Study of Charm Quark Fragmentation into D^* Mesons with the H1 Detector at HERA II

Dissertation
zur Erlangung des Doktorgrades
des Department Physik
der Universität Hamburg

vorgelegt von
Andrej Liptaj
aus Prešov

Hamburg
2008

Study of Charm Quark Fragmentation into D^* Mesons with the H1 Detector at HERA II

Dissertation
zur Erlangung des Doktorgrades
des Department Physik
der Universität Hamburg

vorgelegt von
Andrej Liptaj
aus Prešov

Hamburg
2008

Gutachter der Dissertation:	Prof. Dr. Peter Schleper Dr. Günter Grindhammer
Gutachter der Disputation:	Prof. Dr. Peter Schleper Dr. Hannes Jung
Datum der Disputation:	16.12.2008
Vorsitzender des Prüfunfsausschusses:	Prof. Dr. Dieter Horns
Vorsitzender des Promotionsausschusses:	Prof. Dr. Joachim Bartels
MIN-Dekan des Departments Physik der Fakultät:	Prof. Dr. Arno Frühwald

Kurzfassung

In dieser Arbeit wird die Charm-Quark-Fragmentation in $D^{*\pm}$-Mesonen in der tief unelastischen Elektron Proton Streuung untersucht. Die Arbeit basiert auf dem Datensatz der in den Jahren 2004 bis 2007 vom H1-Detektor bei HERA aufgezeichnet wurde. Die Daten entsprechen einer integrierten Luminosität von 354.1 pb^{-1}.

Differentielle Wirkungsquerschnitte werden als Funktion der drei Observablen z_{jet}, z_{hem} und $z_{\mathrm{hem}}^{(\mathrm{jet})}$, die in unterschiedlichen Näherung den Übertrag des Impulsbruchteiles vom Charm Quark auf das D^*-Meson darstellen, gemessen. Im Falle von z_{jet} entspricht der Impuls des Quarks ungefähr dem Impuls des D^*-Jets, bei den anderen Observablen entspricht er annähernd dem Impuls einer entsprechend definierten D^* Hemisphäre. Der sichtbare Phasenraum wurde durch den kinematischen Bereich $Q^2 > 5\,\mathrm{GeV}^2$ und $0.05 < y < 0.6$ und den Schnitten auf das rekonstruierte D^*-Teilchen, $1.5\,\mathrm{GeV} < p_t(D^*) < 15.0\,\mathrm{GeV}$ und $|\eta(D^*)| < 1.5$ festgelegt. In Ereignissen bei denen die beiden Observablen z_{jet}, $z_{\mathrm{hem}}^{(\mathrm{jet})}$ gemessen werden, wird außerdem ein rekonstruierter D^*-Jet verlangt, der die zusätzliche Bedingung $E_T(D^*\!-\!\mathrm{Jet}) > 3.0\,\mathrm{GeV}$ erfüllt. Innerhalb dieses Phasenraumes werden als Funktion der drei Observablen die normierten, einfach differenziellen Wirkungsquerschnitte extrahiert. Anschließend werden die Wirkungsquerschnitte mit den beiden Monte Carlo Modellen RAPGAP und CASCADE für verschiedene Parametrisierungen der Charm-Fragmentationsfunktion (Peterson und Kartvelishvili) berechnet. Beide QCD Modelle benutzen zur Beschreibung das Lund-String-Modell, wie es im PYTHIA Programm implementiert ist. Der Unterschied im Wirkungsquerschnitt zwichen den Daten und den Monte Carlo Modellen wird anhand eines χ^2-Tests quantifiziert und die optimalen Parameter für die Peterson und Kartvelishvili Parametrisierungen aus den χ^2-Verteilungen extrahiert. Vorhersagen von PYTHIA für e^+e^- Annihilation werden benutzt, um ebenfalls optimale Parameter aus den dazu publizierten Daten von BELLE und ALEPH zu gewinnen und mit den Resultaten dieser Arbeit zu vergleichen.

Die erhaltenen Resultate zeigen, dass die H1 Daten zwar eine Bestimmung der Fragmentationsparameter mit einer Genauigkeit die von Interesse ist erlauben, die Resultate jedoch anscheinend von der zur Verfügung stehenden Energie zur Charm-Quark Produktion abhängen. Die von der z_{hem} Verteilung abgeleiteten Parameter sind nicht konsistent mit jenen, die auf der z_{jet} bzw. $z_{\mathrm{hem}}^{(\mathrm{jet})}$ Verteilung basieren. Ebenso gibt es Inkonsistenzen beim Vergleich mit den Parametern aus der e^+e^- Annihilation. Diese Ergebnisse legen nahe, dass die existierenden Modelle, inklusive der untersuchten einfachen Fragmentationsfunktionen, nicht adequat sind, um die Fragmentation von Charm-Quarks bei verschiedenen Energien, nahe der Schwelle und darüber, zu beschreiben.

Abstract

In this work charm quark fragmentation into D^* mesons is investigated in deep-inelastic electron proton collisions. This work is based on data collected in the years 2004 - 2007 by the H1 detector at HERA, corresponding to a total integrated luminosity of 354.1 pb^{-1}.

Three observables denoted z_{jet}, z_{hem} and $z_{\mathrm{hem}}^{(\mathrm{jet})}$ are measured, each of them meant to approximate the momentum fraction of the charm quark transferred to the D^* meson. In case of z_{jet} the quark momentum is estimated as the momentum of the D^* jet, for the two other observables it is approximated by the momentum of an appropriately chosen D^* hemisphere. The visible range is defined by the phase space requirements on the DIS events: $Q^2 > 5\,\mathrm{GeV}^2$, $0.05 < y < 0.6$ and by the cuts applied on the reconstructed $D^{*\pm}$ particles: $1.5\,\mathrm{GeV} < p_T(D^*) < 15.0\,\mathrm{GeV}$ and $|\eta(D^*)| < 1.5$. An additional constraint $E_T(D^*\mathrm{jet}) > 3.0\,\mathrm{GeV}$ enters the phase space definition in case of z_{jet} and $z_{\mathrm{hem}}^{(\mathrm{jet})}$, where a reconstructed jet containing the D^* meson is required. Within this phase space the normalized single differential cross sections are measured in bins of the three observables. Two Monte Carlo models, RAPGAP and CASCADE, both interfaced with the PYTHIA program for the Lund string fragmentation, are used to make predictions of the respective cross sections for different parametrizations (Peterson and Kartvelishvili) of the charm fragmentation function. The difference in cross sections between data and Monte Carlo model predictions for different values of the fragmentation parameter is quantified by calculating values of χ^2 in order to extract optimal parameters for the Peterson and Kartvelishvili parametrization. Using predictions from PYTHIA for e^+e^- annihilation optimal parameters are extracted also from the published BELLE and ALEPH data.

The obtained results show that the H1 data allow the determination of the fragmentation parameters with a precision which is of interest. The extracted parameters are however found to apparently depend on the charm quark production energy: the z_{hem}-based results are not consistent with those derived from z_{jet} and $z_{\mathrm{hem}}^{(\mathrm{jet})}$, and inconsistencies are also observed when comparing to e^+e^- values. The results suggest that existing models, including the investigated simple fragmentation functions, are not adequate in describing charm fragmentation at different production energies with the same set of fragmentation parameters.

Abstrakt

Táto práca je zameraná na štúdium fragmentácie pôvabného kvarku na $D^{*\pm}$ mezóny v hlbokonepružných elektrón-protónových zrážkach. Je založená na údajoch získaných počas rokov 2004 - 2007 detektorom H1 na urýchľovači HERA, ktoré zodpovedajú 354.1 pb^{-1} integrovanej luminozity.

Boli zadefinované tri pozorovateľné veličiny, označené z_{jet}, z_{hem} a $z_{\mathrm{hem}}^{(\mathrm{jet})}$, z ktorých každá má aproximovať zlomok hybnosti pôvabného kvarku, ktorý je prenesený na D^* mezón. V prípade z_{jet} je hybnosť kvarku odhadnutá pomocou hybnosti jetu, pre zvyšné dve veličiny je aproximovaná hybnosťou vhodne zvolenej D^* hemisféry. Oblasť merania je daná ohraničeniami fázového priestoru $Q^2 > 5\,\mathrm{GeV}^2$, $0.05 < y < 0.6$ a požiadavkami na zrekonštruovaný D^* mezón $1.5\,\mathrm{GeV} < p_T(D^*) < 15.0$ GeV a $|\eta(D^*)| < 1.5$. V prípade z_{hem} a $z_{\mathrm{hem}}^{(\mathrm{jet})}$, kde sa požaduje zrekonštruovaný jet obsahujúci D^* časticu, vstupuje do definície oblasti merania aj požiadavka $E_T(D^*\mathrm{jet}) > 3.0$ GeV. V rámci týchto ohraničení boli namerané normalizované diferenciálne účinné prierezy v spomínaných troch premenných. Ďalej boli využité Monte Carlo modely RAPGAP a CASCADE, oba za účelom modelovania Lund string fragmentácie prepojené s programom PYTHIA, na predpovedanie zodpovedajúcich účinných prierezov pre rôzne parametrizácie (Petersonovu a Kartvelishviliho) fragmentačnej funkcie pôvabného kvarku. Rozdiely v účinných prierezoch medzi nameranými údajmi a Monte Carlo predpoveďami pre rôzne hodnoty fragmentačného parametra boli vyhodnotené za pomoci výpočtu χ^2 hodnôt s cieľom určiť optimálne parametre pre Petersonovu a Kartvilishviliho parametrizáciu. Optimálne parametre boli taktiež získané z publikovaných údajov experimentov BELLE a ALEPH.

Dosiahnuté výsledky ukazujú, že H1 dáta umožnujú určit fragmentačné parametre s dobrou presnosťou. Získané hodnoty parametrov však, zdá sa, závisia na produkčnej energii pôvabného kvarku: výsledky založené na z_{hem} nie sú v zhode s výsledkami získanými pomocou z_{jet} a $z_{\mathrm{hem}}^{(\mathrm{jet})}$ a nezrovnalosti sú pozorované aj pri porovnaní s e^+e^- hodnotami. Výsledky naznačujú, že súčasné modely, vrátane použitých jednoduchých fragmentačných funckií, nepopisujú adekvátne fragmentáciu pôvabného kvarku pri rôznych produkčných energiách s rovnakou sadou fragmentačných parametrov.

Contents

1 Theoretical Overview **7**
1.1 Standard Model and Quantum Chromodynamics . . . 7
1.1.1 Standard Model . . . 7
1.1.2 Quantum Chromodynamics . . . 9
1.2 Deep-Inelastic Scattering . . . 10
1.2.1 Kinematics . . . 11
1.2.2 Structure Functions and Inclusive Cross Sections . . . 12
1.2.3 Quark Parton Model and Evolution Equations . . . 13
1.2.4 Heavy Quark Production in DIS . . . 17
1.3 Fragmentation . . . 18
1.3.1 General Considerations . . . 18
1.3.2 Heavy Quark Fragmentation . . . 19
1.3.3 Fragmentation Models . . . 21
1.4 Simulation Programs . . . 23
1.4.1 Event Generators with LO Matrix Elements . . . 25

2 HERA Accelerator and H1 Detector **27**
2.1 HERA Accelerator . . . 27
2.2 H1 Detector . . . 28
2.2.1 Tracking System . . . 30
2.2.2 Calorimeters . . . 33
2.2.3 Trigger System . . . 36

3 Reconstruction and Measurement Methods **41**
3.1 Central Track Reconstruction . . . 42
3.2 Software Finders and Signal Extraction . . . 43
3.2.1 Electron Finder . . . 43
3.2.2 Energy Flow Algorithm for Hadronic Reconstruction . . . 44
3.2.3 Jet Finder . . . 46
3.2.4 D^* Finder . . . 48
3.2.5 D^* Signal Extraction . . . 49
3.3 Fragmentation Measurement Methods . . . 51
3.3.1 Hemisphere Method . . . 53
3.3.2 Jet Method . . . 54
3.4 Unfolding Methods . . . 55

3.4.1 Bin-by-Bin Method . . . 56
3.4.2 Singular Value Decomposition (SVD) . . . 56
3.4.3 Bayesian Iterative Approach . . . 58
3.5 Reconstruction of the Event Kinematics . . . 59

4 Data Selection 61
4.1 Running Periods and Luminosities . . . 61
4.2 Data Online Selection by Trigger . . . 62
4.2.1 Spacal Trigger Elements . . . 62
4.2.2 Track Trigger Elements . . . 62
4.2.3 VETO Trigger Elements . . . 63
4.3 Offline Selection . . . 64
4.3.1 Z-position of Interaction Vertex . . . 64
4.3.2 Electron Quality Requirements . . . 65
4.3.3 Selection with Respect to Event Kinematic Variables . . . 67
4.3.4 Selection of D^* Events . . . 69

5 Data Treatment, Data Description by Models 75
5.1 Monte Carlo Programs . . . 75
5.1.1 Treatment of QED Radiation and Resolved Processes . . . 76
5.1.2 Treatment of Beauty Component . . . 76
5.2 Trigger Efficiencies . . . 77
5.2.1 Spacal Trigger Efficiency . . . 77
5.2.2 Track Trigger Efficiency . . . 80
5.2.3 CIP VETO Condition . . . 82
5.3 Reweighting of Monte Carlo Distributions . . . 84
5.3.1 Z-position of Interaction Vertex . . . 84
5.3.2 BPC Efficiency . . . 85
5.3.3 Study of Reweighting in $\eta(D^*)$ . . . 86
5.4 Control Distributions for HERA II . . . 87
5.4.1 Event Quantities . . . 88
5.4.2 Detector-Related Track Quantities . . . 90
5.4.3 D^* Variables . . . 90

6 Fragmentation Measurement 95
6.1 Visible Range Definition and Previous Fragmentation Measurements . 95
6.2 Unfolding Hadron Level Distributions . . . 96
6.2.1 Background Treatment, Migrations into Visible Range and Beauty Subtraction . . . 96
6.2.2 Purity and Stability . . . 98
6.2.3 Unfolding Parameters . . . 99
6.2.4 Treatment of QED Radiative Corrections . . . 100
6.2.5 Hadron Level Unfolded Distributions . . . 102
6.3 Systematic Errors . . . 104
6.3.1 Determination of Systematic Uncertainties . . . 105

6.3.2 Summary of Statistical and Systematic Uncertainties 110

7 Results **115**
7.1 Hadron Level Spectra 115
7.2 Extraction of Fragmentation Parameters 115
7.2.1 Reweighting of Monte Carlo Models in z_{gen} 118
7.2.2 χ^2 Method and Extracted Fragmentation Parameters 119
7.3 Results from e^+e^- Experiments 121
7.4 Comparison of Results and Conclusions 126
7.4.1 Observations 126
7.4.2 Interpretation of Results 127

8 Summary and Outlook **129**

A Additional Control Plots **131**

B Plots of Systematic Errors **135**

C χ^2 Fits and Parameter Extraction **139**

Introduction

Important progress in the field of particle physics during the last century lead to the construction of the so-called Standard Model of particle physics. The Standard Model of particle physics is a quantum-field theory describing processes observed in nature at the level of elementary particles. It is one of the most successful physical theories ever developed, and its predictions are in very good agreement with many different experimental data. We know however that the Standard Model is not the ultimate physical theory - it does not contain gravitation, it does not explain fermion masses and has other problems [1]. In spite of this fact it is worth to mention the great successes of the Standard Model. Quantum electrodynamics (QED), a theory which is part of the Standard Model and which describes the interactions between charged particles which are not interacting strongly, is very successful in describing electromagnetic phenomena. The theoretical prediction for the electron magnetic anomaly - one of the best measured quantities in all of physics [2] - is in perfect agreement with the experimental value. The Standard Model also provides an elegant unification of electromagnetic and weak interactions, nowadays merged into one electroweak theory, which was and is still being confirmed with even better accuracy by experiments [3].

In addition to the electroweak interactions, the Standard Model contains a theory of strong interactions - quantum chromodynamics (QCD). It is a dynamical theory of quarks and gluons, and explains phenomena in which strong interactions play a role. This thesis is based on data from the H1 experiment at the HERA accelerator, where electrons collide with protons. Protons are objects made up of strongly interacting quarks and gluons and thus QCD is the theory which applies to our data[1]. Until now the QCD provided many proofs of being the correct theory to describe strong interactions [4]. It has however an unpleasant feature (emerging from the self-interaction of gluons) that in some phase space regions the perturbative calculations of physical observables are not reliable or not applicable at all. This is true for processes with small momentum transfer, especially such as bound states of quarks. It is believed, although not rigorously proved, that QCD has the property of confinement. This refers to the fact that quarks are not observed as free particles, but are always confined in bound states called hadrons. Perturbative methods do not provide an answer to this observation, lattice QCD calculations however support it strongly and suggest an approximately linear confining potential between quarks. For pro-

[1] An electron-proton interaction is in the leading order electromagnetic. In higher orders QCD effects contribute.

cesses with high momentum transfers perturbative calculations provide results on agreement with measurement. This is due to the important property of QCD called asymptotic freedom, which is rigorously proved. It states that the strong running coupling α_s, which is the expansion parameter of a perturbative series decreases with increasing momentum transfer. Thus, for processes where the momentum transfer is high, the coupling α_s becomes small and the series converges. This property allows for QCD predictions that can be tested by experiments in high-energy physics at particle accelerators.

In such a high-energy experiment one however still deals with real particles in the initial and final state and these particles may be hadrons. Since it is not known how to treat them within perturbative QCD (pQCD), it is obvious that calculations of physical observables for real initial and final state particles containing hadrons cannot rely only on perturbative QCD. Different factorization theorems - some rigorously proved, some only assumed - allow us to split the whole process of interaction into different parts. One part concerns the non-perturbative description of the quark dynamics in the initial-state hadron. The second part is provided by the perturbative calculation of the quark and gluon interactions at high momentum transfer, the so-called hard sub-process. The third one, referred to as hadronization or fragmentation, is supposed to describe the formation of hadrons from the final state partons (quarks and gluons). Even though the perturbative calculation is combined with non-perturbative descriptions, the resulting predictions usually have enough predictive power to serve as a test of perturbative QCD. For this test to be as accurate as possible, the precise knowledge of the non-perturbative parts is essential, and this knowledge is to be extracted from experiment.

In this thesis we focus on the fragmentation of the charm quark into the D^* meson, the D^{*+} meson being an excited bound state $c\overline{d}$ ($\overline{c}d$ for D^{*-}) with the mass of $m_{D^*} = 2010.0 \pm 0.5\,\mathrm{MeV}$. The process of fragmentation is described by different non-perturbative phenomenological models. These models usually make use of a non-perturbative fragmentation function, a function which describes the fraction of the initial quark momentum to be transferred to the final state hadron with some probability. We study, in deep-inelastic ep scattering, the charm fragmentation function parameters for different pQCD models and different parametrizations of the fragmentation function. We also investigate the fragmentation function universality with respect to e^+e^- experiments to gain some insight on whether the factorization theorem holds, i.e. the non-perturbative fragmentation is process-independent.

In the first chapter we present the theoretical basis needed to understand the fragmentation process, and we also mention the Monte Carlo programs used in this analysis. In the second chapter we describe the HERA accelerator, the H1 experiment and its different sub-detectors, putting special emphasis on the systems which are important for this study. In the third chapter we present different experimental methods used to study the fragmentation functions. In the fourth chapter we explain our data selection, the fifth chapter is dedicated to description of data by Monte Carlo models and in the sixth chapter the fragmentation measurement is described. Finally, in chapter 7, we present our results and conclusions. The summary and outlook can be found in chapter 8.

Chapter 1

Theoretical Overview

1.1 Standard Model and Quantum Chromodynamics

1.1.1 Standard Model

The Standard Model of particle physics is a relativistic quantum gauge field theory describing the fundamental interactions between fundamental particles. It was developed between 1970 and 1973 and is consistent with both quantum mechanics and special relativity. The relativistic character follows from its Lagrangian, which is a Lorentz scalar and thus Lorentz invariant. The quantum aspects are a consequence of the quantization of classical fields, and therefore the Standard Model has features like probabilistic interpretation, uncertainty principle, energy quantization, etc. The Standard Model is a gauge field theory, i.e. gauge fields (gauge bosons) were introduced using the local $SU(3) \times SU(2) \times U(1)$ gauge symmetry. However the existence of massive gauge fields $W^{\pm}$ and Z^0 would spoil the gauge symmetry. This is avoided by the introduction of the Higgs boson, a massive scalar field which couples to each fermion proportionally to its mass as well as to itself and $W^{\pm}$ and Z^0 bosons. The Higgs field provides non-zero masses to all massive Standard Model particles by the so-called Higgs mechanism, which is based on spontaneous symmetry breaking. The Higgs particle is the only particle of the Standard Model which has not yet been observed.

As we already mentioned in the *Introduction*, the Standard Model describes successfully three out of four known fundamental forces and all elementary particles observed in nature. The particles of the standard model are summarized in table 1.1. They can be divided (with the exception of the Higgs boson) into fundamental fermions and gauge bosons.

Gauge bosons mediate interactions: the photon the electromagnetic, the $W^{\pm}$ and Z^0 bosons the weak and the gluons the strong interactions. Only $W^{\pm}$ have electric charge and only $W^{\pm}$ and Z^0 are massive. They are all vector particles and only gluons have color - in analogy of the electric charge in QED. Fundamental fermions exist in three families (generations). Each particle from a given family has a "partner" particle in each of two remaining families that has identical quantum numbers and the only difference comes from different masses. This mass difference

	Family	I	II	III
Fundamental Fermions J=$\frac{1}{2}\hbar$	Leptons	e m = 511 keV/c^2 Q = -1	μ m = 105.65 MeV/c^2 Q = -1	τ m = 1.777 GeV/c^2 Q = -1
		ν_e m = 0 eV/c^2 Q = 0	ν_μ m = 0 eV/c^2 Q = 0	ν_τ m = 0 eV/c^2 Q = 0
	Quarks	u m ≈ 3 MeV/c^2 Q = 2/3	c m ≈ 1.3 GeV/c^2 Q = 2/3	t m ≈ 175 GeV/c^2 Q = 2/3
		d m ≈ 6 MeV/c^2 Q = -1/3	s m ≈ 100 MeV/c^2 Q = -1/3	b m ≈ 4.5 GeV/c^2 Q = -1/3

	EM	Weak		Strong
Intermediate Bosons J=1$\hbar$	γ m = 0 eV/c^2 Q = 0	$W^\pm$ m = 80.425 GeV/c^2 Q = ±1	Z m = 91.187 GeV/c^2 Q = 0	g m = 0 eV/c^2 Q = 0

Higgs Boson J=0$\hbar$	H^0 m > 114 GeV/c^2 Q = 0

Table 1.1: Elementary particles of the Standard Model. Electric charges are indicated in the absolute value of the electron charge and quarks exist in three colors. All particles have antiparticles, neutral bosons are identical with their antiparticles. The hypothesis of massless neutrinos assumed in the Standard Model is no longer valid, experiments have shown that neutrinos have non-zero masses.

remains unexplained within the Standard Model. Most of the ordinary matter[1] is made up of the particles from the first family, the particles in the two remaining families are unstable. Each family can be further divided into leptons (they do not interact strongly) and strongly interacting quarks. Although in the Standard Model one supposes massless neutrinos, experiments studying neutrino oscillations have shown that all fundamental fermions including neutrinos have non-zero masses. Incorporation of non-zero neutrino masses into the Standard Model however does not represent a major theoretical problem.

1.1.2 Quantum Chromodynamics

Quantum chromodynamics (QCD) is a non-Abelian quantum gauge field theory which is part of the Standard Model. It is based on the local $SU(3)$ color gauge symmetry. QCD is meant to describe the strong interactions between quarks and gluons. It is governed by the $SU(3)$-symmetric Lagrangian (density)

$$\mathcal{L} = \overline{q}(i\gamma^\mu \partial_\mu - m)q - g(\overline{q}\gamma^\mu T_a q)G^a_\mu - \frac{1}{4}G^a_{\mu\nu}G^{\mu\nu}_a,$$

where q is the quark field, γ^μ are the Dirac matrices, G^a_μ are the gauge fields corresponding to 8 gluon states, T_a are the generators of the $SU(3)$ group and $G^a_{\mu\nu} = \partial_\mu G^a_\nu - \partial_\nu G^a_\mu - g f_{abc} G^b_\mu G^c_\nu$ with f_{abc} being the structure constants of the group. The physics of quarks and gluons is obtained from this Lagrangian by applying the principle of least action followed by the quantization of fields. Each quark has one of three possible strong charges - colors (red, green, blue). Each gluon carries two color indices, one corresponding to a color and one to an anti-color. A gluon can couple to a quark pair and to two or three other gluons. The self-coupling of gluons leads to the strong confining potential between quarks. It is believed that quarks can exist only as bound states within colorless objects, the hadrons. Hadrons can be mesons (quark - antiquark pair) or baryons (red quark, green quark and blue quark)[2]. Perturbative calculations are not applicable to low-momentum transfer processes (like bound states) however, it is generally believed that all observed properties of hadrons can be in principle explained by QCD as partially confirmed by lattice calculations. The QCD property of asymptotic freedom allows us to make perturbative calculations for high-momentum transfer processes. These calculations have shown good agreement between predictions and experiments such as:

- the running of the QCD coupling,
- scaling violations in deep-inelastic scattering,
- jet cross sections,
- event shape observables,

[1] Baryonic matter is meant here, origin of the dark matter is unknown.
[2] Experimental evidence for the (non-)existence of pentaquarks is still not conclusive.

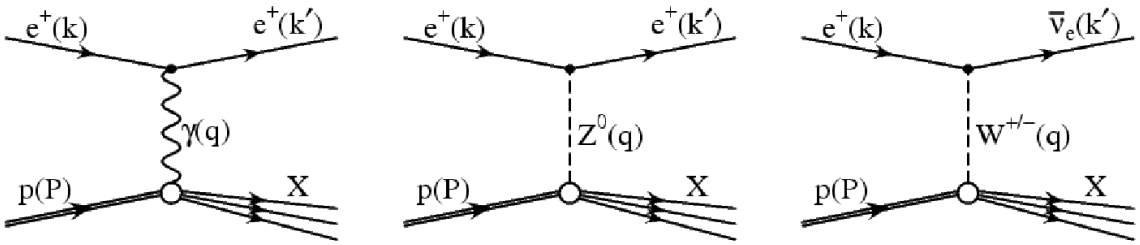


Figure 1.1: Electron - proton interaction. Two neutral current processes and one charged current process.

- heavy-quark production,
- QCD corrections to vector boson production.

Until new and reliable tools for making predictions in the low-momentum region of QCD will be developed, one relies on phenomenological models for describing low-momentum transfer processes. Therefore, if one desires to make QCD predictions with current tools as precise as possible, one needs to understand the parton dynamics within hadrons and the fragmentation of partons into hadrons using non-perturbative phenomenological models.

1.2 Deep-Inelastic Scattering

Generally speaking with deep-inelastic scattering (DIS) one understands a process in which a hadron (usually baryon) is probed by a lepton and which leads to a complete break-up of the hadron. It is this kind of process which allowed to study hadron structure and eventually lead to the discovery of quarks. In the context of the H1 experiment we understand by DIS an electron[3]-proton collision in which the transferred momentum is large enough to resolve the quark structure of the proton ($Q \to \infty$, deep[4]) and in which the invariant mass of the hadronic final state is much higher than the mass of the proton ($m_p \ll m_X$, inelastic). Since electrons do not interact strongly, their interaction with a proton is mediated via a photon, $W^{\pm}$ or Z^0 boson. This is graphically depicted in figure 1.1. When the exchanged boson is neutral, the process is referred to as a neutral current process, when it is charged as a charged current process. In charged current processes the conservation laws require the electron to change into the corresponding neutrino. For the quarks in the proton to be resolved, one needs a short wavelength of the virtual boson and thus a high momentum transfer. The limit is to some extent arbitrary. A process is usually considered to be deep-inelastic when the transferred momentum squared is grater than 1 GeV^2 so that the scattered electron enters into the acceptance of the main H1 detector (operational definition).

[3]Wherever we refer to "electron" in this thesis we refer to an electron or a positron.

[4]The spatial resolution is approximately given by the formula $d \sim \frac{\hbar c}{Q} \simeq \frac{0.197}{Q[\mathrm{GeV}]}[fm]$

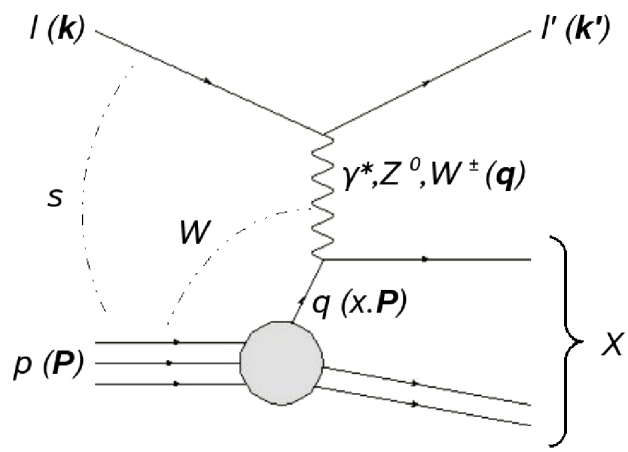


Figure 1.2: Kinematics of the deep-inelastic ep scattering.

1.2.1 Kinematics

The cross sections for unpolarized electron-proton DIS depend on three independent kinematic variables. Since the center-of-mass energy $\sqrt{s}$ is at the accelerator fixed, all predictions can be formulated in terms of two independent variables. In this paragraph we provide a brief overview of some commonly used quantities.

The process of an electron interacting with a proton is, at the lowest order, represented in figure 1.2. The electron radiates a virtual photon, Z^0 or $W^\pm$ which breaks up the proton into the hadronic final state X. Let the proton four-momentum be $\boldsymbol{P}$, the initial electron four-momentum $\boldsymbol{k}$ and the final lepton (electron or neutrino) four-momentum $\boldsymbol{k'}$. They can be related to the center-of-mass energy

$$\sqrt{s} = \sqrt{(\boldsymbol{k} + \boldsymbol{P})^2}$$

and to the negative momentum transfer squared

$$Q^2 = -q^2 = -(\boldsymbol{k} - \boldsymbol{k'})^2 \,,$$

both being Lorentz invariants. Other useful Lorentz invariant quantities are the Bjorken scaling variable

$$x = \frac{Q^2}{2\boldsymbol{P}.\boldsymbol{q}} \,,$$

which can in leading order be interpreted as the fraction of the proton momentum carried by the struck massless quark, and the inelasticity

$$y = \frac{\boldsymbol{P}.\boldsymbol{q}}{\boldsymbol{P}.\boldsymbol{k}} \,,$$

which in the proton rest frame corresponds to the fraction of the electron energy transferred by the exchanged boson to the proton. Finally, one uses the center-of-mass energy of the proton-photon system squared

$$W^2 = (\boldsymbol{P} + \boldsymbol{q})^2 = Q^2 \left(\frac{1}{x} - 1 \right) + m_p^2 \,,$$

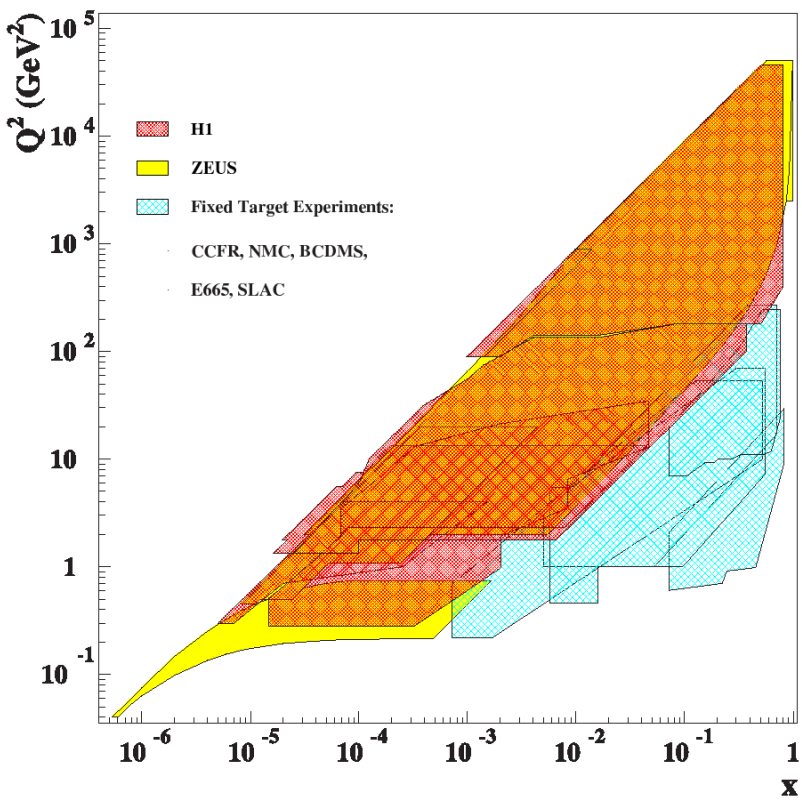


Figure 1.3: Accessible kinematic range at HERA.

which equals the invariant mass squared of the hadronic final state X. Using these formulas, one can relate the introduced quantities in many different ways.

