125]: the lower Q^2 cut was 0.3 GeV^2 instead of 1.0 GeV^2, and the lower W^2 cut was 4 GeV^2 instead of 10 GeV^2, allowing to extend the x range to somewhat lower values than in this work.

In fig. 5.2 the HERMES data are also compared to a parametrisation [120] (determined by the NMC collaboration) of the existing world data, which is based on measurements from NMC, BCDMS and SLAC. In this parametrisation, the average x and Q^2 can be inserted to evaluate σ_d/σ_p in each x bin. Fig. 5.2 shows that rather large corrections are required at low x in order to obtain agreement between the HERMES data and the NMC parametrisation.

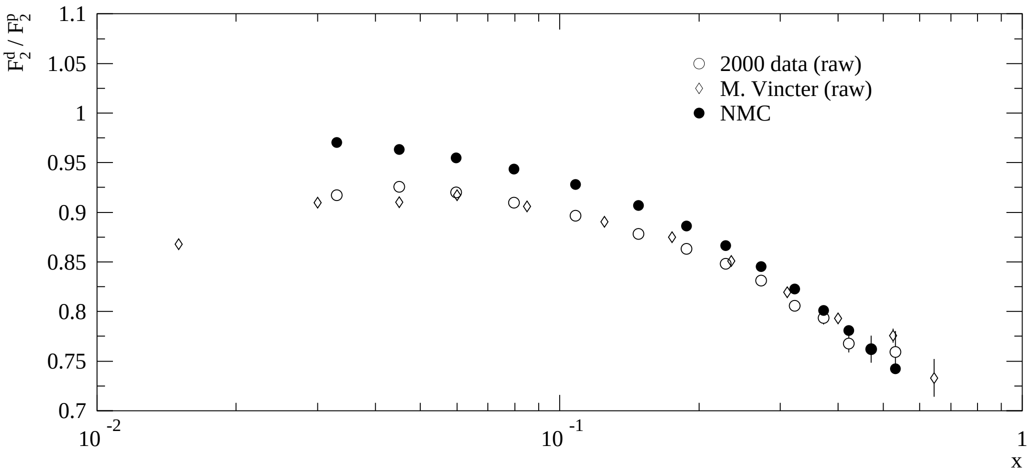

Figure 5.2: The x dependence of $F_2^d(x)/F_2^p(x)$ without any correction is shown in this figure along with the results obtained by M. Vincter. Errors are statistically only. The parametrisation from NMC, calculated at the mean value of x and Q^2 of the experimental 2000 data is also shown.

5.4 Corrections to $F_2^d(x)/F_2^p(x)$

In this section the corrections which were applied to the measured raw yields $Y^{d/p}$ are discussed. These corrections include charge-symmetric background processes, kinematical smearing in the variable x, contributions to the yields from radiative effects and acceptance effects. The hydrogen and deuterium yields were corrected independently from each other.

5.4.1 Background corrections

The sample of DIS events selected by the kinematic cuts described in section 5.2.2 is not a *pure* DIS sample but contains some background. This background generates positrons that pass all the DIS criteria and hence cannot be discriminated directly from the real

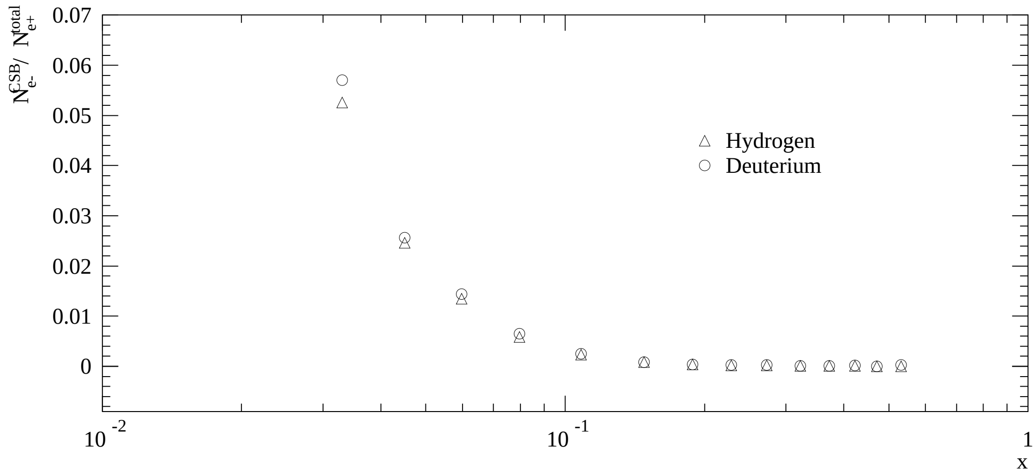

Figure 5.3: Fraction of events which have been subtracted from the DIS sample as originating from charge-symmetric background processes, for hydrogen and deuterium.

DIS positrons. However, since these background processes are charge-symmetric (as will be described below), there exists an indirect way to correct for these processes.

The numbers of DIS events in each x bin were corrected for misidentified positrons from charge-symmetric background processes: photoproduction ($\gamma \to e^-e^+$) can generate positrons that pass all the DIS cuts, especially in the regime of small values of x. Similarly, high energy Bremsstrahlung photons may convert into an e^-e^+ pair. Neutral pions, generated mainly in photoproduction processes, may decay via $\pi^0 \to \gamma\gamma$ with the subsequent conversion of one or both photons into an electron-positron pair $\gamma \to e^-e^+$ within the acceptance of the spectrometer. The beam positron, however, will be scattered under small angles in photoproduction processes and may escape detection in the spectrometer. In these cases, the positron from these processes might be misidentified as the scattered beam positron from the DIS process. On the other hand, both processes also are the dominating sources for detected *electrons*. These electrons can be considered as a tag on the background processes and are used for correcting the sample of identified DIS positrons under the assumption of equal detection efficiencies for both charges. This is accomplished as follows: in each event one searches for the leading lepton (i.e. the one with the highest momentum) that passes all DIS cuts (other than the requirement of positive charge). If this lepton is a positron, the event is considered as a normal DIS event. In the other case, when the lepton is an electron, one DIS positron event with identical kinematics as the electron event that was found, is subtracted from the DIS event sample. Using the notation $N_{e^+}^{\text{total}}(x)$ for the experimentally observed positrons that pas all the DIS cuts, and $N_{e^-}^{\text{CSB}}(x)$ for electrons that pass all the DIS cuts and tag the charge-symmetric background, the "background corrected" number of positron DIS events can be written as:

$$N_{e^+}^{\text{background corrected}}(x) = N_{e^+}^{\text{total}}(x) - N_{e^-}^{\text{CSB}}(x) \tag{5.7}$$

The ratio of the number of electrons to the number of positrons passing the DIS cuts (fig.

5.3) depends on x and decreases rapidly from about 6% for the lowest x bin to zero for the highest x bin. The size of the correction to the F_2-ratio turns out to be negligible (as can be seen from the values listed in table 5.4) since both hydrogen and deuterium show the same behaviour.

As mentioned earlier, the contamination of the electron/positron sample by misidentified hadrons and vice versa is very small and hence has not been corrected for. Apart from being very small, these contaminations are supposed to be very similar for deuterium and hydrogen and cancel out while calculating the F_2-ratio.

5.4.2 Corrections for detector smearing and radiative effects

Partial energy losses and local inefficiencies in detector components can cause a shift in the measured values of the kinematic variables as compared to the original values at the interaction point. This effect is called *detector smearing*. Another effect that can cause a shift in the event kinematics is e.g. the emission of a radiative photon by the initial beam positron. In this case, the initial energy of the incoming positron is not longer the beam energy E_{beam}, so the formula to calculate x ($x = Q^2/(2M(E_{beam} - E')))$ will not give the correct result. This is an example of a *radiative effect*. Although the origin of detector smearing and radiative effects is different and unrelated, both phenomena affect the kinematics of the event. Thus it seems natural to correct for both effects together without unfolding the reason why the kinematics of an event are altered. In the following sections detector smearing and radiative effects will be described separately in more detail and it will be explained how they can be corrected for together. These corrections are sizeable and must therefore be well understood. For that reason two independent correction methods will be applied and the outcome will be compared.

5.4.2.1 Detector smearing

Detector smearing (also called kinematical smearing) describes a systematic difference between measured and true kinematical quantities caused by effects of the apparatus. The main source of kinematical smearing is the interaction of scattered particles with components inside the spectrometer. In particular electrons and positrons lose parts of their energy due to multiple scattering, even in small amounts of material. A second source of systematic shifts of the measured quantities is introduced by the finite detector resolution, which is caused by calibration and alignment offsets of the tracking devices, as well as by the track reconstruction algorithm (HRC) itself.

Previous analyses [126, 127] have reported that the detector smearing is negligible at low x and only significant in the mid x ranges. In these reports the kinematical smearing was shown to be the same for all targets within 0.3%. Hence, one could argue that the effect of kinematical smearing can be neglected as the ratio F_2^d/F_2^p is taken (and include the smearing contribution into the systematic error on the structure function ratio). Nevertheless, in this work it has been decided to take the contribution from smearing into account, for two reasons: it happens at no cost with the method presented below and the influence of the RICH detector on smearing is not yet clear: the cited studies [126, 127] were performed when the RICH detector was not yet installed in HERMES and

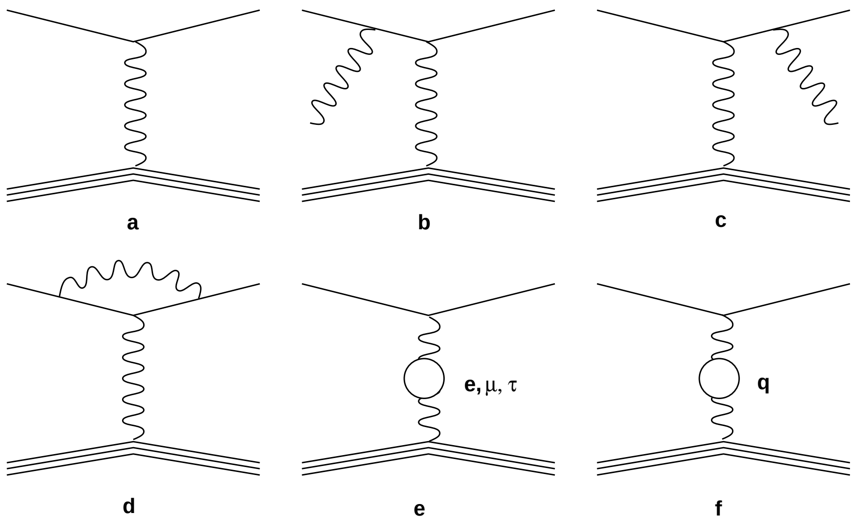

Figure 5.4: Feynman diagrams for radiative effects. The one-photon exchange term (Born term) is shown in (a). The other plots show the diagrams for (b) initial state Bremsstrahlung (i.e. Bremsstralhung radiated by the incoming beam positron), (c) the final state Bremsstrahlung (i.e. Bremsstralhung radiated by the scattered positron), (d) the vertex correction (i.e. photon exchange between the incoming and outgoing positron), and (e),(f) the vacuum polarisation (where the virtual photon fluctuates into a lepton anti-lepton or quark anti-quark pair).

tracking studies [128] have revealed that the aluminium entrance window of the RICH deteriorates the tracking resolution considerably. It is not yet resolved if this effect is target dependent.

5.4.2.2 Radiative effects

Before describing the radiative effects, some words will be spent on the three channels for lepton scattering on nuclei. Apart from the inelastic channel, two other basic channels for lepton scattering exist, namely elastic and quasi-elastic. In the case of elastic scattering, the beam positron is scattered off the nucleus, leaving it in its ground state; in the case of quasi-elastic scattering, there was a direct collision between the beam positron and an individual nucleon inside the nucleus. In the case of quasi-elastic (resp. elastic) scattering (and ignoring radiative effects for the moment), the invariant mass squared of the hadronic final state (W^2) must be equal to the mass of the nucleon (resp. the mass of the nucleus), yielding a value for x of 1 (resp. M_A/M).

Up to now, deep-inelastic scattering has been discussed in the Born approximation of one-photon exchange. But the measured events in the defined DIS region can also result from higher-order processes (fig. 5.4). Since the treatment of radiative effects is somehow contra-intuitive, as will be explained later, it is important to define the terminology rigorously: the "true" kinematics are the kinematics at the lepton-photon vertex, the "observed" kinematics are the kinematics that are affected by radiative effects. The higher-order processes introduce a deviation between the true and observed value of x, as can be understood from the definition of x ($x = 4EE'sin^2(\theta/2)/2M(E-E')$, where E' is the observed energy of the scattered positron and E is the beam energy); e.g. initial and final Bremsstrahlung corrupt the observed values of E or E'. Apart from this problem, the higher-order processes have an additional drawback since one cannot assure anymore that the lepton scattered inelastically: radiative effects alter the observed kinematics and may drive the elastic and quasi-elastic contributions into the defined DIS region. Thus several processes can contribute to the experimentally measured cross section:

$$\sigma_{\text{observed}} = \sigma_{\text{Born}} + \sigma^{\text{inelastic}}_{\text{non-Born}} + \sigma_{\text{quasi-elastic}} + \sigma_{\text{elastic}} \tag{5.8}$$

The cut on the kinematical variable $y > 0.85$ is just a phenomenological cut, only able to reduce the major part of contributions from radiative effects [131]. A simple example can be given by considering only the initial and final Bremsstrahlung. The variable y is measured as $1 - E'/E$. In the case of initial Bremsstralhung, taking the beam energy as the value for E is an overestimation, yielding an observed value for y that is too large compared to the true value. In the case of final Bremsstralhung, taking the observed energy of the scattered positron as the value for E' is an underestimation, again giving a measured value for y that is too large compared to the true value. So the observed values of y are shifted towards higher values compared to the true ones, and an upper cut on y eliminates the kinematical region where the radiative effects are dominant.

The method used to extract the pure Born contribution from σ_{observed} is based on a Monte Carlo program. The application of the radiative corrections is realised by the POLRAD [129, 130] program which integrates the applicable cross sections and contains the expression of the higher-order processes. In principle POLRAD works for polarised DIS, however here the program is used only to correct for the unpolarised kinematics. The Monte Carlo procedure for the generation of events with a possible photon radiation is as follows: the Monte Carlo DIS generator starts off with the generation of the *observed* kinematics (not the true kinematics!) of the scattered positron and the calculation of an event weight from these kinematics and the Born cross section (i.e. a Born-like positron with the same kinematics is considered to calculate the event weight). Then the appropriate scattering channel (non-radiative, elastic, quasi-elastic or inelastic radiative) has to be chosen according to their contribution to the total observed cross section. If a radiative channel is selected the radiated photon has to be generated and the values of the kinematical variables have to be recomputed in order to obtain the true values. For each event the weight has to be recalculated. The new weight is obtained by multiplying the original weight with the ratio of the radiatively corrected and the Born cross section (this factor will be called *sigcor* for convenience). For this new weight, we will use the notation *weight*. After this recalculation the *weight*ed sum of all events (generated originally in accordance with the Born cross section) gives the observed cross section.

The following terminology will be used in the next sections to describe the correction for radiative effects:

- x_{true}: the value of x as it was recalculated by POLRAD, i.e. the true value of x. This value only reflects the true kinematics for events that end up in the observable range. In the case of quasi-elastic scattering POLRAD assigns $x_{true} = 1$, in the case of elastic scattering $x_{true} = M_A/M$.

- x_{obs}: the value of x as it was observed, i.e. the originally generated value of x before POLRAD corrected it.
- x_{meas}: the value of x as it was measured. This definition is artificial and is introduced to include the detector smearing: the value of x_{obs} is changed by the smearing and experimentally one does not measure x_{obs} but x_{meas}.
- *weight*: the weight of an event, including the radiative effects
- *weight/sigcor*: the weight of an event, as it was originally defined during the generation stage on the Born level (the denominator *sigcor* cancels the factor for radiative effects).

From the definitions above, it is obvious that the Born cross section relates to a spectrum where x_{obs} (not x_{true}!) is used in combination with the event weight *weight/sigcor*. Alternatively, one could run only the DIS program with radiative effects turned off, in which case $x_{true} = x_{obs}$ and $sigcor = 1$.

5.4.2.3 The unsmearing matrix method

The goal of the unsmearing matrix method is twofold: it compensates for detector smearing and allows to obtain the Born part of the measured number of positrons in each x bin. The method relies heavily on the Monte Carlo simulation of the HERMES experiment.

To create the unsmearing matrix U, the complete chain GMC_dis (with radiative effects) - HMC - HRC (fig. 4.19) has been run for deuterium and hydrogen separately where about 10^6 events were generated for each target. As both x_{true} and x_{meas} are known in the simulation, the unsmearing matrix for each target can be written as:

$$U_{ij} = \frac{\text{number of positrons generated in } x_{true} \text{ bin } i \text{ and accepted in } x_{meas} \text{ bin } j}{\text{number of positrons accepted in } x_{meas} \text{ bin } j} \tag{5.9}$$

The numerator can be obtained by simply counting the number of DIS positrons with their generated x_{true} in x bin i and the finally measured value x_{meas} in x bin j. The denominator is calculated by adding up all the DIS positrons with a value of x_{meas} that belongs to x bin j. The question which weight must be applied to the Monte Carlo events and if the DIS cuts have to be applied to the measured or true values of the kinematical variables for the calculation of U, can simply be answered by considering the denominator: this is nothing else than a Monte Carlo simulation of the experimentally measured x spectra of DIS positrons. Thus it is obvious that, just like for experimental data, the DIS cuts have to be applied to the measured kinematics, and that one has to use *weight* (and not *weight/sigcor*) to construct the unsmearing matrix. An example of the unsmearing matrix for deuterium is shown in fig. 5.5.

As can be seen in fig. 5.5 the smearing and radiative effects mainly cause migration to smaller values of x. In general, the unsmearing matrix is not a square matrix: there are 14 bins that relate to x_{meas} and 16 bins that relate to x_{true}. The numbering scheme for the rows and columns has been chosen as follows: the values of x_{meas} are limited in the range $0.02 < x_{meas} < 0.70$ and this range is divided in 14 bins ($j = 1, 14$). A very small

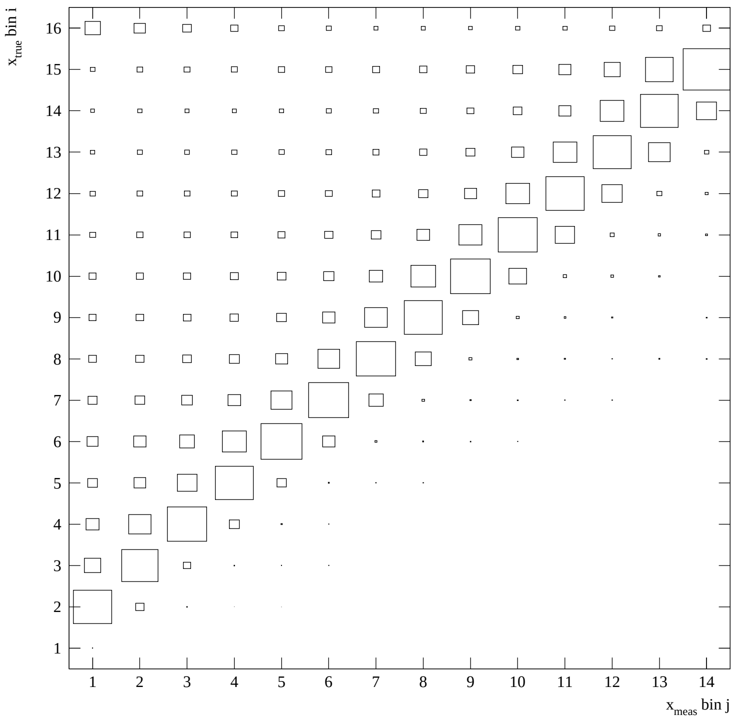

Figure 5.5: The unsmearing matrix U_{ij} for deuterium. The size of each box represents the magnitude of the matrix element.

number of events have a x_{true} below 0.02, these events are collected in the first row of U. The next 14 rows ($i = 2, 15$) correspond with the binning for x_{meas} ($j = 1, 14$). The last row ($i = 16$) collects all events which have a $x_{true} > 0.7$, these are predominantly the generated events belonging to the elastic or quasi-elastic channel.

The unsmearing matrix is applied to the background corrected experimental data (eq. (5.7)) by multiplication

$$[N_{e+}^{\text{unsmeared}}(x_{true})]_i = \sum_j [U(x_{true}, x_{meas})]_{ij} [N_{e+}^{\text{background corrected}}(x_{meas})]_j \tag{5.10}$$

yielding 'unsmeared distributions' as a function of x_{true}. The corrections due to the unsmearing are rather large, as illustrated in fig. 5.6 where the spectra $N_{e+}^{\text{background corrected}}(x_{meas})$ and $N_{e+}^{\text{unsmeared}}(x_{true})$ are compared. For this last spectrum the first bin ($i = 1$) and the last bin ($i = 16$) have been ignored and are not shown in fig. 5.6: the first bin contains very few statistics, the last bin has no physical meaning in the framework of deep-inelastic scattering since it contains the (quasi-)elastic events. The number of positrons ending up in this last bin is considerably large, which explains why both spectra do not contain the same integrated number of positrons.

As will be explained below, the obtained spectra $N_{e+}^{\text{unsmeared}}(x_{true})$ for hydrogen and deuterium are not yet the desired Born spectra, although they allow to calculate again the F_2-ratio. It should be emphasized that this intermediate result, shown in fig. 5.7 (the

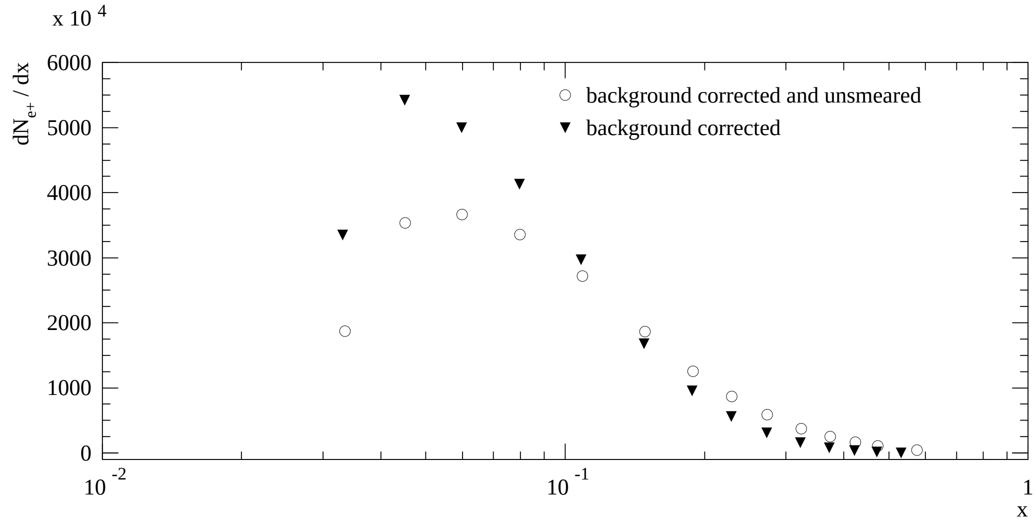

Figure 5.6: The number of DIS positrons, divided by the bin size, for deuterium. The spectrum where only a background correction was applied and the spectrum where the unfolding matrix was applied to the background corrected spectrum are compared.