At the HERA accelerator the electron beam energy $E_e = 27.6\,\text{GeV}$ and the proton beam energy $E_p = 920\,\text{GeV}$, as of 1998, provide a center-of-mass energy of $\sqrt{s} = 318.7\,\text{GeV}$. The kinematic range that can be explored at HERA is shown in figure 1.3.

1.2.2 Structure Functions and Inclusive Cross Sections

A neutral current proton-electron interaction can be mediated by a photon or a Z^0 boson. In the kinematic range of our analysis $Q^2 < 100\,\text{GeV}^2$ the Z^0 contribution to the cross section can be neglected because of the high mass of the Z^0 boson ($m_{Z^0}^2 \gg 100\,\text{GeV}^2$). Thus, in the rest of the work we will refer only to photon, which is the only intermediate boson relevant in the context of this analysis. In the one-photon exchange approximation the squared amplitude for the inclusive unpolarized $ep \to eX$ scattering can be written in the form

$$|M_{ep\to eX}|^2 = \frac{e^4}{Q^4} L^{\mu\nu} W_{\mu\nu} \,,$$

where

$$L^{\mu\nu} = \sum_{\text{spin } e} \bar{u}(k')\gamma^\mu u(k)\bar{u}(k)\gamma^\nu u(k')$$

is the so-called leptonic tensor, whose expression follows from the Feynman rules for the electron-photon vertex, and

$$W_{\mu\nu} = \frac{1}{2} \sum_{\text{spin}\, p} \sum_{X} \langle p \,|J_\mu|\, X\rangle \, \langle p \,|J_\nu|\, X\rangle^*$$

is the hadronic tensor. Our inability to express $|p\rangle$ and $|X\rangle$ in terms of quark fields leads us to the parametrization of the hadronic tensor with respect to its known transformation properties. Requiring the correct Lorentz transformations and taking into account the Dirac equation and the Ward identity one can parametrize $W_{\mu\nu}$ with two scalar functions of x and Q^2

$$W_{\mu\nu} = -\left(g_{\mu\nu} - \frac{q_\mu q_\nu}{q^2}\right) W_1(x, Q^2) + \frac{1}{m_p^2}\left(P_\mu - \frac{\boldsymbol{P}.\boldsymbol{q}}{q^2} q_\mu\right)\left(P_\nu - \frac{\boldsymbol{P}.\boldsymbol{q}}{q^2} q_\nu\right) W_2(x, Q^2).$$

Combining the expressions for $L^{\mu\nu}$ and $W_{\mu\nu}$ one arrives (in the *Bjorken limit* $E_p \gg m_p$) at the double-differential cross section formula

$$\frac{d^2\sigma_{\text{NC}}}{dx\, dQ^2} = \frac{4\pi\alpha_{\text{em}}^2}{xQ^4}\left(xy^2 \cdot F_1(x, Q^2) + (1-y) \cdot F_2(x, Q^2)\right), \tag{1.1}$$

with $F_1 = W_1$ and $F_2 = \frac{\boldsymbol{P}.\boldsymbol{q}}{m_p} W_2$. Functions F_1 and F_2 are so-called structure functions of the proton, and they need to be determined experimentally. One often uses a different linear combination of these function and introduces $F_L = F_2 - 2xF_1$. Then the cross section formula has the form

$$\frac{d^2\sigma_{\text{NC}}}{dx\, dQ^2} = \frac{4\pi\alpha_{\text{em}}^2}{xQ^4}\left((1 - y + \frac{y^2}{2}) \cdot F_2(x, Q^2) - \frac{y^2}{2} \cdot F_L(x, Q^2)\right).$$

The function F_2 dominates the expression, F_{L} plays a role only at large values of y.

1.2.3 Quark Parton Model and Evolution Equations

The proton structure functions introduced in the previous section are well-defined objects within the framework of the Standard Model. However, they have the disadvantage to characterize only $ep \to eX$ scattering, one cannot use them directly in different processes (e.g. proton-hadron scattering). One would thus like to find objects that describe the proton as such, in a process-independent way. Such a description obviously requires some kind of factorization theorem, which would allow to separate the description of the parton behavior in the proton and the parton hard interaction.

In the "naive" parton model proposed by Feynman [5] and as further developed by Bjorken [6] for inelastic ep scattering, the proton consists of three partons identified with the charged, spin $\frac{1}{2}$ point-like quarks of the proton. A high-energy collision with a lepton is described as a photon interacting with one of the quarks (which are regarded as free as explained later), the parton (quark) carrying the momentum fraction ξ of the proton momentum. The proton is regarded as moving very rapidly

(infinite-momentum frame) so that the parton momenta are collinear and the partons with different momenta remain together. Their momentum distributions are described by probability functions $q_i(\xi)$ (so-called parton distribution functions), which give the probability for the parton i to carry the proton momentum fraction between ξ and $\xi + d\xi$. The assumption of free partons is mostly based on the fact that the typical time scale of the hard interaction is much smaller than the typical time scale of parton interactions inside the proton when Q^2 and W are large enough. Supposing that the parton-lepton interaction is described like a fermion interaction in QED one can, within the parton model, calculate the $ep \to eX$ cross section and compare it to the expression (1.1). The comparison leads to

$$F_2(x) = 2xF_1(x) = \sum_{\text{quarks}} \int_0^1 d\xi\, q(\xi)\, x e_q^2 \delta(x - \xi) = \sum_{\text{quarks}} e_q^2 x q(x),$$

x being the Bjorken scaling variable. It is interesting to note that in the parton model the structure functions lose their Q^2 dependence. This property is known as Bjorken scaling. And indeed, scaling is observed at $x \approx 0.1$ (see figure 1.4), but at lower and higher x values scaling is broken. Furthermore, the sum of parton momenta predicted by the parton model reaches only approximately half of the proton momentum. This suggests that the "naive" parton model, in spite of being quite successful, does not provide a description precise enough to explain all measured data and thus needs to be improved.

An improvement can be achieved by considering QCD effects when describing the parton dynamics inside the proton. In addition to the basic leading order photon-quark scattering diagram with the QED vertex (see figure 1.5 a) one can take into account real gluon emission from the quark lines (QCD vertices, see figures 1.5 b and c) and calculate the corresponding contributions to the total scattering amplitude in pQCD. Such an approach intuitively explains the violation of Bjorken scaling at higher x values ($x \lesssim 1$). Whereas a low-virtuality photon cannot resolve a possible gluon radiation from the quark, a high-virtuality photon probes smaller distances and thus can interact with a quark after the quark has radiated a gluon (see figure 1.6). In the latter case, the photon effectively interacts with a quark carrying less momentum. At smaller x values ($x \ll 1$) the scaling violation is related to other QCD effects: virtual gluon radiation off valence quarks and gluon splitting into virtual sea-quarks. These effects also account for the missing proton momentum: only a part of the proton momentum is carried by the valence quarks, a large part is also carried by gluons and the sea-quarks. The pQCD approach, however, makes the model more complex and introduces two new scales. The first one is the renormalization scale μ_R, which is related to the renormalization of the strong coupling α_s appearing in the QCD vertices. The second is the factorization scale μ_F. Since the corrections to the parton dynamics are calculated perturbatively one requires some hard scale for these calculations to be valid. Above this scale the partons dynamics is treated perturbatively, below this scale it is absorbed into the non-perturbative part of the parton density functions (PDFs). The divergences that appear in the perturbative calculations, which are due to collinear and soft gluon radiation, are also absorbed

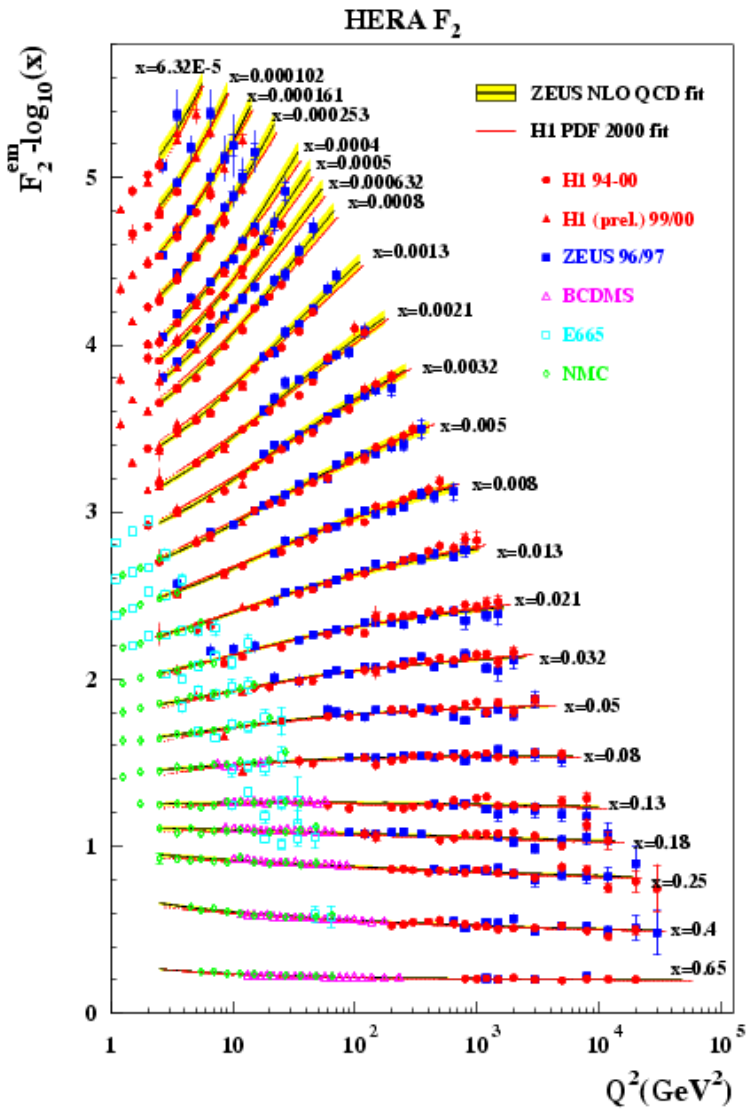


Figure 1.4: Measured values of $F_2(x, Q^2)$ at different experiments. Real gluon radiation and creation of sea-quarks, antiquarks and gluons and can account for scaling violations at high and low x and for the missing proton momentum.

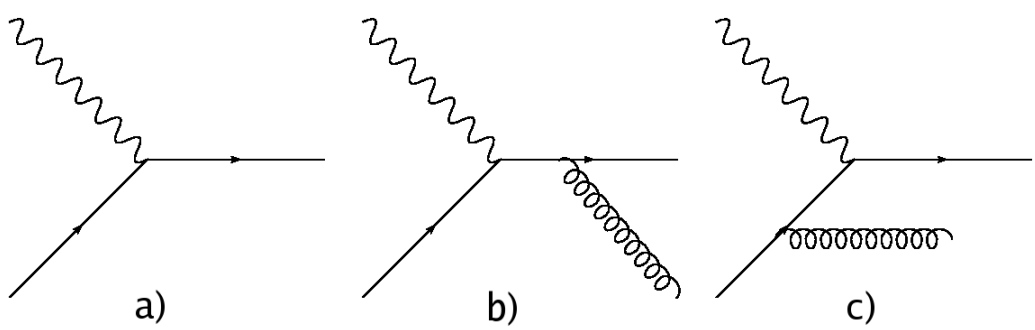


Figure 1.5: Photon-quark interaction and gluon emissions.

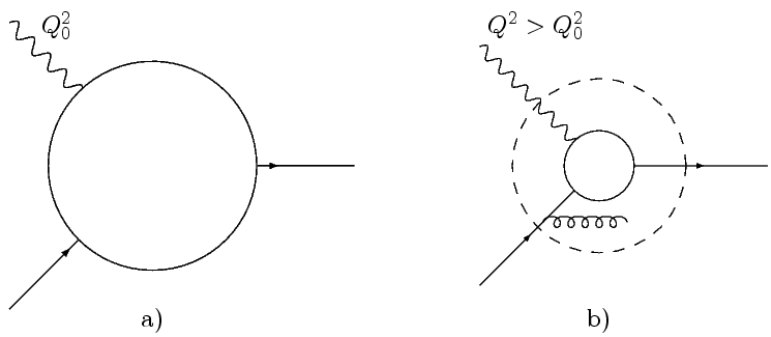


Figure 1.6: Intuitive interpretation of the scaling violation for $x \lesssim 1$.

into the non-perturbative parts of the PDFs. In DIS one usually chooses $\mu_R^2 = \mu_F^2 = Q^2$. The factorization theorem states that the long and short distance interactions factorize, i.e. one obtains, within the improved quark parton model, the following formula for the total inclusive cross section:

$$\sigma(x, Q^2) = \sum_f \int_x^1 \frac{d\xi}{\xi} \hat{\sigma}(\xi, Q^2, \mu_R, \mu_F) q_f(\frac{x}{\xi}, Q^2, \mu_F) \,, \tag{1.2}$$

where $\hat{\sigma}$ is the hard partonic cross section which is perturbatively calculable , q_f are the parton density functions and the summation is done over valence quarks, gluons and sea-quarks and antiquarks.

The parton density functions need to be determined experimentally. It is however enough to measure them at one scale μ_0, their behavior at a different scale can be predicted from evolution equations. These evolution equations are derived from the requirement that the physical cross section should not depend on an (arbitrary) factorization scale μ_F. Depending on how the gluon radiation is treated (approximated) one gets different prescriptions. In the majority of approaches one represents the gluon radiation by a ladder diagram (see figure 1.7), and depending on what suppositions are made on the gluon emissions one obtains different evolution equations:

- The DGLAP[5] formalism [7, 8, 9, 10] predicts the evolution of the PDFs in Q^2. It assumes strong ordering in the virtuality of the exchanged gluons $k_0^2 \ll k_1^2 \ll \ldots \ll k_{n-1}^2 \ll k_n^2 \ll Q^2$, which at small x implies strong ordering in the transverse momenta $k_{t,0}^2 \ll k_{t,1}^2 \ll \ldots \ll k_{t,n-1}^2 \ll k_{t,n}^2 \ll \mu_F^2$. It also requires the longitudinal momenta $x_i \boldsymbol{P}$ to be greater than the transverse momenta (collinear factorization). With these assumptions the DGLAP evolution is expected to be valid at high Q^2 and not too small x. In fact, it describes successfully the measurements of the structure function at HERA down to the smallest x accessible in the experiments.

- The BFKL[6] approach [11, 12] describes the evolution of the PDFs in x. The longitudinal momenta of the parton propagators are supposed to be strongly

[5] Dokshitzer, Gribov, Lipatov, Altarelli, Parisi
[6] Balitsky, Fadin, Kuraev, Lipatov

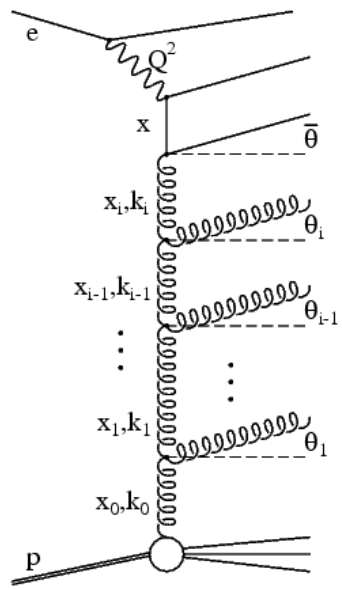


Figure 1.7: Gluon emission ladder diagram.

ordered $z_i = x_{i+1}/x_i \ll 1$, but no restrictions are applied to the transverse momenta k_t. Unlike the PDFs in the DGLAP formalism, the PDFs here depend explicitly on the transverse momentum k_t (so-called unintegrated PDFs) of the gluon in the proton. This approach is supposed to be valid at low values of x.

- The CCFM[7] model [13, 14, 15, 16] has the ambition to describe the evolution of the PDFs at both, small and large x. It provides an evolution in Q^2 and x in the region of large Q^2 and moderate x in agreement with the DGLAP approach and in the small x region according to the the BFKL approximation. It is based on the strong angular ordering of subsequent parton emissions, and it makes use of k_t dependent unintegrated gluon densities.

1.2.4 Heavy Quark Production in DIS

The dominant heavy quark production process is the so-called boson-gluon fusion process (BGF) represented in figure 1.8 a. Non-negligible contributions to heavy quark production at low Q^2 may arise also from resolved-photon processes which are depicted in figures 1.8 b-d. In resolved processes the photon fluctuates into virtual hadronic states, which consequently interact with the gluon coming from the proton. The photon fluctuations can be described by the photon parton density functions. The heavy quarks are produced almost exclusively in the perturbative regime, their production in the fragmentation phase is very improbable.

At the HERA accelerator pairs of charm ($m_c \approx 1.3$ GeV) and bottom quarks ($m_b \approx 4.5$ GeV) can be produced via the BGF process. The top quark ($m_t \approx 175$ GeV) cannot be observed at HERA, because the beam energy is not high enough

[7] Ciafaloni, Catani, Fiorani, Marchesini

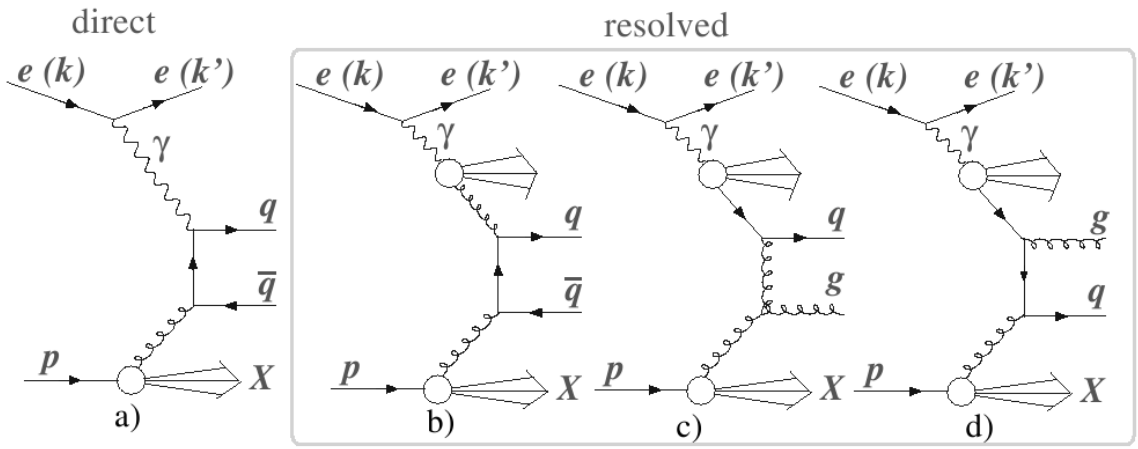


Figure 1.8: DIS charm production in leading order of perturbative QCD, direct process and resolved photon processes.

to allow the pair production of top quarks, and the single top production process has a very low cross section. On the other hand the charm and bottom quark production cross sections represent an important part of the total inclusive cross section. The fraction of charm production vs. inclusive QCD processes is of order 10% in the perturbative QCD regime, the beauty production is with respect to charm suppressed by two orders of magnitude (higher b mass and smaller electromagnetic coupling) [17].

The heavy quarks are interesting objects to be studied. Their high mass provides a hard scale for perturbative calculations so that heavy quark physics can serve as a good test field for perturbative QCD calculations. Since their production is dominated by gluon-induced processes, heavy quark physics also provides information about the gluon density in the proton. In addition, the high mass of heavy quarks implies hard fragmentation, and thus one can expect a good correlation between parton and hadron level, between the heavy quark and the corresponding jet of hadrons.

1.3 Fragmentation

1.3.1 General Considerations

Non-perturbative effects in particle collisions are associated not only with initial state hadrons but also with final state hadrons. If only inclusive or sufficiently inclusive observables are studied, the detailed character of these effects can be neglected. This is why the expression (1.2) does not contain any information about hadron formation from final state partons. On the same basis jet observables can be studied without precise knowledge of the fragmentation, provided that these observables do not depend on details of the jet internal structure.

However, a description of the fragmentation phase is needed, if one is interested in details of the hadronic final state. This description is based on the QCD fac-

torization theorem; the basic idea being that physics at very different scales can be factorized. The hadron formation from partons is modeled using so-called fragmentation functions $D_{i\to h}(z,\mu_f)$ which are formally very similar to parton density functions and which give (if normalized) the probability to observe a hadron h carrying the momentum fraction z of the initial parton i. The differential cross section for the production of hadron h as a function of its transverse momentum p_T can be schematically written in the form of the convolution

$$\frac{d\sigma_h}{dp_T} \sim \sum_{ik} q_i(x,\mu_f) \otimes \frac{d\widehat{\sigma}_{i\gamma\to kX}}{dp_T} \otimes D_{k\to h}(z,\mu_f) + \mathcal{O}\left(\frac{\Lambda_{\mathrm{QCD}}}{p_T}\right),$$

where $q_i(x,\mu_f)$ are scale dependent parton density functions, $\widehat{\sigma}_{i\gamma\to kX}$ is the partonic cross section for parton k to be produced in the photon-quark interaction which can be calculated in pQCD and $D_{k\to h}(z,\mu_f)$ are scale dependent fragmentation functions. Summation is done over initial and final state partons. The fragmentation functions need to be established experimentally at some initial scale μ_0 (usually low), then they can be evolved using DGLAP-inspired evolution equations to any other scale μ_f. The choice of the fragmentation scale is to some extent arbitrary. The evolution of the quark produced in the hard sub-process can be part of the perturbative calculations or it can be absorbed into the fragmentation function. In practice our knowledge of the perturbative quark evolution is limited by our ability to calculate higher orders in perturbation theory. In addition, divergences originating from collinear radiation emerge in calculations. These divergences can be absorbed into the non-perturbative fragmentation functions.

1.3.2 Heavy Quark Fragmentation

Due to the higher masses of the heavy quarks it can be shown [18] that the fragmentation function for a heavy quark Q can be split into two parts: a fully perturbative and process independent fragmentation function $D_Q^{\mathrm{pert}}(z,\mu_f)$ and a non-perturbative but scale independent fragmentation function $D_{Q\to H}^{\mathrm{np}}(z)$, H being the heavy hadron. The function $D_Q^{\mathrm{pert}}(z,\mu_f)$ accounts for the perturbative gluon radiation off the quark and can be evolved via DGLAP-like equations. The collinear logarithms generated by gluon emissions are resumed to all orders and can be absorbed into the perturbative cross section. The non-perturbative fragmentation function $D_{Q\to H}^{\mathrm{np}}(z)$ describes the hadronization of the heavy quark which is, after having radiated gluons, almost on its mass-shell.

Many different parametrizations are available to describe the non-perturbative fragmentation functions. They are based on different models which all agree that heavy quark fragmentation is hard, i.e. the heavy quark looses on average only a small fraction of its momentum when turning into a hadron (in contrast to light quark fragmentation). This is easily explained using a simple kinematic argument (Bjorken [19], Suzuki [20]): the momentum of a massive quark is only slightly influenced when picking up a light quark from the vacuum. Supposing the typical light quark mass to be of the order of Λ, one expects from momentum conservation $m_Q.v_Q \approx$

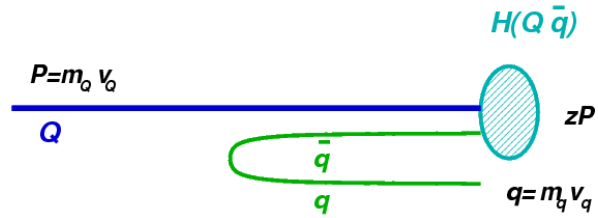


Figure 1.9: Heavy quark Q fragmenting into a heavy hadron H.

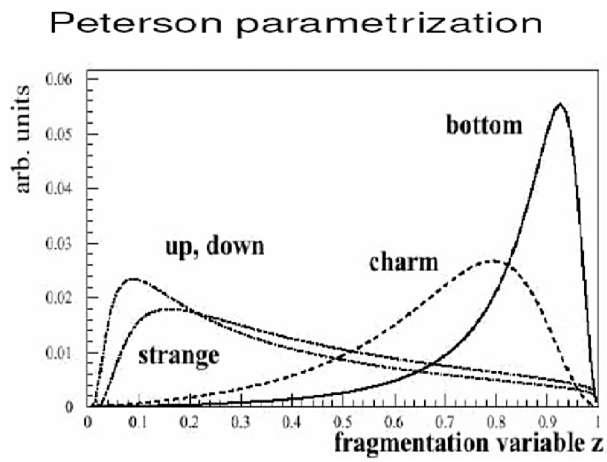


Figure 1.10: Fragmentation functions for quarks with different masses based on the Peterson parametrization. Appropriate parameter values are supposed.

$z(m_Q.v_Q) + \Lambda.v_q$, $v_Q \approx v_q$, from which follows $\langle z \rangle^{\text{np}} \approx 1 - \frac{\Lambda}{m_Q}$ (see figure 1.9). The heavier the quark is, the harder the fragmentation is expected to be. An illustrative example based on the Peterson parametrization of the fragmentation functions is shown in figure 1.10. Two commonly used parametrizations for the non-perturbative fragmentation function which depend on only one parameter are:

- Peterson parametrization [21]. It is derived from quantum mechanical considerations on transition probabilities between two energetically close states with the energy difference $\triangle E$. The considered energy difference is $\triangle E = E_H + E_q - E_Q$, q stands for a light quark forming a hadron with a heavy quark. The predicted parametrization has the form

$$D^{\text{np}}_{Q \to H}(z) = N \times \frac{1}{z}\left(1 - \frac{1}{z} - \frac{\varepsilon}{1-z}\right)^{-2},$$

  where N is a normalization factor, and ε is a parameter related to the hardness of the fragmentation. The parameter ε is supposed to be of the order of $\frac{m_q^2}{m_Q^2}$, but since the uncertainty on light quark masses is big, ε is usually considered as a free parameter which can be fitted in order to describe the data. The Peterson parametrization provides a rather precise prediction for the ratio of the ε parameters for charm and beauty quark fragmentation $\frac{\varepsilon_c}{\varepsilon_b} = \frac{m_c^2}{m_b^2} \approx 0.1$.

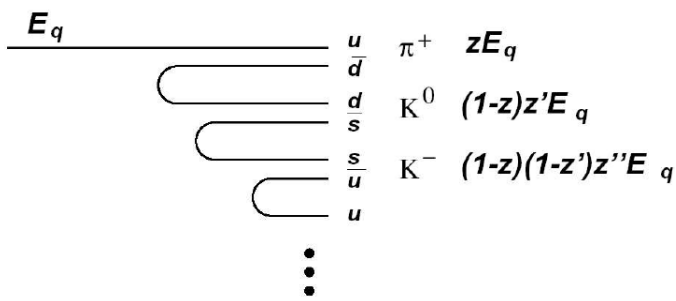


Figure 1.11: Field-Feynman fragmentation model.

- Kartvelishvili parametrization [22]. Here the authors assumed that the fragmentation function $D^{\rm np}_{Q\to H}(z)$ behaves at large values of z like the density function $f^Q_H(z)$ of the heavy quark Q in the hadron H ("reciprocity relation"), the density function being a function of the momentum fraction of the hadron carried by the quark. Using the Kuti-Weisskopf model [23] for calculating the density functions and extrapolating the validity of the "reciprocity relation" to all values of z, they arrived at the parametrization:

$$D^{\rm np}_{Q\to H}(z) = N \times z^{\alpha}(1-z)\,,$$

  where N is a normalization factor and α a parameter which equals 3 for the charm quark and 9 for the bottom quark. When the Kartvelishvili parametrization is nowadays used, the parameter α is regarded as a free parameter related to the hardness of the fragmentation.

1.3.3 Fragmentation Models

Fragmentation models are phenomenological models more or less inspired by insights from QCD which are used in different Monte Carlo programs (we discuss them in the following section) in order to describe the fragmentation process. Some of the models take a fragmentation function in a parametrized form, while other models do not make any use of an explicit fragmentation function at all but rely on different mechanism to produce hadrons from quarks. In this section we briefly summarize the most common models.

Independent Fragmentation Model

Independent fragmentation is based on the idea of a parton fragmenting independently of other final state partons. One of the well-known models assuming independent fragmentation is the so-called Field-Feynman model [24]. In this model the fragmenting quark forms a bound state with one of the quarks of a quark-antiquark pair produced from vacuum fluctuations. The remaining quark pairs up with a quark from a pair again produced by fluctuations... the mechanism continues recursively, as indicated in figure 1.11. The momentum fraction which is transferred from ini-

tiating quark to the produced hadron is described by a fragmentation function; the Peterson or Kartvelishvili parametrizations are commonly used.

The independent fragmentation model is nowadays rarely used. It has the disadvantage of not being Lorentz invariant and requires an *ad hoc* treatment of the remaining quark from the very last quark-antiquark pair production. While it describes many general features of particle production it fails to describe experimental data in detail at the high level of experimental precision which has been reached. The independent fragmentation model is implemented in the PYTHIA Monte Carlo program as a non-default option.