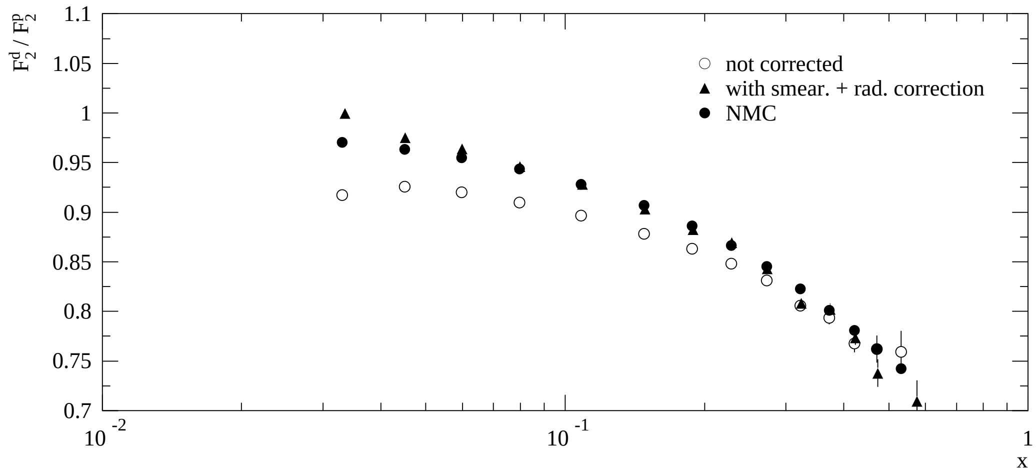

Figure 5.7: The x dependence of $F_2^d(x)/F_2^p(x)$ without any correction and the NMC parametrisation from fig. 5.2 are repeated and compared with the result after application of the unsmearing procedure (triangles). As explained in the text, these triangles present only an intermediate result.

numerical values are listed in table 5.4) has no physical meaning and only illustrates that the magnitude of the unsmearing corrections is considerably different for hydrogen and deuterium, as can be concluded by comparing fig. 5.7 and the results presented by the open dots in fig. 5.2. To facilitate the comparison, the uncorrected results and the NMC parametrisation from fig. 5.2 are repeated in fig. 5.7. The intermediate result after multiplication with U agrees better with the NMC parametrisation compared to the raw uncorrected F_2-ratio.

The application of the matrix U already yields the positron spectra as a function of x_{true}, however these spectra are not yet the desirable Born spectra. This can be understood by considering eq. (5.10): if one would apply the unsmearing matrix to the measured Monte Carlo positron spectra (instead of the measured experimental spectrum $N_{e+}^{\text{background corrected}}(x_{meas})$), eq. (5.9) shows that one would get the Monte Carlo spectrum for x_{true}, where the quantity *weight* is used as event weight (since both the unsmearing matrix and the mentioned measured Monte Carlo spectrum are calculated using *weight*). As explained above, this last spectrum has nothing to do with the Born spectrum. In order to obtain the Born spectra, a last (bin by bin) correction step must be applied to $N_{e+}^{\text{unsmeared}}(x_{true})$

$$[N_{e+}^{\text{Born}}(x_{obs})]_i = \frac{[N_{e+}^{\text{generated}}(x_{obs})]_i}{[N_{e+}^{\text{generated}}(x_{true})]_i} \cdot [N_{e+}^{\text{unsmeared}}(x_{true})]_i \tag{5.11}$$

Fig. 5.8 shows the correction factor

$$C_i^{\text{Born}} = \frac{[N_{e+}^{\text{generated}}(x_{obs})]_i}{[N_{e+}^{\text{generated}}(x_{true})]_i} \tag{5.12}$$

for deuterium. This factor can easily be understood:

- The denominator is calculated by considering the true kinematics of all reconstructed events (i.e. one looks at those GMC events that made it through HMC and HRC), hence the positron spectrum $N_{e+}^{\text{generated}}(x_{true})$ is constructed from the measured events, using x_{true} and *weight*. The kinematical cuts on $N_{e+}^{\text{generated}}(x_{true})$ are applied on x_{meas} because this was also the case for $N_{e+}^{\text{unsmeared}}(x_{true})$.
- For the numerator, one looks at all GMC events (also the ones that are lost in the detector, like HMC and HRC simulated). On the GMC level both x_{obs} and x_{true} are defined. The question which one has to be used can simply be answered by referring to the explanation above: one has to use x_{obs} and $weight/sigcor$, because this gives the Born contribution. The kinematical cuts on $N_{e+}^{\text{generated}}(x_{obs})$ are applied on the observed kinematics, because these kinematics are the relevant ones for the Born contribution.

After all these corrections, the Born distributions for hydrogen and deuterium are known and eq. (5.4) gives immediately the result for the F_2-ratio, as presented in fig. 5.9 (this is again an intermediate result, since acceptance corrections are not yet applied) (the numerical values are listed in table 5.4). In this figure, the F_2-ratio from the 2000 data set, obtained in this work, is compared to the results of an earlier analyis based on the 1996 data set, done by M. Vincter [125]. It is noted that the analyses are founded

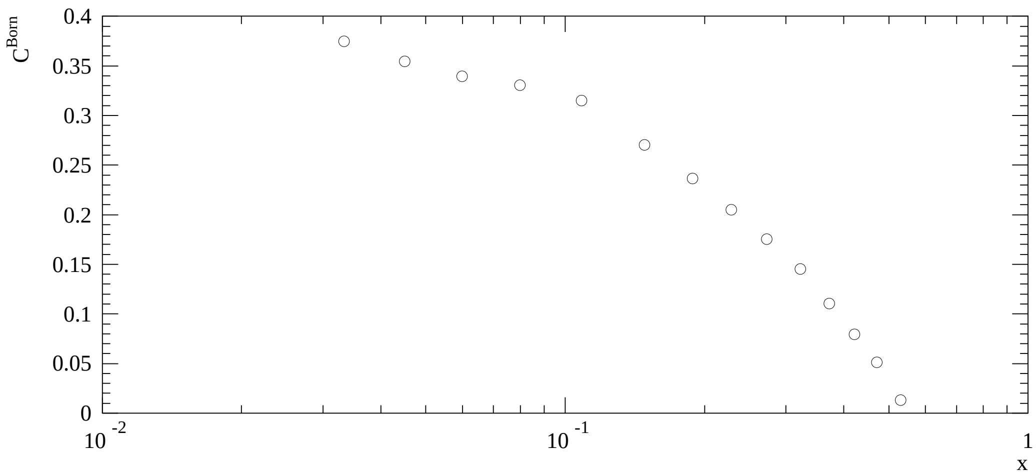

Figure 5.8: The x dependence of the correction factor C_i^{Born} for deuterium.

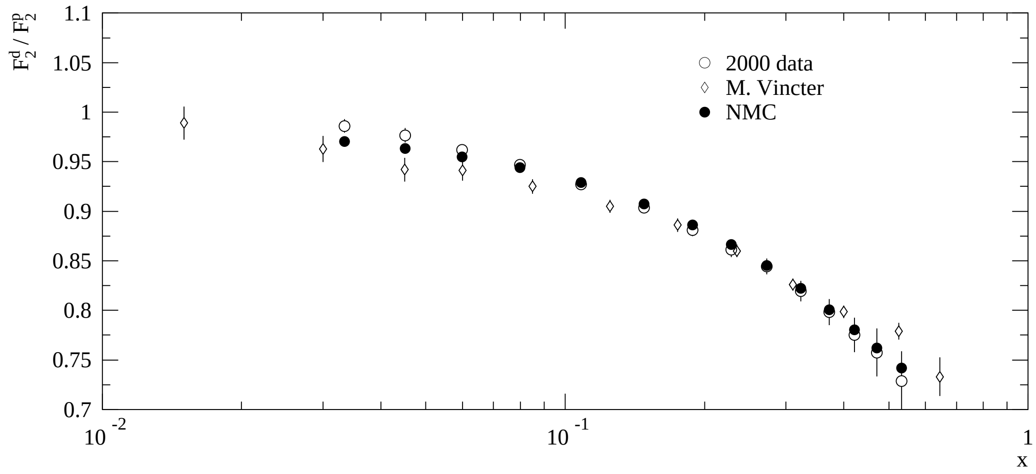

Figure 5.9: The x dependence of $F_2^d(x)/F_2^p(x)$ after correction for charge-symmetric background, detector smearing and radiative effects. The results from the 2000 data are compared with the results obtained by M. Vincter. Errors are statistically only. The parametrisation from NMC, calculated at the mean value of x and Q^2 of the experimental 2000 data is also shown.

on two different versions of the basic event reconstruction and have employed different data quality criteria and different kinematic requirements. In particular the W^2 and Q^2 requirements were different. In the 96 data used by M. Vincter, the reconstruction included the tracks recorded by the vertex chambers, whereas those were not used in this work based on the 2000 data, when the vertex chambers were not longer present in HERMES. The agreement between the two analyses is reasonable, especially if it is realised that the two approaches are rather different: M. Vincter did not correct for charge-symmetric background, the code to calculate the radiative effects was different (M. Vincter used the TERAD code [132]) and an iterative method was used to perform the correction, instead of the unsmearing matrix presented here.

5.4.2.4 Corrections for acceptance

The HERMES detector is a forward spectrometer with a limited angular acceptance that largely limits the topology of the DIS positron. Nevertheless, the geometrical acceptance is not expected to bias the F_2-ratio: the acceptance limits the range of the polar angle θ and the azimuthal angle ϕ that can be detected, but not directly x. Moreover, acceptance effects are expected to cancel out while calculating the F_2-ratio, since the yields for both targets are divided by each other. Nevertheless, some studies about acceptance will be presented in this section.

First of all, it should be mentioned that the Monte Carlo data discussed in the previous sections were 'generated in a box acceptance': to save CPU-time, GMC (with standard settings) feeds only those tracks to HMC that match the angular acceptance of the HERMES spectrometer. In other words: the numerator and denominator of the correction factor in eq. (5.12) were both generated in a box acceptance, thus the final corrected Born spectrum $N_{e^+}^{\text{Born}}(x_{obs})$ is only valid in this box acceptance. To bypass this problem, an independent GMC event generation was done in the 4π solid angle (changing the standard settings of GMC). This GMC was not interfaced to HMC and no radiative effects were calculated. Using the notations of section 5.4.2.2, it means that $x_{obs} \equiv x_{true}$ and $sigcor \equiv 1$ for this GMC production, which can be directly used to obtain the Born positron spectra in the 4π solid angle by simple modification of the numerator of eq. (5.11)[2]

$$[N_{e^+}^{\text{Born }(4\pi)}(x_{true})]_i = \frac{[N_{e^+}^{\text{generated }(4\pi)}(x_{true})]_i}{[N_{e^+}^{\text{generated (box)}}(x_{true})]_i} \cdot [N_{e^+}^{\text{unsmeared}}(x_{true})]_i \tag{5.13}$$

The correction factor from this last step,

$$(C_i^{\text{Born}})_{4\pi} = \frac{[N_{e^+}^{\text{generated }(4\pi)}(x_{true})]_i}{[N_{e^+}^{\text{generated (box)}}(x_{true})]_i} \tag{5.14}$$

is shown in the left panel of fig. 5.10 for deuterium. Comparison with fig. 5.8 shows that the correction factors for a box and 4π acceptance are remarkebly different, especially in

[2] The validitity of introducing $N_{e^+}^{\text{Born }(4\pi)}(x_{true})$ as numerator in equation (5.13) could easily be cross-checked. Therefore, artificial box cuts where applied to the GMC-4π Monte Carlo production, to mimic the geometrical box acceptance of the HERMES spectrometer. In other words: appropriate angular cuts have been applied to $N_{e^+}^{\text{Born }(4\pi)}(x_{true})$ a posteriori. It has been verified that the resulting Born correction factor in this case is exactly the same as shown in fig. 5.8.

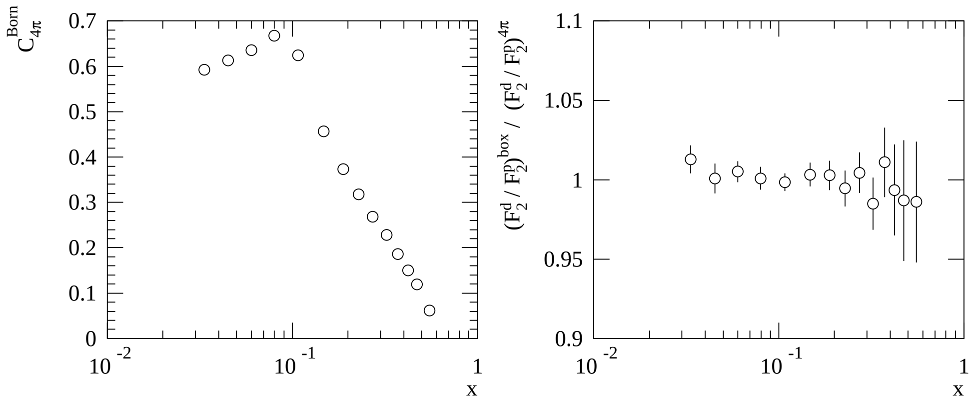

Figure 5.10: Left panel: the x dependence of the correction factor C_i^{Born} for deuterium, in the case of 4π solid angle corrections (eq. (5.13)). Right panel: The x dependence of the ratio of $(F_2^d(x)/F_2^p(x))^{\mathrm{box}}$ (corrected for charge-symmetric background, detector smearing and radiative effects) and $(F_2^d(x)/F_2^p(x))^{4\pi}$ (corrected for charge-symmetric background, detector smearing, radiative effects and acceptance).

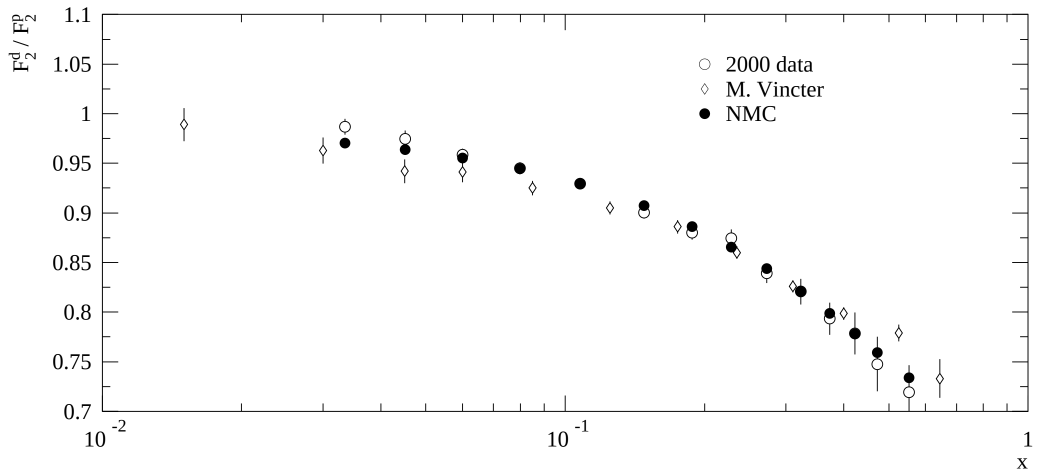

Figure 5.11: The x dependence of $F_2^d(x)/F_2^p(x)$ after correction for charge-symmetric background, detector smearing, radiative effects and detector acceptance. The results from the analysis of the 2000 data are compared with the results obtained by M. Vincter. Errors are statistically only. The parametrisation from NMC, calculated at the mean value of x and Q^2 of the experimental results from the 2000 data is also shown.

<x>	<Q^2>	F_2^d/F_2^p (a)	F_2^d/F_2^p (b)	F_2^d/F_2^p (c)	F_2^d/F_2^p (d)	F_2^d/F_2^p (e)	error
0.0335	1.194	0.917	0.913	0.999	0.986	0.987	0.008
0.0451	1.357	0.926	0.925	0.975	0.977	0.975	0.008
0.0600	1.517	0.920	0.919	0.964	0.962	0.959	0.006
0.0800	1.686	0.910	0.909	0.946	0.947	0.945	0.006
0.1077	2.000	0.896	0.896	0.928	0.927	0.929	0.005
0.1479	2.807	0.878	0.878	0.903	0.904	0.900	0.006
0.1880	3.678	0.863	0.863	0.883	0.881	0.880	0.007
0.2282	4.609	0.848	0.848	0.869	0.861	0.874	0.009
0.2725	5.710	0.831	0.831	0.843	0.844	0.839	0.010
0.3226	7.042	0.806	0.806	0.808	0.820	0.821	0.013
0.3725	8.539	0.793	0.793	0.802	0.798	0.793	0.016
0.4223	10.190	0.768	0.767	0.773	0.775	0.778	0.021
0.4722	12.095	0.762	0.762	0.738	0.758	0.748	0.028
0.5527	15.846	0.759	0.759	0.709	0.729	0.719	0.027

Table 5.4: Binning and numerical values for F_2^d/F_2^p. Column (a): the F_2-ratio from raw experimental data (fig. 5.2). Column (b): the F_2-ratio after correction for charge-symmetric background (CSB). Column (c): the F_2-ratio after correction for CSB and after application of the unsmearing matrix U (fig. 5.7). Column (d): the F_2-ratio after correction for CSB, detector smearing and radiative effects (box acceptance) (fig. 5.9). Column (e): the F_2-ratio after correction for CSB, detector smearing, radiative effects and geometrical acceptance (fig. 5.11). The errors in the last column are the combined statistical errors from experimental data and Monte Carlo data, propagated through the entire analysis code.

the low x region. After application of the acceptance correction, the F_2-ratio has been calculated again and $(F_2^d(x)/F_2^p(x))^{4\pi}$ is compared to the results without the 4π correction in the right panel of fig. 5.10. Although the Born correction factor is surely not the same in the case of a box and 4π acceptance, the acceptance correction has a negligible influence on the F_2-ratio, as expected a priori.

The final result for the F_2-ratio after all aforementioned corrections, including the acceptance correction is shown in fig. 5.11 and the numerical values are listed in table 5.4. The error bars represent the statistical errors from experimental and Monte Carlo data, which have been propagated through the complete analysis chain. A study of the systematic error will be presented in the next section. Although the NMC data was taken at much higher values of Q^2, the agreement between the results obtained in this work and the NMC parametrisation is surprisingly good.

5.4.2.5 Consistency check

As both the radiative correction calculation and the HERMES Monte Carlo generator (GMC) need a model for the structure function F_2 as input, the extraction method for

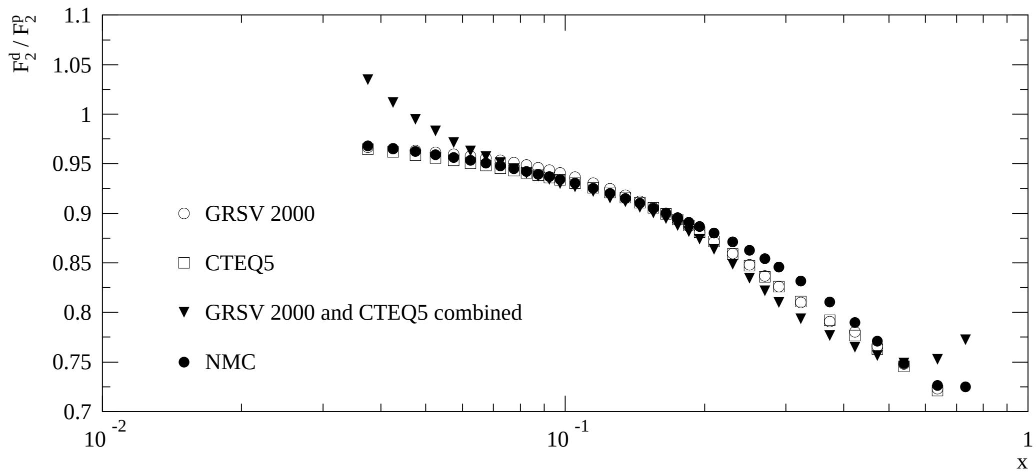

Figure 5.12: The x dependence of $F_2^d(x)/F_2^p(x)$ for different parametrisations, calculated at the mean value of x and Q^2 of HERMES. The triangles represent the ratio of F_2^d from the GRSV 2000 parametrisation and F_2^p from the CTEQ5 parametrisation.

the F_2-ratio could be biased. Also the unsmearing matrix U itself could be sensitive to the event generator. As a crosscheck, the complete analysis has been repeated where instead of the standard GMC parametrisation for F_2, different settings have been chosen: for deuterium, a polarised Monte Carlo production was done, using a model for F_2 that employs polarised quark distributions (GRSV 2000 [133]), while for hydrogen a Monte Carlo production with the unpolarised CTEQ5 [134] parton distributions as a model for F_2 was run. Although the F_2-ratio from both parton density sets agrees rather well with the NMC parametrisation, a significant deviation is observed if the F_2-ratio is calculated from GRSV 2000 for deuterium and CTEQ5 for hydrogen, as can be noticed in fig. 5.12. Combining these models is thus a rather severe crosscheck for the model independence of the complete analysis method.

The complete method to correct for detector smearing, radiative effects and detector acceptance has been repeated with the combination of CTEQ5 and GRSV 2000, yielding $(F_2^d/F_2^p)'$ which is compared in fig. 5.13 with the F_2-ratio (F_2^d/F_2^p) that was obtained with the standard GMC parametrisation as described in section 5.4.2.4 (fig. 5.11). There is no systematic difference between both results within the statistical errors, indicating that the input model for F_2 has a negligible influence on the correction procedure.

5.4.2.6 The modified Monte Carlo method

The unsmearing matrix method (UMM) is a complicated technique from the viewpoint of programming. Therefore another method, the so-called 'modified Monte Carlo method' (MMM) will be used as a crosscheck to perform the same corrections, but using a complete different approach. The method was originally proposed in [135]. The propagation of statistical errors through this new procedure is not straight-forward and will be omitted

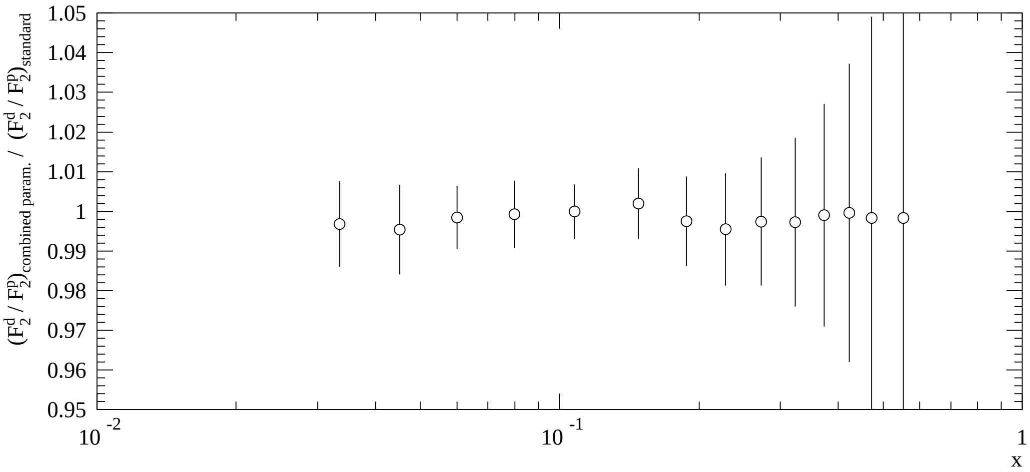

Figure 5.13: The x dependence of the ratio of $(F_2^d/F_2^p)'$ (where the corrections for detector smearing, radiative effects and detector acceptance were determined with the combination of GRSV 2000 as F_2^d parametrisation and CTEQ5 as F_2^p parametrisation) and (F_2^d/F_2^p) (where the standard GMC parametrisation was used for both F_2^d and F_2^p).

in this section because the MMM will only be used as a consistency check for the UMM. On the other hand, the MMM easily permits to extend the corrections to a second dimension: the corrections in the UMM were done only for x, while averaging over Q^2. As the results in this section will show, this is admitted since a 2-D correction procedure gives the same outcome as the 1-D one.