String Fragmentation Model

In string fragmentation model [25, 26] the fragmentation of final state partons depends on the other final state partons. The linear confining potential between partons is modeled via a string - a massless and relativistic object having the form of a color flux tube with typical transverse dimensions of the order of the hadronic size ($\approx$ 1fm). A string is stretched between two quarks, an energetic gluon produces a "kink" in the string. The typical string tension is $\kappa \approx 1$ GeV/fm, and as the quarks move apart the string may break and produce a quark-antiquark pair from the vacuum. The probability of pair creation is modeled by the quantum tunneling process and obeys the formula $\exp(-\pi m_{q,\perp}^2/\kappa)$, where $m_{q,\perp}^2 = m_q^2 + p_{q,\perp}^2$ is the transverse mass. Since this probability depends on the quark mass, heavy quark production in the fragmentation process is strongly suppressed. The string breaking respects the neutral color of daughter strings. The transverse momenta of produced quarks are assumed to follow a Gaussian distribution and are locally compensated between the quark and the antiquark. The string breaking occurs recursively until on-mass-shell hadrons are formed, the hadron transverse momentum being the sum of the transverse momenta of the quarks. The variable[8] $z = \frac{(E+p_z)_h}{(E+p_z)_{\text{string}}}$, indicating which fraction of the quantity $E + p_z$ is transferred from the string to the hadron, is determined by the fragmentation function $f(z)$. Usually the Lund symmetric fragmentation function is used $f(z) = \frac{1}{z}(1-z)^a \exp\left(-\frac{bm_{h,\perp}^2}{z}\right)$, where a and b are free parameters. This function provides independence of the result of the string breaking on the end of the string at which the breaking starts. To account for specific aspects of heavy quark fragmentation (its hardness), often a different fragmentation function (Peterson, Kartvelishvili, Bowler) is used.

Hadron formation is based on spin states counting and some additional model complications must be introduced to describe baryon formation (diquark production). The model also needs to cope with different string topologies which can occur.

In charm production in DIS via the BGF process at least two strings are required in order to ensure color neutrality. One string connects the produced antiquark with a quark from the proton remnant, the other one connects the produced quark with the remnant diquark, as seen in figure 1.12.

[8] This definition of z has the advantage to be Lorentz invariant under the boost along the parton direction.

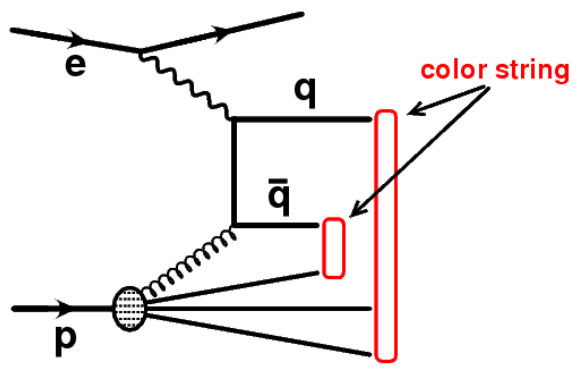


Figure 1.12: Color strings in the BGF process.

The string model is commonly used and is implemented in the Monte Carlo program PYTHIA (and JETSET).

Cluster Fragmentation Model

The cluster fragmentation model, which is realized in the HERWIG Monte Carlo program [27, 28] exploits the property of "pre-confinement" [29]. It states that at the end of the perturbative phase color-connected partons tend to be close in phase space, a local compensation of color occurs. Thus such quarks are merged into color-singlet clusters, but before the cluster formation takes place each gluon is forced to split (non-perturbatively) into a quark-antiquark pair (see figure 1.13). Typically clusters have small mass as of a couple of GeV, and subsequently (with the exception of too heavy or too light clusters) they decay independently directly into hadrons. Unless the cluster involves a perturbatively produced quark, its decay is in its rest frame isotropic with no angular momentum involved. If a cluster is too heavy, it is split into two clusters and too light clusters decay into one hadron only. In the latter case a small rearrangement of energy and momentum with neighboring clusters is needed (the decays might not be completely independent). The choice of the cluster's decay channel is based on the phase space probability and spin degeneracy.

The cluster fragmentation model is in spite of the small number of parameters quite successful and provides a fair description of the data. In general, however, the string model provides an improved description of different data.

1.4 Simulation Programs

Present experiments in high-energy physics need computer-based simulations in order to determine the response of the detector . They allow to obtain different correction factors which are usually related to

- fiducial acceptance of the detector and extrapolations to experimentally non accessible phase space regions,

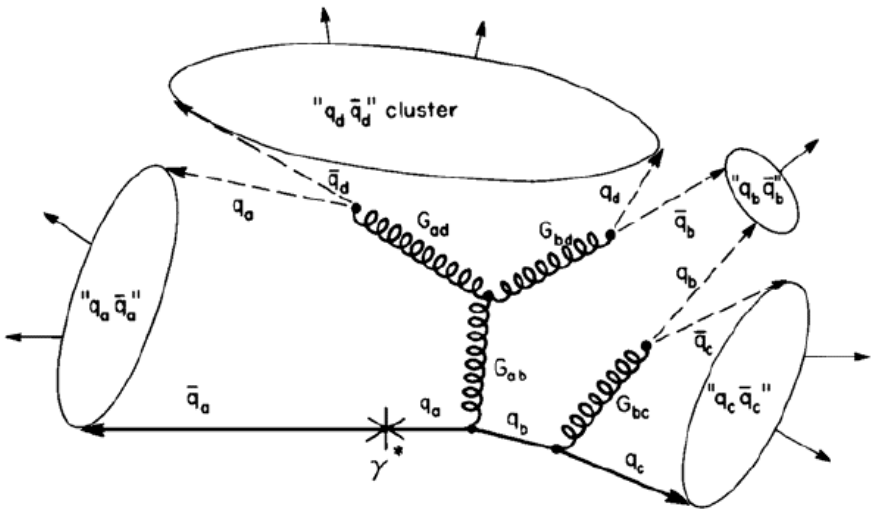


Figure 1.13: Cluster fragmentation model: gluons are split into quark-antiquark pairs and color-connected quarks are merged into color singlet clusters which subsequently decay into hadrons.

- resolution effects and migrations,
- effects of initial and final state QED radiation.

A typical simulation includes an event generator program and a detector simulation program. The event generator simulates the physics process or processes and provides a complete set of final state partons as well as particles with their four-momenta. Such a program is usually based on a random number generator and thus it is called a Monte Carlo program. Once the information on the final state particles is available, the detector response to the physics event is simulated. In this analysis the internal H1 GEANT-based [30] software package H1SIM is used for this purpose. After the detector simulation the data format of the simulated physics is the same as the data format of the real physics data. Thus both can be treated in the same way and be processed by the reconstruction software H1REC.

In the rest of this section we focus on different event generators which we used and the differences between them. An event generator may comprise :

- a calculation of the matrix elements of the hard subprocess (in leading order, no event generator with next-to-leading order matrix elements is so far available for *ep* physics in DIS).
- parton density functions or an interface to a program for them and for their evolution.
- an approximation of higher order contributions to the perturbative processes. This is accomplished using different types of parton shower mechanisms, where

an energetic quark or gluon radiates further gluons and can be associated with the perturbative fragmentation function.

- hadron formation (non-perturbative fragmentation) and hadron decays into stable particles[9].
- additional effects due to QED radiation.
- multiple interactions. Such effects may occur when the initial states consist of several partons (hadron-hadron collisions) and correspond to additional hard or semi-hard scatterings. These effects do not seem to play an important role at HERA in DIS and can be safely neglected in this analysis.

1.4.1 Event Generators with LO Matrix Elements

RAPGAP

RAPGAP [31] is an event generator incorporating leading-order QCD matrix elements. It is based on the collinear factorization scheme (DGLAP evolution) using for that purpose modified routines from the LEPTO 6.1 [32] and PYTHIA 6.2 [33] programs. It includes contributions from higher orders by parton showers in the leading log approximation matched to the LO matrix element such as to avoid double counting. The charm quark is treated as massive ($m_c = 1.5\,\mathrm{GeV}$) and the renormalization scale is chosen to be $\mu_r^2 = Q^2 + p_T^2$ (for heavy quarks $\mu_r^2 = Q^2 + p_T^2 + m_{\mathrm{HQ}}^2$). Resolved-photon processes[10] in heavy quark production are also implemented, resolved-photon events are generated separately from direct events. Effects of real photon emission and virtual QED contributions are simulated by interfacing RAPGAP with the HERACLES [34] event generator. The fragmentation in RAPGAP is done with the Lund-sting model as it is implemented in the PYTHIA program.

In this analysis we use RAPGAP version 3.1, the CTEQ5L [35] parametrization of the PDFs of the proton and SaS-G 2D [36] for the PDFs of the photon in case of resolved processes.

CASCADE

The CASCADE Monte Carlo program [37] differs significantly from RAPGAP. It employs CCFM evolution with an unintegrated (k_t - dependent) gluon density function. In this evolution the gluon which enters the hard sub-process is virtual and thus requires an off-shell matrix element. Only the direct BGF process is considered in charm quark production, but the k_t-factorization scheme partially includes a hadronic photon component (see reference [38]). The renormalization scale is chosen to be $\mu_r^2 = 4m_c^2 + p_T^2$ and like in the case of RAPGAP, CASCADE is interfaced with

[9] The definition of a stable particle is to some extent arbitrary. By a stable particle we understand all particles which do not decay further or which have a high probability of decaying in the tracking detector.

[10] We use them in the evaluation of systematic errors.

the PYTHIA 6.2 program to account for the fragmentation (Lund-string model) and particle decays.

In this analysis we have chosen CASCADE as an alternative program to RAPGAP to correct our data and thereby to study the model dependence of our results. The version CASCADE 1.2 with the A0 [39] gluon density functions is used.

Chapter 2

HERA Accelerator and H1 Detector

2.1 HERA Accelerator

The HERA[1] accelerator (figure 2.1) was[2] a particle accelerator situated in Germany in the city of Hamburg at DESY[3] research institute. It was a unique high-energy collider with asymmetric beams: an electron beam colliding with a proton beam, each beam stored in an independent storage ring. It had approximately circular shape, and it was housed in a 6336 m long tunnel with internal diameter of 5.2 m situated 10 to 25 m under ground.

The first idea of HERA was brought in the early 1970s by the Norwegian physicist Bjørn H. Wiik to DESY and the project study followed in 1980. Agreement for the construction of the facility was signed in 1981. Construction started in 1984 with international support and first operation of HERA started in 1990 (first physics run in 1992). In 2000-2001 HERA underwent a luminosity upgrade to HERA II with a difficult restart of operation in 2002.

The operation of HERA required a chain of pre-accelerators (LINAC II and III, DESY synchrotron II and III and PETRA II and III) in order to provide HERA with electrons and protons at suitable energies. Once the protons and electrons were injected into HERA, HERA took over and accelerated them to their final energy. The designed final energy (after 1998) for the two types of particles was

$$E_p = 920\,\mathrm{GeV},$$

$$E_e = 27.6\,\mathrm{GeV},$$

with 96 ns between two following bunch-crossings. In total 174 colliding bunches were stored in each storing ring and the beam crossing angle at the interaction point was zero. The electron beam was naturally transverse-polarized (Sokolov-Ternov

[1] Hadron Electron Ring Anlage

[2] The HERA collider was shut-down, after 15 years of successful operation, at midnight on 30.06.2007.

[3] Deutsches Electronen-Synchrotron

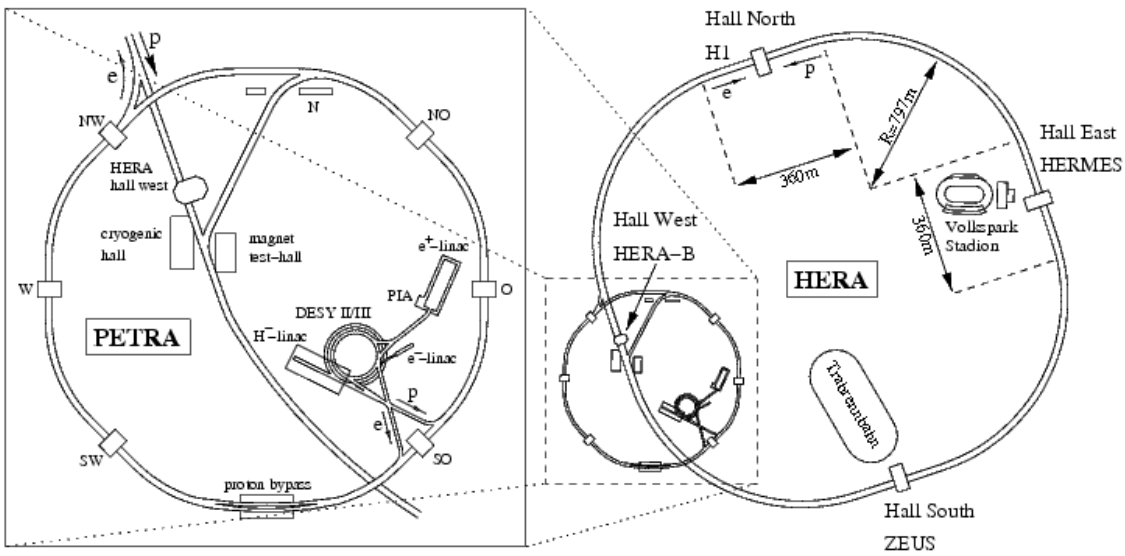


Figure 2.1: HERA accelerator and pre-accelerator subsystems, HERA experiments.

mechanism) and a longitudinally polarized electron beam could be obtained with a system of spin rotators (installed after the upgrade to HERA II).

Initially four experiments were designed to use the HERA beams. The HERA-B experiment was using only the proton beam on a fixed target. It was shut down prematurely because of unexpected and serious problems during the detector construction and the following loss of competitiveness with other world experiments exploring the same physics area. The HERMES experiment was using the unpolarized and polarized electron beam on a fixed unpolarized and polarized target to study mainly the spin structure of the nucleon. The two remaining experiments, H1 and ZEUS, were colliding-beam experiments with multipurpose detectors studying the proton structure and other topics related to proton-electron collisions. Sharing the same beams and using complementary detectors allowed these two experiments to cross-check their physics results.

2.2 H1 Detector

The H1 detector was located in the north hall of the HERA ring. It was a multipurpose detector covering most of the solid angle around the nominal interaction point. Its approximate dimensions were $12 \times 10 \times 15\,\mathrm{m}^3$ with the weight of 2800 t. Because of the asymmetry in beam energy, the center of mass of the electron-proton system was boosted along the proton direction (we will refer to this direction as "forward" or "positive"), which was reflected in an asymmetric design of the detector. The H1 detector consisted of most of the standard detector subsystems used in a high-energy physic experiment: a high-resolution tracking system, a fine-granularity calorimetric system and muon detectors. The tracking system as well as the calorimeter were situated within a magnetic field of strength 1.2 T which was provided by a supercon-

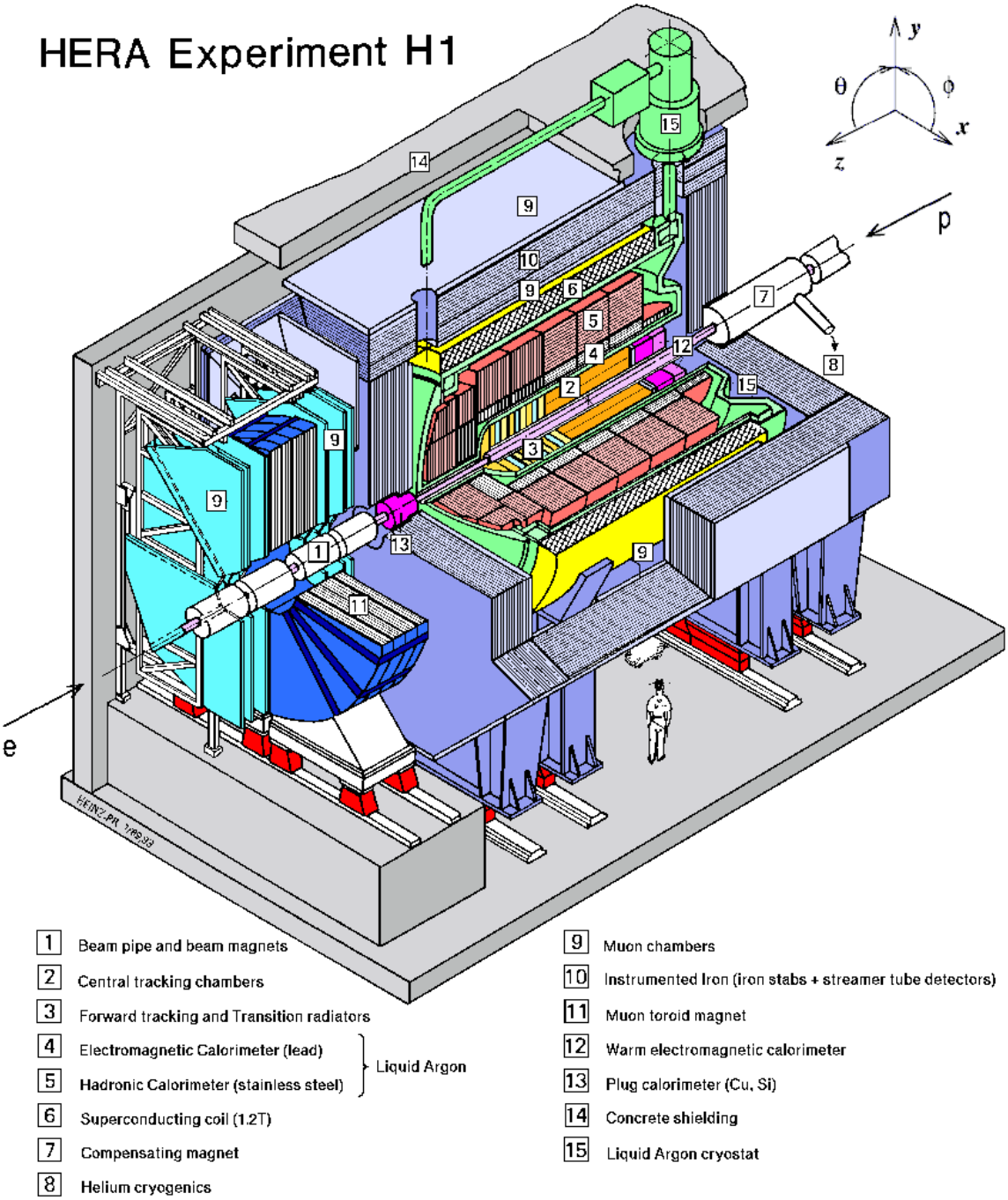


Figure 2.2: The H1 detector and the H1 coordinate system.

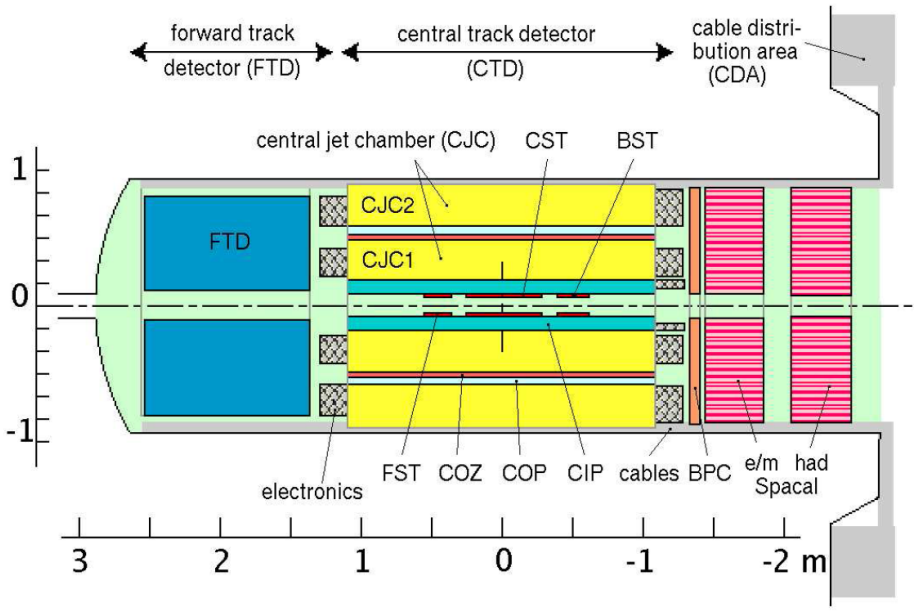


Figure 2.3: Design of the H1 tracking system.

ducting coil. Such a design allowed for less dead material in front of the calorimeter and better particle identification (especially electron identification) in the electromagnetic part of the calorimeter. The main components of the H1 detector and the H1 coordinate system can be identified in figure 2.2. The running of the H1 detector required an intelligent and highly efficient trigger system, since the time window between two successive bunch-crossings was very short and the background was high.

The hardware components of the H1 detector as well as the trigger system underwent important upgrades during the luminosity upgrade of HERA in 2001 and 2002. This analysis is based on the data from the HERA II running period and thus we will describe the detector status in this period, only briefly mentioning the status before.

In our analysis we make no or only a small use of such subsystems as the muon chambers, the plug calorimeter or the tail-catcher system. Thus we will omit their description in the following sections, and we will focus only on those detector subsystems that are related to this analysis. An interested reader may consult references [40, 41] for more information.

2.2.1 Tracking System

The beam asymmetry at HERA was also reflected in the design of the H1 tracking system (figure 2.3). The system consisted of the central and forward tracking detectors (CTD and FTD) and a backward tracking system, which was installed in front of the backward calorimeter to improve the spatial resolution of the calorimeter, reject photoproduction background induced by neutral hadrons and improve γ/e separation. The tracking system provided besides the information for track reconstruction

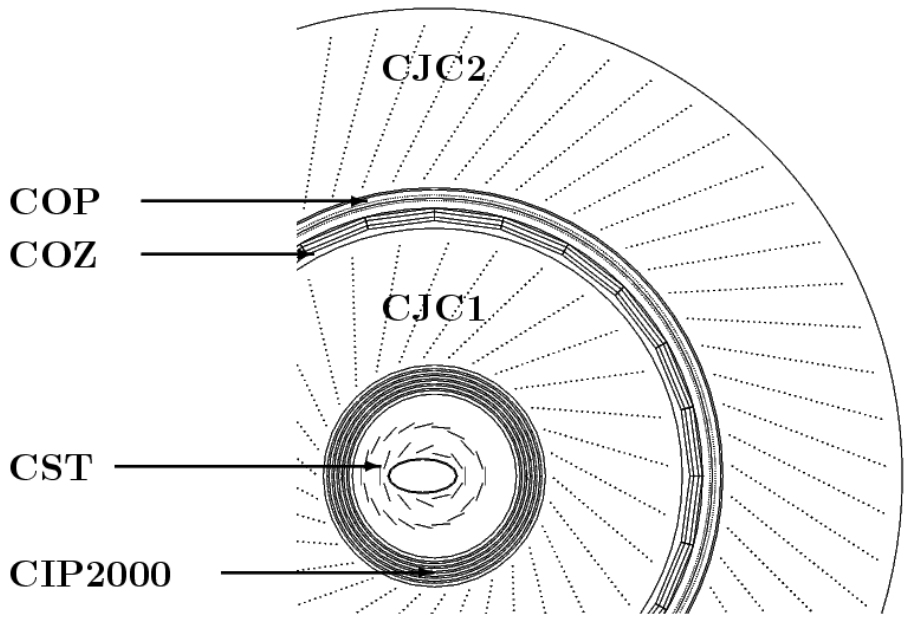


Figure 2.4: Radial view of the central tracking system.

also the capacity for track-based triggering and for particle identification via dE/dx. The H1 tracking system is, for this analysis, of crucial importance, because the tracking information is the main input for the reconstruction of the D^* mesons. Since we apply a constraint on the pseudorapidity[4] of the D^* mesons $|\eta_{D^*}| < 1.5$ we use for the D^* reconstruction only the most precise information from the central tracking detector.

Central Tracking Detector

The H1 central tracking detector in its initial form consisted of two concentric multiwire drift chambers CJC1 and CJC2[5], of two thin concentric drift chambers CIZ and COZ, providing precise measurements of the z-coordinate, and of two concentric proportional chambers CIP and COP. The central and backward silicon trackers (CST and BST) were installed during HERA I running in close proximity of the beam pipe to provide precision spatial measurement. The CJC1 and CJC2 as well as the COZ, the COP, the CST and the BST remained after the detector upgrade for HERA II running. During the upgrade an additional silicon detector was added to cover the forward region (FST). The CIZ and CIP were removed and replaced with CIP2000 - a cylindrical proportional chamber with more layers for improved triggering and vertex reconstruction. The status of the CTD after the upgrade is depicted in figure 2.4. The angular coverage[6] of the CTD was $25° < \theta < 155°$.

The CJC1 and CJC2 were the most important components of the CTD, their main parameters were $-112.5\,\text{cm} < z < 107.5\,\text{cm}$ and $20.3\,\text{cm} < r < 84.4\,\text{cm}$. The

[4]For the definition of the pseudorapidity see section 3.2.3.

[5]Central Jet Chamber 1 and 2

[6]The angle θ is measured with respect to the "positive" (proton beam) direction.

sense wires of these chambers were organized in drift cells and were parallel to the beam axis. The CJC1 contained 30 drift cells with 24 sense wires per cell, the CJC2 comprised 60 drift cells with 32 sense wires each. The drift cells were inclined by 30° with respect to the radial direction so as to optimize the electron drift direction in the magnetic field for nearly straight high momentum tracks. Single hits were reconstructed with a spatial resolution of $\sim 170\,\mu$m in the $r\phi$-plane from the drift-time measured by the sense wires. The information about the z coordinate could be extracted from a comparison of the signals at both ends of the sense wire with a precision of 22.0 mm. The precision of the momentum measurement for reconstructed tracks was $\sigma_p/p^2 < 0.01\,\mathrm{GeV}^{-1}$. After track reconstruction one can also determine the energy loss of particles with established precision of $\sigma_{dE/dx} \approx 6\%$. The knowledge of the energy loss helps - depending on the particle's momentum - in particle identification or at least in the calculation of the probability for a particle to be of a certain type. The chambers CJC1 and CJC2 were also used in the reconstruction of the interaction vertex. The x and y position of the vertex can be deduced from track extrapolation inside the beam pipe, the information about the z position was however determined with better precision from other chambers.

The CIP 2000 [42, 43] was a five layer proportional chamber, which was designed to provide - in comparison to the CIP in HERA I - a more efficient vertex trigger with larger solid angle acceptance and better rejection capabilities. The radius of the chamber was $15\,\mathrm{cm} < r < 20\,\mathrm{cm}$ and the length was 2 m. Each detector layer comprised 120 pads in the z direction and the layers were organized in sixteen ϕ-sectors. Since the detector was designed for trigger purposes it had a short response time (~ 75 ns) and was used for online event selection. The angular coverage of the detector was $11° < \theta < 169°$ and the spatial resolution in z amounted to about 1.5 cm.

The central silicon tracker (CST) [44, 45] provided precise vertex and track information and therefore allowed for precision determination of track parameters. It was the innermost tracking detector ($r \approx 5-10\,\mathrm{cm}$) built up of two 36 cm long layers of silicon strip detectors covering the polar angle region $30 < \theta < 150°$. The strip detectors were organized in ladders, the inner layer consisted of 12 and the outer layer of 20 ladders. The CST hits were measured with precision of $12\,\mu$m in the $r\phi$-plane and $25\,\mu$m in z. Track reconstruction is mainly based on the CJC measurement, however, in combination with precise CST information the resolution is much improved. The interaction vertex position can be determined with a precision of $\sim 40\,\mu$m.

The chambers COP and COZ were situated in-between CJC1 and CJC2. The COP was a proportional chamber with short response time that was in the HERA I period used in combination with CIP for track triggering purposes and was of less importance in HERA II after the CIP upgrade. The COZ was a two meters long drift chamber which improved the z-coordinate measurement with respect to CJC1 and CJC2. It was situated at the distance of $92\,\mathrm{cm} < r < 97\,\mathrm{cm}$ from the beam axis and covered the angular region $25° < \theta < 155°$. The sense wires of COZ were perpendicular to the beam axis which allowed for precision of $350\,\mu$m in determination of the z-coordinate.

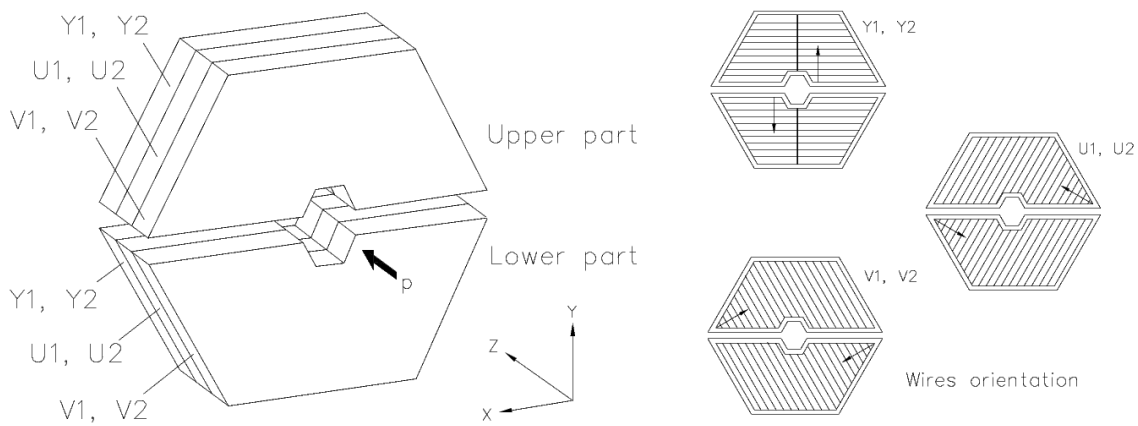


Figure 2.5: Backward proportional chamber.

Backward Tracking Detector

The backward tracking detector fulfilled two important tasks: it provided more precise xy position determination of particles with respect to a measurement by the backward calorimeter, and it allowed to differentiate between charged and neutral particle induced clusters. The latter is important for the identification of the scattered electron and thus a correct reconstruction of the event kinematics. The backward drift chamber (BDC) of HERA I was in HERA II replaced by a new backward proportional chamber (BPC) situated at $z = -146\,\text{cm}$. It consisted of six wire layers with three different orientations inclined by 60° with respect to each other. The whole detector had a hexagonal shape and was divided into two parts with a gap of approx. 80 mm between them (see figure 2.5). The inner radius of the BPC was approx. 140 mm and the outer radius approx. 800 mm. The BPC enabled to measure the angle θ with a precision of $\sigma_\theta = 0.5\,\text{mrad}$, however, the detector alignment precision in HERA II introduced a systematic error of the order $\sigma_\theta^{\text{syst.}} = 0.8\,\text{mrad}$. The horizontal gap of the detector, which included at small radius highly populated regions for the scattered electron did not allow to always require BPC information for the measurement, since this would lead to a big loss of statistics. In the default H1-OO[7] electron finding algorithm the BPC xy measurement is taken into account, if the extrapolated position of the electron candidate in the backward calorimeter is close to the electron candidate cluster (< 4 cm).

2.2.2 Calorimeters

The H1 calorimetric system comprised the liquid argon calorimeter which covered the forward and the central region and the lead/scintillating-fibre calorimeter (spaghetti calorimeter, Spacal) which covered the backward region. The H1 calorimetric system provided identification and measurement of electrons, photons, muons and penetrating neutral particles (interacting strongly).

[7] H1 object-oriented analysis framework.