While for the UMM, one trusts HMC to create the unsmearing matrix U and trusts GMC to obtain the Born correction factors C_i^{Born}, the MMM does not rely on HMC but corrects for possible badly simulated detector behaviour by renormalising HMC to experimental data. Just like for the UMM, one has to rely on GMC to extract the final Born distributions.

The MMM relies on 3 different data sets, for both targets:

- The experimental data set P_{exp}.
- A Monte Carlo production which includes both radiative effects and detector tracking, i.e. the Monte Carlo data from the complete chain GMC_dis (with radiative effects) - HMC - HRC. This production will be quoted as P_{HRC}.
- A Monte Carlo production without radiative effects and detector tracking, i.e. a purely Born production (P_{Born}). As explained in the previous sections, several possibilities for P_{Born} exist: one could take the generated events from P_{HRC}, and use x_{obs} and $weight/sigcor$ to obtain P_{Born}, or one could simply run GMC_dis again with radiative effects turned off ($sigcor \equiv 1$), using $x_{obs} \equiv x_{true}$ to obtain P_{Born}. For the results presented in this section, the last option was chosen and P_{Born} events were generated in the 4π solid angle, allowing us to make an acceptance correction.

The MMM will be explained in the 2-D case. Several choices for the $x - Q^2$ binning were studied with this method, resulting in a negligible difference for the final F_2-ratio. Hence the method will be explained for the 14 bins in x that are listed in table 5.4 and 10 bins in Q^2. The Q^2 bins were chosen in such a way that for every x bin, the 10 Q^2 bins contain about the same number of DIS positrons. This binning was defined on the basis of P_{exp} and is sligthly different for deuterium and hydrogen. The different steps to perform the correction in the MMM scheme are, for both targets separately:

- Extract the inclusive DIS positron spectrum from experimental data, according to the bins defined above ($I_{exp}(x, Q^2)$).
- Extract the inclusive DIS positron spectrum from P_{HRC}. This goal of this spectrum is to mirror the experimental spectrum $I_{exp}(x, Q^2)$, hence the measured kinematics (x_{meas}, Q^2_{meas}) have to be used, along with *weight*. This measured Monte Carlo distribution for the DIS positrons is denoted as $I^{meas}_{HRC}(x, Q^2)$.
- Calculate the ratio of the spectra for each bin:

$$C^{meas}(x, Q^2) = \frac{I_{exp}(x, Q^2)}{I^{meas}_{HRC}(x, Q^2)} \tag{5.15}$$

  This ratio is nothing else than a measure for the quality of the Monte Carlo simulation of the DIS process, detector smearing and radiative effects in each (x, Q^2) bin.

- Process the DIS events from P_{HRC} one by one. For each event, find (x_{true}, Q^2_{true}) and (x_{meas}, Q^2_{meas}). Now execute two steps:
  1. Retrieve $C^{meas}(x, Q^2)$ according to the kinematics (x_{meas}, Q^2_{meas}). Replace the weight of each event by $weight \times C^{meas}(x, Q^2)$.
  2. Consider (x_{true}, Q^2_{true}) of the event and use these true kinematics in combination with the newly calculated weight to obtain another distribution $I^{true'}_{HRC}(x, Q^2)$.

  Intuitively, this last step is hard to interpret: as can be understood from section 5.4.2.2, using true kinematics in combination with *weight* does not have any physical meaning. The easiest way to explain what is going on, is considering what happens in the second step, if one would use (x_{meas}, Q^2_{meas}) instead of (x_{true}, Q^2_{true}) to create $I^{true'}_{HRC}(x, Q^2)$. In this imaginary case, we would have a spectrum that is weighted in such a way that the measured kinematics from P_{HRC} perfectly match the experimentally measured kinematics. In other words: we would mimic the behaviour of a P_{HRC} that perfectly simulates the experimental DIS process, detector smearing and radiative effects. But in reality, the distribution $I^{true'}_{HRC}(x, Q^2)$ was composed using (x_{true}, Q^2_{true}) in the second step. The considerations just mentioned help us to understand that $I^{true'}_{HRC}(x, Q^2)$ is thus a spectrum of true kinematics that would generate a spectrum of measured kinematics that perfectly matches the spectrum of experimentally observed (x, Q^2). The reason why the distribution $I^{true'}_{HRC}(x, Q^2)$ is needed, will become clear later on.

- Extract another inclusive positron spectrum from P_{HRC}. For each event, find (x_{true}, Q^2_{true}) and use *weight* to create the spectrum called $I^{true}_{HRC}(x, Q^2)$. Again, using true kinematics in combination with *weight* does not make any physical sense, but the next step will clarify this issue.

- Calculate the following ratio for each (x, Q^2) bin:

$$C^{true}(x, Q^2) = \frac{I^{true'}_{HRC}(x, Q^2)}{I^{true}_{HRC}(x, Q^2)} \tag{5.16}$$

  The numerator of this ratio is a spectrum of true kinematics that is corrected in such a way that it would generate a spectrum of measured kinematics that perfectly matches the spectrum of experimentally observed (x, Q^2). The denominator is a spectrum of true kinematics that is not corrected for anything. Hence the ratio expresses the correction factor that must be applied to true kinematics in order to reproduce the experimentally observed (x, Q^2) distribution.

- In the last step, the correction factor (5.16) is applied to the true kinematics from P_{Born}. The DIS events from P_{Born} are processed one by one. The final wanted 1-dimensional Born spectrum is created as follows: for each event, x_{true} is taken as abscissa for the positron DIS spectrum. The weight of each event in this spectrum is calculated as *weight* $\times\, C^{true}(x, Q^2)$, where the correction factor $C^{true}(x, Q^2)$ is retrieved according to (x_{true}, Q^2_{true}).

  It is obvious what is happening in this last step: suppose we would not replace *weight* by *weight* $\times\, C^{true}(x, Q^2)$, then we just would extract the Born distribution from P_{Born}. By applying the correction factor $C^{true}(x, Q^2)$, we correct this Born distribution so that it is made sure that the DIS process, the influence of detector smearing, radiative effects and acceptance in the Monte Carlo are modelled in such a way that the measured kinematics agree 100% with the experimentally measured (x, Q^2).

This complete procedure has been repeated for the experimental data on hydrogen and deuterium, resulting in corrected Born yields of DIS positrons. The corrected yields for both targets were used to calculate the F_2-ratio which is compared in fig. 5.14 with the previously presented result obtained with the unsmearing matrix method (fig. 5.11). The agreement between both methods is extremely good.

5.5 Conclusion

In this chapter, the ratio of the structure functions F_2 for deuterium and hydrogen was first extracted from raw experimental data. Comparison with the existing NMC parametrisation showed that sizeable corrections were required. These corrections include radiative effects, kinematical smearing and detector acceptance. After applying these corrections, a good agreement with the NMC parametrisation of F_2^d/F_2^p was obtained. The corrections were calculated with two completely independent methods, the unsmearing matrix method and the modified Monte Carlo method, with an excellent agreement. This gives

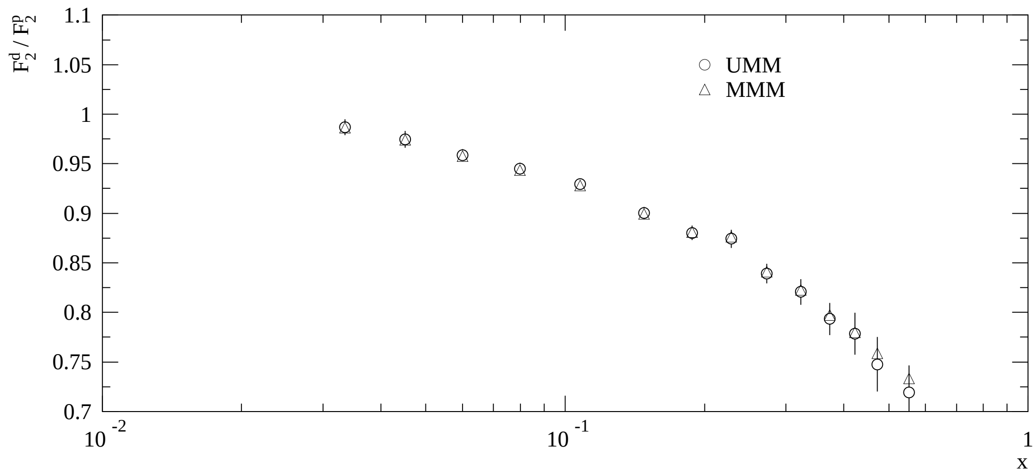

Figure 5.14: The x dependence of $F_2^d(x)/F_2^p(x)$ after correction for charge-symmetric background, detector smearing, radiative effects and detector acceptance. The results from the unsmearing matrix method (UMM) are compared with results from the modified Monte Carlo method (MMM). The error bars were calculated only in the case of the UMM.

confidence for a similar analysis that will be presented in the next chapters, where semi-inclusive data will be studied by considering hadrons in coincidence with the scattered positron.

In general, the structure function F_2 is defined by eq. (2.15) as the flavour sum of the x-weighted quark distributions within the nucleon. Since there are 2 different nucleons, the proton and the neutron, the interpretation of the ratio F_2^d/F_2^p becomes easier if the deuteron structure function is expressed as

$$F_2^d = \frac{1}{2}(F_2^p + F_2^n) \tag{5.17}$$

which allows to derive F_2^n/F_2^p from the cross section ratio obtained in the previous sections. The neutron to proton ratio is shown in fig. 5.15 (on a linear scale) along with a second order polynomial fit to the presented data. At low x $\rightarrow 0$ the DIS process scans the sea quarks in the nucleon. Since the difference between a proton and a neutron is only determined by the valence quarks, F_2^n/F_2^p is expected to be unity in this x-region. For high x $\rightarrow 1$ the contribution of the sea quarks becomes negligible and the valence quarks gain importance. Several predictions exist for F_2^n/F_2^p in this region. Exact SU(6) symmetry predicts $u_v^p(x) = 2d_v^p(x)$ and $d_v^n(x) = 2u_v^n(x)$. This leads to a value of $F_2^n/F_2^p = 2/3$ (completely determined by the charge of the quarks) for large values of x. DIS measurements indicate that SU(6) symmetry is broken (e.g. the u and d quark have a slightly different mass) and that the proton d_v-quark distribution is softer than the u_v-quark distribution. If u-quark dominance is assumed (i.e. only u-quarks contribute in the limit $x \approx 1$), the prediction is $F_2^n(x)/F_2^p(x) \rightarrow 1/4$ as $x \rightarrow 1$. Other assumptions lead to a prediction that the ratio of d_v/u_v at large x is non-zero. Models by Farrar and Jackson [136] and Brodsky et al. [137] predict that $u_v/d_v \rightarrow 5$ as $x \rightarrow 1$, and thus $F_2^n/F_2^p \rightarrow 3/7$. The initial analysis

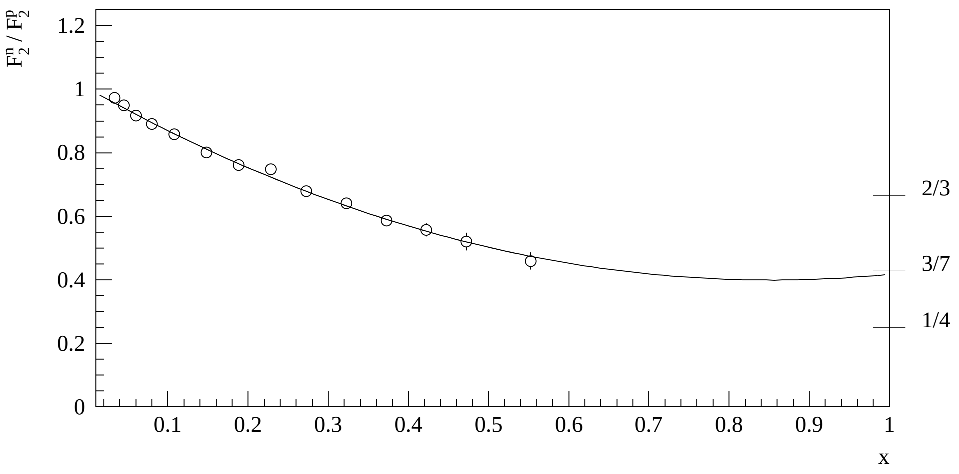

Figure 5.15: The x dependence of $F_2^n(x)/F_2^p(x)$. The solid line shows a second order polynomial fit to the data, extrapolated outside the measured region.

of SLAC and EMC experiments [138, 139] indicated a value close to 1/4 but the extraction of the neutron structure function did not include nuclear effects. A later re-analysis [140], including binding effects, yielded a value closer to the 3/7 prediction. The HERMES data presented in fig. 5.15 also favour a value of 3/7, although the insecure extrapolation does not allow to draw a firm conclusion, moreover no nuclear corrections have been applied here. The only clean way to measure F_2^n/F_2^p is by making use of a recoil dectector that is able to tag DIS events on a neutron in a deuterium target. A recoil detector identifies the recoil protons originating from the target. They can be detected in coincidence with the DIS lepton. A positive identification of a recoil proton in a DIS event indicates that the scattering process took place on the neutron in the deuteron and that F_2^n is being probed directly.

Chapter 6

Semi-inclusive DIS: d_v/u_v, Δs and p_t spectra of hadrons

In the previous chapter an inclusive analysis of the deep-inelastic scattering process was presented: the information about the scattered positrons only was sufficient to determine F_2^d/F_2^p. Inclusive DIS is sensitive only to the sum of the quark and anti-quark distribution functions because the virtual photon couples to the charge of the (anti-)quarks. In order to separate the flavour contributions from quarks and anti-quarks, hadrons must be detected in the final state in coincidence with the scattered positron.

For instance, the presence of a π^+ in the final state indicates that it is very likely that a u-quark was struck in the scattering because the π^+ is a $u\bar{d}$ bound state. Scattering from a u-quark is favoured both by its charge (2/3) and by the fact that the $\bar{d}$-quark is a sea quark and hence has a reduced probability of existing in the proton for the HERMES x-range. The technique of detecting hadrons in the final state to obtain information about specific quark and anti-quark distributions is called *flavour tagging*. Flavour tagging has an additional advantage: requiring at least one hadron in the final state assures that the event was purely DIS and not elastic or quasi-elastic. This simplifies the procedure to correct for radiative effects.

In this chapter several topics related to semi-inclusive DIS will be discussed: the ratio of the valence quark x distributions $d_v(x)/u_v(x)$, the polarisation of the strange quarks $\Delta s(x)$ and the transverse momentum distributions of hadrons. The d_v/u_v analysis relies on identified pions, the extraction of Δs employs identified kaons and the p_t distributions will be extracted for positively and negatively charged pions, kaons and protons separately. It is obvious that the HERMES RICH detector, described in chapter 4 will play a crucial role in these analyses.

As explained in the previous chapter, the extraction of physics quantities from raw detector data requires a good quality of the experimental data and appropriate cuts on kinematical variables. Furthermore, rather large corrections are needed to account for charge-symmetric background, detector smearing, radiative effects and some influences of the geometrical acceptance of the detector. In this chapter the same strategy as for the F_2-ratio analysis will be followed and only the differences or additional requirements will be enlightened; implicitly referring to chapter 5 if the same approach was used.

6.1 Data selection

The employed data set, the data selection criteria and the inclusive cuts on the scattered lepton used for the analyses in this chapter are the same as described in chapter 5. Three additional requirements have been used to constrain the semi-inclusive data sample: in order to select hadrons that predominantly originate from the current fragmentation region and to assure that the RICH determined the hadron type properly, cuts were made on the hadron kinematics:

- $0.2 < z < 0.8$. The upper limit on the fractional hadron energy ($z < 0.8$) discards any hadrons that originate from exclusive events and hence not belong to the semi-inclusive data sample studied here. Exclusive events are characterised by the fact that the complete energy of the virtual photon is transferred to one single hadron ($z \to 1$) so there is no other energy available to create additional particles. The lower cut $0.2 < z$ excludes hadrons that got only a small fraction of the virtual photon energy. These are mainly hadrons that are not created from fragmentation of the struck quark but from fragmentation of the target remnant.

- $0.1 < x_F$. Since for HERMES kinematics z and x_F are strongly correlated, the criterium on the Feynman variable has more or less the same effect as the requirement on z. Although this condition could be omitted, it is kept here for safety reasons and to guarantee compatibility with results from collider experiments, where $x_F > 0$ effectively selects the current fragmentation region.

- $2 < P_h < 15$ GeV. This condition on the hadron momentum enables the RICH detector to identify hadrons efficiently, as explained in section 4.3. Only 5% of the hadrons are lost with this requirement [72]. The lower momentum cut selects at the same time hadrons with a forward momentum that is not too small, helping to ensure they originate from the current fragmentation region.

Using these criteria, the correlation between the struck quark in the event and the detected hadrons is enhanced so the hadron sample will consist mostly of leading hadrons (i.e. hadrons containing the struck quark). Furthermore, Monte Carlo studies [141] have shown that the adopted cut of $W^2 > 10$ GeV^2 provides an optimum between increasing quark-hadron correlation and decreasing statistics since the separation between the target and current fragmentation region increases with the invariant mass squared.

As explained in section 4.9, any hadron that is identified by the RICH detector as either a pion, kaon or proton, is only identified correctly with a certain probability. Hence, the unfolding algorithm must be applied to extract the spectra of the true hadron types from those of the identified hadron types. In the case were the "spectra" are momenta spectra, the basic equation for the unfolding procedure was already shown in eq. (4.63)

$$\vec{N} = \mathsf{Q} \cdot \vec{I} \tag{6.1}$$

where the Q matrix and the vectors $\vec{N}$ and $\vec{I}$ are binned in momentum, charge and number of particle tracks per detector half (= track multiplicity). The Q matrix was determined by the iteration procedure presented in section 4.9.2. In this chapter the hadron spectra will be either x or p_t spectra. Therefore, the unfolding was done on the event level. For a given hadron track, the quantity $Q^i_t(P_h;$ multiplicity; charge) is added to the respective x (or p_t) bin for all true hadron types, instead of adding 1 for the most likely identified hadron type. For instance, in the case of x spectra:

$$N_\pi(x) \to N_\pi(x) + Q^i_\pi(P_h; multiplicity; charge) \quad (6.2)$$
$$N_K(x) \to N_K(x) + Q^i_K(P_h; multiplicity; charge) \quad (6.3)$$
$$N_p(x) \to N_p(x) + Q^i_p(P_h; multiplicity; charge) \quad (6.4)$$

This does not change the total number of particles and it automatically produces the correctly unfolded hadron x (or p_t) distributions.

The algorithm used for RICH identification was IRT. The RPS could not be used yet at the time this semi-inclusive analysis was done since the updated RPS P matrix (which is needed for the hadron unfolding) was not yet available. Both the iteration procedure and the method using the inverse P matrix method have been used for hadron unfolding, yielding negligible differences over the complete kinematical domain.

As a final remark, it has to be mentioned that the selected semi-inclusive sample contains some additional background from vector mesons decay. For example, in the case of pions the cross section for diffractive production $\gamma^* \to \rho^0 \to \pi^+\pi^-$ dominates largely over the cross section for DIS. Due to the limited detector resolution, a small fraction of the scattered beam positrons from such diffractive processes passes the kinematical cuts for DIS processes and the decay pions may mimic semi-inclusive DIS events. In [142] the effect of the background of vector meson decay on the semi-inclusive pion asymmetry defined by eq. (2.49) was studied, using Monte Carlo techniques. The corrections turned out to be small and the associated systematic error was larger than the correction itself. Furthermore, the HERMES Monte Carlo does not describe all vector mesons decay well. Hence, it was decided not to correct for contamination of vector meson decay in the DIS sample. Moreover, in e.g. the d_v/u_v analysis presented later in this chapter, *ratios* of positive and negative pions will be calculated. Since the decays produce a π^+ and π^- at the same time, the influence of this background is expected to cancel out finally.

6.2 The valence ratio $d_v(x)/u_v(x)$

6.2.1 Basic formalism

Using the cross section for semi-inclusive pion production (eq. (2.46)) together with the isospin symmetry between the parton density functions in the neutron and proton (eq. (2.35)-(2.37)), the positive and negative pion multiplicities on hydrogen and deuterium can be written as:

$$\frac{1}{N_p^{e^+}(x)} \cdot \frac{dN_p^{\pi^+}(x,z)}{dz}dz = \frac{x}{F_2^p(x)}[4u(x)D_u(z)^{\pi^+} + 4\bar{u}(x)D_{\bar{u}}^{\pi^+}(z) + d(x)D_d^{\pi^+}(z)$$
$$+\bar{d}(x)D_{\bar{d}}^{\pi^+}(z) + s(x)(D_s^{\pi^+}(z) + D_{\bar{s}}^{\pi^+}(z))]dz \quad (6.5)$$

$$\frac{1}{N_p^{e^+}(x)} \cdot \frac{dN_p^{\pi^-}(x,z)}{dz}dz = \frac{x}{F_2^p(x)}[4u(x)D_u^{\pi^-}(z) + 4\bar{u}(x)D_{\bar{u}}^{\pi^-}(z) + d(x)D_d^{\pi^-}(z)$$

$$+\bar{d}(x)D_{\bar{d}}^{\pi^-}(z) + s(x)(D_s^{\pi^-}(z) + D_{\bar{s}}^{\pi^-}(z))]dz \tag{6.6}$$

$$\begin{aligned}\frac{1}{N_d^{e^+}(x)} \cdot \frac{dN_d^{\pi^+}(x,z)}{dz}dz = \quad & \frac{x}{2F_2^d(x)}[4(u(x)+d(x))D_u^{\pi^+}(z) + 4(\bar{u}(x)+\bar{d}(x))D_{\bar{u}}^{\pi^+}(z) \\ & +(d(x)+u(x))D_d^{\pi^+}(z) + (\bar{d}(x)+\bar{u}(x))D_{\bar{d}}^{\pi^+}(z) \\ & +2s(x)(D_s^{\pi^+}(z) + D_{\bar{s}}^{\pi^+}(z))]dz \end{aligned} \tag{6.7}$$

$$\begin{aligned}\frac{1}{N_d^{e^+}(x)} \cdot \frac{dN_d^{\pi^-}(x,z)}{dz}dz = \quad & \frac{x}{2F_2^d(x)}[4(u(x)+d(x))D_u^{\pi^-}(z) + 4(\bar{u}(x)+\bar{d}(x))D_{\bar{u}}^{\pi^-}(z) \\ & +(d(x)+u(x))D_d^{\pi^-}(z) + (\bar{d}(x)+\bar{u}(x))D_{\bar{d}}^{\pi^-}(z) \\ & +2s(x)(D_s^{\pi^-}(z) + D_{\bar{s}}^{\pi^-}(z))]dz \end{aligned} \tag{6.8}$$

where $s = \bar{s}$ and factorisation have been assumed. Introducing the notations

$$\mathrm{N}^\pi = \int_{0.2}^{0.8} \frac{dN^\pi(z)}{dz}dz \tag{6.9}$$

for the z-integrated measured number of pions, a ratio R can be defined as

$$R = \frac{(\mathrm{N}_d^{\pi^+} - \mathrm{N}_d^{\pi^-})/(L^{\Delta t})_d}{(\mathrm{N}_p^{\pi^+} - \mathrm{N}_p^{\pi^-})/(L^{\Delta t})_p} - 1 \tag{6.10}$$

where the subscript p/d refers to the target. Integration of equations (6.5)-(6.8), in combination with eq. (5.4) and the following assumptions for the z-integrated fragmentation functions ($D \equiv \int_{0.2}^{0.8} D(z)dz$)

$$D_u^{\pi^+} + D_d^{\pi^+} - D_u^{\pi^-} - D_d^{\pi^-} = 0 \tag{6.11}$$

$$D_{\bar{u}}^{\pi^+} + D_{\bar{d}}^{\pi^+} - D_{\bar{u}}^{\pi^-} - D_{\bar{d}}^{\pi^-} = 0 \tag{6.12}$$

$$D_d^{\pi^+} + D_{\bar{d}}^{\pi^+} - D_d^{\pi^-} - D_{\bar{d}}^{\pi^-} = 0 \tag{6.13}$$

$$D_s^{\pi^+} + D_{\bar{s}}^{\pi^+} - D_s^{\pi^-} - D_{\bar{s}}^{\pi^-} = 0 \tag{6.14}$$

gives after some simple algebra the ratio of the d and u valence quark distributions as a function of the measured ratio R:

$$\frac{d_v(x)}{u_v(x)} = \frac{4R(x)+1}{4+R(x)} \tag{6.15}$$

Thus by measuring the number of positive and negative pions for the deuterium and hydrogen target and knowing the luminosity of the 2000 data set for these two targets, one can derive $d_v(x)/u_v(x)$. It has to be remarked that the knowledge of the F_2-ratio is not explicitly required for this analysis, as was previously assumed in other analyses [125, 131]. Of course, this statement is only correct if eq. (5.4) is valid for the experiment, as confirmed in the previous chapter. Another important remark is that the formula for d_v/u_v only holds if the symmetry relations (6.11)-(6.14) between the different fragmentation functions hold. It has to be stressed that these relationships are sufficient to derive the formula for d_v/u_v and that it is not necessary to require full charge conjugation symmetry and isospin invariance between the fragmentation functions:

$$D_u^{\pi^+} = D_d^{\pi^-} = D_{\bar{u}}^{\pi^-} = D_{\bar{d}}^{\pi^+} \equiv D^f \tag{6.16}$$

$$D_u^{\pi^-} = D_d^{\pi^+} = D_{\bar{u}}^{\pi^+} = D_{\bar{d}}^{\pi^-} \equiv D_L^{uf} \tag{6.17}$$

$$D_s^{\pi^+} = D_s^{\pi^-} = D_{\bar{s}}^{\pi^+} = D_{\bar{s}}^{\pi^-} \equiv D_s^{uf} \tag{6.18}$$

as was wrongly assumed in previous HERMES analyses [123, 125, 143, 131] or an older EMC analysis [144]. Thus the number of independent fragmentation functions was limited to three: one *favoured* fragmentation function D^f, and two *unfavoured* fragmentation functions, one for the light quarks (D_L^{uf}) and one for the strange quarks (D_s^{uf}).

While isospin symmetry is a natural assumption on the level of the quark scattering and on the level of the string breaking, the interference of these processes does not guarantee that the final overall fragmentation process is symmetric.

This can be studied by the HERMES Monte Carlo: a GMC production was done, in the 4π solid angle and without radiative effects (to avoid all possible complications) and the charged pion fragmentation functions were extracted from the struck quark information. As can be seen in fig. 6.1, the symmetry assumptions (6.16)-(6.18) are not valid at HERMES kinematics. Fortunately, the prerequisites for the derivation of the formula that expresses d_v/u_v as a function of R are fulfilled within the statistical errors of the Monte Carlo, as can be seen in fig. 6.2, where the x-dependence of the assumptions (6.11)-(6.14) is shown.

The observation that the fragmentation functions shown in fig. 6.1 exhibit a slight x dependence should not be interpreted as a breaking of the factorisation assumption, but as a consequence of correlations between kinematical variables: since the invariant mass squared of the hadronical final state can be written as $W^2 = M^2 + Q^2(1/x - 1)$, a smaller value of x accounts for a higher W^2 (the fact that x and Q^2 are also correlated does not cancel this behaviour). A higher value of W^2 means that there is more energy available for the fragmentation, thus the multiplicity is higher which leads (on average) to smaller fractional hadron energies z. As a crosscheck all charged pion fragmentation functions from MC data, integrated over z, as a function of x have been studied in different W^2 bins, yielding a flat behaviour in x [145]. This proves that the factorisation assumption is valid. The statistics in the experimental data sample do not allow for a W^2 binning, hence the complete range $W^2 > 10$ GeV^2 has been taken for the d_v/u_v analysis. Looking at fig. 6.1, it is obvious that all fragmentation functions show qualitatively the same behaviour in function of x. Since the calculation of the ratio R implies a division of fragmentation functions, the slight x dependence can be neglected. Furthermore, it has been pointed out that the real assumptions needed for the d_v/u_v analysis are the criteria shown in fig. 6.1 which show almost no x-dependence. Hence neglecting the slight x (or W^2) dependence is permitted.

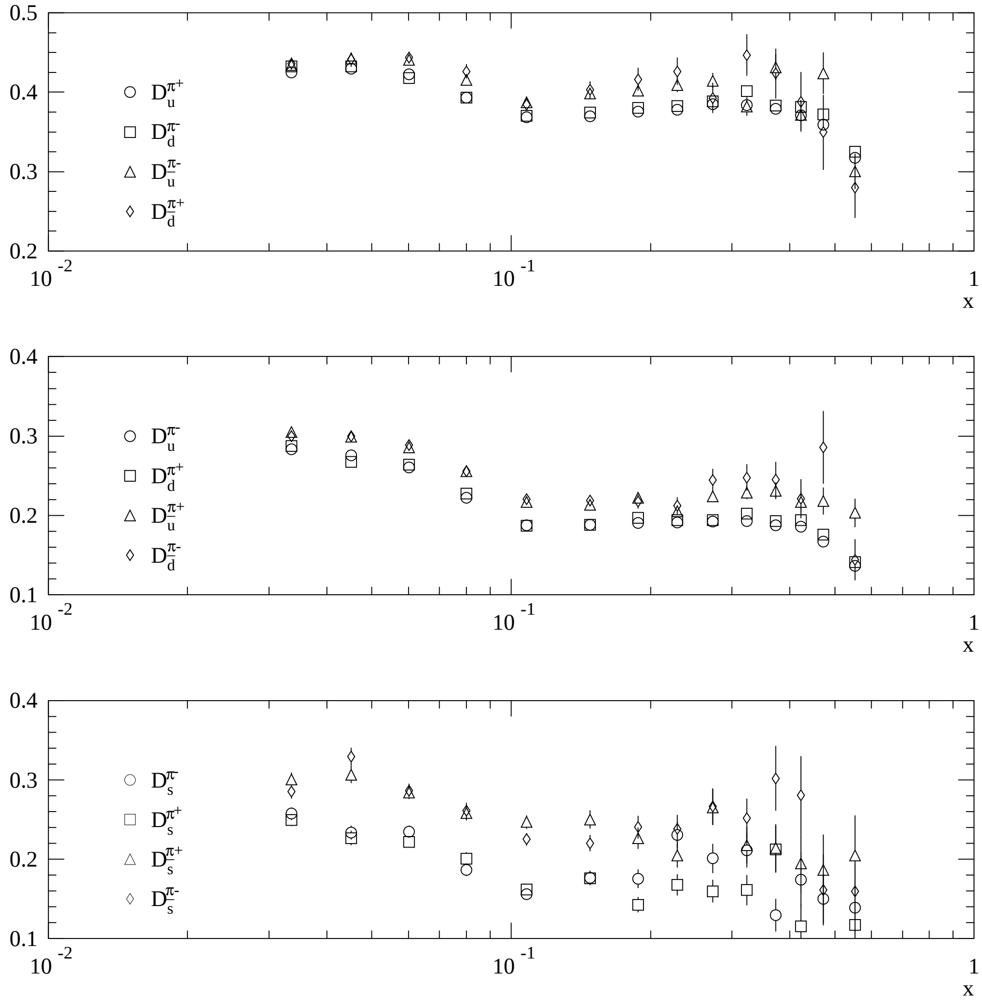

Figure 6.1: The favoured, unfavoured and strange pion fragmentation functions integrated over z from 0.2 to 0.8 as a function of x for Monte Carlo data. See text and equations (6.16)- (6.18) for more explanation.

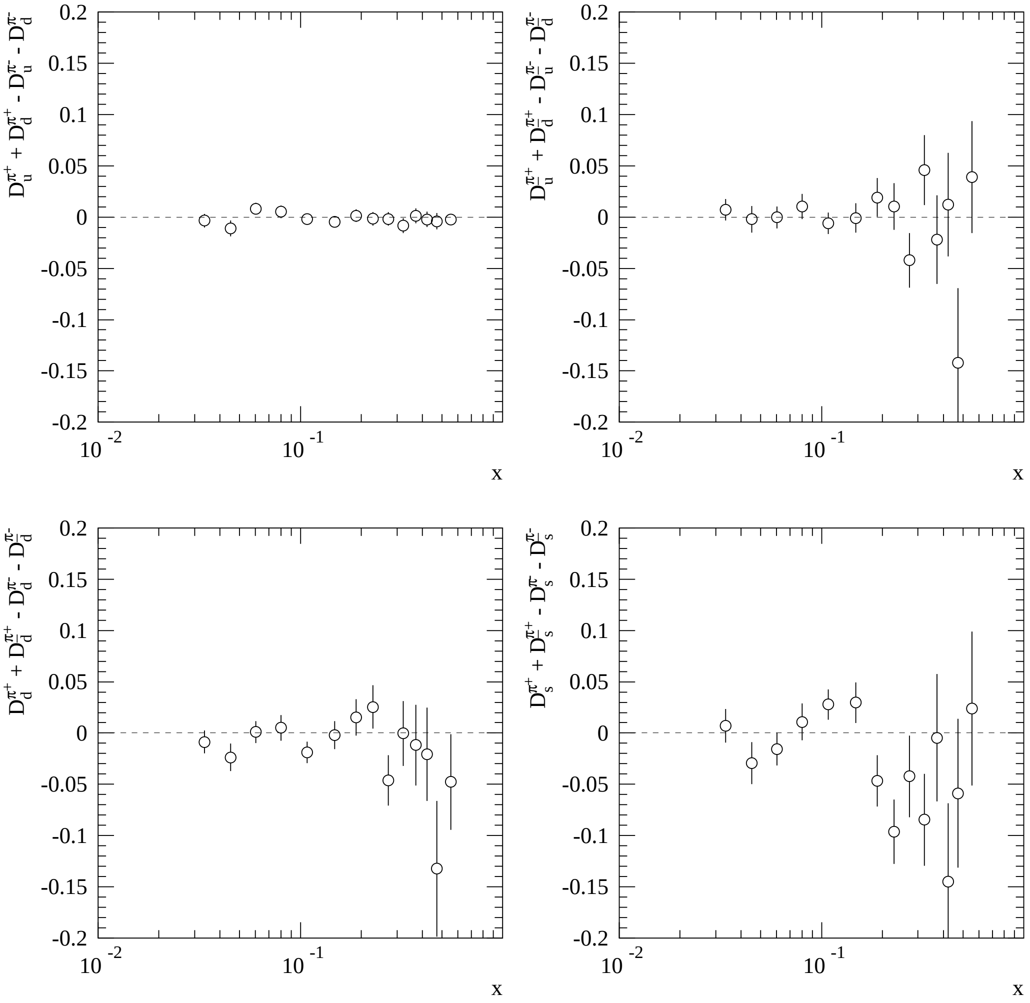

Figure 6.2: The x dependence of the assumptions (6.11)-(6.14). The fragmentation functions are integrated over z from 0.2 to 0.8.

6.2.2 Result for $d_v(x)/u_v(x)$

After correction of the π^+ and π^- yields for charge-symmetric background, detector smearing, radiative effects and acceptance, the ratio R can be obtained and d_v/u_v can be calculated. The result is shown in fig. 6.3 and is compared with two different parametrisations of the valence quark distributions (GRSV 2000 [133] and CTEQ5 [134]) in the upper panel and with other experimental results in the middle panel. The latter include an older HERMES analysis by M. Vincter, based on the 96 data [125], the d_v/u_v ratio from the European Muon Collaboration [144] where the physics probe was deep-inelastic muon-nucleon scattering, and the CERN West Area neutrino experiments WA 21/25 [146, 147]. Both the 1996 HERMES analysis and the EMC analysis used the same method as presented here, but applied an additional correction for the fact that the data sample contained all hadrons and was not a pure pion sample as in this analysis. Although the Q^2 range for HERMES is rather different compared to the kinematical range of EMC ($2 < Q^2 < 250$ GeV^2) and WA 21/25 ($1 < Q^2 < 100$ GeV^2), the data points agree well which confirms that the flavour tagging method works for the HERMES experimental conditions.

The rather large (statistical) error on the d_v/u_v-ratio is at first sight surprising since the 2000 HERMES data set was characterised by an excellent luminosity. Applying Gaussian error propagation shows that the major part of the error is due to the ratio R defined in eq. (6.10): the subtraction of the term -1 in the right hand side of this equation vanishes in the calculation of the derivatives while calculating the Gaussian error, blowing up the relative error σ_R/R as R is of the order of 0.2. One can easily verify that increasing the statistics with e.g. a factor of 4 would only reduce the statistical error by 50%. The large statistical errors are an inherent consequence of the applied method and are expected to overwhelm the systematic error which was not determined in this analysis.

In the high x-region where the sea quark contributions to the structure function ratio F_2^n/F_2^p are negligibly small, one can easily show that d_v/u_v can be approximated as

$$\frac{d_v}{u_v} \approx \frac{4\frac{F_2^n}{F_2^p} - 1}{4 - \frac{F_2^n}{F_2^p}} \tag{6.19}$$

Taking the fit from fig. 5.15 to calculate F_2^n/F_2^p, the bottom panel of fig. 6.3 shows that this approximation is valid in the region $x > 0.3$ where the valence quarks dominate. In the limit $x \to 0$ the ratio d_v/u_v tends to 0.5, indicating that *up* valence quarks have twice the number density compared to the *down* valence quarks, as expected in the quark parton model.

Finally, the ratio $d_v(x)/u_v(x)$ characterises the structure of the nucleon with respect to the dynamical distribution of the valence quarks. The good agreement between the results coming from deep-inelastic lepton-nucleon scattering, muon-nucleon scattering and the results from neutrino scattering shows that the charged and neutral currents are probing the same substructure of the nucleon.

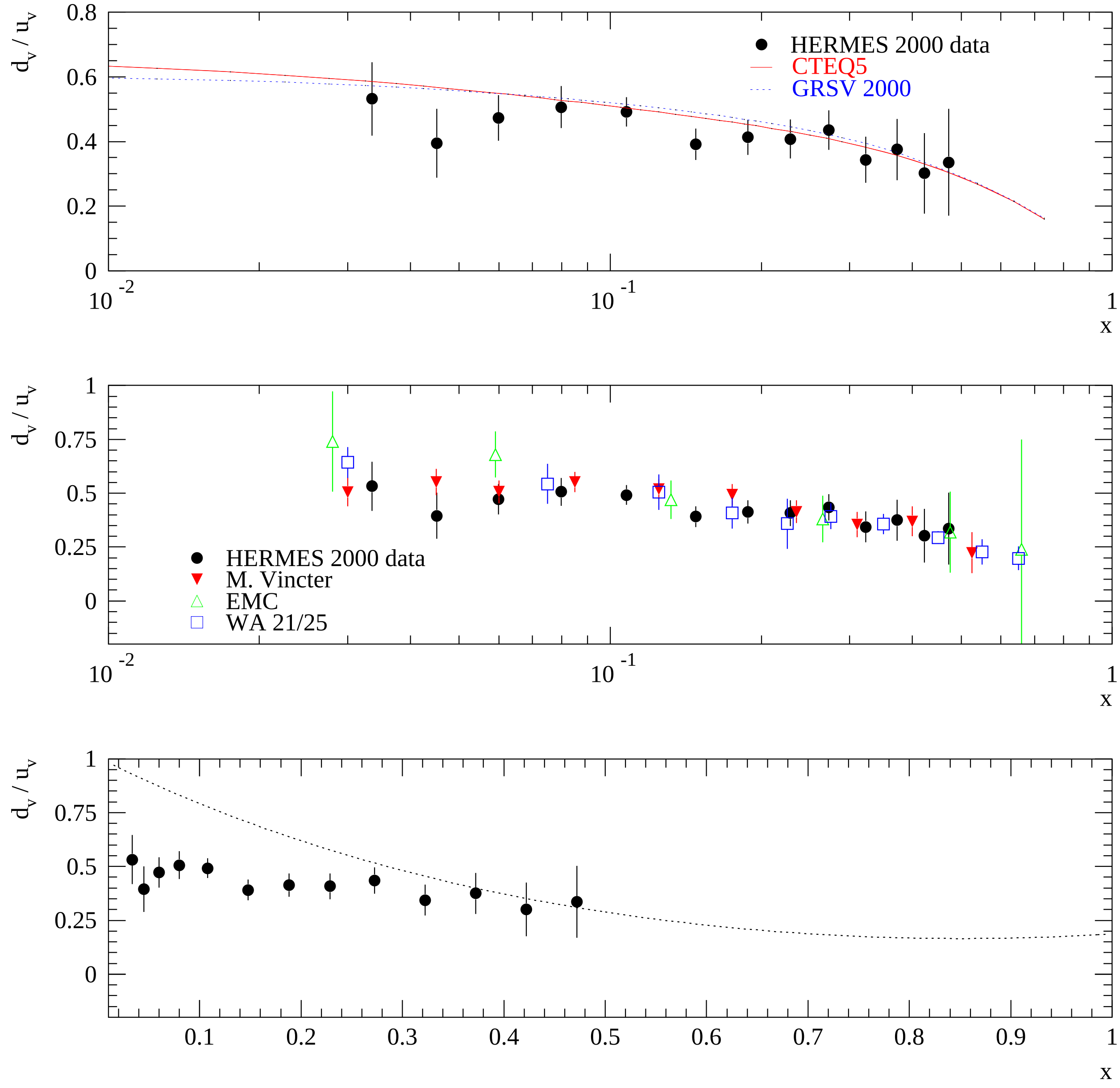

Figure 6.3: The x dependence of the valence quark distribution function ratio. The upper plot shows a comparison between this analysis and the CTEQ5 and GRSV 2000 parametrisations of the valence quark distributions. The middle panel compares the results from this analysis with an earlier HERMES analysis by M. Vincter, the EMC result and WA 21/25 data. The error bars of the 2000 data analysis are statistically only. The lower panel shows the comparison with $d_v(x)/u_v(x)$ calculated according to eq. (6.19), where F_2^n/F_2^p has been taken as the fit from fig. 5.15.

6.3 The strange quark polarisation $\Delta s(x)$

Whereas both the ratios F_2^d/F_2^p and d_v/u_v were examples of unpolarised deep-inelastic scattering, the determination of the strange quark contribution to the proton spin employs the unique feature of HERMES that both the beam and target are polarised. However, the below presented extraction method for Δs relies on *purities* that are again unpolarised quantities. These purities are closely related to the fragmentation functions for kaons, as will be laid out in the next sections. It will be shown that, analogous to the pions, the symmetry relationships between the different kaon fragmentation functions are broken, leading to the impossibility of extracting Δs without the use of any Monte Carlo simulation to generate the purities, as was originally the goal of this analysis.

6.3.1 Two component purity analysis

In section 2.4.5 it has been pointed out that HERMES recently extracted the up, down and strange quark polarisations from a five parameter fit, using inclusive and semi-inclusive asymmetries, as well as a set of purities that was determined from a Monte Carlo simulation. Especially the polarisation of the *strange* sea in the proton is a complete novel result, hence an independent crosscheck is desirable. For this reason H. Jackson presented [148] a simple method which requires only the inclusive asymmetry $A_d(x)$ and the semi-inclusive kaon charge sum asymmetry $A_d(x)^{K^++K^-}$ from the deuterium data. Using the purity formalism, only a two by two equation needs to be solved in this case:

$$\begin{bmatrix} A_d(x) \\ A_d^{K^++K^-}(x) \end{bmatrix} = \begin{bmatrix} P_{ns}(x) & P_{s+\bar{s}}(x) \\ P_{ns}^{K^++K^-}(x) & P_{s+\bar{s}}^{K^++K^-}(x) \end{bmatrix} \begin{bmatrix} \frac{\Delta q_s(x)}{q_s(x)} \\ \frac{\Delta q_{ns}(x)}{q_{ns}(x)} \end{bmatrix} \tag{6.20}$$

where the sum of the non-strange quarks is denoted by $q_{ns} = u + \bar{u} + d + \bar{d}$ and the sum of strange and anti-strange quarks by $q_s = s + \bar{s}$. Since the rows are normalised, only two elements of the purity matrix are independent and the knowledge of the purity elements for the non-strange quarks P_{ns} and $P_{ns}^{K^++K^-}$ is sufficient to fully determine the purity matrix. The two relevant purities can be calculated as

$$P_{ns}(x) = \frac{5q_{ns}(x)}{5q_{ns}(x) + 2q_s(x)} \tag{6.21}$$

$$P_{ns}^{K^++K^-}(x) = \frac{q_{ns}(x)\left[4\int D_u^{K^++K^-}(z)dz + \int D_d^{K^++K^-}(z)dz\right]}{q_{ns}(x)\left[4\int D_u^{K^++K^-}(z)dz + \int D_d^{K^++K^-}(z)dz\right] + 2q_s(x)\int D_s^{K^++K^-}(z)dz} \tag{6.22}$$

where isoscalarity and the following symmetry relationships between the kaon fragmentation functions have been used:

$$D_u^{K^++K^-} = D_{\bar{u}}^{K^++K^-} \tag{6.23}$$

$$D_d^{K^++K^-} = D_{\bar{d}}^{K^++K^-} \tag{6.24}$$

$$D_s^{K^++K^-} = D_{\bar{s}}^{K^++K^-} \tag{6.25}$$

There exist several methods to determine the required non-strange purities (6.21) and (6.22). One could use a Monte Carlo method for the fragmentation process, as was done in the recently completed HERMES analysis that determined all the quark polarisations from a five parameter fit. Since the goal of the two component method is to verify the latter analysis, this is not desired. Alternatively, one could use the existing parametrisations for the unpolarised quark distributions $q(x)$ and calculate $P_{ns}(x)$ and $P_{ns}^{K^++K^-}$ directly from (6.21) and (6.22), if the fragmentation functions are known. Based on the data from e^+e^- collider experiments, the charge summed kaon fragmentation functions for non-strange quarks were parametrised [149] as a function of z and Q^2. However, there are two problems associated with the use of these parametrisations. First, the asymmetries that were measured at HERMES are not corrected for acceptance effects, while the parametrised fragmentation functions are only valid in the 4π solid angle. This is a minor problem, since the effects of acceptance are shown to be small while measuring asymmetries at HERMES: since their calculation involves a ratio, most acceptance effects cancel. A more important issue is that the light quarks $q = u, d, s$ were assumed to be massless in the analyses of the e^+e^- collider data. This is justified in view of the high center of mass energies available in these experiments. However, this approximation is not valid for the lower energies available at fixed target experiments. Here fragmentation of non-strange quarks into kaons is suppressed because of the extra mass which must be generated in forming a $s\bar{s}$ pair at the string break. The lower probability to generate a $s\bar{s}$ pair compared to the lighter quark anti-quark pairs is accounted for by an extra *strangeness suppression factor* $\lambda_s < 1$ that connects the fragmentation functions needed for the purity $P_{ns}^{K^++K^-}$

$$D_s^{K^++K^-} \equiv \lambda_s D_u^{K^++K^-} \tag{6.26}$$

The value for λ_s that is normally assumed is 0.3 [150] but this factor could be strongly dependent on the total energy available in the scattering, and at HERMES kinematics the suppression can be even stronger. Indeed, the tuning of HERMES Monte Carlo to the observed hadron yields has suggested that $\lambda_s = 0.16$.