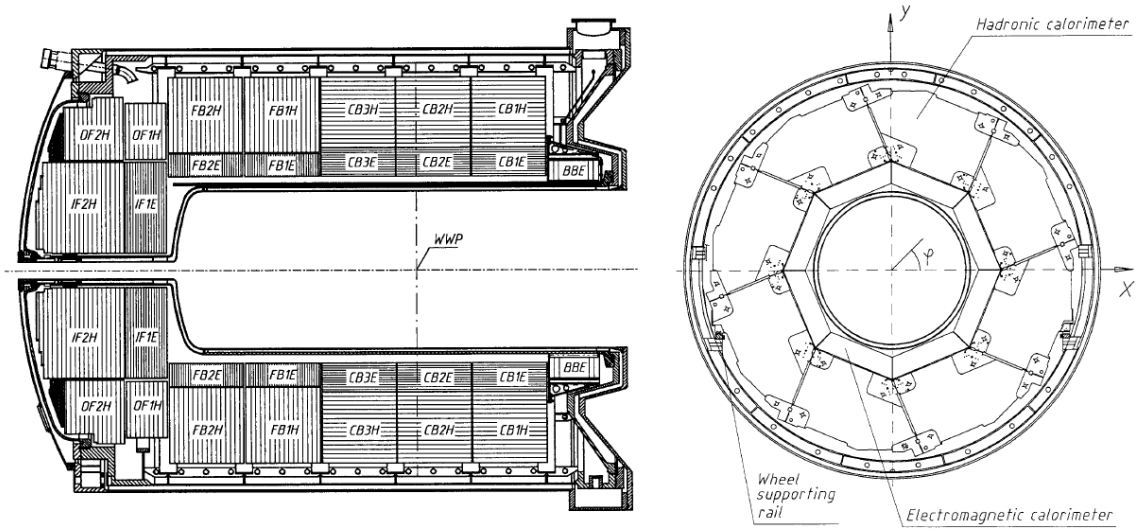


Figure 2.6: Liquid argon calorimeter, the longitudinal and the radial cross section.

Liquid Argon Calorimeter

The design of the liquid argon (LAr) calorimeter [46] is shown in figure 2.6. It was a non-compensating[8] sampling calorimeter with a total weight of 450t containing 53m^3 of liquid argon at the temperature of 90 K. Its angular coverage was $4° < \theta < 154°$. The calorimeter was segmented in z into eight wheels, and each wheel consisted of eight ϕ segments - octants. All wheels except the very forward one contained an electromagnetic part, optimized for the measurement of electromagnetic showers from electrons and photons, and a hadronic part, optimized for the measurement of hadronic showers.

The absorber material for the electromagnetic part consisted of 2.4 mm thick lead plates. They were organized into "sandwiches", the gap between two plates comprised the liquid argon and the charge collection and read-out structure and was 2.35 mm thick. Electromagnetic showers were well contained in the electromagnetic part of the calorimeter, since its lateral dimensions corresponded to 20 - 30 radiations lengths (X_0), depending on the polar angle.

The absorber material in the hadronic part of the calorimeter consisted of stainless steel plates. They were 19 mm thick with a double gap of 2.4 mm liquid argon and the charge collection and read-out structure in between them. The depth of the hadronic calorimeter was over 5 to 8 interaction lengths (λ), depending on the polar angle.

The orientation of the absorber plates in the whole calorimeter was such that the incident angle of a particle coming from the nominal interaction vertex was always less than 45°. Since the calorimeter itself was non-compensating, an offline software compensation algorithm was developed to correct for this effect.

The H1 liquid argon calorimeter was highly segmented and had about 65 000

[8]Response of a non-compensating calorimeter differs for electromagnetic and hadronic particles of the same energy.

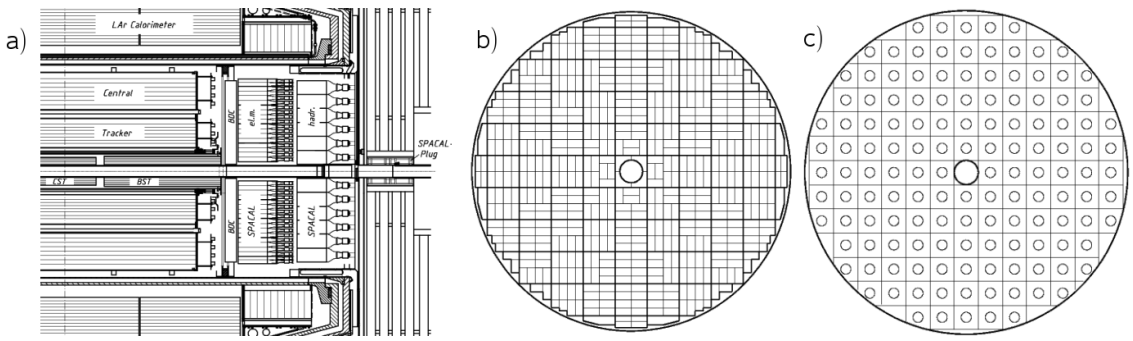


Figure 2.7: Spacal calorimeter in HERA I: a) emplacement in the H1 detector b) segmentation of the EM part into 16-cell modules and 2-cell modules c) segmentation of the hadronic part, circles symbolize photomultipliers.

electronic read-out channels. Even though the charge-collection time was rather long, thanks to the information pipeline system a level 1 trigger decision could be derived using a special electronics chain for the trigger readout and the "constant fraction technique" [47]. The energy resolution of the electromagnetic part was in the test beam determined to be $\sigma_E^{\mathrm{em}}/E \approx 11\%/\sqrt{E[\mathrm{GeV}]} \oplus 1\%$ and that of the hadronic part to be $\sigma_E^{\mathrm{had}}/E \approx 46\%/\sqrt{E[\mathrm{GeV}]} \oplus 2\%$.

Spaghetti Calorimeter - Spacal

The Spacal calorimeter [48, 49, 50] was a lead/scintillating-fiber sampling calorimeter situated at $z = -160$ cm, which replaced the original BEMC calorimeter already during the HERA I running period. The detector had an overall cylindrical shape with radius of approx. 80cm. Its main purpose was to measure the scattered electron in DIS in the backward detector region with good spatial and energy resolution and thereby allow for a precise event kinematics reconstruction. It consisted of an electromagnetic and a hadronic section. Its placement inside the H1 detector as well as the electromagnetic and hadronic $r\phi$ segmentation are depicted in figure 2.7 (status in HERA I). Both, the electromagnetic and the hadronic part were situated in a magnetic field which was taken into account for a correct functioning of the photomultipliers which converted the scintillation light into electric signals. The original angular coverage of $153° < \theta < 178°$, which allowed to measure the scattered electron in the kinematic range $2\,\mathrm{GeV}^2 < Q^2 < 100\,\mathrm{GeV}^2$, had to be reduced because of space requirements for superconducting quadrupoles needed for the luminosity upgrade for HERA II running period. The innermost cells had to be removed, and thus the angular coverage and the Q^2 range were reduced to approximately $153° < \theta < 173°$ and $4\,\mathrm{GeV}^2 < Q^2 < 100\,\mathrm{GeV}^2$. The innermost region not covered by the detector acceptance had after the upgrade an elliptical shape oriented in the horizontal direction.

The electromagnetic part consisted of approximately 1500 cells organized into 2-cell and 16-cell modules. A drawing of a 2-cell module is shown in figure 2.8. The cross section of one cell was $40.5 \times 40.5\,\mathrm{mm}^2$, its active length was $250\,\mathrm{mm} \approx 28\,\mathrm{X}_0$.

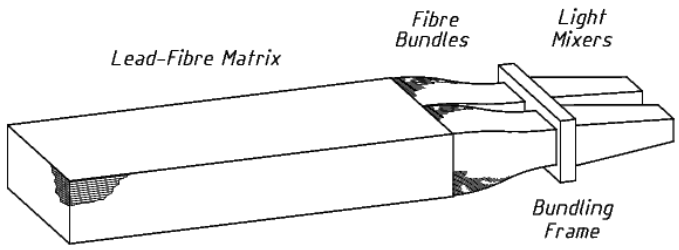


Figure 2.8: Spacal electromagnetic module comprising 2 cells.

This design guaranteed - with respect to electron shower properties ($X_0 = 9.0\,\mathrm{mm}$, Molière radius $= 25.5\,\mathrm{mm}$) - a good containment of the shower in a small number of cells without leakage. The energy resolution was determined to be $\sigma_E^{\mathrm{em}}/E = 7.1\%/\sqrt{E[\mathrm{GeV}]} \oplus 1\%$, and the spatial resolution as a function of the electron energy to be $\sigma_{xy}^{\mathrm{em}}/E = 4.4\,\mathrm{mm}/\sqrt{E[\mathrm{GeV}]} + 1.0\,\mathrm{mm}$.

The hadronic part of the Spacal had a less fine granularity, the cell size was $119.3 \times 119.0\mathrm{mm}^2$, and its active length was 250mm. The latter is of the order of only one interaction length, and thus the energy measurement of hadrons was less precise. The energy resolution measured in test beams in the energy range $1\,\mathrm{GeV} - 7\,\mathrm{GeV}$ for pions yielded $\sigma_E^{\mathrm{had}}/E \sim 38\%$ for a shower length of the order of one interaction length (first interaction required to be in the hadronic section) and $\sigma_E^{\mathrm{had}}/E \sim 29\%$ for shower lengths up to two interaction lengths (first interaction in the electromagnetic section; electromagnetic and hadronic information are combined).

The Spacal was a device suitable to provide fast trigger signals. It had a fast response with excellent time resolution - better than 0.4 ns, see reference [49]. This feature allowed the Spacal to serve also as a time-of-flight veto, i.e. to reject background originating outside the interaction time window.

2.2.3 Trigger System

The event rate at HERA was high and therefore the H1 data acquisition system could not read-out every event that occurred in the H1 detector. Even if it could, it would be a huge waste of storage space, because the rates of background events were more than an order of magnitude higher than the rates of electron-proton interaction events. This "non-*ep*" background was mainly related to interactions of beam particles with atoms of the remaining gas in the beam pipes (beam-gas events), to interactions of beam particles which were too far away from their nominal orbit with the walls of the beam-pipe (beam-wall events) and to synchrotron radiation of electrons. The bunch-crossing time at HERA was 96 ns which corresponds to a rate of 10.4 MHz. However, only ~ 1000 electron-proton collisions which are of interest were expected per second because of small *ep* cross section. The estimates [40] for different background and physic rates assuming the designed HERA I luminosity of $\mathcal{L} = 1.5 \times 10^{31}\mathrm{cm}^{-2}\mathrm{s}^{-1}$ are shown in table 2.1. After the upgrade to HERA II, the beam-related background rates scaled approximately with the intensity of beam currents (which remained similar to HERA I) and the estimates for physics-related

Beam gas interactions	50 kHz
Untagged photoproduction	1 kHz
Cosmic muons	700 Hz
Tagged photoproduction	25 Hz
DIS $Q^2 < 100\,\mathrm{GeV}^2$	2.2 Hz
DIS $100\,\mathrm{GeV}^2 < Q^2$	$1.4\,\mathrm{min}^{-1}$
Charged current DIS ($25\,\mathrm{GeV} < p_T$)	$3.0\,\mathrm{h}^{-1}$
W production	$0.5\,\mathrm{d}^{-1}$

Table 2.1: Rate estimates for HERA I.

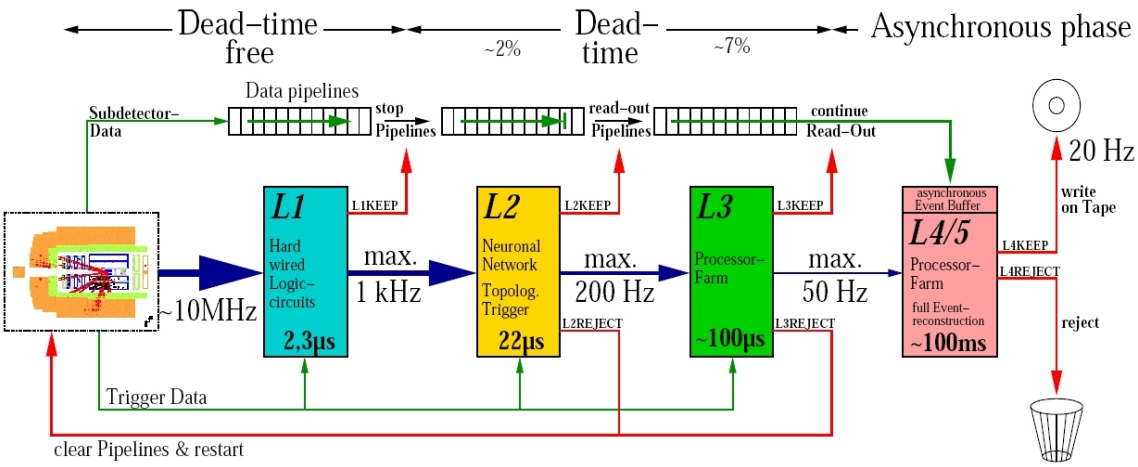


Figure 2.9: H1 trigger system.

event rates increased by the luminosity factor (~ 2.5). Stronger focusing magnets installed during the upgrade lead to important rise of the synchrotron radiation.

To cope with this situation one needed a highly efficient trigger system that is able to reject background, filter the physics and save those events which are valuable for physics analysis. These requirements were reflected in the design of the H1 trigger system (figure 2.9). It comprised a four-level system with an input rate of 10 MHz and an output rate of about 20 Hz (events written to tape). Each trigger level had more time than the previous one and thus could analyze and reconstruct the event in more detail. The dead-time[9] of the H1 detector during the readout of one event was approx. 1.4 ms and so one needed to find a compromise between minimizing the overall dead-time and allowing for a sufficiently high trigger rate. The H1 trigger was run such as to keep the overall dead-time under 10%. The four trigger levels are described in the following sections.

[9]It is time needed to process an event, during which the detector is not sensitive to register new events.

Level 1 Trigger

On level 1 a trigger decision was made every 96 ns, because an interesting event could have in principle occurred in any bunch-crossing. It implies that level 1 triggering did not lead to detector dead-time. Since the detector components could not be read-out within 96 ns, this problem was solved by feeding the detector information into pipelines whose length varied depending on the read-out time of the subdetector. The whole detector information needed to be stored at least for the period corresponding to the time interval between the occurrence of the event and the level 1 trigger decision for which a time of 2.3 μs was foreseen. Different detector components (LAr calorimeter, Spacal calorimeter, CIP2000, CJC and others) provided so-called trigger elements, a trigger element being an information bit. These bits were generated using fast electronics from very aggregate detector information concerning timing, tracks and energy depositions (e.g. the total calorimetric energy compared to a threshold) and were sent to the central trigger logic (CTL). Here they were combined using logical operators into so-called subtriggers. If a certain subtrigger "fired", an L1Keep signal was generated, the pipelines were stopped and the detector read-out started. A subtrigger might have been prescaled with a factor n, meaning that the read-out started only in one of n cases where the subtrigger fired. A prescale factor was set to be different from one, if a certain subtrigger induced a large detector dead-time. This was the case for some physics processes with a large cross section, like photoproduction or very low Q^2 inelastic scattering, but may also have happened when high backgrounds occurred for not immediately understood reasons. Even thought the level 1 trigger used only a very aggregate information, it decreased the rate under 1 kHz what was acceptable for the level 2.

Level 2 Trigger

The level 2 trigger was built up of two independent systems running in parallel: a neural network system (L2NN) and a topological trigger system (L2TT) based on event topologies. They combined the available information from different subsystems in order to study events in more detail. Their decision was delivered within 20 μs and sent to the CTL. If the event was rejected, the detector read-out was stopped and the detector was put back into the state where it could accept new events. If however the event was accepted, the L2Keep signal was generated and read-out of the whole detector was initiated. Before the level 3 system became operational during the HERA II running, the output rate of the level 2 trigger was 50 Hz at maximum. In combination with the level 3 trigger the output rate could be raised up to 200 Hz.

Level 3 Trigger

The level 3 trigger was implemented during the year 2006. It was based on the fast track trigger[10] (FTT, see appendix B and C in reference [51]) and provided a decision within less than 100 μs. Depending on the level 3 decision, the L3Reject

[10] After being implemented, the FTT also triggered on the level 1 and the level 2.

or the L3Keep signal was generated. In the first case the read-out of the detector was stopped, in the second case the read-out continued and after being finished the event was passed to the level 4, the last trigger system. The implementation of the FTT allowed to trigger on many interesting events, where the tracking information plays a major role by performing a fast reconstruction of tracks using a subset of "hits" in the CJC. It made it possible to study different explicit channels, where a suitable trigger was missing before, for example the production of D^* mesons in the photoproduction regime. The maximum output rate of the level 3 trigger which could be tolerated by the next trigger level was 50 Hz.

Level 4 Trigger

The level 4 trigger corresponded to a full event reconstruction and classification of the event on a computer farm. It did not contribute to the dead-time of the detector, since it worked in an asynchronous mode. The level 4 trigger system first checked the decisions of the previous trigger levels with improved resolution. Then different software finders classified the reconstructed event into one of many predefined event classes. If the event looked like a background event[11] (class 0), then it was highly prescaled so that only a small part of these events was kept, mainly for trigger monitoring purposes. If the event was recognized as an electron-proton event but did not fit into any other class, it was classified as "soft physics" and prescaled with respect to its Q^2. If the event was labeled to belong to any other class, then it was kept as were the prescaled events and they were stored on tape. The row event information as well as the reconstructed data are stored on so-called production output tapes (POT) and the reconstructed information was written in compact form to the so-called data summary tape (DST), the starting point for analyses in H1. The output rate of this trigger level was limited to ~ 20 Hz.

[11] The term "not classified junk" was used for these events.

Chapter 3

Reconstruction and Measurement Methods

The detector hardware components described in the previous chapter provide basic information consisting of signals in tracking detectors and of ionization-charge or scintillating light measured in calorimeters. These data are used by smart software algorithms to reconstruct the kinematics of an event at the detector level in order to allow the analysis, in our case, of charm fragmentation.

In this chapter we summarize the most important reconstruction and measurement methods for our analysis topic. Since the reconstruction of D^* mesons relies almost exclusively on the tracking, we consider it necessary to briefly mention the track reconstruction in H1. We further quickly explain the different particle finders that are used and which play an important role: the electron finder, which allows for identification of the scattered electron and thus for the correct kinematics reconstruction, the finder of hadronic final state particles and the jet finder which are necessary in our fragmentation study, where we use jets and event topologies, and finally the D^* finder. One subsection is dedicated to the extraction of the D^* signal from the background. Next, we explain two different observables which are sensitive to the fragmentation of a charm quark into a D^* meson and we provide a brief description of the unfolding methods we use to correct for detector effects. The last subsection describes the reconstruction of the event kinematics. Although all these aspects of reconstruction and measurement methods are not necessarily directly related to each other, we prefer to give their compact description in one dedicated chapter rather than try to describe them in the appropriate places in the measurement chapter, which would be, we believe, less clear.

One should also keep in mind that the afore mentioned topics represent a large amount of information, which cannot be covered in all detail. Thus we only briefly present the main ideas; an interested reader should consult the given references.

3.1 Central Track Reconstruction

The track reconstruction in the central detector region [52, 53] is based on the CJC1 and the CJC2. It proceeds in two steps: a fast track reconstruction followed by an optimized reconstruction of all tracks.

At the beginning of the track reconstruction the initial T_0 of the event is estimated (timing information). The estimate comes from the leading edge of the drift time spectrum for all wires. Next the TGV[1] software package is called. This program combines different wires within a drift cell and searches for three successive wires having a hit, so-called triplets. Thanks to the inclination of wire planes ($\approx 30°$) with respect to the radial direction, the drift side ambiguity can be solved using the criterion of a straight-line track coming from the interaction point. Then a circular fit going through the origin is performed for each triplet and the fitted circle is characterized by its curvature κ (inverse of the radius) and its angle φ at some reference radius ($\sim$ middle of the chamber, in each chamber CJC1/CJC2 separately). Afterward triplets are collected which cluster in the $\kappa\varphi$ plane. Such triplet clusters are regarded as track candidates (for each chamber separately) and once more a circular fit is performed taking into account all hits belonging to a cluster in consideration. This fit does not require the nominal interaction point, so in addition to its curvature κ and its angle at the origin φ_0, it is also characterized by its distance of closest approach to the nominal interaction point (*dca*). Finally, the matching of CJC1 and CJC2 fits is done, the fits are combined in order to find the tracks that go through both chambers. In this way the "fast tracks" are reconstructed in the xy-plane, their reconstruction in z is done separately. The fast track reconstruction, however, works well only for tracks with small curvature (medium or high transverse momentum) and originating near the primary vertex.

For the final reconstruction of all tracks (in H1REC) the already found fast tracks are used as seeds. The algorithm finds hits in roads (1 cm wide) around each fast track and repeats the fit. The fit is performed in the xy-plane, the z-development of the track is fitted separately. Tracks crossing anode or cathode wire planes allow to determine a track T_0. This T_0 information from different tracks is histogrammed, and an improved T_0 of the event is estimated from the peak position of the histogram. The found tracks are then once more improved using this T_0 information. The track reconstruction continues by removing all hits used up to now and by searching for further tracks with the aim to also reconstruct tracks with big curvature and big *dca*. Again a triplet search is done in each cell using wires with remaining hits and chains of triplets are formed. On found triplet chains a road search is done, and the ambiguity due to mirror hits is solved by using of the two possible solutions the one with the longer chain. The procedure is achieved by a circular fit and in this way the non-vertex fitted tracks are reconstructed.

Further requirements and the knowledge of the run-dependent interaction vertex region are applied to non-vertex fitted tracks in order to identify the tracks originating from the primary vertex. After repeating track fits with a common interaction

[1] "Tracks à Grande Vitesse"

vertex constraint, the vertex-fitted track are obtained. The information from other tracking detectors (COZ, CIP2000[2], CST[3]) is taken into account in order to improve the z information and the vertex reconstruction.

3.2 Software Finders and Signal Extraction

The different software finders are part of the H1 object-oriented analysis framework (H1-OO). This framework is based on the C++ programming language and the ROOT software package. All finders which we will describe have been developed inside the H1 collaboration, and the most precise information about them can be obtained directly from the source codes of the appropriate classes.

3.2.1 Electron Finder

The Q^2 range studied, $5\text{GeV}^2 < Q^2 < 100\text{GeV}^2$, implies that the electron is measured with the Spacal calorimeter, and therefore we describe only the electron finding algorithm for this detector component. The high Q^2 electrons, detected in the liquid argon calorimeter, are identified using a different software algorithm[4].

The electron finding in Spacal is done by looping over all reconstructed clusters[5]. The clusters which have too low energy or are situated at too small Spacal radius, in comparison to given energy and radius thresholds, are filtered out. For each remaining cluster the θ and φ coordinates are calculated with respect to the actual interaction vertex position of the event. Next, the BPC hits are considered. In the case where candidate hits are found, the track position extrapolated into the Spacal is determined. If the distance between this position and the cluster barycenter is smaller than 4 cm then the BPC measurement of the track is taken into account and θ and φ are recalculated. Then CJC tracks which can be associated to the cluster are searched for. Preferentially vertex-fitted tracks of high quality[6] are associated, then vertex-fitted tracks of lesser quality[7] and finally also non-vertex-fitted tracks are considered. In the final step a correction for the beam-tilt[8] is applied to the θ and φ coordinates of the electron candidates.

The electron candidates found with this procedure are further studied with respect to different criteria like cluster isolation, cluster energy in the hadronic calorimeter, etc., and the resulting information is provided to the user. If more than one electron candidate is found, then the candidate with the highest transverse momen-

[2]At the time of writing this chapter the CIP2000 is not used for the track reconstruction.

[3]CST becoming operational during 2006.

[4]The corresponding software class is called *H1CreateLArEm* and is part of the standard H1-OO framework.

[5]A cluster is an object reconstructed by software algorithms from detector signals that reflects the position and spatial distribution of the energy deposit induced in the calorimeter by an energetic particle.

[6]So-called "Lee West" tracks, a high quality subset of DTRA tracks.

[7]So-called DTRA tracks.

[8]The beam-tilt refers to a non-zero angle between the z axis and the beam direction.

tum is labeled as the scattered electron. The electron identification efficiency is close to 100% [54, 55]. The electron-finding algorithm as described is implemented in *H1CreateSpacalEm* and *H1CreatePartEm* classes of the H1-OO framework.

3.2.2 Energy Flow Algorithm for Hadronic Reconstruction

The reconstruction of the hadronic final state[9] is based on the *Hadroo2* energy flow algorithm [56]. An energy flow algorithm is characterized by the combination of information coming from different detector components, in our case tracking and calorimetric information. Before the algorithm is applied, the input objects - tracks and clusters - need to be preselected.

The tracks are required to be of good quality ("Lee West tracks", see reference [57]), and only central ($20° < \theta < 160°$) and combined ($0° < \theta < 40°$) tracks are accepted. The central tracks are reconstructed using information from the central tracking detector only, while the combined track reconstruction relies on both, the central and the forward tracking system. In addition, further requirements related to different track quantities like transverse momentum, starting radius, radial length, etc., are applied so to select only those tracks whose reconstruction and measurement is well understood within H1. These requirements differ for central and combined tracks. A more detailed list of the essential track cuts can be found in reference [56].

The calorimetric clusters considered are those of the Spacal and of the liquid argon calorimeter. Since the liquid argon calorimeter is non-compensating, a weighting algorithm is applied to correct for the on average lower response to hadrons in comparison to electrons or photons of the same energy. The weighting is done in the H1REC package, however, the classification of what are hadronic or electromagnetic clusters in the electromagnetic part of the LAr calorimeter is modified in *Hadroo2*. An important issue is also the noise suppression since a relatively large amount of noise is present in the liquid argon calorimeter (several GeV per event). First, one-cell only clusters and clusters with energy smaller than 0.2 GeV are removed. Then a set of noise and background finders is applied. These finders (FSCLUS, HALOID, HNOISE, NEWSUP) reject low energy isolated clusters and also clusters which are due to either beam halo particles or cosmic ray muons. Their description can be found in reference [58].

The remaining clusters and selected tracks enter the *Hadroo2* algorithm which constructs hadron candidates (or hadronic objects) by combining the tracks and clusters, taking their respective resolution and geometric overlap into account, without double counting of energy. The algorithm is based on the comparison of the relative errors of the track-based energy measurement $\frac{\sigma_{E_{\text{track}}}}{E_{\text{track}}}$ and the calorimeter-based energy measurement $\left(\frac{\sigma_E}{E}\right)_{\text{LAr}}$. The tracking measurement is better at low transverse momenta, at high transverse momenta the calorimetric measurement becomes more precise. It is however not straightforward to compare these two quantities, since the energy measured in the calorimeter can have a contribution from neutral particles. Thus the estimate of $\left(\frac{\sigma_E}{E}\right)_{\text{LAr expectation}}$ is based on the energy of the track, i.e.

[9] The hadronic final state refers here to all final state particles except the scattered electron.

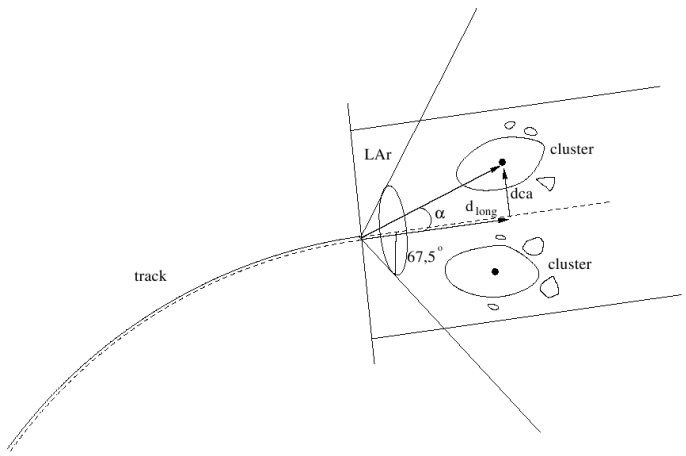


Figure 3.1: The geometry used in the calculation of E_{cylinder}.

$\frac{\sigma_{\text{E LAr expect.}}}{E_{\text{track}}} = \frac{0.46}{\sqrt{E_{\text{track}}[\text{GeV}]}}$, and only then the quantities are compared.

If $\frac{\sigma_{E_{\text{track}}}}{E_{\text{track}}} < \frac{\sigma_{\text{E LAr expect.}}}{E_{\text{track}}}$ then the track measurement is preferred. In this case one needs to take into account possible calorimetric contributions from neutral particles. For this purpose the track is extrapolated to the surface of the calorimeter and the energy E_{cylinder} is computed as the sum of all clusters in the overlapping volume of a $67,5°$ cone and two cylinders of radius 25 cm in the electromagnetic part and 50 cm in the hadronic part of the liquid argon calorimeter (figure 3.1) . In order not to misidentify the hadronic energy fluctuation as a neutral particle, one does not compare directly E_{cylinder} to E_{track} but rather to the quantity

$$\widetilde{E}_{\text{track}} = E_{\text{track}} \times \left[1 + 1.96 \sqrt{\left(\frac{\sigma_{E_{\text{track}}}}{E_{\text{track}}}\right)^2 + \left(\frac{\sigma_E}{E}\right)^2_{\text{LAr expectation}}} \right],$$

which should exclude a hadronic energy fluctuation at 95% C.L. with respect to the estimated error. If $E_{\text{cylinder}} < \widetilde{E}_{\text{track}}$ then the whole energy E_{cylinder} is subtracted from the calorimetric measurement. If $E_{\text{cylinder}} > \widetilde{E}_{\text{track}}$ then only the energy E_{track} is subtracted, the rest being regarded as an energy deposit induced by a neutral particle.

If $\frac{\sigma_{E_{\text{track}}}}{E_{\text{track}}} > \frac{\sigma_{\text{E LAr expect.}}}{E_{\text{track}}}$ three different possibilities are considered

- If $(E_{\text{cylinder}} - 1.96\,\sigma_{E_{\text{cylinder}}} < E_{\text{track}} < E_{\text{cylinder}} + 1.96\,\sigma_{E_{\text{cylinder}}})$ then the two measurements are interpreted as compatible, and the calorimetric measurement is used to define a hadron.

- If $(E_{\text{track}} < E_{\text{cylinder}} - 1.96\,\sigma_{E_{\text{cylinder}}})$ then neutral particle in addition to the track is supposed. The tracking information is used, and the energy E_{cylinder} is subtracted from the calorimetric measurement.

- If $(E_{\text{cylinder}} + 1.96\,\sigma_{E_{\text{cylinder}}} < E_{\text{track}})$ then the track is discarded and the calorimetric measurement alone is used to define a hadron.

The clusters and tracks not matching each other become particle candidates on the basis of measurement from one detector system only. The whole procedure was checked in detail, and it was shown that the noise suppression algorithms work properly and do not suppress the signal for analyses of exclusive final states. The comparison of the *Hadroo2* algorithm with previous algorithms used by H1 demonstrated that the *Hadroo2* algorithm effectively improves the hadronic final state reconstruction and resolution, especially in the high transverse momentum region.

3.2.3 Jet Finder

The QCD confinement implies that free partons are never directly observed, they are always bound within hadrons. For a high-momentum parton produced in a hard interaction one however expects the hadronization effects to be small in comparison with the parton energy, such that the individual particles corresponding to a given parton are expected to be confined within a rather small angular region. Thus one expects to observe streams of particles, so-called jets that originate from (high-energy) partons. A high-energy jet is also expected to well approximate for example the energy and the angle of the initiating parton. Jets were for the first time visually observed in e^+e^- collisions at the PETRA collider at DESY and their definition was originally more or less intuitive. Soon the intuitive approach became insufficient. It was not clear when close particle streams should be considered as separate jets and when they should be merged into one jet. In addition, comparison between experiment and theory and between different experiments was requiring a more rigorous approach. Thus different jet-finding algorithms were proposed and continue to be proposed until today. Most algorithms used nowadays can be split into two groups: cone algorithms and clustering algorithms.