Since these parametrised fragmentation functions are not adapted to the HERMES acceptance and since their use requires the knowledge of the strangeness suppression factor λ_s, it would be very advantageous if one could extract the kaon fragmentation functions from experimental data, since this would allow to determine Δs immediately from the measured asymmetries in the HERMES acceptance, without any model dependence.

6.3.2 Negative fragmentation functions

Since the asymmetries are known in the HERMES acceptance, a consistent solution of eq. (6.20) requires the purities and thus the fragmentation functions to be known in the same acceptance. As will be shown later in this section, the attempt made to determine the fragmentation functions yielded unphysical negative values. Detailed studies revealed that this behaviour is not due to the acceptance. Therefore, to avoid complications due to the acceptance while formulating the algorithm to determine kaon fragmentation functions, a 4π acceptance will be assumed.

Assuming charge conjugation allows to write the following symmetry relationships between the different kaon fragmentation functions (where the superscripts f and uf denote the favoured and unfavoured fragmentation functions respectively).

$$D_u^{K^+} = D_{\bar{u}}^{K^-} \equiv D_u^f \tag{6.27}$$

$$D_d^{K^+} = D_{\bar{d}}^{K^+} = D_d^{K^-} = D_{\bar{d}}^{K^-} = D_u^{K^-} = D_{\bar{u}}^{K^+} \equiv D_L^{uf} \tag{6.28}$$

$$D_s^{K^+} = D_{\bar{s}}^{K^-} \equiv D_s^{uf} \tag{6.29}$$

$$D_s^{K^-} = D_{\bar{s}}^{K^+} \equiv D_s^f \tag{6.30}$$

The fragmentation functions needed for the purity $P_{ns}^{K^++K^-}$ can then be written as a linear combination of the single kaon fragmentation functions:

$$D_u^{K^++K^-} = D_u^{K^+} + D_u^{K^-} = D_u^f + D_L^{uf} \tag{6.31}$$

$$D_d^{K^++K^-} = D_d^{K^+} + D_d^{K^-} = 2D_L^{uf} \tag{6.32}$$

$$D_s^{K^++K^-} = D_s^{K^+} + D_s^{K^-} = D_s^{uf} + D_s^f \equiv \bar{D}_s \tag{6.33}$$

As for the pions, we can use the cross section for semi-inclusive kaon production (eq. (2.46)) together with the isospin symmetry between the parton density functions in the neutron and proton (eq. (2.35)-(2.37)), to write the positive and negative kaon multiplicities on hydrogen and deuterium in each x bin as:

$$\left(\frac{N^{K^++K^-}}{L^{\Delta t}}\right)_p = C \cdot [4(u+\bar{u})D_u^f + 2(2u+2\bar{u}+d+\bar{d})D_L^{uf} + 2s\bar{D}_s] \tag{6.34}$$

$$\left(\frac{N^{K^++K^-}}{L^{\Delta t}}\right)_d = C \cdot [4(u+d+\bar{u}+\bar{d})D_u^f + 6(u+d+\bar{u}+\bar{d})D_L^{uf} + (4s)\bar{D}_s] \tag{6.35}$$

$$\left(\frac{N^{K^-}}{L^{\Delta t}}\right)_d = C \cdot [(4\bar{u}+4\bar{d})D_u^f + (5u+5d+\bar{u}+\bar{d})D_L^{uf} + 2s\bar{D}_s] \tag{6.36}$$

where again $s = \bar{s}$ and factorisation have been assumed. In the above equations, all quantities are integrated over z from 0.2 to 0.8 and only depend on x. The constant $C \equiv C(x)$ in each x bin is defined as $C = (x \cdot N_{e^+}^p)/(F_2^p \cdot (L^{\Delta t})^p)$, and eq. (5.4) has been used to derive the above three equalities. The constant C is just a normalisation factor and will cancel while calculating the purities or ratios of fragmentation functions from these three equations with three unknowns.

The choice of these three equations is based on the following arguments: measuring K^+ and K^- simultaniously maximises the statistics for the first two equations, compared to measuring only one of the charged kaons. Picking K^- on the deuteron as the third reaction channel enhances the sensitivity to the strange quark fragmentation since the negative kaon ($\bar{u}s$) is an all sea object, weakening the fragmentation from the u-quark. The positive kaon ($u\bar{s}$) gives less sensitivity to the strange quark fragmentation since this channel has a strong favoured contribution from u-quark fragmentation.

For this analysis only experimental kaons with momentum $4.0 < p < 13.8$ GeV have been used. The reason is twofold: the available asymmetries were measured in this range and secondly, the kaon identification with the RICH is the most efficient in this momentum region as can be seen in fig. 4.31. After correcting the experimental kaon data for RICH inefficiencies, charge-symmetric background, detector smearing, radiative effects and acceptance effects, eq. (6.34)-(6.36) could be used to extract ratios of fragmentation functions, as shown in fig. 6.4 where the unpolarised quark distributions needed to solve the system of linear equations have been taken from the CTEQ5 parametrisation. Obviously, this result does not make any sense: the overall fluctuations are large and in some x-bins the fragmentation functions are negative and the error bars are as large as the values of the ratios. The large error bars are not due to limited statistics, but to the nature of the equations (6.34)-(6.36) that are highly unstable: studies indicated that the introduction of minimal variations in the kaon yields gives rise to outrageous variations in the ratios of fragmentation functions.

The reason for this behaviour can, like for the pions, be examined by the HERMES Monte Carlo: a GMC production was done in the 4π solid angle and without radiative effects (to avoid all possible complications) and the charged kaon fragmentation functions were extracted from the struck quark information. As can be seen in fig. 6.5, the Monte Carlo simulation indicates that the symmetry assumptions (6.27)-(6.30) are not valid at HERMES kinematics, especially in the case of D_L^{uf}. Hence the proposed method to determine the fragmentation functions and thus the purities from experimental data is not valid, if the Monte Carlo simulation predicts the kaon fragmentation correctly.

As a fall-back solution, one could try to determine the purity $P_{ns}^{K^++K^-}$ from the earlier mentioned parametrisations of the kaon fragmentation functions based on the data from e^+e^- collider experiments. Of course this is only valid if the assumptions (6.23)-(6.25) hold. The Monte Carlo simulation indicates that these relationships for the light quarks are more or less valid, but strongly violated for the strange quarks, as illustrated in fig. 6.6. Moreover, the introduction of the strangeness suppression factor as defined in eq. (6.26) seems not be adequate as the ratio $D_u^{K^++K^-}/D_s^{K^++K^-}$ is clearly not constant. Therefore, the attempt to determine Δs from an independent method to crosscheck the result from the five parameter fit did not succeed. On the other hand, the fact that naturally assumed symmetry relationships are violated, can be seen as a source of new information and needs further exploration. A number of tests has already been done, not one of them able to explain the observed asymmetries in the kaon fragmentation functions presented in fig. 6.5. Several options and parameters in the Lund Monte Carlo simulation program were changed:

- Releasing or tightening the cuts on z and x_F has a minor influence.
- An artificial requirement that the kaons have their origin in the struck quark and not in target fragments resulting from string breaks had no qualitative impact.
- Turning off the switches that control photon-gluon fusion, hard gluon radiation or soft colour interactions did not make any difference.
- A recalculation of the kaon fragmentation functions for only those DIS events with no K_L^0 in the final state or K_s^0 in a decaying intermediate state has an impact on the x-dependence of the fragmentation functions, but does not yield the expected symmetry relations.

Further communications with the authors of the Lund model may shed more light on the broken symmetry relations.

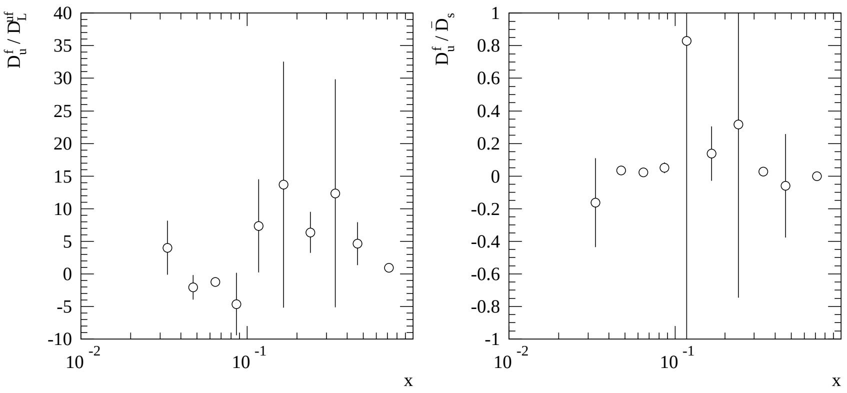

Figure 6.4: The x dependence of the ratio of the kaon fragmentation functions D_u^f/D_L^{uf} and $D_u^f/\bar{D}_s$ as determined from equations (6.34)-(6.36). The fragmentation functions are integrated over z from 0.2 to 0.8.

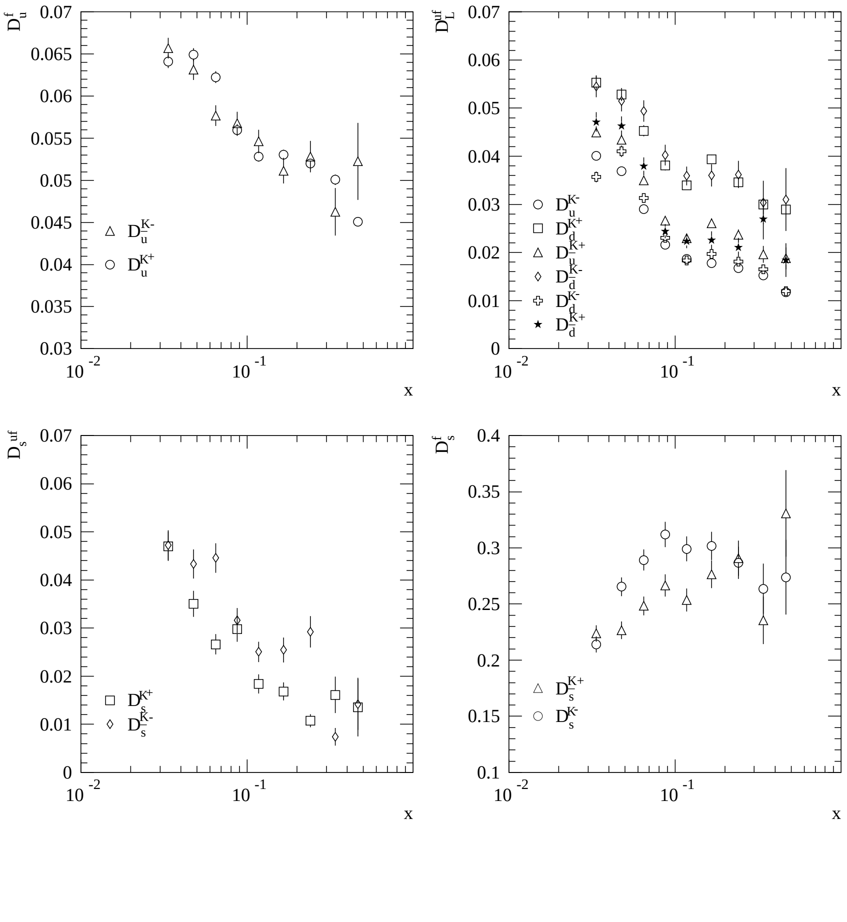

Figure 6.5: The favoured and unfavoured kaon fragmentation functions for the light and strange quarks, integrated over z from 0.2 to 0.8, as a function of x for Monte Carlo data. See text and equations (6.27)-(6.30) for more explanation.

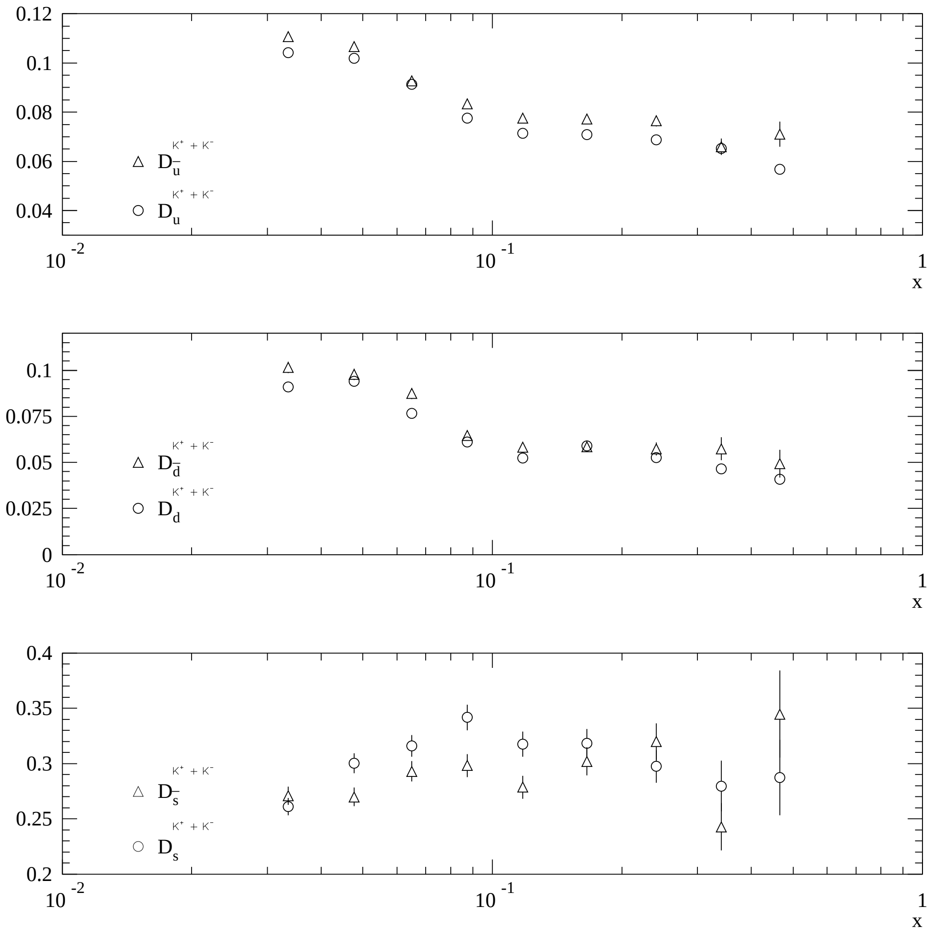

Figure 6.6: Fragmentation functions for the charge summed kaons from Monte Carlo data, integrated over z from 0.2 to 0.8, as a function of x. See text and equations (6.23)-(6.25) for more explanation.

6.4 Transverse momentum spectra of pions, kaons and protons

As pointed out in section 2.3.2, the naive quark parton model (formulated in the Breit reference frame) only explains longitudinal fragmentation and does not leave any room for transverse momentum of the quarks. In the Lund model however, quarks have intrinsic transverse momentum that is transferred to the hadrons during the fragmentation process. Moreover, the process of (gluon and quark) fragmentation itself can also contribute to the transverse momentum of the observed hadrons and other QCD processes as photon-gluon fusion and gluon emission can also contribute. Hence, an experimental study of transverse momentum distributions of semi-inclusively produced hadrons can provide insight in quark fragmentation and associated QCD-effects in the framework of the Lund fragmentation model. In this section the transverse momentum distributions of pions, kaons and protons will be measured and fitted with a phenomenological function.

6.4.1 Experimental and Monte Carlo spectra

The transverse momentum P_h^T of a certain hadron h (also denoted as p_t) is defined in eq. (2.11) and illustrated in fig. 2.2. The transverse momentum spectrum or multiplicity is defined as:

$$\frac{1}{N_e}\frac{dN_h}{dp_t^2} \qquad (h = \pi^+, K^+, p, \pi^-, K^-, p^-) \tag{6.37}$$

where N_e is the total number of positrons satisfying all DIS cuts, dp_t^2 is the width of each p_t^2 bin and dN_h is the number of hadrons h in the concerning p_t^2 bin. In this analysis 14 p_t^2 bins were used, ranging from 0 to 1.4 GeV^2.

The p_t^2-spectra from (uncorrected) experimental data and Monte Carlo data (HRC, i.e. fully tracked) are shown in fig. 6.7. The (dis)agreement between the simulation and the experimental data is related to the settings for the parameters controlling the fragmentation in the Lund model: during the tuning process of these parameters [31] it was only required that the p_t^2-multiplicities and $< p_t >$ for *all* hadrons from the Monte Carlo match those from experimental data. No distinction was made between charges and different types of hadrons. Since the pions dominate the semi-inclusive sample, the agreement between experimental and simulated pions in fig. 6.7 is to be expected, while for the kaons and protons, the (dis)agreement is due to coincidence. The rather large disagreement for K^+ and protons should not be interpreted as a problem because the Monte Carlo simulation will only be used to correct the raw experimental p_t^2-spectra for kinematical smearing, radiative effects and detector acceptance; the independency of the correction method with respect to the quality of the simulation has been proven in section 5.4.2.5. These corrections, using the unsmearing matrix method (cfr. section 5.4.2.3) will be described in the next section and multiplicities on the Born level will be presented.

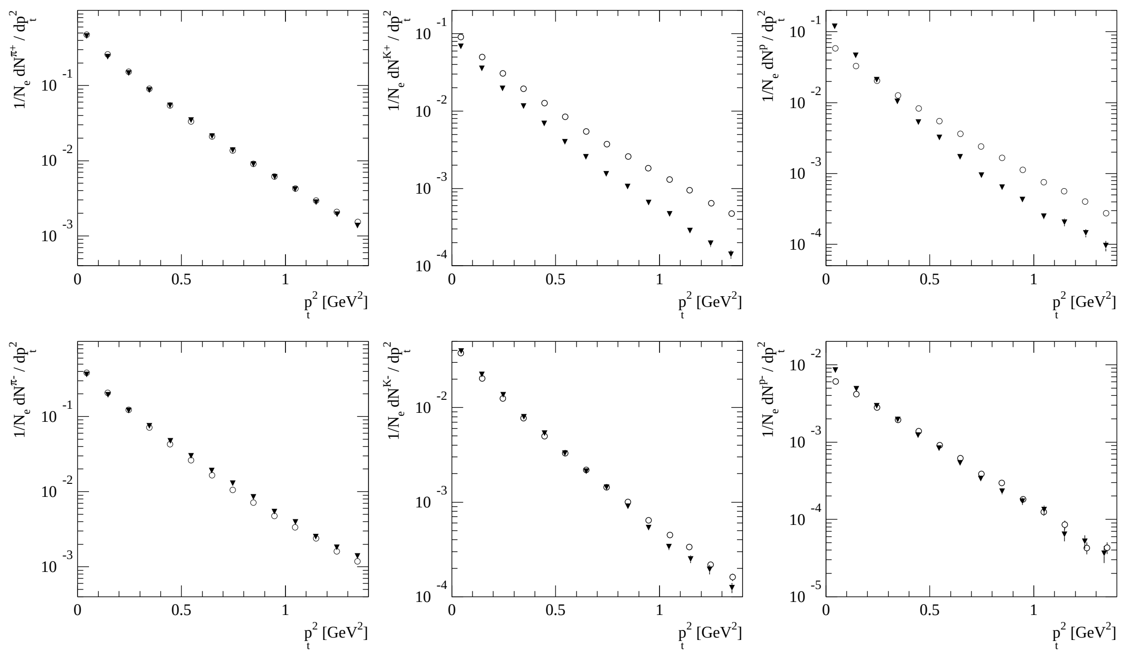

Figure 6.7: Comparison of semi-inclusive transverse momentum multiplicities between experimental data, only corrected for charge-symmetric background (dots) and fully tracked Monte Carlo (HRC) data (triangles) for positive and negative pions, kaons and protons (for a deuterium target).

6.4.2 Corrections to the experimental spectrum

In contradiction to the inclusive process that can be described by 2 variables, the (semi-inclusive) study of p_t^2-spectra must be defined by 5 independent variables. Here the set of variables (y, Q^2, p_t, z, Φ^γ) will be chosen, where Φ^γ is defined (fig. 2.2) as the angle between the lepton scattering plane (spanned by the lepton momenta vectors $\vec{k}$ and $\vec{k'}$) and the hadron production plane (defined by the vectors $\vec{q}$ and $\vec{P_h}$). The angle Φ^γ can be calculated as:

$$\phi^\gamma = \arccos\left(\frac{(\vec{q}\times\vec{k})\cdot(\vec{q}\times\vec{P_h})}{|\vec{q}\times\vec{k}||\vec{q}\times\vec{P_h}|}\right)\frac{\vec{P_h}\cdot(\vec{q}\times\vec{k})}{|\vec{P_h}\cdot(\vec{q}\times\vec{k})|} \tag{6.38}$$

Having 5 independent kinematical variables implies an extension of the unsmearing matrix to 5 dimensions. For instance, in the case of only two dimensions (p_t^2 and Q^2) the unsmearing matrix (5.9) needs to be modified as:

$$U_{ij}^{kl} = \frac{\text{hadrons generated in } (p_t^2 \text{ bin } i; Q^2 \text{ bin } k) \text{ and accepted in } (p_t^2 \text{ bin } j; Q^2 \text{ bin } l\ )}{\text{hadrons accepted in } (p_t^2 \text{ bin } j; Q^2 \text{ bin } l)} \tag{6.39}$$

An extension of the above equation to 5 dimensions would make the analysis code unnecessary complicated and the limited statistics in the experimental data sample would give a small population in some of the (5-D) bins. Hence another approach was followed, where the unsmearing matrix is kept 1-dimensional in the variable p_t^2 and the 4 other variables are taken into account by including them into the systematic error. The procedure for e.g. positive kaons proceeds as follows, where only the variables (p_t^2,Q^2) are illustrated:

1. Extract experimental p_t^2 multiplicities for K^+ and subtract the charge-symmetric background (fig. 6.7).
2. Calculate the unsmearing matrix (5.9) for the variable p_t^2 and use this matrix to correct for kinematical smearing and radiative effects, using equation (5.10) where the variable p_t^2 is used instead of x and the particles are kaons instead of positrons. (For each Monte Carlo event, the event *weight* is used while calculating the matrix U.)
3. Use eq. (5.13) to obtain K^+ multiplicities as a function of p_t^2 on the Born level, that are valid in the 4π solid angle.
4. Extract experimental Q^2 multiplicities for K^+ and subtract the charge-symmetric background (fig. 6.8).
5. Extract Q^2 multiplicities for K^+ from fully tracked Monte Carlo data (HRC) (fig. 6.8).
6. For each Q^2 bin, divide the experimental and HRC multiplicities from the last two steps by each other and call this ratio r_{Q^2}.
7. Find the Q^2 bin of each Monte Carlo K^+ event and replace the original event *weight* by a new weight that is calculated as $weight \cdot r_{Q^2}$. Use this last weight instead of the original one to recalculate the unsmearing matrix as explained in step 2. The unsmearing matrix U' that is obtained now, is thus modified in such a way that the disagreement between the experimental and simulated Q^2 spectra is compensated for.
8. Repeat step 3 and consider the difference between the p_t^2 spectra in 4π obtained with U' and U as a systematic error on the latter.

The steps 4-8 are repeated for the variables Q^2, y, z and Φ^γ and the resulting systematic uncertainties on the p_t^2 multiplicities are added in quadrature. This procedure was repeated for positive and negative hadrons and for the hydrogen and deuterium target. The corrected p_t^2 multiplicities with statistical and systematic errors are shown in fig. 6.9 and 6.10. Although the discrepancies between the experimental and HRC spectra for the variables (Q^2, y, z,Φ^γ) are rather large as illustrated in fig. 6.8, the total systematic error arising from the reweighting method is small as can be seen in the previous two figures. This indicates that the extraction of p_t^2 spectra from experimental data is independent of the quality of the Monte Carlo simulation (with respect to the other 4 variables) that was used to correct these spectra.

As a crosscheck a 2-D unfolding in the variables p_t^2 and Q^2 has been done according to eq. (6.39), yielding no difference for the presented p_t^2 spectra.

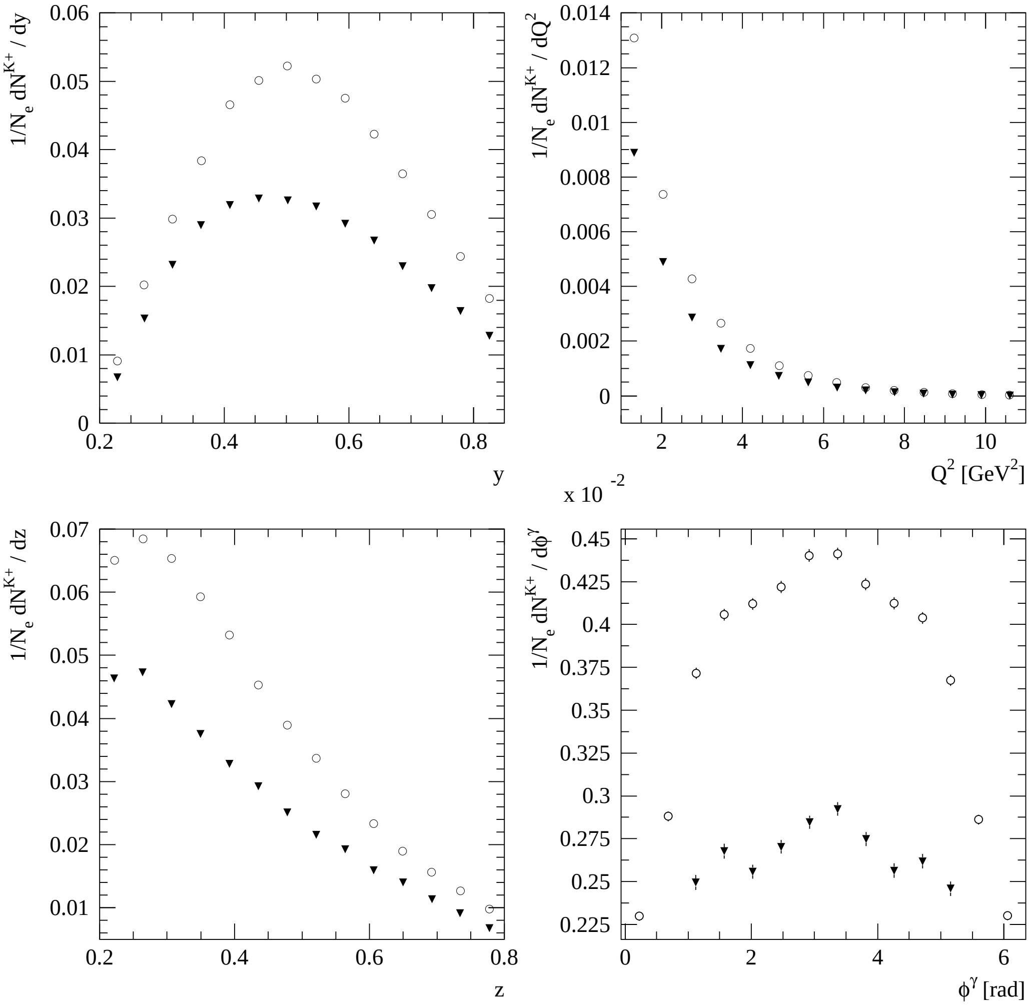

Figure 6.8: Comparison of multiplicities for positive kaons between experimental data (dots) and fully tracked Monte Carlo (HRC) data (triangles) for the variables y, Q^2, z and Φ^γ (for a deuterium target).

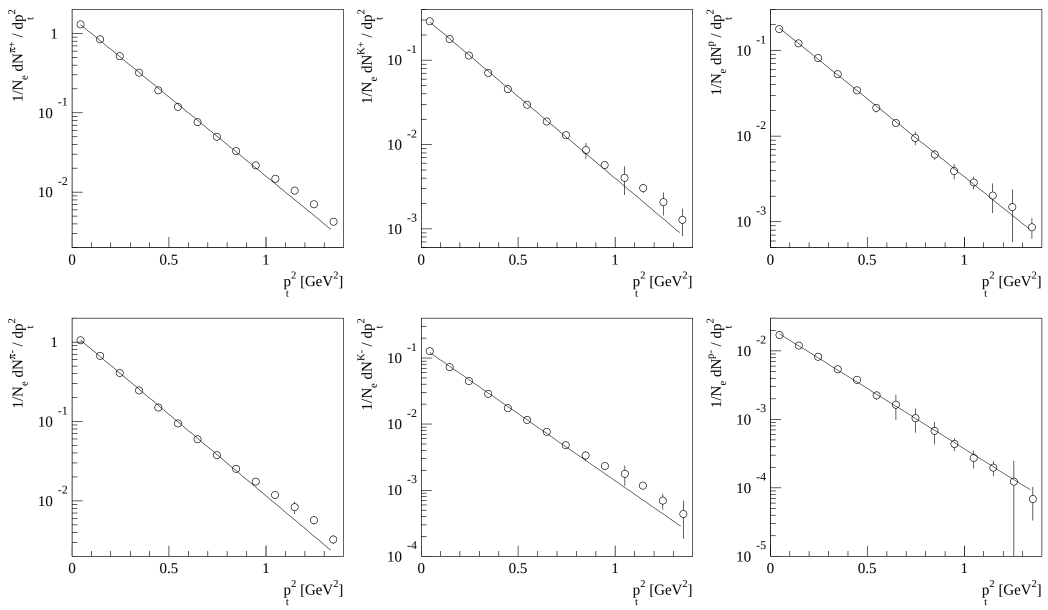

Figure 6.9: Semi-inclusive transverse momentum multiplicities for positive and negative pions, kaons and protons (for a deuterium target). The spectra were corrected for charge-symmetric background, detector smearing, radiative effects and acceptance. The error bars represent the combined statistical and systematic error.

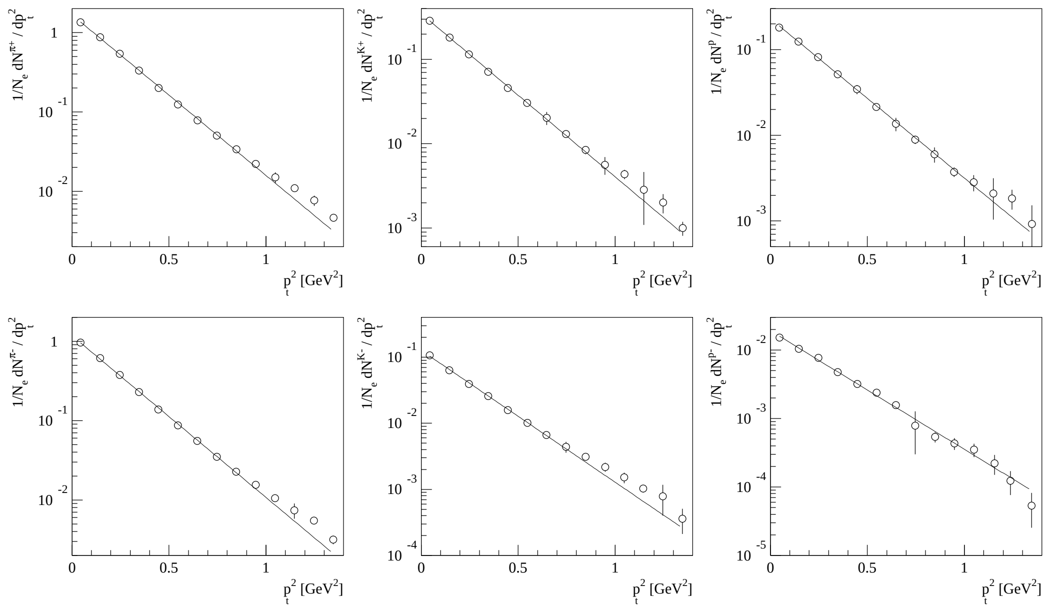

Figure 6.10: The same as fig. 6.9, but for a hydrogen target.

6.4.3 Phenomenological fit

The corrected spectra presented in fig. 6.9 and 6.10 have been fitted with an exponential shape:

$$\left(\frac{1}{N_e}\frac{dN_h}{dp_t^2}\right)^{\text{corrected}} = A \cdot \exp\left(-bp_t^2\right) \tag{6.40}$$

and the results from the fit are listed in table 6.1 for all charged hadron types and both the deuterium and the hydrogen target. A graphical view of the numerical values is shown in fig. 6.11. The stability of the fit was examined by changing the binning (using 10 and 20 bins in p_t^2 instead of 14), giving results for A and b that are compatible within the error bars. The values mentioned in table 6.1 are compatible with the world data, where values of b for pions are about 4 GeV^{-2} [151, 152], for kaons 3 to 5 GeV^{-2} [151, 153, 154, 155] and for protons 4 to 5 GeV^{-2} [151, 156]. Since the Q^2 range of these experimental data is not the same as for HERMES, this indicates that the hadronisation process is independent of the scattering process, confirming the hypothesis of factorisation. The choice of the exponential shape is governed by the arguments given in section 2.4.4: if QCD effects can be ignored, the only remaining processes contributing to p_t are hadronisation (described by the exponential (2.47)) and the intrinsic transverse momentum k_T of the struck quark, which follows a Gaussian distribution. Hence the p_t^2 distribution from these two processes is expected to have an exponential shape. Gluon Bremsstrahlung and photon-gluon fusion cause [157] a deviation from an exponential function for higher values of p_t which can be described by the three-parameter function $B(1+p_t^2/m^2)^{-\alpha}$ instead of eq. (6.40). The fits shown in fig. 6.9 and 6.10 do not show an explicit deviation from an exponential function, indicating that first order QCD processes do not contribute significantly to the p_t^2 spectra. To confirm this conclusion, the Lund model has been used to generate two different Monte Carlo productions (without detector tracking, 4π geometry) with first order QCD effects once turned on and once turned off. Between these two cases the difference for the slope parameters was less than 2% for all hadron types, indeed confirming that photon-gluon fusion and gluon Bremsstrahlung do not influence the p_t^2 spectra substantially in the range 0-1.4 GeV^2 (at HERMES energies).

While the intercept parameter A reflects the magnitude of the cross section for the production of different hadron types, the slope parameter b can be interpreted in terms of the mean value $< p_t^2 >$ which can be calculated from eq. (6.40) as

$$< p_t^2 >= \frac{\int_0^{1.4} A \cdot \exp\left(-bp_t^2\right) p_t^2 \mathrm{d}p_t^2}{\int_0^{1.4} A \cdot \exp\left(-bp_t^2\right) \mathrm{d}p_t^2} \approx \frac{1}{b} \tag{6.41}$$

The lower values of b for protons and anti-protons compared to pions and kaons can now easily be understood: since baryons contain 3 quarks, more string breaking processes occur compared to the mesons that have a two-quark configuration. As more transverse momentum is acquired during these string breaks, a larger value for $< p_t^2 >$ or a lower value for b is expected. Furthermore, the fact that for all hadron types no difference in $< p_t^2 >$ is observed between the hydrogen and deuterium target, can be related to a flavour independence of the internal transverse quark momentum k_T. Indeed, if e.g. the u and the d quark would have a different k_T, the $< p_t^2 >$ value for the pions from the hydrogen and deuterium target would be notably different, which is clearly not the case.

	Deuterium				Hydrogen			
	A	$\sigma(A)$	b	$\sigma(b)$	A	$\sigma(A)$	b	$\sigma(b)$
π^+	1.5838	0.0122	4.604	0.021	1.6748	0.0095	4.661	0.019
π^-	1.3026	0.0081	4.724	0.019	1.1747	0.0094	4.694	0.024
K^+	0.3467	0.0038	4.466	0.043	0.3491	0.0034	4.456	0.048
K^-	0.1486	0.0023	4.675	0.054	0.1267	0.0015	4.596	0.055
p	0.2216	0.0039	4.191	0.074	0.2296	0.0038	4.283	0.075
p^-	0.0215	0.0009	4.057	0.112	0.0191	0.0007	3.990	0.097

Table 6.1: Fit parameters and their errors for corrected p_t^2 spectra on hydrogen and deuterium (fig. 6.9 and 6.10). The fit function is given by eq. (6.40).

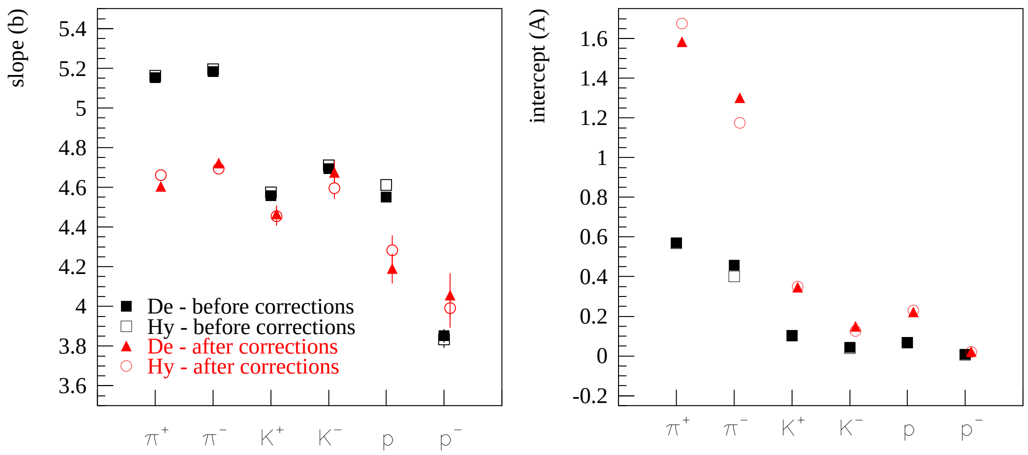

Figure 6.11: Graphical representation of the fitting parameters listed in table 6.1. The fitting results for the uncorrected data are shown for comparison.

Chapter 7

Summary

This section summarises the results obtained during the work for this Ph.D. thesis. The work was split in two parts, which comprise a hardware contribution and the analysis of data taken with the HERMES detector during the year 2000.

The first part of this thesis described the hadron identification with the RICH detector. The first test measurements after the installation in 1998 revealed that the RICH in principle functioned, although the design resolution was not reached. An alignment procedure (based on experimental data only) solved this problem and enabled the RICH to identify pions, kaons and protons. Two different PID algorithms were optimised and tested: Indirect and Direct Ray Tracing, both having their strengths and weaknesses. The RICH PID Scheduler decides which reconstruction method gives the more reliable result. The performance of the hadron identification was summarised in the P matrix and could be partly studied by means of decaying particles. The complete characterisation of the P matrix required the development of the RICH Monte Carlo program, which described the experimental data well. A detailed study of the effects included in the simulation was described in the text, as well as a motivation for omitted phenomena. The adjustable parameters in the simulation were tuned to reach an accurate agreement with the experimental data. The simulation provided the full P matrix relevant to the DIS process, which is indispensable for the unfolding of hadron momentum distributions.

In the second part, deep-inelastic positron nucleon scattering in the inclusive and semi-inclusive channel was studied from events taken on hydrogen and deuterium targets. The structure function ratio $F_2^d(x)/F_2^p(x)$ was extracted from carefully checked experimental **inclusive** data. Important corrections were required to reproduce the true F_2-ratio in nature. After applying these corrections and crosschecking them with an independent method, the resulting F_2-ratio agreed with a parametrisation of the existing world data, which was thus proven to be valid in the HERMES kinematical domain. The limits $x \to 0, 1$ of the F_2-ratio were compatible with expectations from the Quark Parton Model.

In the last chapter, several topics related to **semi-inclusive** DIS were discussed: the ratio of the valence quark distributions $d_v(x)/u_v(x)$, the kaon fragmentation functions and the transverse momentum distributions of pions, kaons and protons. The RICH detector was invaluable for these analyses.
The obtained valence ratio d_v/u_v was compatible with existing parametrisations and experimental results from neutrino experiments, showing that the charged and neutral cur-

rents are probing the same substructure of the nucleon.
The identification of semi-inclusive kaons aimed for an independent verification of the recently obtained HERMES result for $\Delta s(x)$, bypassing the use of any DIS Monte Carlo simulation. Broken symmetry relations between kaon fragmentation functions prevented this analysis. These violations are not yet understood and need further exploration.
Transverse momentum distributions for pions, kaons and protons on hydrogen and deuterium targets were interpreted in the framework of the Lund fragmentation model. It could be concluded that p_t is mainly due to fragmentation and that the intrinsic quark transverse momentum k_T is flavour independent.

Bibliography

[1] M. Anselmino, A. Efremov and E. Leader, Phys. Rep. **261** (1995) 1, Erratum: Phys. Rep. **281** (1997) 399

[2] M. Arneodo et al., Nucl. Phys. **B483** (1997) 3

[3] L. W. Whitlow et al., Phys. Lett. **B250** (1990) 193

[4] A. C. Benvenuti et al., Phys. Lett. **B237** (1990) 599

[5] V. N. Gribov and L N. Lipatov, Sov. J. Nucl. Phys. **15** (1972) 438

[6] G. Altarelli and G. Parisi, Nucl. Phys. **B126** (1977) 298

[7] Y. L. Dokshitzer, Sov. Phys. JETP **46** (1977) 641

[8] M. Tytgat, *Diffractive Production of ρ^0 and ω Vector Mesons at HERMES*, PhD thesis, University of Gent (2000)

[9] M. Dueren, *The HERMES Experiment: From the Design to the First Results*, Hermes Internal Note 95-002 (1995)

[10] P. L. Anthony et al., The E155 Collaboration, Phys. Lett. **B458** (1999) 529

[11] K. Gottfried, Phys. Rev. Lett. **18** (1967) 1174

[12] K. Ackerstaff et al., Phys. Rev. Lett. **81** (1998) 5519

[13] J. D. Bjorken, Phys. Rev. **148** (1966) 1467; Phys. Rev. **D1** (1970) 1376

[14] K. Abe et al., The E143 Collaboration, Phys. Rev. **D58** (1998) 112003

[15] J. Ellis and R. L. Jaffe, Phys. Rev. **D9** (1974) 1444; Phys. Rev. **D10** (1974) 1669

[16] C. Adloff et al., Phys. Lett. **B406** (1997) 256

[17] J. Breitweg et al., Eur. Phys. J. **C11** (1999) 251

[18] M. Arneodo et al., Z. Phys. **C31** (1986) 1

[19] C. T. Jones et al., Z. Phys. **C51** (1991) 11

[20] J. Binnewies, B. A. Kniehl and G. Kramer, Z. Phys. **C65** (1995) 471

[21] J. J. Aubert et al., Phys. Lett. **B114** (1982) 373

[22] J. Ashman et al., Z. Phys. **C52** (1991) 361

[23] A. Airapetian et al., Eur. Phys. J. **C21** (2001) 599

[24] M. Arneodo et al., Nucl. Phys. **B321** (1989) 541

[25] P. Geiger, *Measurement of the fragmentation functions at HERMES*, PhD thesis, Ruprecht-Karls University Heidelberg (1998)

[26] R. D. Field and R. P. Feynman, Nucl. Phys. **B136** (1978) 1

[27] G. Marchesi et al., Comp. Phys. Comm. **67** (1992) 465

[28] B. Andersson et al., Phys. Rept. **97** (1983) 33

[29] T. Sjöstrand, Comp. Phys. Commun. **82** (1994) 74

[30] H. Tallini, *Measurement of the quark spin distributions of the nucleon at Hermes*, PhD thesis, Liverpool University (1998)

[31] F M. Menden, *Determination of the gluon polarization in the nucleon*, PhD thesis, Albert-Ludwigs University Freiburg (2001)

[32] R. Odorico, Phys. Lett. **B89** (1979) 89

[33] ZEUS Collaboration, Phys. Lett. **B511** (2001) 19

[34] J. Huston et al., Phys. Rev. **D51** (1995) 6139

[35] L. Apanasevich et al. (E706), Phys. Rev. Lett. **81** (1998) 2642; Phys. Rev. **D59** (1999) 074007

[36] M. Zielinski, hep-ph/9811278

[37] L. Apanasevich et al., Phys. Rev. **D59** (1999) 074007

[38] M. Ruh, *Measurement of the Polarized Quark Distributions of the Nucleon at HERMES*, PhD thesis, Albert-Ludwigs University Freiburg (1999)

[39] M. Glück et al., Phys. Rev. **D63** (2001) 094005

[40] J. Blümlein and H. Böttcher, Nucl. Phys. **B636** (2002) 225

[41] A. A. Sokolov and I .M. Ternov, Sov. Phys. Comm. **82** (1964) 1203

[42] J. D. Jackson, Rev. Mod. Phys. **48** (1976) 417

[43] J. Buon, K. Steffen, Nucl. Instr. and Meth. **A245** (1986) 248

[44] D. P. Barber et al., Nucl. Instr. and Meth. **A338** (2000) 21

[45] M. Beckmann et al., Nucl. Instr. and Meth. **A479** (2002) 334

[46] C. Baumgarten et al., Nucl. Instr. and Meth. **A496** (2003) 277

[47] D. De Schepper et al., Nucl. Instr. and Meth. **A419** (1998) 16

[48] T. Weiland and R. Wanzenberg. *Wake Fields and Impedances*, Technical Report DESY M-91-06 (1991)

[49] F. Stock et al., Nucl. Instr. and Meth. **A343** (1994) 334

[50] C. Baumgarten et al., Nucl. Instr. and Meth. **A496** (2003) 263

[51] C. Baumgarten et al., Nucl. Instr. and Meth. **A482** (2002) 606

[52] K. Ackerstaff et al., Nucl. Inst. and Meth. **A417** (1998) 230

[53] G. Charpak et al., Nucl. Instr. and Meth. **97** (1970) 377

[54] W. Wander, *Reconstruction of High Energy Scattering Events in the HERMES Experiment*, PhD thesis, University of Erlangen-Nürnberg (1997)

[55] A. Andreev et al., Nucl. Instr. and Meth. **A465** (2001) 482

[56] J. T. Brack et al., Nucl. Instr. and .Meth. **A469** (2001) 47

[57] S. Bernreuther et al., Nucl. Instr. and Meth. **A416** (1998) 45

[58] R. Kaiser, *Measurement of the Spin Structure of the Neutron using Polarised Deep Inelastic Scattering*, PhD thesis, Simon Fraser University Vancouver (1997)

[59] J. D. Jackson, Classical Electrodynamics, J. Wiley & Sons (1975)