A good jet algorithm should fulfill certain criteria. It should be easily applicable at different levels, e.g. at parton level (the domain of theoretical predictions by pQCD), hadron level (theoretical predictions including models for hadronization, hadron level corrected experimental data) or at detector level (tracks, clusters). The algorithm should lead to good correlation between parton and detector level and should also be collinear and infrared safe. This last property refers to the fact that the result of a jet-finding algorithm should not depend on radiation of soft particles or on a particle splitting into two collinear particles. In the experiment, this property reflects as a small dependence of the result on the detector granularity.

In this work we use the so-called k_t-clustering algorithm [59, 60, 61] which fulfills the previous criteria. In addition, it has the nice feature of being invariant under boosts along the beam axis, since it is based on quantities ϕ (azimuthal angle) and $\widetilde{y}$ (rapidity[10]) that transform simply under such boosts. The azimuthal angle does not transform at all and the rapidity defined as $\widetilde{y} = \frac{1}{2}\ln\left(\frac{E+p_z}{E-p_z}\right)$ has a simple transformation rule $\widetilde{y} \to \widetilde{y} - \tanh^{-1}\beta$, so that the shape of the rapidity distribution $dN/d\widetilde{y}$ is invariant. For high momenta $m \ll p$ one often approximates the rapidity

[10]The common notation is y but in order to avoid a confusion with the inelasticity we prefer to use $\widetilde{y}$.

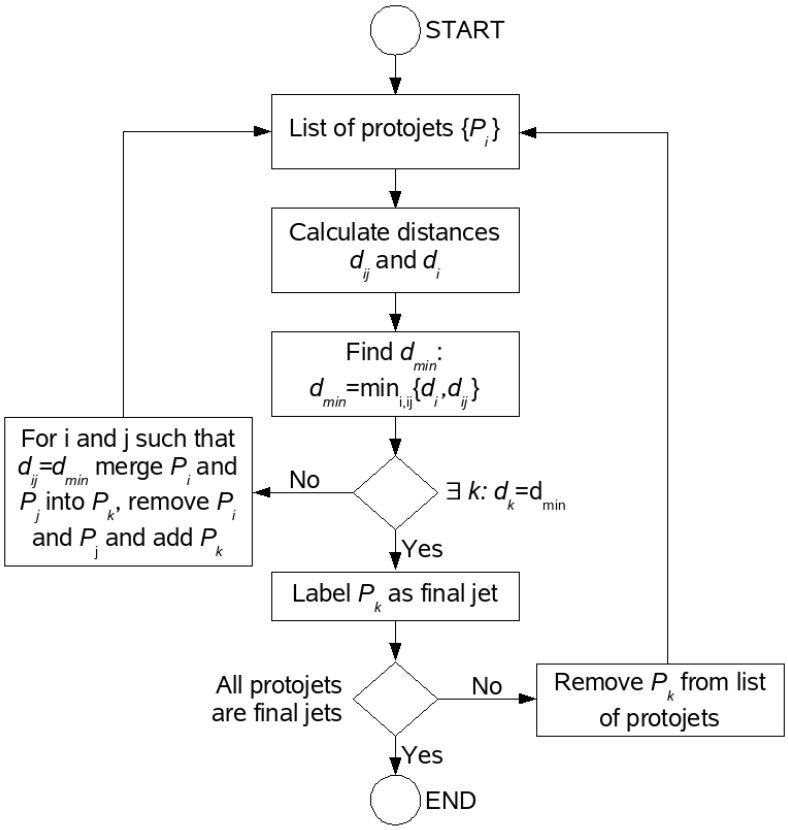


Figure 3.2: Diagrammatic representation of the jet-finding k_t clustering algorithm.

by the pseudorapidity $\widetilde{y} \approx \eta = -\ln\tan(\theta/2)$ which can be directly related to the measured angle θ of the particle in the detector. The k_t-clustering algorithm is represented by the chart in figure 3.2. An object entering the algorithm - a protojet - can be of different nature: a parton, a stable particle or a reconstructed detector object. From among different possibilities, we have chosen the E_T-recombination scheme that treats protojets as massive objects. The distances d_i and d_{ij} mentioned in the chart are calculated for each possible protojet pair (including the distance of a protojet to itself) according to the formulas

$$d_{ij} = \min(E_{T,i}^2, E_{T,j}^2).[(\widetilde{y}_i - \widetilde{y}_j)^2 + (\phi_i - \phi_j)^2]/R_0^2,$$

$$d_i = E_{T,i}^2,$$

with R_0 being an adjustable parameter related to the opening angle between jets. We use the default value $R_0 = 1$. The merging of two protojets is done by summing up their four-vectors

$$(p_{x,k}, p_{y,k}, p_{z,k}, E_k) = (p_{x,i} + p_{x,j}, p_{y,i} + p_{y,j}, p_{z,i} + p_{z,j}, E_i + E_j).$$

All other necessary information to understand the algorithm is contained in the presented chart.

3.2.4 D^* Finder

In this work we consider for the D^* meson reconstruction only the "golden" decay channel $D^{*\pm} \to D^{0(\overline{0})}\pi_s^\pm \to K^\mp\pi^\pm\pi_s^\pm$. The mass difference $m_{D^*} - (m_{D^0} + m_\pi)$ is small, and so only little kinetic energy is available in the D^* meson decay. Thus, in the D^* rest frame the D^0 meson and the pion are produced almost at rest. This implies that in the laboratory frame the D^0 meson carries most of the D^* energy (because of its high mass) and the π_s only a small fraction. Therefore, the pion is referred to as "slow" and is given the index "s". The reconstruction of the D^* meson relies on found tracks, with the slow pion expected to be reconstructed as a track with large curvature and small transverse momentum p_T. At HERA energies the produced D^* and D^0 mesons do not have enough energy to live long enough such that a secondary vertex cannot be reconstructed in most of decays[11]. Therefore in the D^* reconstruction one uses only tracks originating from the primary vertex.

The decay channel used has a rather small branching ratio $BR(D^{*\pm} \to K^\mp\pi^\pm\pi_s^\pm)$ $= BR(D^{*\pm} \to D^0\pi_s^\pm) \times BR(D^0 \to K^\mp\pi^\pm) = (2.546 \pm 0.064)\%$ which could be seen as a drawback. On the other hand this channel allows for clean signal reconstruction and the combinatorial background[12] is within reasonable limits. In addition, instead using of the D^* invariant mass spectrum the so-called ΔM tagging technique [62], where $\Delta M = M(K^\mp\pi^\pm\pi_s^\pm) - M(K^\mp\pi^\pm)$, is applied. In the ΔM spectrum the signal peak position is near the pion-mass threshold and thus the combinatorial background is suppressed. Furthermore, the ΔM technique allows for partial cancellation of several systematic errors.

The finder algorithm itself is rather simple. It uses all good-quality primary-vertex fitted tracks and runs over them in three mutually nested loops so that every three-track combination is considered. In the outermost loop every track is regarded as being a kaon, only a transverse-momentum requirement $p_T(K) > 0.25$ GeV is applied. In the following loop a pion-candidate track is associated. This track is required to have a transverse momentum $p_T(\pi) > 0.25$ GeV, and only the correct charge combination with respect to the kaon candidate is accepted[13]. The mass of the reconstructed D^0 meson candidate is required to fulfill $|m(D^0_{\text{candidate}}) - m(D^0)| <$ 0.45 GeV. In the innermost loop the slow-pion candidate track is added. Once more one requires the right charge combination with respect to the kaon, and the requirement on the transverse momentum is $p_T(\pi_s) > 0.07$ GeV. Further cuts on the reconstructed D^* candidate are $m(D^*_{\text{candidate}}) - m(D^0_{\text{candidate}}) < 0.17$ GeV, $p_T(D^*) > 0.7$ GeV and the pseudorapidity cut $|\eta(D^*)| < 1.5$ is introduced in order to restrict the reconstruction to the region of the central tracking detector. Finally, a common requirement on the D^* and the D^0 candidates $[|\, m(D^0_{\text{candidate}}) - m(D^0)| <$ $0.1\,\text{GeV}\ OR\ m(D^*_{\text{candidate}}) - m(D^0_{\text{candidate}}) < 0.152\,\text{GeV}]$ is to discard such candidates that have rather badly reconstructed both, the D^0 mass and the D^*D^0 mass differ-

[11] The D^* mesons decay strongly and thus are short-lived. The D^0 mesons decay weakly, but their mean lifetime $\tau = (410.3 \pm 1.5) \times 10^{-15}$s is short.

[12] The combinatorial background arises from three tracks that do not come from the golden-channel D^* decay but accidentally fulfill the criteria of the D^* finder.

[13] The wrong-charge combinations $K^\mp\pi^\mp\pi_s^\pm$ are also separately reconstructed. They can be used to estimate the combinatorial background.

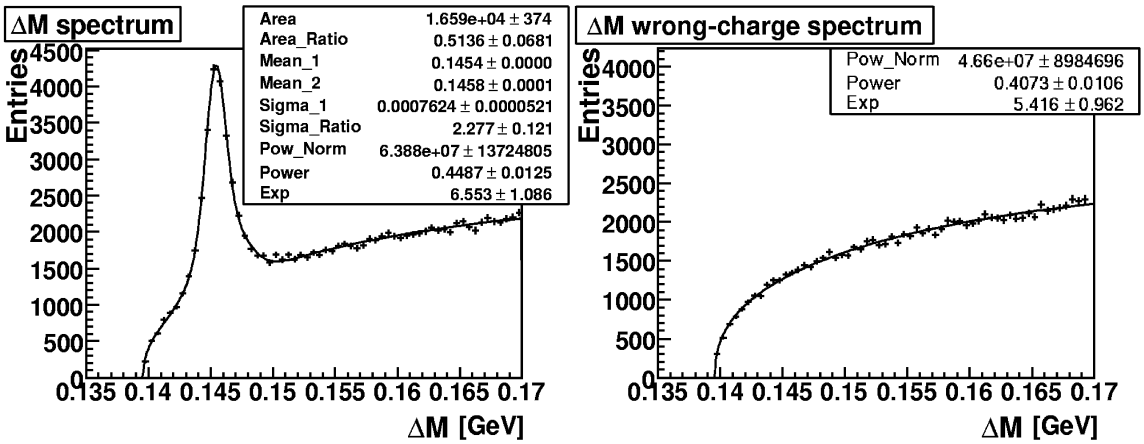


Figure 3.3: Right-charge (left) and wrong-charge (right) ΔM spectra for all HERA II data displayed together with corresponding fits.

ence. If the track combination fulfills all mentioned criteria, a D^* meson candidate is found. Since all three-track combinations are considered, it is possible to have more than one D^* candidate per event.

3.2.5 D^* Signal Extraction

A typical ΔM spectrum resulting from the D^* finder with additional requirements on the event (explained later in chapter 4) is presented in figure 3.3-left. The signal peak appears near the pion threshold around $\Delta M \approx 0.1455$ GeV and its approximate width is $\sigma_{\Delta M} \approx 0.001$ GeV. The signal is superposed on a smooth and rising combinatorial background. The most appropriate way of extracting the signal (= the number of D^*s) would probably be done by comparing the measured ΔM spectrum with the one from a Monte Carlo simulation. If the Monte Carlo model would describe the measured data well, then the signal extraction would be rather straightforward, since in a Monte Carlo simulation the true D^* events can be identified. This method is however almost outside the range of our technical possibilities, because the combinatorial background does not originate from charm physics only but originates also from many other physics channels. Therefore, this method would require a fully inclusive Monte Carlo simulation with very high statistics. An enormous CPU time would be needed to generate it, and one could not guarantee that the result would describe the measurement well[14]. Thus, the signal is extracted by fitting the measured ΔM spectrum with an appropriate function that is a sum of a background and a signal function.

The behavior of the background can be studied with a wrong-charged ΔM spec-

[14]The tracking in the MC simulation actually does not reproduce the data behavior exactly, the signal peaks in the simulation are usually slightly narrower than in the data.

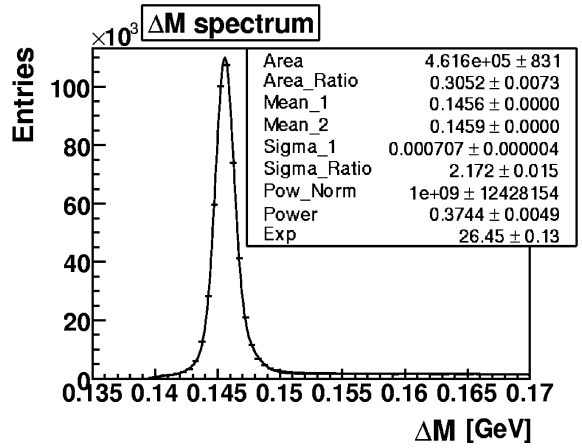


Figure 3.4: Right-charge $\triangle M$ distribution for the radiative RAPGAP signal Monte Carlo simulation corresponding to all of HERA II running. The combinatorial background is very small but present, the signal asymmetry and non-Gaussian tails can be observed.

trum[15], where the signal is not present (figure 3.3 - right). The function

$$f_{\mathrm{bg}}(x) = N_{\mathrm{bg}}(x - m_\pi)^\alpha \exp(-\beta x),$$

where N_{bg}, α and β are free parameters, describes the wrong-charged spectrum well. It also describes well the right-charged spectrum outside the signal region and thus is chosen to be the background function in our fit. The signal peak might be described by a Gaussian-like function; a simple Gaussian, however, does not work satisfactorily. It is due to the fact that the signal has larger tails which are not properly described by a simple Gaussian and that the signal shape is asymmetric , the asymmetry being also observed in the Monte Carlo simulation. The smooth behavior of the wrong-charged distribution in the tail regions as well as non-Gaussian tails of the peak in a signal Monte Carlo simulation (figure 3.4) suggest that the tails indeed contain signal and thus should be included in the fit. The double-Gaussian function

$$\begin{aligned} f_{\mathrm{sig}}(x) &= A_{\mathrm{tot}}(1 - A_2/A_{\mathrm{tot}})\frac{1}{\sigma_1\sqrt{2\pi}} \exp\left(-\frac{(x-\mu_1)^2}{2\sigma_1^2}\right) + \\ &\quad + A_{\mathrm{tot}}(A_2/A_{\mathrm{tot}})\frac{1}{\sigma_1(\sigma_2/\sigma_1)\sqrt{2\pi}} \exp\left(-\frac{(x-\mu_2)^2}{2[\sigma_1(\sigma_2/\sigma_1)]^2}\right) \end{aligned}$$

provides a good description of the signal region in the right-charged spectrum and the sum

$$f(x) = f_{\mathrm{sig}}(x) + f_{\mathrm{bg}}(x)$$

describes properly the whole ΔM spectrum. The free parameters of the signal function are

[15] A wrong-charged D^* candidate is a fake particle created by combining a supposed kaon track with a pion track candidate that has an inappropriate (i.e. the same) charge. The two particles therefore cannot come form a D^0 decay and so the spectrum of fake (wrong-charged) D^* particles does not contain the signal and allows to study the behavior of the combinatorial background.

- $A_{\text{tot}} = A_1 + A_2$ - total area (sum of areas of the two Gaussians) or number of D^*s,
- A_2/A_{tot} - ratio of the area of the second Gaussian to the total area of the two Gaussians (a single parameter in the fit),
- μ_1, μ_2 - means of the two Gaussian functions,
- σ_1 - width of the first Gaussian,
- σ_2/σ_1 - ratio of widths (a single parameter in the fit).

When extracting the number of D^* mesons as a function of a variable V, then for all events in a given bin of the V-distribution the ΔM fit is performed, and the signal is extracted. Since in a single bin of the V-distribution the statistics becomes usually low, the parameters A_2/A_{tot}, μ_1, μ_2 and σ_2/σ_1 are fixed to the values from the fit to all events. The parameters to be fixed were determined studying the behavior of fits for different distributions; fixing the presented set of parameters allows to avoid wrong fits caused by statistical fluctuations and still allows for enough freedom to fit the ΔM distribution well. The signal extraction for a signal Monte Carlo simulation (which also contains a small amount of background) is done in a similar way.

3.3 Fragmentation Measurement Methods

The main aim of this thesis is to study the fragmentation of the charm quark into a D^* meson in electron-proton collisions in the DIS regime. Such a study requires to define an appropriate variable that is sensitive to the fragmentation process. Since fragmentation functions are related to the momentum fraction of the c quark that is transferred to the produced D^* meson, it seems appropriate to study the fragmentation with respect to this momentum fraction. Within an experiment it is however impossible to access directly the momentum of the initial quark, and therefore approximations need to be done.

Charm quark fragmentation has already been studied in several e^+e^- experiments [63, 64, 65]. The commonly used variables are momentum or energy fractions

$$x_E = \frac{E_{D^*}}{\sqrt{s}/2} \,,$$

$$x_P = \frac{|\overrightarrow{p}_{D^*}|}{\sqrt{s/4 - m_{D^*}^2}} \,,$$

where s denotes the center-of-mass (CMS) energy of the electron-positron system. Such definitions follow straightforwardly from the lowest order charm quark production diagram (see figure 3.5) , where a charm quark from the quark-antiquark pair is in lowest order carrying one half of the CMS energy. The results obtained by e^+e^- experiments allowed to study charm fragmentation with quite some precision and led to standard parametrizations of fragmentation functions. Assuming

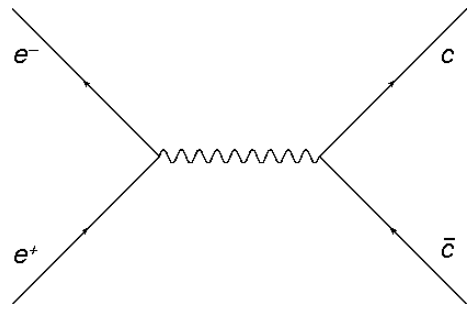


Figure 3.5: The lowest order $c\overline{c}$ production diagram in e^+e^- collisions.

universality of fragmentation functions, these parametrizations were later used also in electron-proton collisions without studying in all detail the correctness of such an extrapolation.

In electron-proton collisions at HERA we can hardly compete with some of the mentioned experiments from the point of view of statistics. On the other hand, we have the unique possibility to study the fragmentation universality by checking the compatibility of our results with those of e^+e^- experiments. The definition of an observable is however less straightforward compared to e^+e^- experiments.. In the BGF process (see figure 1.8) the CMS energy of the produced quark-antiquark pair is not fixed and thus it is not possible to compare the measured D^* momentum with some fixed momentum. It would appear desirable to make the study in the gluon-photon rest frame. The CMS energy in this frame would not be fixed either but otherwise the situation would be similar to the situation in e^+e^-collisions, because the produced charm quarks would have back-to-back oriented momenta. Unfortunately, the experimental constraints do not allow us to get precise enough information to perform a boost into this frame. Thus, we do our study in the photon-proton center-of-mass frame (often referred to as the γP frame), where at least for the leading order direct BGF process, neglecting any transverse momentum of the initial gluon and final-state gluon radiation, the transverse momenta of the c and $\overline{c}$ quarks are balanced. The available CMS energy W can be calculated from the event kinematics, see section 1.2.1. The quark pair is mostly produced in the photon direction since, supposing the resolved-photon contribution to be minor, the photon enters the interaction with its whole momentum, whereas the proton interacts via a gluon which typically carries only a small fraction of the proton momentum.

In this work we define two observables, both in the γP frame, in order to cross-check the compatibility of the results coming from them. We replace in the fragmentation analysis the four-vectors of the three decay particles $K^{\mp}\pi^{\pm}\pi_s^{\pm}$ by the corresponding four-vector of the D^* candidate. This is done in order to avoid situations where the decay particles are found in different event hemispheres or different jets.

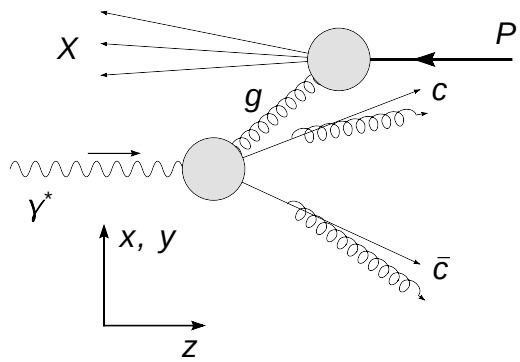


Figure 3.6: Schematic representation of the charm production in the γP frame via the BGF process. Arrows denote particle momenta.

3.3.1 Hemisphere Method

The hemisphere method is based on the global event topology in the γP frame (figure 3.6). We use the fact that the $c\bar{c}$ quark momenta are in some approximation transversely balanced, allowing to divide the whole momentum space into two hemispheres. To do so we consider the momenta of all reconstructed particles and firstly exclude those which point in the direction of the proton (i.e. the z component of their momentum is negative). In this way we discard particles which are usually not originating from the $c\bar{c}$ quark pair, but which are part of the proton remnant and the initial parton shower. A Monte Carlo study using the RAPGAP event generator was done in [66][16] and demonstrated the correctness of this requirement. Since we expect a back-to-back topology in the transverse direction, we consider the projections of all remaining particle momenta onto a plane perpendicular to the z-axis and define[17] a two-dimensional thrust variable with respect to an arbitrary axis in this plane

$$T = \frac{\sum_i \left| p^*_{i\parallel} \right|}{\sum_i |p^*_i|}.$$

Then the axis maximizing $T = T_{\max}$ is found. A completely isotropic event would have $T_{\max} = 0.5$, an event with an ideal back-to-back topology would have $T_{\max} = 1$ (see figure 3.7). With the thrust axis found one can define a plane orthogonal to it and which can be used to split the original plane containing the particle projections into two hemispheres. We refer to the hemisphere containing the D^* as the "D^* hemisphere", to the second hemisphere as the "other hemisphere". We finally construct our observable

$$z_{\text{hem}} = \frac{(E^* + p^*_L)_{D^*}}{\sum_{i\in\text{hem}} E^*_i + \left| \sum_{i\in\text{hem}} \overrightarrow{p^*_i} \right|},$$

where the summation in the denominator is done over all particles of the D^* hemisphere that propagate in the photon direction and p^*_L is the longitudinal component

[16] Measurement of charm fragmentation at HERA using 2000 data.

[17] We use the superscript * to design variables which are defined in the γP frame.

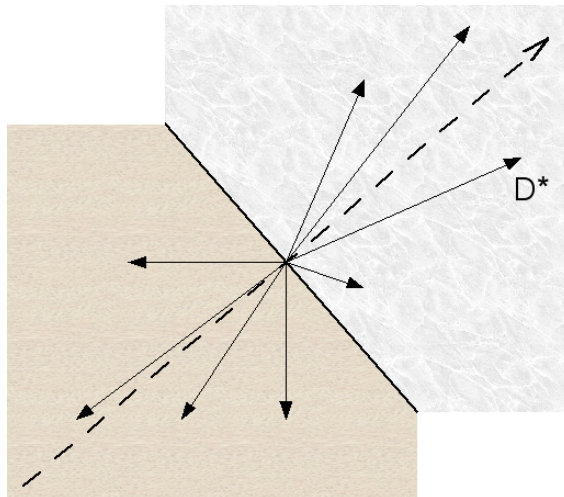


Figure 3.7: Plane orthogonal to the z-axis in the γP frame with projections of particle momenta. The thrust axis (dashed line) and the plane orthogonal to the thrust axis (full line) are indicated.

of the D^* momentum with respect to the momentum of the whole hemisphere. This light cone definition of the z_{hem} observable has the advantage of being invariant under a boost along the D^* hemisphere direction. The hemisphere method has by its construction the tendency to sum up all gluon radiation in the D^* hemisphere.

In this thesis the distribution of the z_{hem} observable will be studied for two data samples. For one sample we ask the D^* to be contained within a jet having $E_T^*(D^* \text{ jet}) > 3$ GeV, for the other sample this requirement is not made[18]. This study is motivated by the results of a previous fragmentation analysis [66], which suggest that the MC models fail to describe the data at charm production threshold. Demanding the presence of a D^* jet with an E_T^* above some minimal E_T^* is expected to select events where the charm quarks are produced further away from the production threshold compared to when no D^* jet is required. The minimum E_T^* of 3 GeV is chosen to be the same as in the jet method (see the following section). To distinguish between the two data sets we will use two different notations, z_{hem} when referring to the inclusive sample and $z_{\text{hem}}^{(\text{jet})}$ when referring to the sample with a D^* jet.

3.3.2 Jet Method

In the jet method the charm quark momentum is approximated by the momentum of the jet that contains the D^* meson. The observable is naturally defined[19] as

$$z_{\text{jet}} = \frac{(E^* + p_L^*)_{D^*}}{(E^* + p^*)_{\text{jet}}} ,$$

[18]The data samples are not independent, one is a subset of the other.

[19]This definition is driven by the definition of fragmentation variable in the Lund string fragmentation model, see section 1.3.3.

where E^* and p^* are energies and momenta in the γP frame and p_L^* is the D^* momentum component parallel to the jet axis. Jets are reconstructed using the k_t-clustering algorithm (section 3.2.3). In order to ensure a good quark-jet correlation the jet is required to have a transverse energy $E_T^*(D^* \text{jet}) > 3\,\text{GeV}$.

Unlike the hemisphere method, the jet method sums up only gluon radiation with small transverse momentum with respect to the jet axis, gluon radiation with high p_T may be reconstructed as a separate jet. One thus expects the hemisphere observable to be more analogous to the e^+e^- observables than the jet observable. One should also notice that unlike the hemisphere observable, the jet observable may not be defined for every event with a D^* candidate, since there might not be a reconstructed jet with a transverse energy high enough.

3.4 Unfolding Methods

The aim of experimental work in the area of elementary particle collisions is to measure observable quantities of particles as they emerge from collisions; in our case the quantities are z_{hem} and z_{jet}. This task is not straightforward, because measuring devices are not ideal and generally distort the original distribution that is to be measured. We will refer to the original "true" distribution as the "hadron level" distribution and to the actually measured distribution by the detector as the "detector level" distribution. The imperfections of the detector are due to its final granularity, to the inactive materials inside the detector (supporting structures, electronics), to the imperfect calibration, limited geometric acceptance and other effects. Some of these effects can be corrected only with help of a Monte Carlo program where they are simulated. In such a simulation the different effects can be quantified and correction factors for the measured data can be extracted. Let us suppose we measure a distribution d in bins labeled by an integer. Thought-out this section we will use the following notations

- $d_i^{\text{HL, Data}}$ - number of entries in the ith bin of the measured distribution at hadron level. The extraction of these numbers is our aim.
- $d_i^{\text{DL, Data}}$- number of entries in the ith bin of the measured distribution at detector level. These numbers are obtained in the actual experimental measurement by the detector.
- $d_i^{\text{HL, Mc}}$- number of entries in the ith bin of the Monte Carlo distribution at hadron level.
- $d_i^{\text{DL, Mc}}$- number of entries in the ith bin of the Monte Carlo distribution at detector level.

Neglecting a possible background, the response of the detector can be, with the help of Monte Carlo simulation, expressed in the form of a response matrix

$$d_i^{\text{DL, Mc}} = \sum_{j=0}^{N} R_{ij} d_j^{\text{HL, Mc}}.$$

With the response matrix obtained from MC simulations the hadron level distribution can be determined using data

$$d_i^{\mathrm{HL,\,Data}} = \sum_{j=0}^{N} \left(R^{-1}\right)_{ij} d_j^{\mathrm{DL,\,Data}}.$$

This straightforward method may work in certain cases, in practice, however, the results are usually unsatisfactory. This is due to the fact that matrix inversion is often close to an ill-defined problem, i.e. it is sensitive to the statistical fluctuations of the measured distribution and may lead to a strongly fluctuating resulting distribution with huge error bars in each bin. This "unfolding problem" is to be solved by an appropriate procedure, and since one cannot get "something for nothing" an additional *a priori* knowledge must be used. One often requires a certain level of "smoothness" for the unfolded distribution. In the following paragraphs we present three procedures that are commonly used for correcting the measured data to the hadron level. It is worth to mention that all correction methods rely entirely on the correct description of the detector effects by the Monte Carlo simulation.

3.4.1 Bin-by-Bin Method

The bin-by-bin method is a very basic method of correcting the measured data to hadron level. One may be reluctant to call it unfolding, since it takes into account only efficiency related effects of the detector. For each bin a correction factor C_i is extracted from the Monte Carlo simulation $C_i = d_i^{\mathrm{DL,\,Mc}}/d_i^{\mathrm{HL,\,Mc}}$ and this correction factor is applied to the data $d_i^{\mathrm{HL,\,Data}} = d_i^{\mathrm{DL,\,Data}}/C_i$. It can be regarded as an unfolding procedure with $R_{ij} = C_i\delta_{ij}$.

The bin-by-bin procedure usually provides smooth hadron level distribution without big error bars, but its use is justified only when migrations between different bins are small or perfectly described by the Monte Carlo program. This is usually studied using the Monte Carlo simulation by considering *purity* and *stability* distributions. Their definitions are as follows

- purity in bin i: $P_i = \frac{d_i^{\mathrm{DL\&HL,\,Mc}}}{d_i^{\mathrm{DL,\,Mc}}}$,
- stability in bin i: $S_i = \frac{d_i^{\mathrm{DL\&HL,\,Mc}}}{d_i^{\mathrm{HL,\,Mc}}}$,

where $d_i^{\mathrm{DL\&HL,\,Mc}}$ denotes the number of those events that were both generated (hadron level) and reconstructed (detector level) in bin i. One usually requires the purity to be above 40% in each bin. One might adjust the bin limits in order to improve the purity and stability in individual bins in order to make the bin-by-bin method valid.

3.4.2 Singular Value Decomposition (SVD)

This approach to data unfolding is based on the singular value decomposition of the response matrix. It can be regarded as a decomposition of the unfolded distribution

into a basis of vectors of different frequencies. Implementing an *a priori* requirement of smoothness of the unfolded distribution by introducing a regularization parameter τ allows highly oscillating components of the solution to be suppressed, and thus the result becomes less sensitive to statistical fluctuations of the measured distribution. The method is described in detail in reference [67].

In singular value decomposition a real matrix R is factorized into a product of matrices

$$R = USV^T,$$

where U and V are orthogonal matrices $U^TU = UU^T = V^TV = VV^T = I$ and S is a diagonal matrix $S_{ij} = s_i\delta_{ij}$ with $s_i > 0$. The numbers s_i are called "singular values" of the matrix R, and the columns of the matrices U and V are called left and right "singular vectors" of the matrix R respectively. As already mentioned, a non-regularized unfolding of an experimental distribution is equivalent to solving the linear system

$$Rd^{\mathrm{HL,\,Data}} = d^{\mathrm{DL,\,Data}},$$

written now in the matrix form. If the matrix or the right-hand side (r.h.s) of the system of equations have a certain level of uncertainty and some of the singular values of R are significantly smaller than others, then the system is ill-defined and difficult to solve. For this reason regularization is needed. The whole SVD procedure is done in several steps.

- The system is normalized, a change of the variable $w_i = d_i^{\mathrm{HL,\,Data}}/d_i^{\mathrm{HL,\,Mc}}$ is performed. The system can be rewritten in the form

  $$\widetilde{R}w = d^{\mathrm{DL,\,Data}},$$

  where $\widetilde{R}_{ij} = R_{ij}d_j^{\mathrm{HL,\,Mc}}$. This step is motivated by two reasons. Firstly, the ratio $d_i^{\mathrm{HL,\,Data}}/d_i^{\mathrm{HL,\,Mc}}$ expresses the deviation of the distribution to be unfolded from the hadron level Monte Carlo distribution. If we believe that the initial Monte Carlo distribution is close to the reality then the migrations in the variable w should be small. Secondly, the matrix elements $\widetilde{R}_{ij}$ are not interpreted as probability, but rather contain the actual number of events that migrate from the bin i to the bin j. This allows to suppress the influence of poorly populated matrix elements, if there are any.