[60] H. Avakian et al., Nucl. Instr. and Meth. **A417** (1998) 69

[61] H. Avakian et al., Nucl. Instr. and Meth. **A378** (1996) 155

[62] R. Kaiser, *Particle Identification at HERMES*, Hermes Internal Note 97-025 (1997)

[63] J. Wendland, *Improved Particle Identification at HERMES and Polarised Valence Quark Distributions in the Proton*, Master thesis, Simon Fraser University Vancouver (1999)

[64] H. Tanaka, *Laser-based Gain Monitoring System for HERMES Experiment*, Master thesis, Tokyo Institute of Technology (2002)

[65] G. Schnell, *Longitudinal Polarization of the Lambda in Deep Inelastic Scattering of Polarized Positrons from Nucleons*, PhD Thesis, New Mexico State University (1999)

[66] T. Benisch et al., Nucl. Instr. and Meth. **A471** (2001) 314

[67] T. Benisch, *Polarisierte Bhabha-Streuung und Luminositätsmessung im HERMES-Experiment*, PhD thesis, University of Erlangen-Nürnberg (1998)

[68] U. Elschenbroich, *Bhabha-Streuung and Luminositätsmessung in HERMES-Experiment*, Diplomarbeit, Heinrich Heine University, Düsseldorf (2001)

[69] C. Weiskopf, *Untersuchung der BHABHA-Asymmetrien mit dem HERMES-Luminositätsmonitor*, Diplomarbeit, University of Erlangen-Nürnberg (1998)

[70] T. Ypsilantis and J. Seguinot, Nucl. Instr. and Meth. **A343** (1994) 30

[71] Particle Data Group, Phys. Rev. **D54** (1996) 1

[72] E. Cisbani et al., *Proposal for a Dual Radiator RICH for HERMES*, Hermes Internal Note 97-003 (1997)

[73] I. Adachi et al., Nucl. Instr. and Meth. **A355** (1995) 390

[74] D. Ryckbosch, on behalf of HERMES Collaboration, Nucl. Instr. and Meth. in Phys. Research **A433** (1999) 98

[75] N. Akopov et al., *The HERMES Dual-Radiator Ring Imaging Čerenkov Detector*, DESY 00-190, Nucl. Inst. and Meth. **A479** (2002) 511

[76] L. Zhang, *Ring Imaging Cerenkov Counter with Aerogel Radiator for HERMES*, Master thesis, Tokyo Institute of Techology (1999)

[77] A. Buzykaev et al., Nucl. Instr. and Meth. **A379** (1996) 465

[78] R. De Leo et al., Nucl. Instr. and Meth. **A440** (2000) 338

[79] P. Carter, *The HERMES Experiment: I. Analyzing Powers in Pion Electroproduction II. The Aerogel Radiator of the HERMES RICH*, PhD thesis, CalTech (1999)

[80] E. Aschenauer et al., Nucl. Instr. and Meth. **A440** (2000) 338

[81] D. De Schepper et al., *Particle Identification with the HERMES RICH Detector: Description of the Different Approaches*, Hermes Internal Note 98-008 (1998)

[82] K. Negodaeva, *Inclusion of the Aerogel Boundary Refraction into the HERMES RICH Reconstruction*, Hermes Internal note 99-022 (1999)

[83] D. De Schepper, B. Hommez, K. Van de Kerckhove, *Report on RICH Mirror Alignment*, HERMES RICH Internal note (1999)

[84] P. Jung, *Pion-Kaon Separation mit dem HERMES RICH Detektor*, Diplomarbeit, Humboldt University Berlin (2000)

[85] R. Kaiser, *Likelihood Analysis for the RICH IRT Method*, Hermes Internal Note 00-026 (2000)

[86] E. Cisbani, *La coppia di rivelatori RICH per lo studio di processi semi-inclusivi ad HERMES*, PhD thesis, Universita di Roma 'La Sapienza' (1998)

[87] A. Maas, *Particle Identification with RICH Detectors: Algorithms and their Optimization*, Diploma Thesis, TU Darmstadt (2000)

[88] The HERMES RICH webpage, http://hermes.desy.de/~kai/rich/

[89] B. Hommez, R. Kaiser, A. Maas et al. HERMES RICH Software Workshop Zeuthen, July 1999

[90] R. Kaiser, *Conventions for the HERMES RICH Particle Identification*, Hermes Internal Note 99-010 (1999)

[91] K. Negodaeva, *The PID efficiency of the HERMES RICH detector for the IRT method*, Hermes Internal Note 00-028 (2000)

[92] Particle Data Group, Eur. Phys. J. **C15** (2000) 1

[93] G. Ingelman et al., *LEPTO 6.5 - A Monte Carlo Generator for Deep Inelastic Lepton-Nucleon Scattering*, Comp. Phys. Comm. **101** (1997) 108; DESY 96-057 (1996); http://www3.tsl.uu.se/thep/lepto/

[94] CERN Computing and Network Division, *GEANT Detector Description and Simulation Tool*, CERN Program Library Long Writup W5013 (1993)

[95] CERN Programming Techniques Group, *ADAMO-Entity-Relationship Programming System, Version 3.3*, CERN (1993); http://adamo.web.cern.ch/Adamo/

[96] K. Suetsugu, *Performance of pion, kaon, and proton identification with the Ring Imaging Čerenkov of HERMES*, Master thesis, Tokyo Institute of Techology (2001)

[97] Y. Sakemi, private communications

[98] B. Hommez, *The inclusion of the RICH in the HERMES Montecarlo*, HERMES RICH Internal note (2002)

[99] S. Bernreuther, B. Hommez, Y. Miyachi, *New RICH MC Implementation*, HERMES RICH Internal note (2002)

[100] S. Henning and L. Svensson, Physica Scripta **23** (1981) 697

[101] D. Gray, AIP Handbook, McGraw-Hill, New York, 2nd edition, 1963

[102] D. R. Lide, CRC Handbook of Chemistry and Physics, 74th Edition, CRC Press, 1993-1994

[103] R. Kaiser, Y. Miyachi, private communications.

[104] T. Oest, HERA-B internal note (1996)

[105] D. Ryckbosch et al., *Photon collection efficiency in the HERMES RICH*, HERMES RICH Internal note (1998)

[106] D. Ryckbosch, *Some additional thoughts about the HERMES RICH funnels*, HERMES RICH Internal note (1999)

[107] F. Sato, *Study of the RICH Mirror Reflectivity*, HERMES RICH Internal note (2000)

[108] R. Kaiser, *Relative RICH Mirror Reflectivities*, HERMES RICH Internal note (2001)

[109] K. Negodaeva, *Inclusion of the Aerogel Boundary Refraction into the HERMES RICH Reconstruction*, HERMES RICH Internal note (1999)

[110] S. Bernreuther and Y. Miyachi, *Study of the influence on different models for the RICH aerogel refraction index in the rich.c program and on the effiency and contamination for kaons*, HERMES RICH Internal note (2002)

[111] B. Hommez, *Edges of aerogeltiles in the RICH: Effects on experimental data*, HERMES RICH Internal note (2001)

[112] Y. Miyachi et al., *Aerogel tile edge effect study in the rich program using experimental data*, HERMES RICH Internal note (2002)

[113] B. Hommez, *Tuning the RICH parameters in the montecarlo*, HERMES RICH Internal note (2002)

[114] R. Kaiser, *Unfolding of Hadron Distributions and Asymmetries*, Hermes Internal Note 00-027 (2000)

[115] B. Hommez and R. Kaiser, *Using the RICH Detector for Physics Analysis* http://www-hermes.desy.de/groups/richgrp/rich/richanalysis/index.html

[116] R. Schwitters, *Rings upon rings: A new approach to pattern recognition in Ring-Imaging Čherenkov Detectors*, HERA-B internal note (1998)

[117] B. Hommez, *Track independent ring fitting for the HERMES RICH: a mathematical approach*, HERMES RICH Internal note (1999)

[118] L. W. Whitlow et al., Phys. Lett. **B250** (1990) 193

[119] BCDMS Collaboration, A. C. Benvenuti et al., Phys. Lett. **B237** (1990) 592

[120] NMC Collaboration, M. Arneodo et al., Nucl. Phys. **B487** (1997) 3

[121] C. Weiskopf, *Messung der spinabhängigen Strukturfunktion g_1^d des Deuterons mit dem HERMES-Experiment*, PhD thesis, University of Erlangen-Nürnberg (2002)

[122] B. E. Tipton, *Measurement of polarised parton distributions with spin-dependent deep-inelastic scattering*, PhD thesis, Massachusetts Institute of Technology (1999)

[123] M. Vincter, *A determination of $F_2^n(x)/F_2^p(x)$, $d_v(x)/u_v(x)$, and the light quark sea flavor asymmetry using the 1995 HERMES unpolarised data*, HERMES Internal Note 97-013 (1997)

[124] L. De. Nardo, *Measurement of the Structure Function g_1^d at HERMES and Extraction of Polarized Parton Distributions*, PhD thesis, University of Alberta (2002)

[125] M. Vincter, *A Determination of $F_2^n(x)/F_2^p(x)$, $d_v(x)/u_v(x)$, and $(\bar{d}(x) - \bar{u}(x))/(d(x) - u(x))$ from the 1996 HERMES Unpolarised Data*, HERMES Internal Note 97-028 (1997)

[126] T. Shin, *The Nuclear Dependence of Structure Function Ratios from Unpolarized Deep Inelastic Scattering*, PhD thesis, Massachusetts Institute of Technology (2000)

[127] A. Bruell et al., *Report on the F_2^N/F_2^D and F_2^{He}/F_2^D Structure Function Analysis*, HERMES Internal Note 99-062 (1999)

[128] E. Aschenauer, private communications

[129] I. Akushevich et al., Comp. Phys. Comm. **104** (1997) 201

[130] I. Akushevich and N. Shumeiko, J. Phys. **G20** (1994) 513

[131] K. Ackerstaff, *First Results from the HERMES Experiment using Unpolarized Targets*, PhD thesis, University of Hamburg (1996)

[132] D. Yu. Bardin, C. Burdik, P. Ch. Christova, T. Riemann, Z. Phys. **C44** (1989) 149

[133] M. Gluck, E. Reya, M. Stratmann, W. Vogelsang, Phys. Rev. **D63** (2001) 094005

[134] H. L. Lai et al., Eur. Phys. J. **C12** (2000) 375

[135] J. Ouyang, *Report on the HERMES Charged Pion Fragmentation Function Analysis*, Hermes Internal Note 99-014 (1999)

[136] G. R. Farrar and D. R. Jackson, Phys. Rev. Lett. **35** (1975) 1416

[137] S. J. Brodsky, M. Burkardt and I. Schmidt, Nucl. Phys. **B441** (1995) 197

[138] J. J. Aubert et al., Nucl. Phys. **B293** (1987) 740

[139] A. Bodek et al., Phys. Rev. **D20** (1979) 1471

[140] W. Melnitchouk and A. W. Thomas, Phys. Lett. **B377** (1996) 11

[141] M. Beckmann, *Extraction of Polarised Quark Distributions of the Nucleon from Deep Inelastic Scattering at the HERMES Experiment*, PhD thesis, Albert-Ludwigs-Universität Freiburg im Breisgau (2000)

[142] M. Beckmann et al., *Update on the Quark Polarization Extraction from 1996-2000 data on polarized Hydrogen and Deuterium Targets*, preliminary HERMES release report (2002)

[143] M. Vincter, *A Possible Cause for the Unexpected Behavior of $d_v(x)/u_v(x)$*, Hermes Internal Note 99-004 (1999)

[144] J. Ashman et al., Z. Phys. **C52** (1991) 361

[145] B. Mahieu, private communications

[146] G. T. Jones et al., Z. Phys. **C62** (1994) 601

[147] D. Allasia et al., Phys. Lett. **135** (1984) 231

[148] H. Jackson, *Strange Sea Polarization from a Two Component Purity Analysis of D_2 Data*, Hermes Internal Note 00-039 (2000)

[149] J. Binnewies, B. Kniehl and G. Kramer, Phys. Rev. **D52** (1995) 4947

[150] J. P. Albanese et al., Phys. Lett. **B144** (1984) 302

[151] J. F. Martin et al., Phys. Rev. **D20** (1979) 5

[152] C. Berger et al., Phys. Lett. **B70** (1977) 471

[153] R. G. Hicks et al., Phys. Rev. Lett. **45** (1980) 765

[154] I. Cohen et al., Phys. Rev. Lett. **40** (1978) 1614

[155] V. V. Ammosov et al., Nucl. Phys. **B162** (1980) 205

[156] H. Ackermann et al., Nucl. Phys. **B120** (1977) 365

[157] D. Trayner, *Hadronic fragmentation studies in diffractive deep inelastic scattering at HERA*, PhD thesis (H1 experiment), University of London (2001)

Samenvatting

Inleiding en motivering

De zoektocht naar de elementaire, ondeelbare bouwstenen van de materie is één van de meest fundamentele uitdagingen voor zowel de experimentele als de theoretische natuurkunde. In het begin van de vorige eeuw ontdekten fysici dat het atoom geen ondeelbaar deeltje is, maar dat het opgebouwd is uit een kern waarrond electronen cirkelen. Vervolgens werd ontdekt dat deze kern eigenlijk geen basisbouwsteen is, maar op zijn beurt bestaat uit dicht opeengepakte nucleonen (neutronen en protonen). In de jaren 1930 begonnen fysici met het bouwen van versnellers die bundels van hoog-energetische electronen gebruikten als projectiel om de kern te splitsen. Ongeveer 20 jaar later werden die machines zo krachtig dat ze in staat waren om zelfs de protonen en de neutronen uit elkaar te laten spatten. Maar het resultaat van deze pogingen was chaos: wanneer het proton of het neutron gebombardeerd wordt met zeer energetische projectielen, splitst het niet meer op in kleinere subdeeltjes. De krachtige botsingen produceren echter een brede waaier aan nieuwe, exotische deeltjes d.m.v. de omzetting van kinetische energie in massa. Bij gebrek aan een betere naam werden deze deeltjes 'elementaire' deeltjes genoemd. In de voorbije 3 decennia vonden fysici overtuigende bewijzen dat protonen, neutronen en andere elementaire deeltjes opgebouwd zijn uit compacte ondeelbare puntdeeltjes, de zogenaamde quarks. Volgens het standaardmodel van de hedendaagse deeltjesfysica bestaan er zes types quarks, waarvan alleen de "smaken" *up*, *down* en *strange* quarks belangrijk zijn voor dit doctoraatswerk. Zo bestaat het proton uit twee u-quarks en 1 d-quark en het neutron uit twee d-quarks en 1 u-quark. Dit zijn de drie zogenaamde *valentie*quarks die instaan voor de meeste "statische" eigenschappen van deze nucleonen, zoals hun massa. Daarnaast zijn er de electrisch neutrale gluonen die de sterke wisselwerking overdragen welke de quarks aan elkaar bindt. Tenslotte worden er voortdurend virtuele quark anti-quark paren gevormd in het veld van de sterke wisselwerking en ook deze *zee*quarks maken een inherent deel uit van het proton en neutron. De nucleonen (proton en neutron) zijn voorbeelden van baryonen (opgebouwd uit drie (valentie)quarks), daarnaast bestaan er ook de zogenaamde mesonen (bestaande uit één quark en één anti-quark). Belangrijke voorbeelden van mesonen zijn het pion ($\pi^+ = u\bar{d}$, $\pi^- = \bar{u}d$) en het kaon ($K^+ = u\bar{s}$, $K^- = \bar{u}s$). De term hadronen wordt gebruikt als verzamelnaam voor baryonen en mesonen.

Verstrooiingsexperimenten hebben steeds een cruciale en doorslaggevende rol gespeeld in de vooruitgang van de kennis over de substructuur van de materie. De quarks werden ontdekt via *diep-inelastische* verstrooiing: de term inelastisch verwijst naar het uit elkaar spatten van het doelwit en de terminologie *diep*-inelastisch houdt verband met de 'resolutie' van het verstrooiingsproces. Bijvoorbeeld aan het Stanford Linear Accelerator Center (SLAC) worden electronen verstrooid aan protonen. De interactie kan theoretisch beschreven worden d.m.v. de uitwisseling van een virtueel foton tussen het electron en (een quark in) het proton. Een maat voor de resolutie van het experiment wordt gegeven door het zgn. vier-moment van de virtuele foton probe, genoteerd als Q^2. De resolutie is omgekeerd evenredig met Q^2. De SLAC experimenten in de jaren 1960 werden gekenmerkt door een waarde van Q^2 van 7.4 GeV^2, wat overeenkomt met een resolutie die veel kleiner is dan de afmetingen van een proton, vandaar de naam *diep-inelastische* verstrooiing.

Met behulp van *gepolarizeerde* diep-inelastische verstrooiing kan ook de spin van het proton onderzocht worden. Op het eerste zicht zou het logisch lijken dat de drie valentiequarks, elk met spin 1/2, op zijn minst het leeuwendeel van de spin kunnen opbrengen. Uit diverse experimenten is echter gebleken dat dit naïef beeld niet opgaat: het aandeel van alle quarks samen (dus valentie- en zeequarks samen) bedraagt slechts 12% van de totale nucleonspin. Dit probleem is bekend geraakt als de spincrisis en werd aan het licht gebracht door *inclusieve* experimenten -waarbij slechts het verstrooide electron waargenomen wordt- die enkel de totale bijdrage van alle quarks tesamen geven, maar niet de individuele bijdrages van de verschillende quarksmaken.

De nood aan een beter inzicht in de complexe nucleonspinstructuur vormde de aanleiding voor een reeks nieuwe metingen, waarbij in 1995 aan het HERA (Hadronen Elektronen Ring Anlage) complex te DESY (Deutsches Elektronen SYnchrotron) het HERMES (HERa MEasurement of nucleon Spin) experiment van start ging. HERMES was ontworpen als een precisie-experiment met als doel een zo compleet mogelijk beeld te geven van de spinstructuur van nucleonen. Dit wordt bewerkstelligd via de verstrooiing van de gepolarizeerde electron/positron bundel van HERA aan een (on)gepolarizeerd gasvormig doelwit. De bijdragen van de individuele quarks tot de nucleon spin kunnen bepaald worden via gepolarizeerde *semi-inclusieve metingen* van hadron productie. Daarbij wordt naast het verstrooide electron ook een deel van de gevormde reactieproducten waargenomen. Meestal gaat het dan om de hoogstenergetische deeltjes, waarbij er een grote kans is dat deze rechtstreeks gevormd zijn rond de quark waaraan het electron verstrooid is.

In de herfst van 1998 werd het HERMES experiment uitgebreid met een nieuw type RICH (**R**ing **I**maging **Ch**erenkov) detector. Met deze detector werd het mogelijk kaonen (die strange quarks bevatten) eenduidig te identificeren. Daardoor kon HERMES onlangs de tot nog toe volledig onbekende bijdrage van de *strange* quarks tot de nucleon spin bepalen. Een belangrijk deel van dit doctoraatswerk werd besteed aan het ontwikkelen en optimaliseren van algoritmes om pionen, kaonen en protonen te identificeren met deze RICH detector. Dit zal samengevat worden in de volgende paragraaf. Na het op punt stellen van de RICH detector en het ontwikkelen van een Monte Carlo simulatie programma ervoor, werden twee types diep-inelastische processen bestudeerd: inclusieve en semi-inclusieve verstrooiing. Aan de hand van het inclusieve proces werd de F_2 structuurfunctie van waterstof en deuterium onderzocht. Voor de studie van het semi-inclusieve kanaal werden de capaciteiten van de RICH extensief benut om drie verschillende *fragmentatie* processen te bestuderen, zoals verder uitgebreid zal beschreven worden. De term fragmentatie verwijst hier naar het vormen van een hadron wanneer een bundelelectron verstrooit aan een quark.

De RICH detector

De technische doelstelling van dit doctoraatswerk hield rechtstreeks verband met de studie van de performantie van de RICH en de algoritmes om hadronen te identificeren met deze RICH. De RICH detector werd in de eerste maanden van 1998 opgebouwd in Hamburg en in de herfst van 1998 werden de eerste testmetingen uitgevoerd waarmee bewezen werd dat de RICH in principe werkt. De nodige software om pionen, kaonen en protonen te identificeren op basis van de respons van de RICH stond echter nog niet op punt en een groot deel van dit doctoraatswerk werd besteed aan het ontwikkelen en verbeteren van de benodigde algoritmes. In deze thesis werd duidelijk gemaakt hoe de methode voor hadronidentificatie geöptimaliseerd en uitgebreid werd, zodat deze zich op heden in een gefinaliseerde staat bevindt en zeer goede resultaten oplevert. Tevens werd een Monte Carlo beschrijving van de RICH detector ontwikkeld, hetgeen een verre van eenvoudige taak was, gezien zich heel wat gecompliceerde optische verschijnselen voordoen in dit type detectoren. Toch werd een betrouwbare Monte Carlo simulatie geprogrammeerd die een uitstekende overeenkomst met het experiment oplevert. Ook deze Monte Carlo simulatie werd in dit proefschrift beschreven en de goede overeenkomst met de werkelijke experimentele resultaten werd geïllustreerd.

In hoofdstuk 4 werd allereerst het algemene werkingsprincipe van RICH detectoren beschreven. Kort samengevat komt dit neer op het volgende: in een materiaal met een brekingsindex n groter dan 1, is de lichtsnelheid kleiner dan die in het vacuüm: $c_m = c/n$. Als door zo' n materiaal een deeltje trekt met een snelheid v die groter is dan deze lichtsnelheid c_m, wordt Čerenkovstraling uitgezonden in een kegel, waarvan de openingshoek (de zgn. Čerenkovhoek θ) afhankelijk is van de rustmassa van het deeltje. Dus door meting van θ kan men het type deeltje bepalen. In de HERMES RICH detector worden twee media gebruikt waarin Čerenkovlicht opgewekt kan worden: aerogel en C_4F_{10} gas. Daardoor is het mogelijk om hadronen te identificeren over vrijwel het gehele kinematische gebied dat van belang is in HERMES. Hierbij zijn slechts 3 types hadronen relevant: pionen, kaonen en protonen. In fig. 4.3 en fig. 4.4 ziet men een schematische voorstelling van de RICH. Een deeltje (aangegeven met de pijl) komt binnen in de detector, en afhankelijk van zijn impuls produceert het een lichtkegel in de aerogel of in het C_4F_{10}. De Čerenkovfotonen worden door de sferische spiegel weerkaatst en vallen in op de fotondetector, die bestaat uit een matrix met een hexagonale stapeling van 1934 photomultipliers (PM). In dat vlak wordt de kegel afgebeeld als een ring. De grootte van de ring houdt rechtstreeks verband met de Čerenkovhoek θ en laat dus toe om hadronen te identificeren. Hoe dit precies in zijn werk gaat, werd uitgebreid beschreven in deze thesis en in principe waren er twee toepasbare algoritmes: Indirecte Ray Tracing (IRT) en Directe Ray Tracing (DRT).

De doelstelling van IRT was om de invloed van het afbeeldingssysteem te elemineren: aangezien het PM-vlak niet sferisch is, komt het niet overeen met het echte focale vlak van de sferische spiegel. Daarom zijn de gedetecteerde ringen niet echt cirkels maar eerder asymmetrisch gedeformeerde ellipsen. Voor de reconstructie van de Čerenkovhoek θ was het dus niet mogelijk om één of ander algoritme te gebruiken dat zou gebaseerd zijn op het fitten van cirkels. De IRT methode liet toe om θ te bepalen voor elke geraakte PM. Alvorens deze methode kon gebruikt worden om pionen, kaonen en protonen te identificeren, moest eerst een alignering van de RICH detector uitgevoerd worden. De korte tijd die kon besteed worden aan de installatie van de RICH detector liet niet toe om gedetailleerde metingen uit te voeren van de ganse detector en van de spiegel in het bijzonder. Deze alignatie werd uitgevoerd met een vereenvoudigde versie van het IRT algoritme op basis van experimentele data. Zowel deze vereenvoudigde versie van IRT als de alignatieprocedure werden in de tekst in detail beschreven. Dit was de eerste stap van een slecht geäligneerde RICH naar een bruikbare en begrijpbare detector. Pas daarna kon de RICH detector functioneren als “hadron separator” en kon IRT verder ontwikkeld worden. De efficiëntie van de IRT methode om pionen, kaon en protonen te identificeren werd bepaald en het bleek dat de IRT methode moeilijkheden had met deeltjes die een lage impuls hebben en dus was het de moeite waard om een ander algoritme voor PID te bekijken, de DRT.

De DRT methode simuleerde, analoog aan een Monte Carlo simulatie, de “hits” in de fotondetector voor elke mogelijke hypothese voor het betreffende deeltje (pion, kaon of proton). De waarschijnlijkheid om een bepaalde PM te vuren werd uit deze hits berekend. De gecombineerde hit waarschijnlijkheden voor het ganse PM vlak wordt een “hit waarschijnlijkheidsdistributie” (HWD) genoemd. Deze HWD werd berekend voor elk van de drie verschillende deeltjeshypotheses en werd vergeleken met het werkelijk geobserveerde hitpatroon in de RICH. De best overeenkomende hypothese bepaalde het deeltjestype.

Beide methodes bleken hun sterktes en zwaktes te hebben: DRT was tijdsintensiever dan IRT; in het algemeen scheen DRT beter te werken voor deeltjes met een lage impuls, en IRT leek beter te werken voor hoge impuls. Daarenboven werd algauw duidelijk dat de performantie van beide sterk afhankelijk kon zijn van de topologie in de RICH (bv. als twee deeltjes tegelijkertijd een ring patroon geven in de RICH heeft de overlap ervan een invloed op de identificatie efficiëntie van beide deeltjes). Daarom werd de RPS (RICH PID Scheduler) ontwikkeld, die besliste welke van de twee methodes het meest betrouwbare resultaat geeft voor een bepaalde topologie en impuls.

Nadat de identificatie-algoritmes volledig op punt stonden, kon onderzocht worden wat de performantie ervan was. Er werden een aantal grootheden ingevoerd om deze performantie te karakteriseren, die verzameld werden in een efficiëntie matrix, de zgn. P matrix. Elk deeltje heeft een *echt* deeltjestype (pion, kaon of proton) dat in principe alleen gekend is in een Monte Carlo simulatie, maar niet in experimentele data. Ieder deeltje heeft naast zijn echt deeltjestype ook een *geïdentificeerd* (of gereconstrueerd) type, dit is het type deeltje zoals het door het algoritme bepaald wordt. De P matrix geeft het verband tussen echte en geïdentificeerde deeltjestypes. Deze P matrix is een belangrijke karakteristiek van de RICH en speelt ook een grote rol voor de HERMES fysica, zoals verder zal blijken. Dus was het van fundamenteel belang om de P matrix nauwkeurig te bepalen en goed te begrijpen. Dit was niet evident, aangezien de RICH de enige detector in het HERMES experiment is die een onderscheid kan maken tussen de verschillende types hadronen, dus bestond er maar één mogelijkheid om het echte type deeltje te weten te komen, namelijk door gebruik te maken van vervalproducten van bepaalde mesonen. Zo vervalt het Φ meson bijvoorbeeld in 2 kaonen, dus liet een studie van dit verval toe om na te gaan hoe goed een echt kaon inderdaad als een kaon geïdentificeerd werd en in hoeverre het verkeerdelijk als een pion of een proton geïdentificeerd werd. Hoewel op die manier bepaalde elementen van de P matrix bepaald konden worden, was het jammer genoeg onmogelijk om de volledige matrix op die manier te karakteriseren. Daarom moest een andere strategie toegepast worden, gebaseerd op een Monte Carlo simulatie.

De Monte Carlo beschrijving van de RICH werd volledig ontwikkeld in het kader van deze doctoraatsthesis en in de tekst werd extensief toegelicht welke effecten in de simulatie opgenomen werden. Tevens werd er ruime aandacht besteed aan welke processen konden worden verwaarloosd, nadat ze eerst grondig theoretisch en experimenteel bestudeerd werden. Het design en de ontwikkeling van het simulatie programma nam heel wat tijd in beslag. Een illustratie van de complexiteit van dit proces kan gegeven worden door de computerkracht die hiervoor nodig is. Eén enkele simulatie werd uitgevoerd op een twintigtal CPU 's en nam dan ongeveer 3 dagen rekentijd in beslag, waarbij ongeveer een 15 Gigabyte aan output geproduceerd werd, die moest geänalyseerd worden. Elke wijziging aan de software betekende dat een nieuwe simulatie moest uitgevoerd worden. In totaal werden een 35-tal verschillende versies van dergelijke simulaties berekend, waarbij gebruik gemaakt werd van parallelle cluster-computers in Gent en Hamburg. Hoewel de RICH op het eerste zicht misschien conceptueel een eenvoudige detector lijkt, doen er zich complexe optische fenomenen in voor. Enkele voorbeeldjes zijn: dispersie in aerogel en gas, fotonabsorptie en Rayleigh verstrooiing in aerogel, reflectie en refractie van Čerenkov fotonen, foton detectie efficiëntie, enz. Al deze effecten en nog andere werden geïncludeerd in de simulatie. Daarnaast werd een uitgebreide studie gemaakt van de eigenschappen van aerogel als Čerenkov radiator: de aerogel in de RICH is immers geen massieve “muur”, maar is opgebouwd uit afzonderlijke tegels. Deze tegels hebben geen regelmatige vorm en hun brekingsindex varieert van tegel tot tegel, hetgeen een belangrijke impact heeft op de reconstructie van de Čerenkovhoek θ. De specifieke eigenschappen van de aerogeltegels werden grondig onderzocht en er werd bepaald welke aspecten al dan niet moesten opgenomen worden in de simulatie. Tenslotte bevatte de simulatie-software een aantal vrije parameters die niet eenduidig bepaald waren. De geschikte waarden voor deze variabelen werd bepaald aan de hand van experimentele data, in een iteratieve zgn. *tuning* procedure.

De uiteindelijke overeenkomst tussen experimentele data en de Monte Carlo simulatie werd gecontroleerd aan de hand van heel wat grootheden en deze vergelijking toonde aan dat de overeenkomst tussen het ‘echte’ experiment en de simulatie heel bevredigend was.

Het opnemen van de RICH in de HERMES Monte Carlo was niet alleen van belang om deze detector beter te leren kennen en de identificatie-algoritmes uitgebreid te testen, maar ook om de eerder vermelde P matrix te berekenen. De kennis van de P matrix is essentieel voor tal van fysica analyses in HERMES, die eigenschappen van verschillende hadrontypes onderzoeken. Bijvoorbeeld voor de bepaling van Δs, de strange quark bijdrage tot de protonspin, is het uitermate belangrijk om kaonen correct te identificeren. Nu is de kaon efficiëntie niet perfect, dus is er een zekere kans dat een door de RICH geïdentificeerd kaon in werkelijkheid geen kaon is, maar een pion of een proton. Toch is het mogelijk om het werkelijke totaal aantal kaonen te bepalen uit de meting van de pion, kaon en proton impulsdistributies. Daartoe werd een zgn. *unfolding* algoritme ontwikkeld, dat toelaat om uit de gemeten impulsdistributies de werkelijke distributies af te leiden. Met andere woorden: aan de hand van de verdeling van geïdentificeerde pionen, kaonen en protonen werd het mogelijk om bv. het echte, werkelijke aantal kaonen te bepalen. Het daartoe ontwikkelde algoritme, gebaseerd op de P matrix werd uitgebreid getest en wordt nu veelvuldig gebruikt door de volledige HERMES collaboratie. Het speelde ook een belangrijke rol voor de semi-inclusieve fysica analyse die beschreven werd in deze thesis en verder in deze samenvatting aan bod komt.

Diep-inelastische verstrooiing

Na de behandeling van de eerder technische resultaten in hoofdstuk 4, werd de inclusieve en semi-inclusieve diep-inelastische verstrooiing (DIS) bestudeerd in respectievelijk hoofdstuk 5 en hoofdstuk 6. In het DIS proces wordt een positron verstrooid aan een doelwit, en in deze thesis werd zowel een waterstof als een deuterium doelwit gebruikt. Voor de inclusieve analyse werd enkel het verstrooide positron beschouwd, terwijl voor de semi-inclusieve studie ook de geproduceerde pionen, kaonen en protonen van belang waren.

De doelstelling van de **inclusieve** analyse was de bepaling van de structuurfunctie verhouding F_2^d/F_2^p van deuterium (d) en waterstof (p). De structuurfunctie F_2 is gedefinieerd als:

$$F_2(x) = \sum_f e_f^2\, x q_f(x)$$

In deze vergelijking is x een schaalvariabele (de Bjorkenvariabele), gelegen tussen 0 en 1, die weergeeft welke fractie van de nucleonimpuls gedragen wordt door de quark waaraan verstrooid werd. De quarkimpulsdistributie wordt voorgesteld door $q_f(x)$, d.w.z. $q_f(x)dx$ is de waarschijnlijkheid om een quark (met smaak f en electrische lading e_f) te vinden in het nucleon met impulsfractie x in het interval $[x, x + dx]$. De structuurfunctie F_2 is dus de ladingsgewogen som van de quarkimpulsdistributies. Aangezien het deuteron bestaat uit een neutron en een proton, geeft de verhouding $F_2^d(x)/F_2^p(x)$ informatie over de verschillen tussen de quarkimpulsdistributies in deze twee nucleonen.

In principe was de bepaling van $F_2^d(x)/F_2^p(x)$ vrij eenvoudig: enkel het aantal verstrooide positronen aan een proton en deuterium doelwit moest geteld worden en deling van deze twee getallen gaf dan onmiddellijk de gewenste structuurfunctieverhouding. In de praktijk waren er echter een aantal belangrijke complicaties, enerzijds eerder experimenteel van aard en anderzijds waren er fundamentele fysische problemen. De technische moeilijkheden hielden verband met de experimentele HERMES opstelling: omwille van onvoorziene omstandigheden kon niet alle geregistreerde data gebruikt worden voor een wetenschappelijke analyse. De eerste belangrijke stap was dus een selectie van de beschikbare data, waarbij bijvoorbeeld vereist werd dat de positronbundel stabiel was en dat nergens in de detector een hoogspanningsdoorslag voorkwam. Nadat een verzameling van analyseerbare data gedefinieerd was, bepaalde een complexe set van kinematische voorwaarden hoe het DIS proces kon herkend worden. Op die manier werd een groot deel van de Gigabytes aan geregistreerde data verworpen en met de resterende gegevens kon de F_2-verhouding berekend worden. Uit een eerste naïef resultaat bleek al vlug dat belangrijke correcties nodig waren om de echte F_2-verhouding (zoals die voorkomt in de natuur) af te leiden. Vier verschillende oorzaken verdoezelden immers de berekende F_2-verhouding:

- De collectie DIS data bevatte nog altijd wat achtergrond. Deze achtergrond was namelijk in staat om positronen te genereren die door alle opgelegde voorwaarden heen sluipen en dus niet konden onderscheiden worden van de echte DIS positronen. In de tekst werd uitgelegd hoe gedetecteerde electronen een eenvoudig middel bieden om te corrigeren voor deze achtergrond.

- De HERMES spectrometer heeft een eindige resolutie en dus de meting van bv. de variable x kan eventueel niet de echte waarde van x weerspiegelen. De systematische verschillen tussen de echte en gemeten waarde van kinematische variabelen werden gecorrigeerd aan de hand van een Monte Carlo simulatie van het DIS proces en de volledige HERMES spectrometer.

- Stralingseffecten veroorzaakten ook een belangrijk verschil tussen gemeten en geregistreerde waarden van x. Een voorbeeld van een dergelijk stralingseffect is het onstaan van een radiatief foton dat afgestraald wordt door het bundelpositron. Dergelijke stralingsfotonen zijn niet detecteerbaar en veranderen de energie van het inkomend bundelpositron en dat heeft een rechtstreekse invloed op heel wat kinematische variabelen. Door een simulatie van dergelijke radiatieve effecten kon berekend worden hoe belangrijk ze waren en er werd d.m.v. een matrix methode voor gecompenseerd.

- Tenslotte heeft de HERMES spectrometer een beperkte angulaire acceptantie en dekt ze niet de volledige 4π ruimtehoek. Er werd berekend welke angulaire extrapolatie nodig was om dit effect teniet te doen.

Al deze correcties werden nauwkeurig beschreven in dit proefschrift en een alternatieve methode werd toegepast om na te gaan of er geen fouten in de redenering of de implementatie geslopen waren. Tenslotte kon het bekomen resultaat voor $F_2^d(x)/F_2^p(x)$ vergeleken worden met een bestaande parametrisatie van experimentele wereld data, hetgeen een excellente overeenkomst opleverde. Dit bevestigde de geldigheid van de parametrisatie die tot op heden nog niet getest was in het kinematische domein van HERMES. Tenslotte kon de F_2-verhouding vertaald worden naar de verhouding $F_2^n(x)/F_2^p(x)$ die makkelijker te interpreteren is. Dit werd voorgesteld in fig. 5.15. Men ziet dat bij lage waarden van x het verschil tussen een neutron en een proton verdwijnt, waaruit kon besloten worden dat hier alleen de zeequarks van belang zijn. Voor $x \to 1$ wordt de bijdrage van de zeequarks verwaarloosbaar en spelen alleen de valentiequarks een rol. We zien dan ook een merkbare afwijking van 1 en een extrapolatie liet toe verschillende theoretische modellen te testen.

Terwijl de F_2-analyse alleen informatie gaf over de som van de quarks en anti-quarks, leerde de **semi-inclusieve** analyse meer over de rol van de individuele quark smaken in het nucleon. De aanpak was grotendeels analoog als hierboven beschreven, behalve dat een aantal extra voorwaarden vereist waren om de correlatie tussen het gedetecteerde hadron en de quark waaraan verstrooid werd, te maximaliseren. Tevens werd het RICH unfolding algoritme gebruikt om het werkelijk aantal pionen, kaonen en protonen in de semi-inclusieve DIS reactie af te leiden uit de experimentele data. Eenmaal de set van semi-inclusieve gegevens zorgvuldig geselecteerd was en nadat alle kinematische voorwaarden toegepast waren, werden drie verschillende onderwerpen onderzocht: de verhouding van de (proton) valentiequark distributies $d_v(x)/u_v(x)$, de bijdrage van de strange zeequarks tot de protonspin (Δs) en transversale impulsdistributies van verschillende hadrontypes. De resultaten zullen in hetgeen volgt kort samengevat worden. De drie studies hadden een gemeenschappelijk onderwerp, nl. *fragmentatie*, i.e. het evolutieproces van een geraakte quark tot een hadron in de finale toestand. Het fragmentatieproces kan beschreven worden met fragmentatiefuncties D_f^h, die de waarschijnlijkheid weergeven dat een bepaald type hadron h gevormd wordt indien een quark met smaak f geraakt wordt in het verstrooiingsproces.

De d_v/u_v analyse maakte gebruik van pionen die met de RICH geïdentificeerd werden, opnieuw voor een waterstof en een deuterium doelwit. In de tekst werd het formalisme uitgelegd dat toeliet om uit de geregistreerde pionen rechtstreeks d_v/u_v af te leiden. Een aantal theoretische veronderstellingen (die verband houden met de pion fragmentatiefuncties) werden grondig getest op basis van Monte Carlo gegevens. De bekomen resultaten werden voorgesteld in fig. 6.3, ze toonden een uitstekende overeenkomst met theoretische parametrisaties en experimentele resultaten op basis van neutrino data. Dit bewees dat zowel (neutrale) neutrino probes als (geladen) positron probes dezelfde substructuur van het nucleon aftasten.

In de inleiding werd reeds vermeld dat HERMES er onlangs in gelaagd is om Δs te bepalen. De gebruikte methode steunt echter heel sterk op een simulatie van het DIS proces. De oorspronkelijke bedoeling was om het bekomen HERMES resultaat in dit doctoraatswerk te verifiëren met een onafhankelijke methode die alleen vertrok van experimentele data en niet hoefde te steunen op een simulatie. Een solide theoretische methode werd uitgewerkt en toegepast, waaruit bleek dat het bekomen resultaat weinig of geen fysische betekenis had. Daarom diende gezocht te worden naar de oorzaak van wat er precies fout liep. Op basis van het Lund model dat het fragmentatie proces beschrijft, kon besloten worden dat zeer aannemelijke symmetrierelaties tussen de kaon fragmentatiefuncties blijkbaar niet geldig waren. Deze symmetrierelaties gaan uit van fundamentele principes en worden al jaren als vanzelfsprekend aangenomen. Ze konden echter nog niet experimenteel gecontroleerd worden, aangezien HERMES als één van de eerste experimenten in staat is om semi-inclusieve DIS kaonen te identificeren. Wat nu precies de oorzaak is van de schending van de onderliggende principes kon vooralsnog niet bepaald worden, hoewel een aantal redenen reeds kon uitgesloten worden.

Tenslotte werden transversale impulsdistributies van semi-inclusieve pionen, kaonen en protonen bestudeerd. Het transversale impuls van een hadron (p_t) kan intuïtief het best ingevoerd worden door de analogie met een biljartbal: als een versnelde biljartbal een stilstaande bal raakt, komt die laatste in beweging, met een richting die niet noodzakelijk in het verlengde ligt van de versnelde bal, maar afhankelijk is van het impactpunt van beide ballen bij de botsing. In het DIS proces interageert een virtueel foton met een quark. Aangezien zowel een virtueel foton als een quark *punt*deeltjes zijn, verwacht men dat de geraakte quark gewoon de richting van het virtuele foton volgt. Bij de vorming van een hadron pikt de geraakte quark andere (anti-)quarks op uit de zee waardoor een afwijking in impuls t.o.v. de oorspronkelijke richting kan onstaan: transversale impuls. Daarnaast kunnen ook andere factoren bijdragen tot p_t, zoals gluonen en de Fermi beweging van quarks in het proton. Door de studie van p_t distributies van verschillende types hadronen en verschillende doelwitten kon besloten worden dat voornamelijk het fragmentatieproces bijdraagt tot p_t en dat de intrinsieke component te wijten aan de Fermi beweging niet smaakafhankelijk is.

Besluit

Tijdens dit doctoraatswerk werden twee onderwerpen grondig aangepakt: de RICH detector en de studie van (semi-)inclusieve diep-inelastische verstrooiing.

Voor de RICH detector werd heel wat software ontworpen, waarbij allereerst de RICH geäligneerd moest worden. Daarna kon aandacht besteed worden aan de belangrijkste doelstelling: de identificatie van pionen, kaonen en protonen. Verschillende algoritmes werden op punt gesteld en hun efficiëntie werd bepaald. Aangezien deze efficiëntie niet perfect was, werd de “unfolding” methode ontwikkeld die toeliet om uit de gemeten aantallen pionen, kaonen en protonen de werkelijke aantallen te bepalen. Deze unfolding steunde sterk op de P matrix die alleen kon bepaald worden uit een simulatie. Daarom werd een Monte Carlo programma voor de RICH ontworpen, dat uitstekend overeenkwam met de experimentele data.

Het DIS proces werd allereerst bestudeerd in het inclusieve geval: de structuurfunctie verhouding $F_2^d(x)/F_2^p(x)$ werd afgeleid uit experimenele data waarbij belangrijke correcties toegepast moesten worden. Het finale resultaat stemde goed overeen met een bestaande parametrisatie, die nog niet geverifieerd was in het kinematische gebied van HERMES. Uit de studie van limietgevallen voor x kon informatie afgeleid worden over de quarkdistributies in het nucleon.

Het semi-inclusieve kanaal leverde drie resultaten. De verhouding $d_v(x)/u_v(x)$ was consistent met theoretische berekeningen en resultaten van neutrino experimenten. De bepaling van Δs werd onmogelijk gemaakt door het falen van theoretische onderstellingen, hetgeen op zich een bron is van nieuwe informatie. De studie van p_t spectra liet toe om het fragmentatieproces beter te begrijpen op een fundamenteel niveau en de resultaten werden geïnterpreteerd in het kader van het Lund fragmentatie model.

Acknowledgements

This work could not have been completed without the help of many people, some whom I will mention here.

Firstly I thank my advisor, Dirk Ryckbosch, for his everlasting support and encouragement and for allowing me the freedom to explore many different topics at the HERMES experiment. Although he was busy commuting between Gent and Hamburg during the last 2 years, every friday he was willing to lend an ear to me and his suggestions made me to obtain a better grasp of physics. Moreover he gave me the opportunity to visit some of the nicest places on the globe, one of the many advantages of being involved in particle physics. I believe that his advice helped me to develop my skills in decision making, improved my writing skills, and aided my ability to give public presentations.

I also want to thank our former group leader Robert Van de Vyver together with the other colleagues Ulrike Elschenbroich, Bino Maiheu, Tigran Rostomyan, Michael Tytgat, Arne Vandenbroucke, Yves Van Haarlem, Luc Van Hoorebeke and former group members Natalie Degrande, Heidi Holvoet, Bob Lannoy, Dominique Marchand, Katty Van de Kerckhove and Avtandil Tkabladze for maintaining a great working atmosphere and remembering me the meaning of physics during the too few Gent "offsite meetings". I am especially indebted to Avto: he was always patient enough to satisfy my demand for another crosscheck, he could always be counted on to listen to a problem and give valuable advice. My work with him has been immensely rewarding. It would be more than fair to say that he has been one of those people who helped me along the way and I wish him (and his family) all the best in his career. And Avto, remember: Duvel is and remains the best beer in the world!

Of the many helpful HERMES collaborators, there are a few without whom this work would have been truly impossible; in particular Hal Jackson guided me through the complex world of RICH detectors and his profound experience was very valuable to me. His encouragement and cordial interest has been a constant source of insight and motivation. My appreciation also goes to Elke Aschenauer, who always had time to answer any question, large or small, day or night. Naomi Makins introduced me to the wonderful world of HDB and without her knowledge I would not have dared to get my teeth into the RICH Monte Carlo.

It has been a great pleasure to work with the RICH analysis subgroup for more than three years. I am particularly grateful to Stefan Bernreuther, Dirk De Schepper, Patricia Liebing, Katerina Lipka and Yoshi Miyachi. Ralf Kaiser I owe a great deal for all our discussions about hadron identification at HERMES.

Tenslotte wens ik mijn ouders te bedanken voor de jarenlange steun tijdens mijn studies en doctoraat. Hoewel ik veel te weinig uitgelegd heb waarmee ik nu juist bezig was en waarom ik maar weer es naar Hamburg afreisde, kan ik jullie verzekeren dat het besef dat jullie achter me stonden voor mij van onschatbare waarde was. Elke terugkeer naar Ieper betekende een blij weerzien en het gevoel van thuiskomen gaf (en geeft) me energie om door te gaan.

Evelien, jou wil ik gewoon bedanken voor wie je bent: een zonnemeisje. Deze lente is wat hectisch geweest, met te weinig tijd voor ontspanning. Maar nu de zomer daagt, bloei ik alleen nog maar voor jou. Dit doctoraat was er zonder jou niet geweest. Jouw lach is mijn motivatie, jouw enthousiasme is mijn inspiratie in dit leven.