- The equations are rescaled with respect to the correlation matrix of the measured variable so that each equation in the system has the same weight. If the errors of the measured distribution are purely statistical, then the correlation matrix is diagonal, and the rescaling is equivalent to dividing each equation of the system by the corresponding error $\Delta d_i^{\mathrm{DL,\,Data}}$. If the correlation matrix is not diagonal, the rescaling is slightly more complicated; it is described in reference [67]. The procedure affects the response matrix and the r.h.s of the system. The new system can be formally rewritten as

  $$\widehat{R}w = \widehat{d}^{\mathrm{DL,\,Data}}.$$

- Regularization is introduced by redefining the system of equations

$$\begin{bmatrix} \widehat{R}C^{-1} \\ \sqrt{\tau} \times I \end{bmatrix} Cw = \begin{bmatrix} \widehat{d}^{\mathrm{DL,\,,Data}} \\ 0 \end{bmatrix}, \tag{3.1}$$

  where C is a matrix approximating the second derivative operator, and τ is the regularization parameter. If $\tau = 0$ then the exact, non-regularized solution is obtained, if $\tau \to \infty$, then the input hadron-level Monte Carlo distribution comes out as solution. The regularized system can be solved using the singular value decomposition of the matrix $\widehat{R}C^{-1}$, and the solution of the regularized system can be expressed in terms of the "non-regularized" solution. This expression (see reference [67]) makes explicit the role of the parameter τ; it acts as a (smooth) cut-off parameter to suppress quickly oscillating singular vectors in the solution. Solving 3.1 yields the unfolded experimental distribution $d^{\mathrm{HL,\,Data}}$.

Before applying the SVD method to the measured data one needs to choose an appropriate value for the parameter τ. We studied the statistical significance of the coefficients with which the singular vectors are combined into the solution and set τ to be the square of the singular value that corresponds to the least oscillating singular vector, whose coefficient is compatible with zero taking into consideration its error (see reference [67]).

3.4.3 Bayesian Iterative Approach

The Bayesian iterative approach is based on probability considerations and the Bayes theorem. It is described in reference [68]. Let us use the notation "b_i" to express that an event belongs to the ith bin of a given distribution. Keeping the notations from the previous section one can write down the Bayes theorem relevant for our purposes

$$P(b_i^{\mathrm{HL,\,Data}}|b_j^{\mathrm{DL,\,Data}}) = \frac{P(b_j^{\mathrm{DL,\,Data}}|b_i^{\mathrm{HL,\,Data}}) \times P(b_i^{\mathrm{HL,\,Data}})}{\sum_l P(b_j^{\mathrm{DL,\,Data}}|b_l^{\mathrm{HL,\,Data}}) \times P(b_l^{\mathrm{HL,\,Data}})}, \tag{3.2}$$

with P designating probabilities. The aim is to unfold the distribution $P(b^{\mathrm{HL,\,Data}})$ and, since one relies on Monte Carlo simulation for the detector effects, the assumption $P(b_j^{\mathrm{DL,\,Data}}|b_i^{\mathrm{HL,\,Data}}) \equiv P(b_j^{\mathrm{DL,\,Mc}}|b_i^{\mathrm{HL,\,Mc}})$ is made for the whole unfolding procedure. The approach is iterative, and thus one needs to choose an initial probability distribution $P_0(b^{\mathrm{HL,\,Data}})$. An obvious choice is to take $P_0(b^{\mathrm{HL,\,Data}}) = P(b^{\mathrm{HL,\,Mc}})$, but in principle any initial distribution can be used (for example uniform distribution). The various steps at each (kth) iteration are:

- Using formula 3.2 and the probability distribution $P_k(b^{\mathrm{HL,\,Data}})$ calculate $P_k(b_i^{\mathrm{HL,\,Data}}|b_j^{\mathrm{DL,\,Data}})$.
- Calculate the number of events in bins at hadron level

$$d_{(k)\,i}^{\mathrm{HL,\,Data}} = (1/\varepsilon_i).\sum_j d_j^{\mathrm{DL,\,Data}}.P_k(b_i^{\mathrm{HL,\,Data}}|b_j^{\mathrm{DL,\,Data}}) \ ,$$

where $\varepsilon_i = \sum_j P(b_j^{\mathrm{DL,\,Data}}|b_i^{\mathrm{HL,\,Data}}) \equiv \sum_j P(b_j^{\mathrm{DL,\,Mc}}|b_i^{\mathrm{HL,\,Mc}})$.

- Compute the total number of events at hadron level $d_{(k),\,\mathrm{tot}}^{\mathrm{HL,\,Data}} = \sum_i d_{(k)\,i}^{\mathrm{HL,\,Data}}$ and the new probability distribution at hadron level

$$P_{k+1}(b_i^{\mathrm{HL,\,Data}}) = d_{(k)\,i}^{\mathrm{HL,\,Data}} / d_{(k),\,\mathrm{tot}}^{\mathrm{HL,\,Data}}.$$

- From $P_{k+1}(b^{\mathrm{HL,\,Data}})$ determine the distribution $d_{(k+1)}^{\mathrm{HL,\,Data}}$ (see two first steps) and, on the basis of a χ^2 comparison between the distributions $d_{(k)}^{\mathrm{HL,\,Data}}$ and $d_{(k+1)}^{\mathrm{HL,\,Data}}$, decide to continue with another iteration or stop with $d_{(k+1)}^{\mathrm{HL,\,Data}}$ as the unfolded distribution.

It can be shown that in each iterative step the distribution $P_{k+1}(b^{\mathrm{HL,\,Data}})$ lies between $P_0(b^{\mathrm{HL,\,Data}})$ and the true distribution. The error treatment is described in reference [68] and, since also in this method huge statistical fluctuations appear after a big number of iterations, one needs to stop the iterative procedure at the right time; the number of iterative steps can be regarded as an integer regularization parameter. We determine the number of iterative steps by computing the χ^2: the unfolded distributions obtained in the iterative process tend to converge quickly and after few iterations the differences between $d_{(k)}^{\mathrm{HL,\,Data}}$ and $d_{(k+1)}^{\mathrm{HL,\,Data}}$ become tiny. Calculating $\chi^2[d_{(k)}^{\mathrm{HL,\,Data}}, d_{(k+1)}^{\mathrm{HL,\,Data}}]$, we stop to iterate when further iterations start to have small impact on the unfolded distribution.

3.5 Reconstruction of the Event Kinematics

A good reconstruction of the kinematic variables of the event (see section 1.2.1) is important for our study. It is due to obvious reasons (e.g. correct measurement within the visible range) but also due to the fact that fragmentation observables are determined in the γP frame, where the precision of the boost into this frame is given by the precision of the kinematics reconstruction. Even though the proton remnant cannot be measured, the measurement of electron and final state hadron quantities in the detector provides enough information to reconstruct the kinematics. The kinematics is actually over-constrained by the available measured data, and therefore several different reconstruction methods [69] are possible. They are based on the measured energy E_e and angle θ_e of the scattered electron, or on the energy E_X and the angle θ_X of the hadronic final state (see figure 3.8). The most common methods are the electron method (using E_e and θ_e), the method by Jacquet-Blondel (using E_X and θ_X) or the double angle method (using θ_e and θ_X). In this analysis the conventional electron method is chosen which provides a very good Q^2 and y resolution, it becomes less precise at low y ($y < 0.1$) where other methods become competitive. The kinematic variables, expressed in terms of E_e and θ_e, are

$$y_e = 1 - \frac{E_e}{2E_e^i}(1 - \cos\theta_e),$$

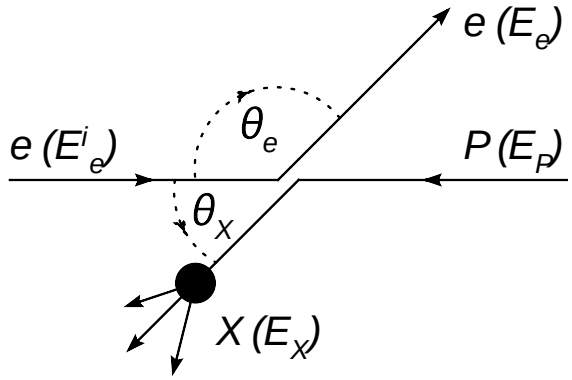


Figure 3.8: Definition of variables used in the reconstruction of the event kinematics.

$$Q_e^2 = 2E_e^i E_e(1 + \cos\theta_e)$$

and

$$x_e = \frac{E_e^i}{E_P} \cdot \frac{E_e(1 + \cos\theta_e)}{2E_e^i - E_e(1 - \cos\theta_e)}.$$

Chapter 4

Data Selection

In this chapter the data selection is presented, starting from general considerations and followed by more detailed technical requirements. The focus here lies on the selection and not on the description of data by Monte Carlo models. With the selection criteria described in the present chapter, approximately 16 600 D^* particles are reconstructed for the whole HERA II running period (see figure 3.3-left).

4.1 Running Periods and Luminosities

The study of charm quark fragmentation into D^* mesons is based on data collected by the H1 detector during the HERA II running period, i.e. during the years 2004 - 2007. For the purpose of more detailed investigations this period is divided into five subperiods. The division is naturally driven by changes of the detector (major shutdowns and upgrades) and of running conditions (positron vs. electron) that may have influenced the data collected. Showing that all these effects are sufficiently well simulated by Monte Carlo simulations for the five subperiods, one can correct the measured data for a possible period dependencies.

The D^* mesons were identified via the "golden" decay channel into $K^{\mp}\pi^{\pm}\pi_s^{\pm}$. The events were selected by the subtrigger *s61* - a level 1 subtrigger described in section 4.2. The total integrated luminosity collected by the H1 detector needs to be corrected for trigger effects, since the subtrigger was disabled for certain runs and had occasionally a small prescale factor applied.

In addition, in our run selection we include only those runs which are labeled as being of a "good" or "medium" quality[1] and during which all major subdetectors were operational (CJC1, CJC2, Spacal, Liquid Argon Calorimeter, Time-of-Flight and Veto Systems, Luminosity System).

The established division into running periods as well as the trigger corrected integrated luminosities that have been collected in the selected runs are presented in table 4.1.

[1] A run at H1 is an artificial time period ($\sim$ tens of minutes) during the HERA accelerator fill ($\sim$ many hours) when all detector and data taking settings remain unchanged. The quality of a run takes into account the general detector performance and data taking conditions.

Running Period	Integrated Luminosity
2004 e^+	50.6 pb^{-1}
2004 e^- and 2005 e^-	112.3 pb^{-1}
2006 e^-	59.2 pb^{-1}
2006 e^+	86.2 pb^{-1}
2007 e^+	45.8 pb^{-1}
All HERA II	354.1 pb^{-1}

Table 4.1: Running periods with corresponding luminosities corrected for high voltage efficiency and trigger prescales.

4.2 Data Online Selection by Trigger

The definition of the subtrigger *s61* which is used to trigger on D^* events was not stable during the whole HERA II running. It can, however, be schematically written in the form

$$s61 = (Spacal\ Part)\ \&\&\ (Track\ Trigger\ Part)\ \&\&\ (VETO\ Part)\ ,$$

where a logical AND is denoted by &&.

4.2.1 Spacal Trigger Elements

The Spacal electron trigger at level 1 [70] is based on a "sliding window" technique. In this method overlapping trigger towers are defined, each trigger tower consisting of 4×4 electromagnetic cells. The trigger towers overlap in both x and y directions so that an electron cluster is well contained within at least one trigger tower. The trigger condition requires that at least one trigger tower energy is above a given threshold; three thresholds have been implemented. The Spacal trigger element combination entering the *s61* subtrigger was stable during the whole HERA II running. It combines two trigger elements using a logical OR

$$Spacal\ Part\ =\ SPCLe_IET > 2\ ||\ SPCLe_IET_Cen_3\ ,$$

where $SPCLe_IET$ refers to a logical OR performed on all Spacal trigger tower energies and it is coded in two bits in order to provide the threshold information. The trigger element $SPCLe_IET_Cen_3$ is a logical OR from Spacal trigger towers in the central IET region[2] with the highest threshold condition. During the HERA II running was the highest trigger threshold set to 9 GeV.

4.2.2 Track Trigger Elements

The track trigger elements related to the subtrigger *s61* were based on trigger information from CJC1 and CJC2. However, the trigger design changed considerably in

[2]Spacal is for triggering purposes divided into 25 so-called IET (Inclusive Electron Trigger) regions, each IET region containing a certain number of trigger towers.

$Track\,Trigger\,Part =$

Trigger Element	Time Period
DCRPh_THig	2004 - January 2005
FTT_mul_Tc>2	January 2005
FTT_mul_Tc>1	January-February 2005
FTT_mul_Td>0	February 2005 - 2007

Table 4.2: Track trigger elements included in the subtrigger *s61* during the HERA II running period.

January 2005, when the DCrϕ trigger [71] was replaced by the more powerful FTT. In addition, the trigger elements from the FTT, included in *s61* were modified twice at the beginning of the FTT running, but only short time periods are affected, after which the FTT trigger elements remained unchanged until the end of the HERA II running. The changes in *s61* related to the track trigger elements are summarized in table 4.2.

- The DCrϕ trigger was based on signals coming from seven wire layers of the CJC1 and three layers of the CJC2. These signals were compared with approximately 10 000 pre-defined signal patterns expected for tracks originating from the origin. The $DCRPh_THig$ trigger element was set, if a track with transverse momentum $p_T > 800\,\text{MeV}$ was found.

- The FTT trigger at level 1 used information from twelve wire layers of the CJC1 and CJC2. These wire layers were organized by three into so-called trigger layers, three trigger layers were situated in the CJC1 and one in the CJC2. The wires of a given trigger layer within one drift cell were called a trigger cell. The trigger decision at level 1 used only rϕ information from hits measured in trigger layers. Signals coming from individual trigger cells were compared with pre-defined masks, and if a match was found so-called track segment together with its $\kappa\phi$ information was considered. Coincidences between different track segments were searched for using the "sliding window" technique in the $\kappa\phi$ plane. If a coincidence between at least two out of four possible segments was found, a track candidate was constructed. The trigger decision was based on the number of tracks above a given threshold. In our case the trigger element definitions were

  1. FTT_mul_Tc>2: At least 3 tracks with $p_T > 400\,\text{MeV}$.
  2. FTT_mul_Tc>1: At least 2 tracks with $p_T > 400\,\text{MeV}$.
  3. FTT_mul_Td>0: At least one track with $p_T > 900\,\text{MeV}$.

4.2.3 VETO Trigger Elements

The VETO trigger elements included in the subtrigger *s61* kept changing during the HERA II period. Their aim was to filter out background, mainly beam-gas and beam-wall events. In the case of *s61* these trigger elements were based on the time-of-flight

(TOF) measurements, i. e. events occurring outside of the interaction time window were not accepted. Since the interaction time window provides a clear and simple rejection criterion, the TOF VETO is considered to be safe and efficient. On the other hand it may happen that a good event is rejected because of a different event (presumably a background event) happening near in time and triggering the TOF VETO. These effects[3] need to be taken into account in a cross section measurement, however, in this analysis we are only concerned with the shape of distributions and not their absolute normalization. Since one does not expect the pile-up background to be correlated with the physics of the event, the results we obtain are not affected by a possible TOF VETO "inefficiency".

In August 2006 a CIP VETO condition was added to *s61*. This step was taken, because of large backgrounds caused by late proton satellite bunches[4] in the proton beam. The condition $!(\,CIP_mul > 11\ \&\&\ CIP_sig == 0\,)$ rejects all events that have a high track multiplicity (>11), and yet the significance of the central peak in the z-vertex histogram is small (as reconstructed from the CIP).

The study of trigger efficiencies for different parts of the subtrigger *s61* is presented in section 5.2.

4.3 Offline Selection

By definition the offline selection cannot improve the measured data. It can, however, reject background events and select events of good quality that fulfill our phase space requirements and correspond to the decay channel under study.

4.3.1 Z-position of Interaction Vertex

The longitudinal size of the interaction region was driven by the proton bunch length ($\sigma_z \approx 13$ cm), the electron bunch length was much smaller ($\sigma_z \approx 2$ cm). For fixed beam parameters the z_{vertex} distribution was approximately Gaussian and the mean position varied by about 1 cm from fill to fill. The beams were focused so as to collide close to nominal interaction vertex and therefore it is meaningful to restrain the z-position of the reconstructed event vertex to the central detector area. Firstly, this way one can reject background events that do not originate from electron-proton collisions. Secondly, one expects the outgoing particles to be well contained in the detector and thus well measured and reconstructed, which might not be the case for events occurring outside of the central detector region. In this analysis the requirement $-35\,\text{cm} < z_{\text{vertex}} < 35\,\text{cm}$ is applied.

[3] Typically the "inefficiency" induced by the TOF VETO reaches couple of percents ($\sim 1-3\%$), see references [72, 54].

[4] The proton bunches were designed to have approximately Gaussian shape. Proton satellite bunches were additional protons situated in tail regions of the main bunches causing deviations from the overall Gaussian shape of the bunches.

Quantity	Selection Criteria
Electron energy	$E_e > 11\,\mathrm{GeV}$
z-Position of the electron cluster barycenter	$z_{\mathrm{cluster}} > -180\,\mathrm{cm}$
Energy in the hadronic part of the Spacal	$E_e^{\mathrm{Had.\ Spacal}} < 0.5\,\mathrm{GeV}$
Fraction of the electron energy in the hadronic part of the Spacal	$\frac{E_e^{\mathrm{Had.\ Spacal}}}{E_e} < 0.03$
Relative difference between the energy of the isolation cone and the energy of the electron cluster	$\frac{E_{\mathrm{Iso.\ Cone}} - E_e}{E_e} < 0.1$

Table 4.3: Requirements on an electron candidate. For more detailed explanation refer to the text.

4.3.2 Electron Quality Requirements

The scattered electron candidate as provided by the electron finder (section 3.2.1) fulfills only the very basic quality criteria. In order to reject photoproduction background[5], accept only events that allow for reliable kinematics reconstruction and avoid a wrong energy measurement or trigger inefficiencies, further quality criteria on the reconstructed electron are applied. Some of these criteria apply to each electron candidate as such, some of them are detector-related and take into account the presence of areas in the Spacal which suffered from a bad energy measurement due to dead or miscalibrated cells and/or from trigger inefficiencies. In addition to the Spacal also the BPC efficiency was low for some time periods during the HERA II running which also affects the reconstruction of the scattered electron.

General Criteria

The criteria related to measured quantities of the scattered electron are summarized in table 4.3. The cut on the electron energy is mainly driven by the Spacal trigger efficiency, which decreases towards low energies. The following three criteria, the z-position of the electron cluster barycenter and the energy and the fractional energy measured in the hadronic Spacal in the direction of the electron candidate, reflect the expectation for an electron cluster to be well contained in the electromagnetic part of the Spacal. The last criterion involves an isolation cone, a suitably defined cone around the electron cluster, that may, in addition to the electron cluster, contain some energy deposition. For a precise electron energy measurement one expects the electron cluster to be separated in space - an energy deposit nearby introduces an ambiguity on whether it is indeed originating from the electron and should contribute to the electron energy or not.

[5]By photoproduction background one understands events, where the electron scatters under a very small angle and escapes down the beam pipe, and a hadron measured in the detector is misidentified as the scattered electron.

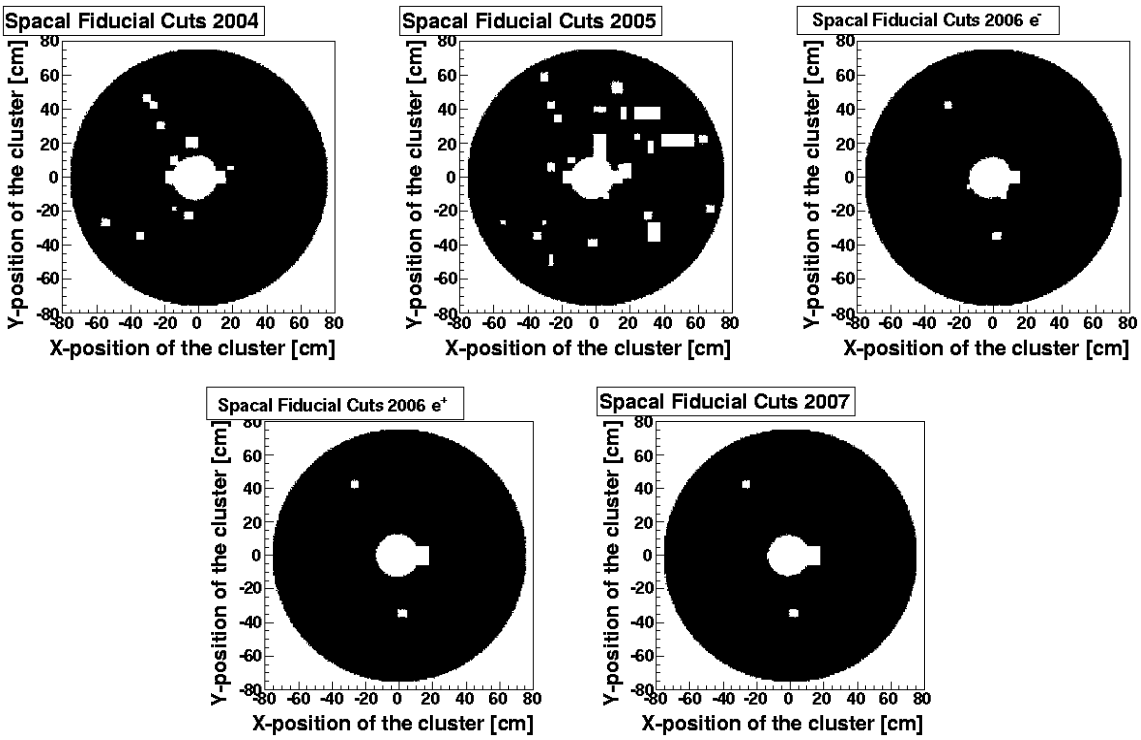


Figure 4.1: Spacal fiducial cuts in five different run periods. The excluded areas are white.

Spacal Fiducial Cuts

The Spacal areas suffering from miscalibrated, dead or trigger inefficient cells are excluded by applying fiducial cuts. The measurement of trigger efficiencies will be explained in section 5.2. The areas leading to wrong energy measurements are identified by detector experts (list of dead cells) and by comparing the measured energy of the electron in the Spacal with the electron energy computed with the "double angle" method (see section 3.5). In general, the areas with biased energy measurements were usually situated near the beam pipe, where leakage of the cluster energy out of the inner Spacal modules is to be expected. Therefore, for each period (see table 4.1) an appropriate circular or elliptical radial cut on the innermost Spacal region is applied. The other dead, miscalibrated or trigger inefficient areas were usually also situated in the inner Spacal region, since this region suffered most from high synchrotron radiation. The Spacal performance was rather bad during the year 2004, became better in 2005, and the detector was in a very good shape during the years 2006 and 2007. The performance depended on both, running conditions and hardware repairs, the major repairs took place during the winter 2005/2006 shutdown. The fiducial cuts are defined with respect to run numbers and so can change during a run period. However, this is the case for only a few of them, the stable cuts are visualized in figure 4.1.

BPC Treatment

Information from the BPC is, if available, used when defining an electron candidate, however, requiring it for every event would lead to an important loss of statistics because of insufficient spatial acceptance of the BPC with respect to the Spacal. Thus, the question of BPC efficiency arises. We define the BPC efficiency to be the ratio of the number of events with a scattered electron candidate within the BPC acceptance and with BPC information to the total number of events with a scattered electron candidate within the BPC acceptance

$$\varepsilon_{\mathrm{BPC}} = \frac{\#\,\text{events(BPC acceptance AND BPC information)}}{\#\,\text{events(BPC acceptance)}}\,.$$

This efficiency is presented for different run periods in figure 4.2. Each histogram bin contains a given number of events that is fixed for each period; the bins are filled progressively starting from the first one respecting the time sequence of the events. One notices that the BPC efficiency was low in the year 2004 and at the beginning of 2005 as well as at the end of the year 2007. In order to avoid the usage of BPC information for periods with doubtful BPC performance, all electron quantities are recalculated for these periods using the Spacal measurement only[6]. We did this for all runs with Run Number < 411288 (2004 and 2005) and for all runs having Run Number > 499811 (2007).

4.3.3 Selection with Respect to Event Kinematic Variables

Energy and Longitudinal Momentum

Four-momentum conservation implies that the quantity $E - p_z$ is conserved in each event. In the initial state only the proton and the electron are present. Neglecting the particle masses ($E \approx p_z$) one can write

$$(E - p_z)_{\mathrm{initial}} = (E - p_z)_p + (E - p_z)_e = 2E_e \approx 55\ \mathrm{GeV}.$$

In the final state one needs to sum over all final-state particles

$$(E - p_z)_{\mathrm{final}} = (E - p_z)_{e,\,\mathrm{final}} + \sum_{i \neq e} (E - p_z)_{i,\,\mathrm{final}}.$$

From the conservation law $(E-p_z)_{\mathrm{initial}} = (E-p_z)_{\mathrm{final}}$ one expects $(E-p_z)_{\mathrm{final}} = 2E_e$. The measurement, however, suffers from resolution effects and, in general, not all particles are measured in the detector and therefore a value differing from the expectation can be obtained when reconstructing $(E - p_z)_{\mathrm{final}}$ from the detector information. The particles that are usually not measured in the detector are those originating from the proton remnant, which in most cases scatter under small angles and go down the beam pipe. The small scattering angle means that the z-component of their momenta is dominant and thus their contribution to the sum $\sum_{i \neq e}(E -$

[6]The kinematic variables reconstructed from the electron quantities were also recomputed.

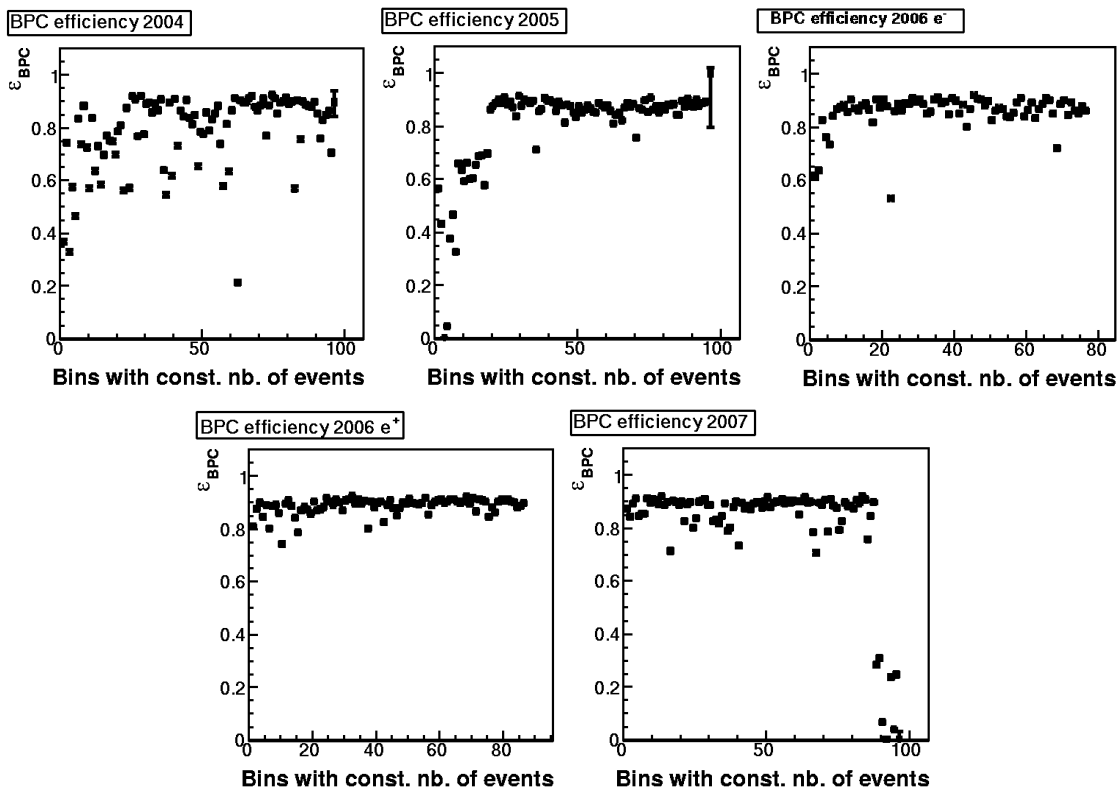


Figure 4.2: BPC efficiency in HERA II. Bins of histograms contain a constant number of events that have the scattered electron within the BPC acceptance.

$p_z)_{i,\,\mathrm{final}}$ is small. Therefore, to a good approximation one still expects $E - p_z \approx 55\,\mathrm{GeV}$, where the electron is supposed to be measured in the backward calorimeter.

The quantity $E - p_z$ is interesting from the point of view of background suppression. If, in a photoproduction event, the scattered electron leaves the interaction under a small angle disappearing in the beam pipe and a hadron misidentified as scattered electron is reconstructed in the Spacal, then $E - p_z$ is strongly shifted towards small values allowing for a good separation between DIS and photoproduction. For a small fraction of events is $E - p_z$ significantly above 55 GeV. Such values cannot be explained by particles not measured in the detector or by resolution effects, but rather point to a poor event reconstruction. To avoid badly reconstructed events and suppress the photoproduction background, we select only events that fulfill

$$40\,\mathrm{GeV} < E - p_z < 75\,\mathrm{GeV}.$$

Phase Space Requirements

In this work the study of charm fragmentation is restricted to the phase space region

$$5\,\mathrm{GeV}^2 < Q^2 < 100\,\mathrm{GeV}^2$$

and

$$0.05 < y < 0.6.$$

Quantity	Requirement
$p_T(D^*)$	$> 1.5\,\text{GeV}$ and $< 15.0\,\text{GeV}$
$p_T(\pi_s)$	$> 0.12\,\text{GeV}$
$p_T(K) + p_T(\pi)$	$> 2.0\,\text{GeV}$
$\lvert m(D^0_{\text{candidate}}) - m(D^0)\rvert$	$< 0.07\,\text{GeV}$

Table 4.4: Selection criteria on D^* candidates with respect to reconstructed momenta and invariant masses.

The first condition - the DIS regime - is driven by the Spacal geometrical acceptance, since Q^2 and the electron scattering angle are strongly correlated. One could think of studying charm fragmentation also in the photoproduction regime. However, before the FTT was implemented there were no appropriate triggers that would allow to trigger on D^* events with high efficiency.

The lower limit of 0.05 on the inelasticity y is due to deteriorating resolution of the "electron method" (section 3.5); this method becomes less precise at lower y. The upper limit in y is related to the minimal electron energy requirement. A higher y limit would lead to a poorly populated high-y phase space region, since y and the electron energy are strongly correlated. In addition, this cut suppresses events which suffer from large corrections due to QED effects.

4.3.4 Selection of D^* Events

The selection of D^* events is based on the D^* finder described in section 3.2.4. All events with at least one D^* candidate are considered, and further requirements are applied so as to improve the signal to background ratio. Since only the D^*-daughter particles ($K^{\mp}\pi^{\pm}\pi_s^{\pm}$) are directly observed in the detector, most of the following criteria are related them.

Requirements on Reconstructed Momenta and Invariant Masses

We mention here only the requirements that are tighter than those in the D^* finder. They are summarized in table 4.4. The tighter cut on the reconstructed invariant mass of the D^0 candidate straightforwardly rejects additional background. Also the cuts on the transverse momenta are related to signal enhancement and background suppression and have been used in most of the H1 D^* analyses. In principle they could be lowered; the corresponding effect was studied on all HERA II data. The cuts were lowered in a consistent way, respecting their mutual correlations[7], down to $p_T(D^*) > 0.8\,\text{GeV}$, $p_T(\pi_s) > 0.07\,\text{GeV}$ and $p_T(K) + p_T(\pi) > 1.1\,\text{GeV}$, which is close to the limits imposed by the D^* finder. One observes that the expected increase of the signal (approximately 20%) is strongly penalized by a much larger rise of the combinatorial background (approximately 180%) with the final effect of a bigger relative error, i. e. $(\sigma_{\#D^*}/\#D^*)_{\text{Higher Cuts}} = 0.0225 < 0.0281 = (\sigma_{\#D^*}/\#D^*)_{\text{Lower Cuts}}$ (see also figures 3.3-left and 4.3). Since the relative error is the relevant quantity

[7]For correlation plots of these quantities see for example reference [51].

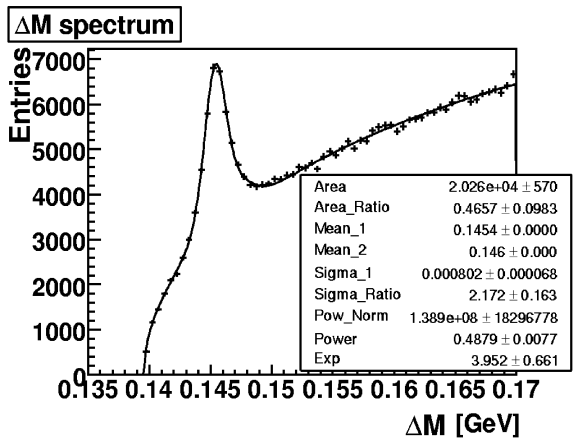


Figure 4.3: $\triangle M$ spectrum for HERA II data, corresponding to lowered p_T cuts on momenta of the D^* meson and its daughter particles (see the text).

when studying the shapes of the $z_{\text{jet/hem}}$ distributions, the lowered p_T cuts were abandoned.

An additional condition is applied when studying z_{jet} and $z_{\text{hem}}^{(\text{jet})}$ (see sections 3.3.2 and 3.3.1). These two observables are determined only for those events where a D^* jet with $E_T > 3.0\,\text{GeV}$ is found. This requirement is considered as an additional constraint and is not taken into account in chapter 5, when control distributions or other plots are presented in which events "from the $\triangle M$ distribution" enter.

Track Quality Requirements

Detector-related quality criteria are applied to the reconstructed tracks of the D^* daughter particles. They are meant to ensure a selection of well measured tracks. The limits on different quantities were studied and defined already in HERA I (see references [73, 66]), and since the relevant tracking detector (the CJC 1 and 2) remained the same in HERA II, the use of the same criteria seems to be justified. They are identical for the kaon and the pion, because comparable transverse momenta for both particles are expected on average. The criteria differ for the slow pion whose typical transverse momentum tends to be much smaller. The requirements are summarized in table 4.5. The quantity d'_{ca} refers to the distance of closest approach to the reconstructed interaction vertex in the $r\varphi$-plane, and $\Delta z_0 = |z'_{\text{ca}} - z_{\text{vtx}}|$, where z'_{ca} is the z-coordinate of the point of closest approach in the $r\varphi$ plane, as indicated in figure 4.4.

Energy Loss and Particle Identification

Further improvement of the signal to background separation is achieved by considering the energy loss of charged particles in the gas of the central tracking chambers. The energy loss differs for different particles and can be determined from reconstructed tracks. However, particle identification based on the energy loss is reliable

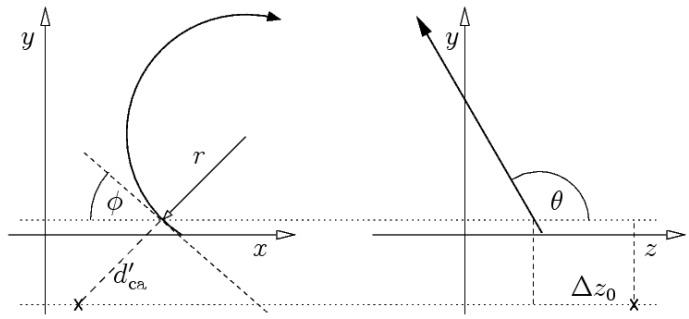


Figure 4.4: Definition of the track quantities d'_{ca} and z_0 that enter the selection criteria. The reconstructed interaction vertex is denoted by "x".

Quantity	Requirement for K and π	Requirement for π_s
Track length	> 18.9 cm	> 10 cm
Starting radius	< 30 cm	< 30 cm
Number of CJC hits	> 10	> 10
$\lvert d'_{\mathrm{ca}} \rvert$	< 1 cm	< 1 cm
Δz_0	< 20 cm	< 20 cm
$\lvert d'_{\mathrm{ca}} \sin\theta \rvert$	< 0.5 cm	< 0.7 cm
$\lvert \Delta z_0 \sin\theta \rvert$	< 18 cm	< 18 cm

Table 4.5: Quality criteria on reconstructed tracks of the D^* daughter particles $K^{\mp}\pi^{\pm}\pi_s^{\pm}$.

only for particles[8] with relatively small momenta ($\sim 2\,\mathrm{GeV}$ and smaller) at H1, for higher momenta the energy losses are very similar and thus the separation becomes difficult, if not impossible. The mean energy loss is predicted by the Bethe-Bloch formula [2]

$$-\frac{dE}{dx} = Kz^2 \frac{Z}{A}\frac{1}{\beta^2}\left[\frac{1}{2}\ln\frac{2m_e c^2\beta^2\gamma^2 T_{\max}}{I^2} - \beta^2 - \frac{\delta}{2}\right]$$

with

$$K = 4\pi N_A r_e^2 m_e c^2,$$

where z is the charge of the incident particle (in units of the elementary charge). Z and A are the atomic number and the atomic mass of the absorber and $\beta = p/m$ and $\gamma = E/m$ are the standard relativistic factors of the incident particles, I is the mean excitation energy, N_A is Avogadro's number, r_e is the classical electron radius, and $T_{\max}$ is the maximum kinetic energy that can be imparted to a free electron in a single collision. The letter δ denotes the density effect correction - a complicated expression, if expressed in terms of material and kinematic variables.

In the H1 software a parametrized formula that mimics the Bethe-Bloch equation is used. It allows to tune the parameters such that possible deviations from the Bethe-Bloch prediction arising from detector and track reconstruction effects are taken into account. Its expression[9] is

$$-\frac{dE}{dx} = \frac{P_1}{\beta^2}\left[P_2 + \ln\left(\beta^2\gamma^2\right) - \beta^2 - \delta\right],$$

with the parameter values $P_1 = 0.45879$ and $P_2 = 9.6433$. The measured energy-loss together with predictions from the H1 parametrization is shown in figure 4.5. The figure is based on all HERA II data and is done for all selected tracks in events with a D^* candidate that satisfies all imposed D^* criteria.

The energy-loss requirements used in this work are inherited from previous D^* analyses ([73, 66]) and are based on likelihood considerations. For each daughter particle of the D^* the normalized likelihood of being the correct type is computed. The calculation uses the H1 parametrization and the measurement errors $\sigma\left(\frac{dE}{dx}\right)$ and $\sigma\left(p_{\mathrm{track}}\right)$. The criteria for a good signal to background separation are chosen depending on the momentum and are shown in table 4.6. The error $\sigma\left(\frac{dE}{dx}\right)$ strongly depends on the number of measured hits for a given track and thus, for the K and π particle tracks, a minimum of 10 so-called dE/dx-hits is required (hits used in the $\frac{dE}{dx}$ calculation). Furthermore, the computation of the energy-loss can in some cases be unreliable or technically impossible, and in these cases $\frac{dE}{dx}$ is set to a negative value. Therefore the cut $\frac{dE}{dx} > 0.01$ is applied for the K and π particle candidates.

[8] By "particles" we refer here to charged hadrons that appear in the detector (p, K, π); electrons and muons are identified by other means.

[9] The correction $\delta = \delta(P_3, P_4)$ depends on two additional parameters and has a rather complicated expression that can be found in the H1-OO class *H1Dedx*. The values of the parameters are $P_3 = 0.022817$ and $P_4 = -1.1995$.

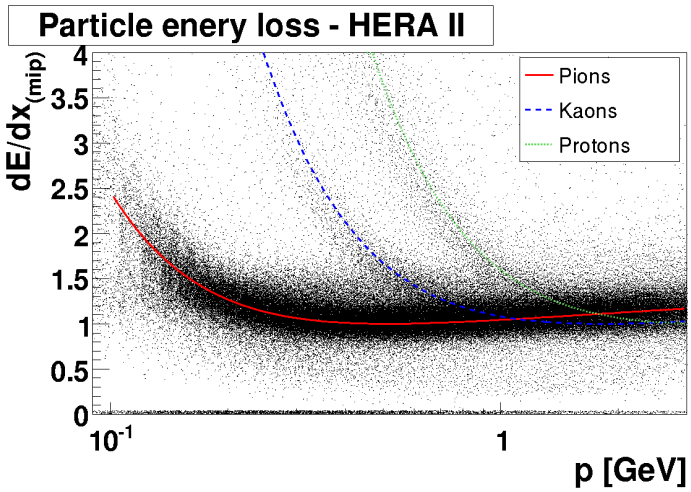


Figure 4.5: The energy loss as function of the particle momentum as measured by the H1 detector together with the H1 parametrizations of energy loss for different particle types. The energy loss is scaled with respect to that of minimum ionizing particle (mip).

Momentum range	LH(K)	LH(π)	LH(π_s)
$p < 0.7$ GeV	$> 20\%$	$> 5\%$	$> 5\%$
0.7 GeV $< p <$ 1.2 GeV	$> 5\%$	$> 5\%$	$> 5\%$
1.2 GeV $< p$	-	-	$> 5\%$

Table 4.6: Criteria imposed on normalized likelihoods (LH) for the D^* daughter particles to be of the correct type.

Chapter 5

Data Treatment and Description of Data by Monte Carlo Models

The structure of this chapter will in many points follow the structure of the previous one. The exception is the first section where some additional information about the used MC programs is provided. In the remaining sections the trigger efficiencies and selection and event quantities are studied in a year-dependent way in order to determine which data corrections and Monte Carlo reweightings are needed. Finally, plots for the whole HERA II running period are presented.

5.1 Monte Carlo Programs

We use the RAPGAP and CASCADE Monte Carlo (MC) event generators together with detector simulation software to correct the measured data for detector effects. The description of RAPGAP and CASCADE was given in section 1.4.1 and we use both of these programs so that the systematic error related to the MC model can be estimated. The description of the data by the MC programs is studied for each running period separately using detector simulation software version appropriate for the period. A good description of the data by the MC model is achieved by applying necessary corrections to both data and MC models, so as to be able to perform the unfolding to the hadron level. The MC events are treated - with the exception of triggers - in the same way as data, identical requirements and cuts are applied. The trigger efficiencies are determined from appropriate data and corresponding corrections are applied to the data to be analyzed, as will be described in section 5.2.

In addition to trigger inefficiency corrections applied to data, other corrections are applied to the MC simulations. The correction procedures are defined in a period-dependent way and figures shown, based on all of HERA II data, take into account period-dependent correction factors. The MC distributions for different time periods are luminosity reweighted so that luminosity fractions corresponding to different running periods in the MC simulations are the same as those in the data.

5.1.1 Treatment of QED Radiation and Resolved Processes

MC models, which are typically based on QCD calculations involving different approximations often do not include QED radiative effects. The measured data however contain these effects and one needs to correct for them, since they are usually not of interest and in order to compare the data to the theory. This correction needs to be done by experimentalists, since it depends on the detector hardware, its resolution and granularity. The correction procedure is possible due to the availability of some MC simulations which contain QED radiation.

In the case of the RAPGAP MC program we have two options: initial and final state QED radiation can be switched on or switched off. We expect the data to be described when QED radiation is turned on. Together with simulations with QED radiation off, it allows us to extract correction factors for QED radiative effects to be applied to the measured data.

In the CASCADE MC program QED radiation is not implemented. Thus, at least for the bin-by-bin unfolding method, QED correction factors from RAPGAP are used. The procedure will be discussed in more detail in the chapter dedicated to the data unfolding to hadron level.

The contribution from resolved processes is omitted for the purpose of this analysis. This can be justified by the fact that the resolved component tends to be suppressed at higher Q^2. Our requirement for the transferred four-momentum is $5\,\mathrm{GeV} < Q^2$, which is a higher limit than the one applied in previous HERA I studies, where the resolved contribution was considered. Furthermore, we observe that the direct contribution of RAPGAP alone describes the data well. This observation is also confirmed by other D^* analyses of HERA II data [51]. In addition, a MC simulation for resolved processes with QED radiation is not available, so including a resolved component in our study would introduce new complications and uncertainties related to properly correcting the MC distributions for QED radiative effects. In the evaluation of systematic errors, an error related to the resolved component is however taken into account.

5.1.2 Treatment of Beauty Component

A small fraction of the reconstructed D^* particles originates from decays of beauty hadrons. Since we study the fragmentation of the charm quark, the beauty component is regarded as background and needs to be subtracted. We subtract the expected beauty contribution as predicted by the RAPGAP MC simulation directly from the measured data, normalizing the data and the beauty MC distributions to the luminosity. All figures which include presentations of data are data after the beauty subtraction.

A D^* particle originating from a beauty hadron takes a certain fraction of the hadron momentum, and the beauty hadron itself inherits a certain fraction of the b quark momentum. These consecutive momentum transfers lead to z_{jet} and z_{hem} spectra with softer fragmentation as will be shown in chapter 6. The beauty-induced D^*s represent approximately 1% - 2% of the total number of measured D^*s.

5.2 Trigger Efficiencies

When studying the measured data one usually tries to describe the measurements by Monte Carlo simulations which include all measurement effects. In this way one aims to achieve a good description of the data by the MC model so as to be able to correct the data to the hadron level. This is the approach we follow with respect to all detector effects like acceptance, efficiencies, resolution etc., however with the exception of trigger efficiencies. The main reason is that the official Monte Carlo production does not yet include the FTT simulation at the time of writing this thesis. Thus the trigger efficiencies for both, the track trigger and the Spacal, are obtained from data. We apply different procedures for the two triggers.

In the case of the Spacal trigger the inefficiency is propagated only into the measurement error, and we do not correct for it. The trigger efficiency of the Spacal depends on

- the electron energy,
- the position of the electron induced shower in the Spacal (detector defects) and
- the background (firing a VETO trigger or a charged hadron firing a Spacal trigger).

One does not expect the fragmentation observables to be correlated with the background, and one does not expect a strong correlation with respect to the electron energy or the electron position in the Spacal. Thus a small inefficiency has an effect on the absolute normalization of our distributions but should not have a significant impact on their shape. In addition, even if some correlation is present, the Spacal trigger is highly efficient and efficiencies are relatively flat with the exception of poorly populated regions (low electron energy, low electron theta) whose statistical significance is suppressed. One should also stress that the error due to Spacal trigger inefficiency has small contribution to the total systematic error.

In the case of the track trigger we correct for the inefficiency and make an estimate for the uncertainty of this correction. The decision to correct for the track trigger inefficiency follows from an observed dependence of the inefficiency on the observables (z_{jet}, z_{hem} and $z_{\mathrm{hem}}^{(\mathrm{jet})}$) and from the fact that the inefficiency becomes rather large (5% - 8%) for some run periods and histogram bins.

We study the trigger efficiencies in a period-dependent way, since important trigger-related upgrades occurred during the HERA II running, e.g. the upgrade of the track trigger at the beginning of 2005 and the repair of the Spacal detector during the shutdown 2005/2006.

5.2.1 Spacal Trigger Efficiency

The Spacal trigger efficiency was determined using a set of triggers based on the Liquid Argon calorimeter. Most of them are meant to trigger on high Q^2 events (electron in the LAr calorimeter) or on charged-current events (neutrino in the final

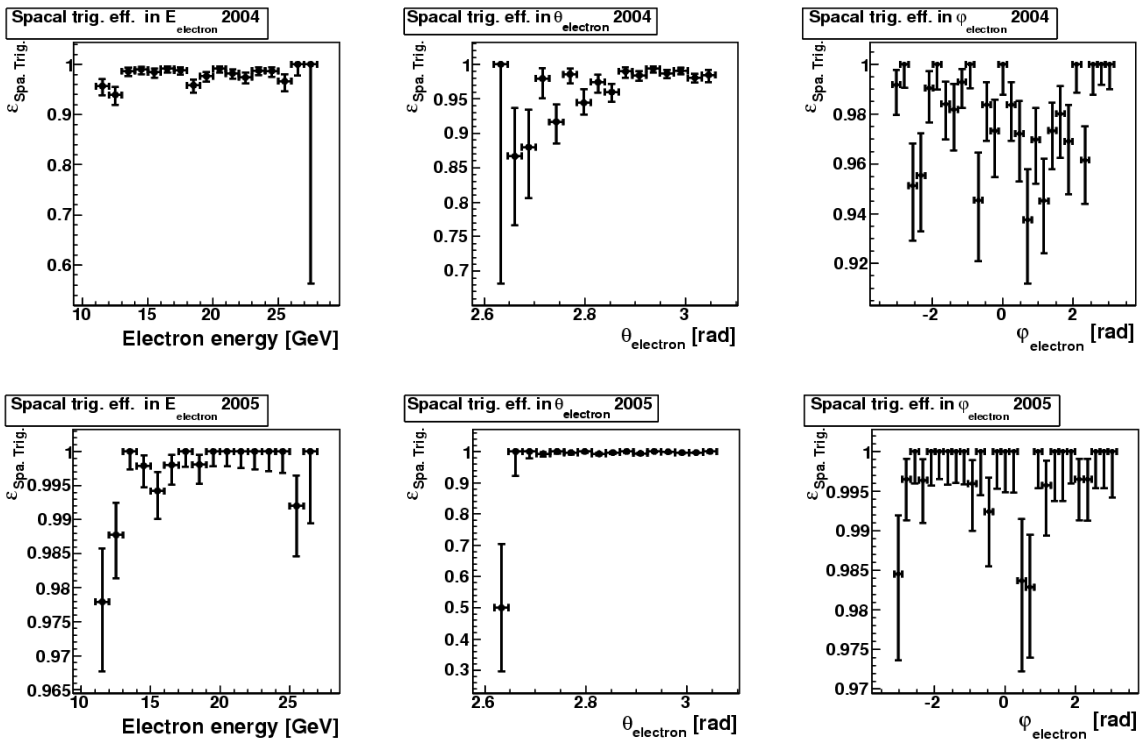


Figure 5.1: Spacal trigger efficiencies as a function of E_{electron}, θ_{electron} and $\varphi_{\text{electron}}$ for the running periods 2004 and 2005.

state and thus missing p_T). We however require the scattered electron to be reconstructed in the Spacal calorimeter and thus no high Q^2 or charged-current events enter our study, the related triggers being fired by background[1]. The actual trigger set chosen to monitor the *s61* trigger contains the subtriggers *s67, s24, s25, s75, s64, s77, s83, s84* and *s68*.

In addition to an appropriate choice of monitor triggers one also needs to make a correct choice of events. If the trigger efficiency would depend only on the electron energy and the electron position in the Spacal, one could study it with all measured events and in this way significantly increase the statistics. However, the trigger efficiency is also related to the background which is reduced by our analysis cuts. So, in order to avoid a possible bias caused by our analysis requirements, we study the Spacal trigger efficiency only for events that enter the ΔM distribution.

Note that by the Spacal trigger efficiency we here understand the ratio of the number of events that were triggered by a monitor trigger (at least one included in the set) and had the "Spacal part" ($= SPCLe_IET > 2 \,||\, SPCLe_IET_Cen_3$) set "on" (the Spacal part of the *s61* fired) to the total number of events triggered by the independent trigger mix

$$\varepsilon_{\text{Spacal trigger}} = \frac{\#\text{ events (monitor trigger AND Spacal part of s61)}}{\#\text{ events (monitor trigger)}}.$$

[1]Background with respect to the selected triggers. Events with D^* in the visible range of this analysis can be accidentally triggered by high Q^2 or charged-current triggers.

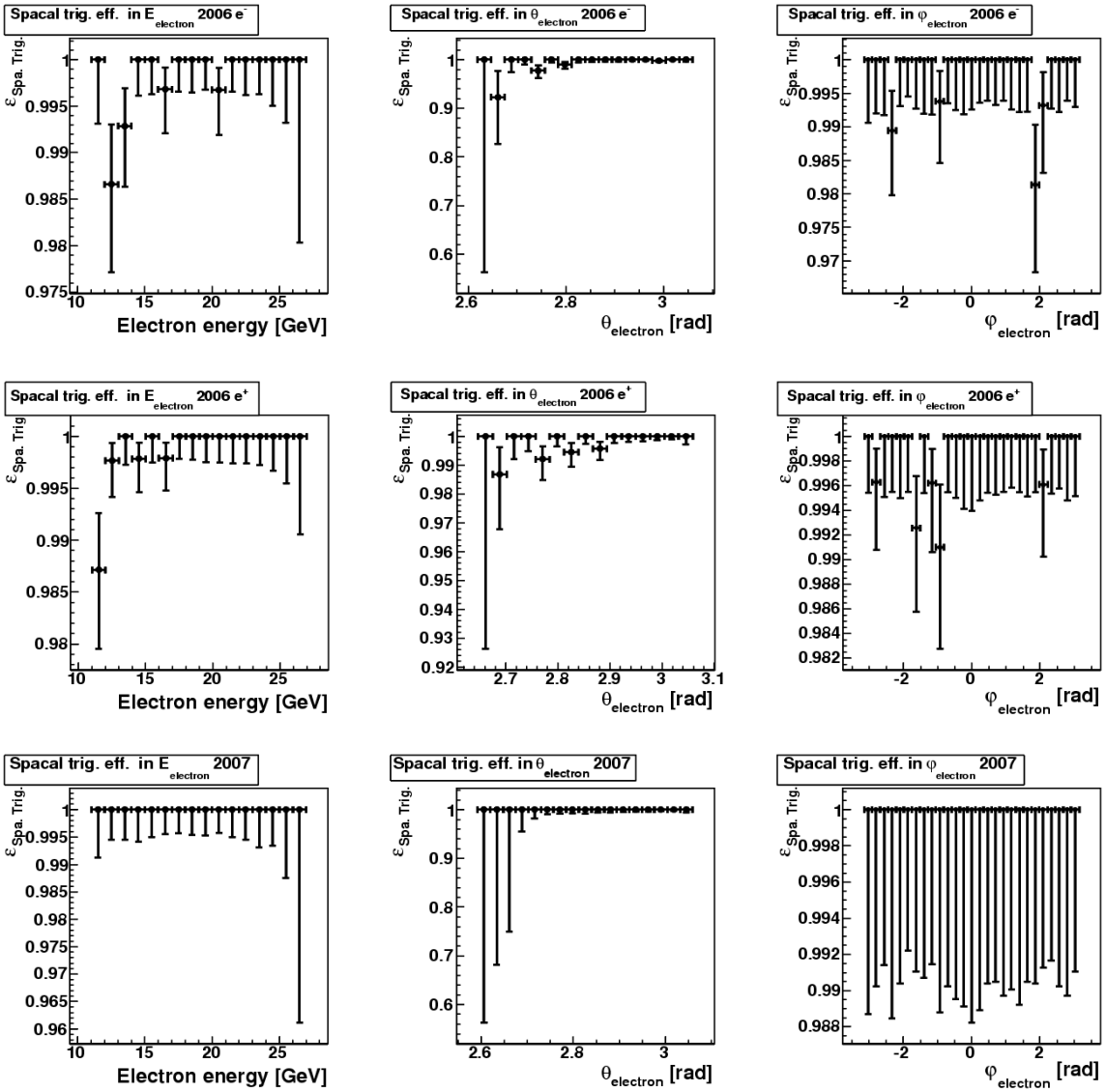


Figure 5.2: Spacal trigger efficiencies as a function of E_{electron}, θ_{electron} and $\varphi_{\text{electron}}$ for the running periods 2006 e^-, 2006 e^+ and 2007.

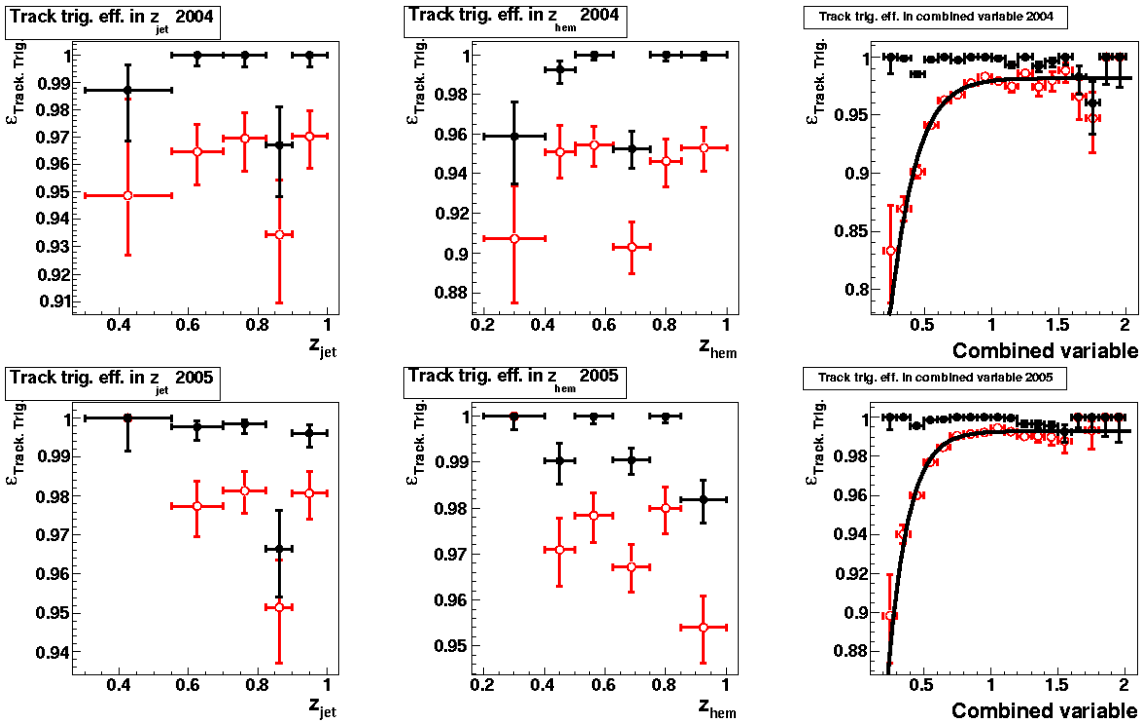


Figure 5.3: Track trigger efficiencies as a function of z_{jet}, z_{hem} and the *combined variable* for the running periods 2004 and 2005. The open circles (red) correspond to the status before the reweighting, the full markers (black) to the situation after the reweighting was applied.

The Spacal trigger efficiencies as a function of electron energy, electron θ and electron φ are shown in figures 5.1 and 5.2 for the five running periods of HERA II. One observes that, with the exception of the year 2004, the efficiencies are high - reaching a level of 99%. The regions of lower efficiencies correspond to low electron energy and high electron scattering angle (low θ_{electron}) and are, as already mentioned, statistically less significant.

5.2.2 Track Trigger Efficiency

The track trigger efficiency was determined using the Spacal subtriggers *s1, s2, s3, s4, s35* and *s36*. The definitions of these subtriggers were changed several times such that it was not possible to use all of them for the entire HERA II running period. The choice of monitor triggers was thus run-dependent; for a given period an appropriate subset of the mentioned subtriggers was chosen in order to use subtriggers that depend on the Spacal only. As in the previous case we considered only those events that enter the ΔM distribution in order to avoid possible biases. Furthermore, for the determination of trigger efficiencies as a function of the variables z_{jet}, z_{hem} and $z_{\text{hem}}^{(\text{jet})}$ the signal was extracted in bins of these variables in order to eliminate possible differences in trigger efficiency between the signal and the combinatorial background.

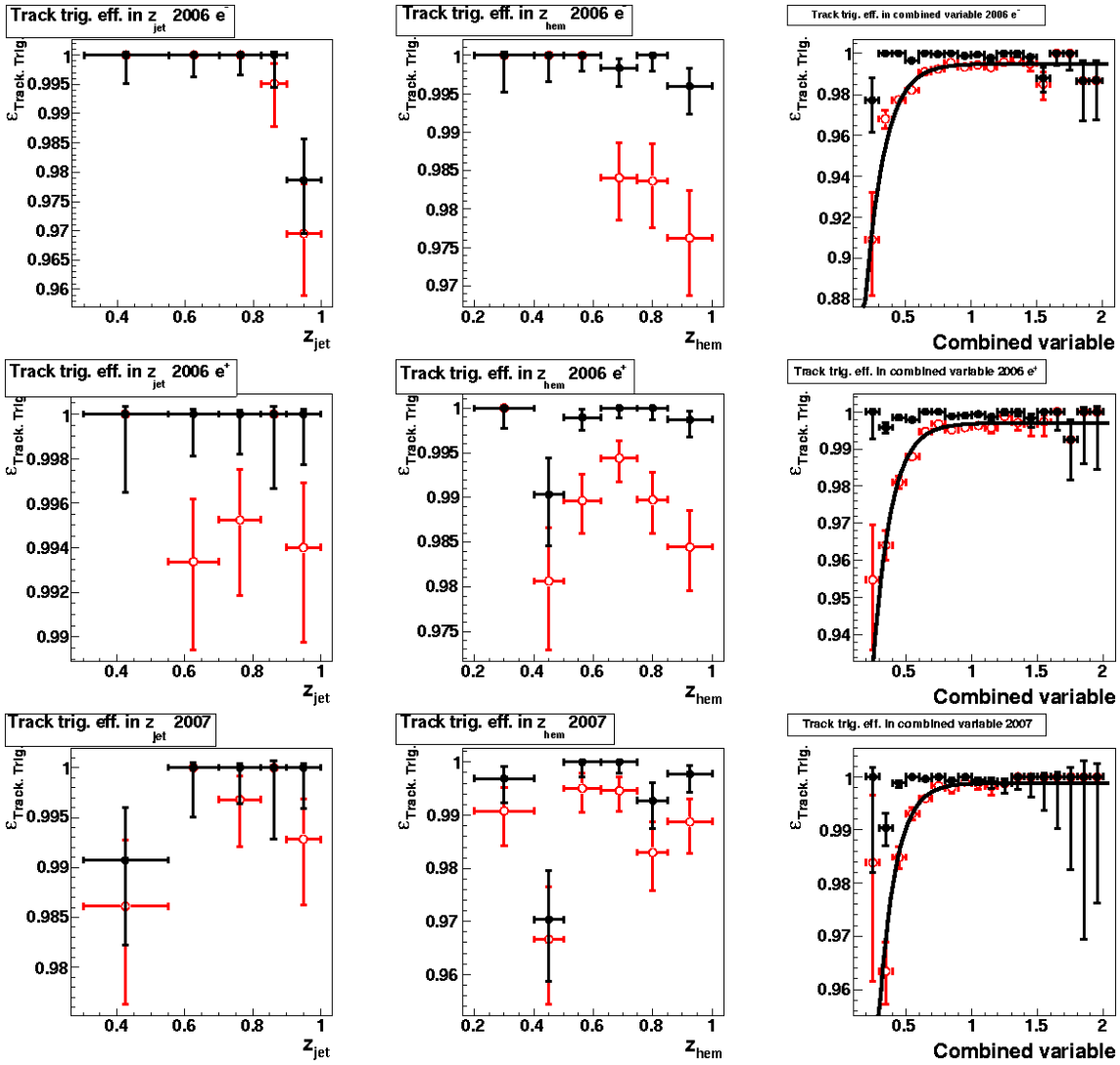


Figure 5.4: Track trigger efficiencies as a function of z_{jet}, z_{hem} and the *combined variable* for the running periods 2006 e^-, 2006 e^+ and 2007. The open circles (red) correspond to the status before the reweighting, the full markers (black) to the situation after the reweighting was applied.

The track trigger efficiency is defined as

$$\varepsilon_{\text{track trigger}} = \frac{\#\text{ events (monitor trigger AND track trigger part of s61)}}{\#\text{ events (monitor trigger)}},$$

where the meaning of "*track trigger part of s*61" is given in table 4.2.

When correcting for the track trigger inefficiency one needs to find an appropriate way when a trigger simulation is not yet available. We chose to reweight each event according to a variable sensitive to the track trigger efficiency. This procedure is more consistent than correcting each individual distribution with different correction factors. Reweighting on an event basis propagates the corrections in a consistent way into each measured distribution. On the contrary, when correcting individual distributions the statistical uncertainties of the correction factors for two correlated distributions might lead to an inconsistent procedure of treating these distributions, which would not respect their mutual correlations.

The track trigger is a complicated device and one cannot expect its efficiency behavior to be fully described by a single-variable distribution. Thus one cannot expect a reweighting depending on a single variable to provide a perfect correction. However, as the results show, this procedure improves the trigger efficiency distributions significantly. The choice of the sensitive variable was driven by the expected behavior of the track trigger. The track trigger is supposed to trigger on high-p_T tracks. In addition, the track trigger is based on the comparison of a measured signal patterns with pre-defined patterns. Hence one expects it to depend on the track multiplicity of the event - the more tracks in the event, the bigger is the probability for one to fit a pattern. Typical transverse momenta of particles are usually below 8 GeV and the typical track multiplicities (for "selected" tracks) are below 40. Thus we define the *combined variable*

$$\textit{combined variable} = \frac{\textit{Highest track } p_T}{8\,\text{GeV}} + \frac{\textit{Event track multiplicity}}{40},$$

with typical values between 0 and 2. The track trigger efficiency as a function of this variable is determined and then fitted with the function

$$f(x) = \frac{A}{1 + \exp[-B(x + C)]},$$

where A, B and C are free parameters. Then each event is weighted by $1/f(x)$ so as to make the trigger efficiency in the *combined variable* flat and close to one. The result of this procedure as well as its impact on the trigger efficiencies for z_{jet} and z_{hem} can be seen in figures 5.3 and 5.4, the efficiencies for $z_{\text{hem}}^{(\text{jet})}$ can be found in appendix A. One observes that the reweighting noticeably improves the trigger efficiencies for the observables z_{jet}, z_{hem} and $z_{\text{hem}}^{(\text{jet})}$.

5.2.3 CIP VETO Condition

The CIP VETO condition introduced in section 4.2.3 contains tracking information and thus one cannot exclude variations of the trigger efficiency as a function of the

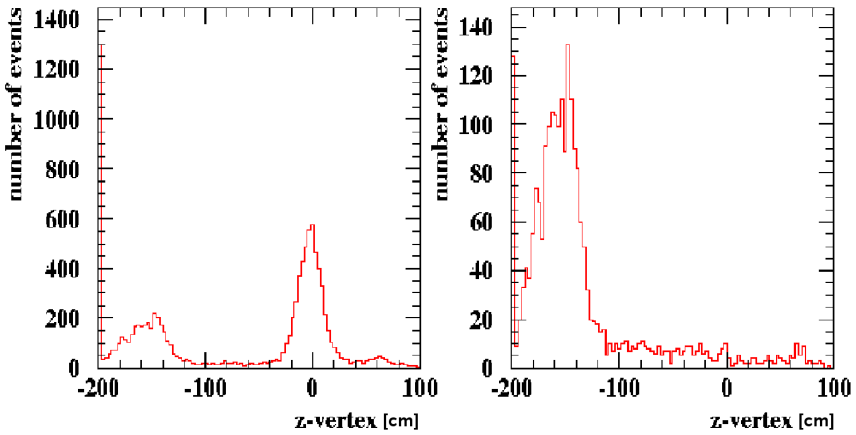


Figure 5.5: Vertex z-position for all triggered events (left) and for events rejected by the CIP VETO (right). The figures are from the Central Trigger presentation given by Z. Rúriková on 21/09/2006.

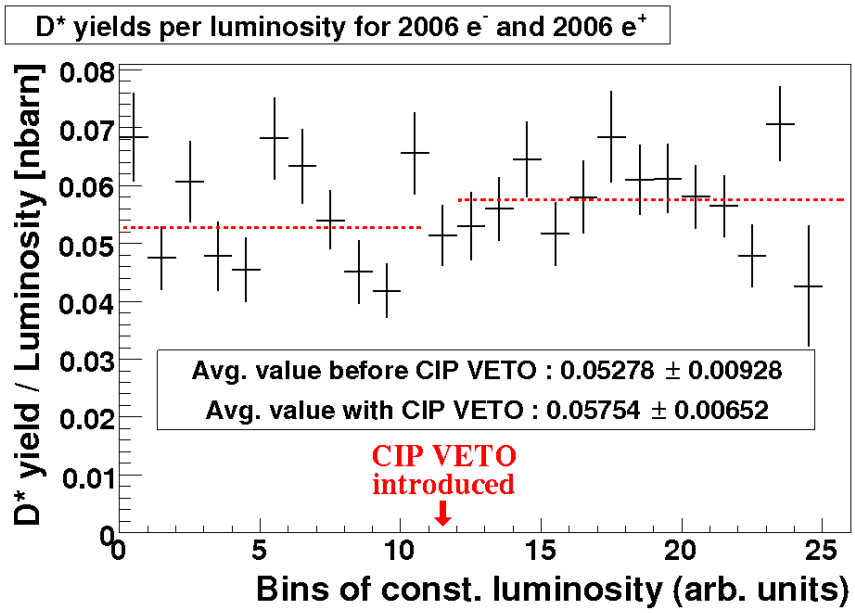


Figure 5.6: Production of D^* mesons in bins of constant luminosity for periods before and after the CIP VETO activation (whole year 2006).

fragmentation observables. Although one does not expect a big correlation (the CIP VETO condition requires a small z-vertex significance, whereas require a well reconstructed interaction vertex) the subject needs a more detailed study to be reported below.

Several investigations were done by central trigger experts after the CIP VETO was introduced in August 2006. In figure 5.5-left the z-position of the interaction vertex for all triggered events is shown. The peak in the region $z \approx -150$ cm originates from interaction of late proton satellite bunches and was the reason for the introduction of the CIP VETO. In figure 5.5-right one can see the $z-$vertex position of events which were rejected by the CIP VETO. One observes that the CIP VETO condition efficiently rejects events with shifted interaction vertex (well out of our selection criterion), the fraction of rejected events having the interaction vertex in the central detector region is at the level of 1% of all triggered events. This suggests that the effect of the CIP VETO should not affect our measurement much.

To specifically study the impact of the CIP VETO on D^* meson production, the total D^* production yield as function of the luminosity was studied in periods before and after the CIP VETO was introduced. The study was done for all events that fulfill the analysis requirements and the results are depicted in figure 5.6. The yields per luminosity are in artificial units that do not correspond to cross sections, since the yields were not corrected for branching ratios and various other effects. The averages were calculated for the two periods to make a quantitative comparison. One concludes that no decrease in the D^* production is observed in relation with the CIP VETO, and the small increase is of no statistical significance.

5.3 Reweighting of Monte Carlo Distributions

If a measured quantity is supposed to be simulated in the MC model, and if the description of the data by the MC simulation is unsatisfactory, then one usually reweights the MC model in order to achieve a good description. We applied reweighting in two cases: the interaction z-vertex position and the BPC efficiency, both quantities being not well simulated in the MC program. Since the correction factors depend on the run period, we study both quantities in a period-dependent way. In addition, we study the option of reweighting the pseudorapidity distribution of the D^* mesons for the whole HERA II running period, but this reweighting is finally not done.

5.3.1 Z-position of Interaction Vertex

A correct reweighting method requires to reweight the MC model at the hadron level so that once the reweighting is done the MC simulation at the reconstructed level describes the data well. Such an approach usually leads to a rather complicated and time-consuming iterative procedure. In the case of z-vertex reweighting we are in a lucky situation because the physics is independent on the spatial position. Thus the z-vertex reweighting factorizes from any physical observable, and it is safe to perform

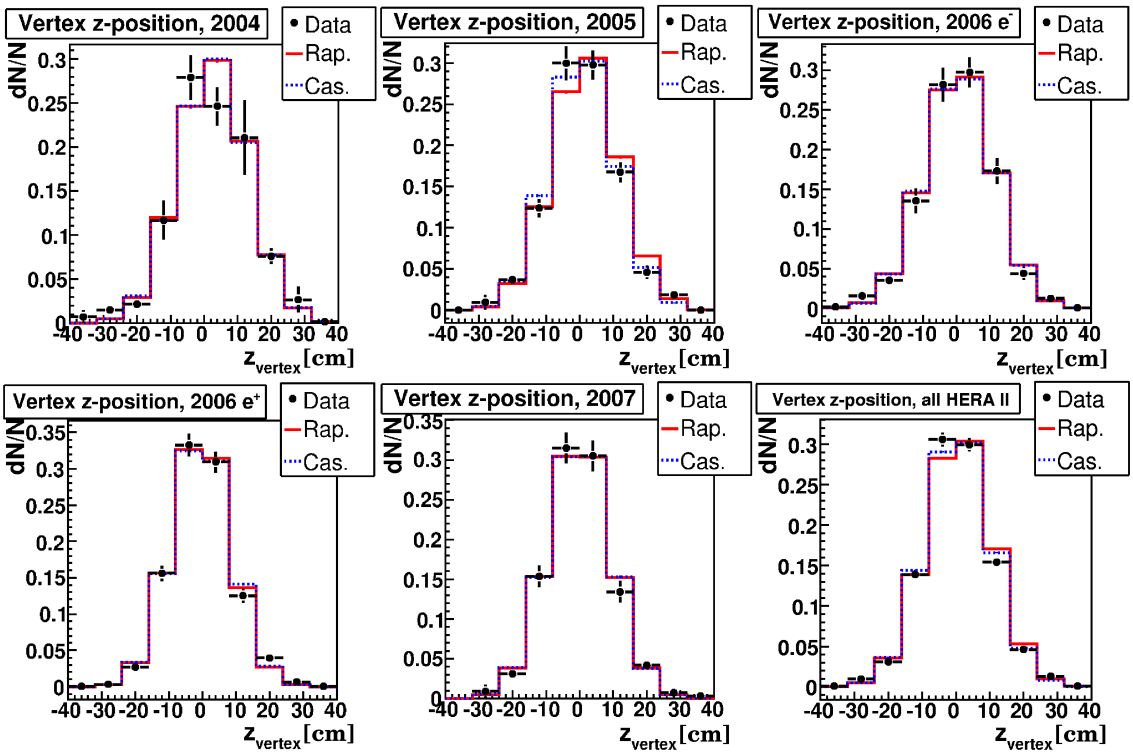


Figure 5.7: Comparison of area-normalized "z-vertex" distributions for data with RAPGAP and CASCADE Monte Carlo programs after the reweighting procedure. The distributions are shown for each running period separately as well as for all of HERA II.

it on the reconstructed level. Reweighting the vertex z-position has its importance - the detector acceptance as seen by particles leaving the interaction point depends on the position of the interaction in the detector. The z-vertex reweighting of the MC model thus corrects for these acceptance effects.

The reweighting as performed in this work is done in a period-dependent way, so as to account for detector changes (track trigger upgrade, Spacal fiducial cuts, etc.). A bin-wise reweighting was avoided in order not to depend on bin limits and rather a smooth reweighting procedure was adopted. For each run period the z-vertex distribution was fitted with a Gaussian both for the data and the MC prediction, and the ratio of the two normalized Gaussian functions was taken as the reweighting factor for a given MC model. The shapes of the distributions in the data were not exactly Gaussian (especially in the year 2004), and this explains small remaining deviations after the reweighting, see figure 5.7.

5.3.2 BPC Efficiency

The MC simulation of the BPC for the HERA II running period is not perfect. The BPC efficiency (as defined in section 4.3.2) is too high - it significantly overshoots the efficiency extracted from the data. A plot produced using the RAPGAP and

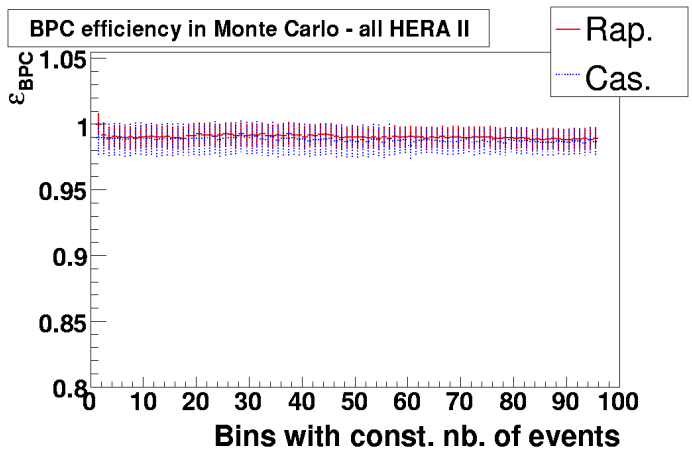


Figure 5.8: BPC efficiency of the RAPGAP and CASCADE MC and detector simulations shown for the whole HERA II running period.

CASCADE MC programs is shown in figure 5.8. This figure is to be compared with results in figure 4.2. Unlike in the data, the efficiency in the MC simulation is close to 100%. To reconcile the MC behavior with the data the following strategy was adopted:

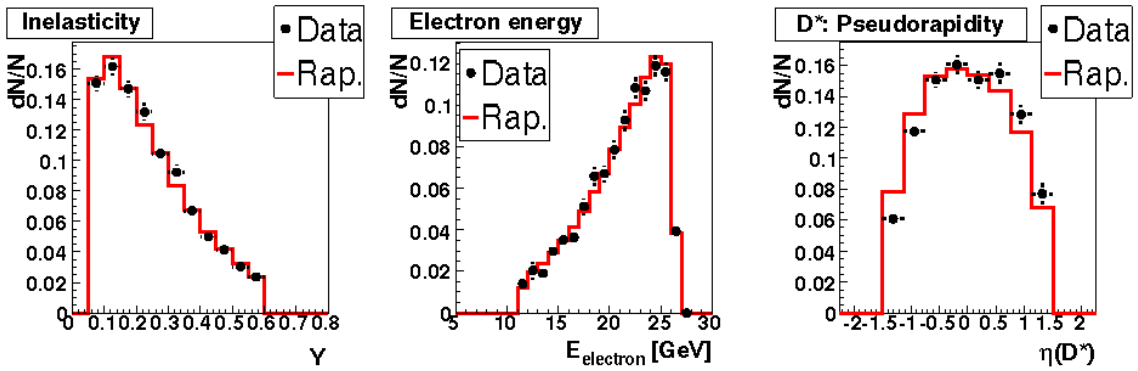
- For periods with low BPC efficiency in the data, electron quantities are recalculated from the Spacal information alone, and the same is done for MC simulation events.
- For periods with high BPC efficiency in the data we fit both, the data and the MC prediction with a constant, and we use the ratio of these constants for correcting the BPC efficiency in MC model. Since events with no BPC information are very rare in the MC simulation a standard reweighting procedure cannot be applied. Thus a different method is adopted: in the MC program we randomly choose to recalculate electron quantities from the Spacal in such a way that the fraction of events with BPC information is the same in the MC simulation as in the data.

5.3.3 Study of Reweighting in $\eta(D^*)$

RAPGAP, unlike CASCADE, does not describe the $\eta(D^*)$ distribution well. The observed differences are related to the physics model of RAPGAP rather than the detector simulation, because major detector effects (trigger and BPC efficiencies) have been taken into account already and the discrepancies appear in every running period. One thus considers reweighting the $\eta(D^*)$ distribution in RAPGAP for the whole HERA II data. Such reweighting was studied and its impact is shown in figure 5.9.

One observes, that although the reweighting improves (by definition) the description of the pseudorapidity distribution, it spoils the description of the basic kinematic variables such as the electron energy and the inelasticity. This correlation strongly

a) No $\eta(D^*)$ reweighting.



b) Reweighting in $\eta(D^*)$ applied.

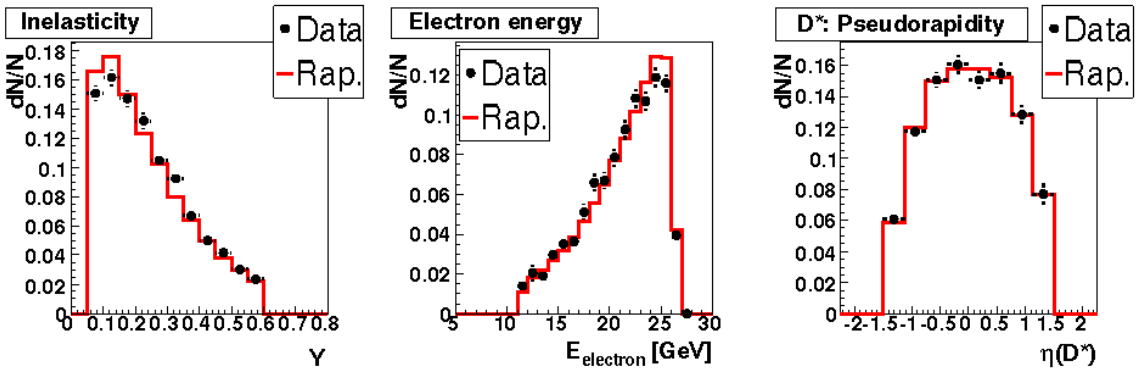


Figure 5.9: Effects of the reweighting in $\eta(D^*)$ for the RAPGAP MC model.

disfavors the reweighting procedure, and since we prefer the basic kinematic variables to be described properly, we do not apply reweighting in $\eta(D^*)$. The remaining discrepancies originate in the physics model and thus will be covered by the systematic error related to the model dependence (RAPGAP vs. CASCADE) of our results.

5.4 Control Distributions for HERA II

The comparison of data and reconstructed MC distributions for various kinematic, detector and D^*-related quantities was studied for the different running periods. No prominent deviations of the MC predictions from the data were observed. In cases where small deviations exist, they were found to be rather observable- than period-related, and thus they were found in all of the HERA II periods. For this reason and for reasons of space we present here only distributions for all of the HERA II periods combined. As explained in the previous paragraphs,

- the small contribution due to beauty production is subtracted from the data distributions,
- the data distributions are corrected for the track trigger inefficiency,

- the reweighting of the z-vertex position of the interaction point is applied for the MC distributions, and
- the MC distributions are corrected for the BPC inefficiency.

The figures are naturally divided into three sets: figures related to the event kinematics, detector-quantities (tracking) and to the D^* kinematics. In all presented distributions the D^* signal was extracted for each individual bin and all distributions are area-normalized. The data are compared to RAPGAP and CASCADE MC predictions. In addition to the figures presented in this section additional control distributions are shown in appendix A.

5.4.1 Event Quantities

The distributions for the chosen event quantities are presented in figure 5.10. One observes a fairly good description for most of the observables. Especially Q^2, inelasticity, electron energy, electron φ and the transverse momentum of all selected tracks are well described. In the remaining plots one can see some deviations.

The electron scattering angle θ is well described except for a single bin at $\theta \approx 3.03$ rad. Since this behavior is observed for both MC models (RAPGAP and CASCADE), was also observed in other D^* HERA II analyses (e.g. [51]) and was not observed in HERA I ([66, 73]) where similar MC programs were used, one expects it to be due to an imperfect detector simulation rather than to the wrong physics model. The scattering angle in question corresponds to the inner Spacal region that was modified during the upgrade to HERA II.

In the pseudorapidity distribution for all selected tracks one observes that the data have the tendency to lie above the MC prediction in the central region ($\eta \approx 0$) and below the MC values in the forward region ($\eta \approx 1.5 - 2$). The selected tracks in the forward region (last two bins) are likely to be "combined" tracks that use information from both, the CTD and the FTD. The performance of the FTD and its simulation were never completely reliable and an imperfect description in this region affects the whole region, since the distributions are area-normalized. The combined tracks might not be the only source of discrepancy, though. An overestimated beam-drag effect or too much gluon radiation in the MC simulations would also explain the excess of tracks in the forward region.

The double-ratio plot $\frac{(E_{\text{measured}}/E_{\text{double angle}})_{\text{Data}}}{(E_{\text{measured}}/E_{\text{double angle}})_{\text{MC}}}$, which compares the direct electron energy measurement with the double-angle method in data and MC simulations, shows acceptable deviations of the order of 1-2% for the upper three bins. Less well described is the lowest-energy bin where the deviation reaches 4%. This can be at least partially explained by the fact that the double-angle method is precise only for high electron energies (or low inelasticity $y < 0.1$), for smaller electron energies the reconstructed hadronic angle θ_X suffers from bad resolution (see reference [74], Section 7.2.2). Moreover, the Spacal calibration at small electron energy is in general more difficult, since the photoproduction background starts to play a role and statistics is significantly smaller in comparison to regions near the kinematic peak.

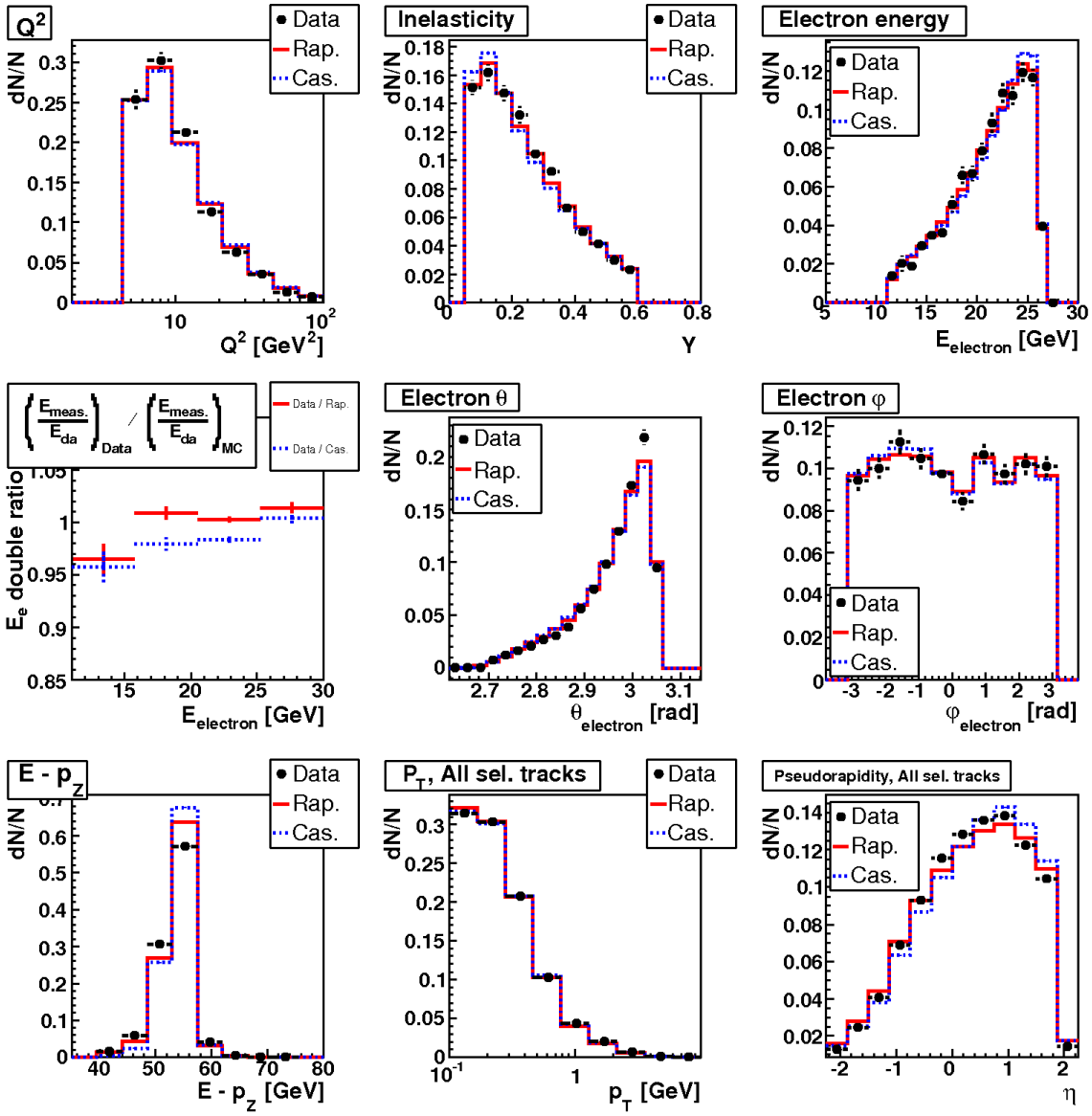


Figure 5.10: Event kinematic variables for all of HERA II data.

We expect the influence of this deviation to be suppressed by its small statistics, and we take the observed deviations in the electron energy measurement into account when studying the systematic errors.

Also in the $E-p_z$ distributions a difference between the data and MC models can be seen, the MC histograms indicating a better resolution. This observation is not specific to this study (see once more reference [51]), our analysis cut is however far away from the region where the disagreement is observed. The calculation of $E-p_z$ is based on all particles reconstructed in the detector and unless the MC simulation is performing well for the whole detector in all its details, one might expect some differences in this quantity. The cause of this discrepancy might be due to the energy calibration of hadronic final state particles for the HERA II data, which at the time of the writing of this thesis is still not finalized.

We consider the observed deviations to be within tolerable limits and do not correct for them. A systematic error will cover the uncertainty in the electron energy measurement.

5.4.2 Detector-Related Track Quantities

The detector-related track quantities (number of CJC hits, track length and pseudorapidity) for K, π and π_s of the reconstructed D^* meson are presented in figure 5.11. One observes that the description is (with the exception of the last bin in the "number-of-hits" histogram) quite good, even in the case of the slow pion, where the measurement is the most difficult.

Since in our selection we also apply requirements related to the particle energy loss, we study in figure 5.12 the agreement between data and simulations using the RAPGAP MC model. As dE/dx is a reconstruction-related quantity (it is primarily related to the detector simulation and not to the physics of the MC model), one expects a similar behavior for RAPGAP as for CASCADE. The data are visualized as black scatter plot, the MC distribution is visualized in form of a contour plot. A rescaling of the MC distribution was needed for an easier comparison with the data. One observes that the contour lines of the MC model follow well the structure of the data scattered plot.

5.4.3 D^* Variables

The D^* variables are presented in figure 5.13. One observes a fairly good description of the D^* transverse momentum, the D^* azimuthal angle φ and, in the case of CASCADE, the D^* pseudorapidity distribution.

The latter distribution is less well described by RAPGAP, the difference might come from different parton evolutions or from the resolved component (partially included in CASCADE). The model dependence of our measurement is however reflected in the systematic error that is introduced in chapter 6. A possible reweighting in y of the D^* meson was abandoned as explained in section 5.3.3.

Both models differ from each other and from the data in the case of the pseudo-

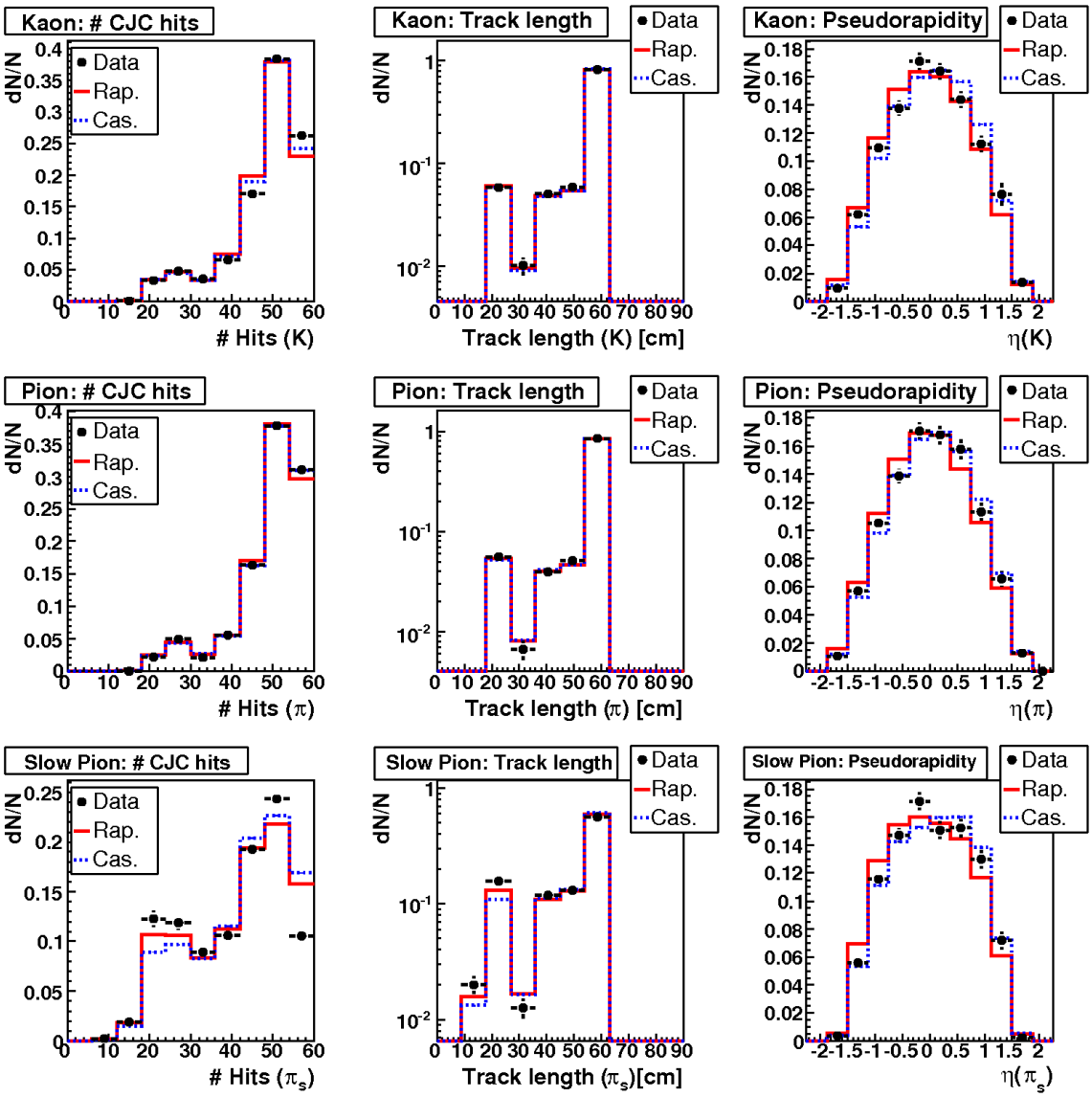


Figure 5.11: Track quantities for D^* daughter particles for all of HERA II data.

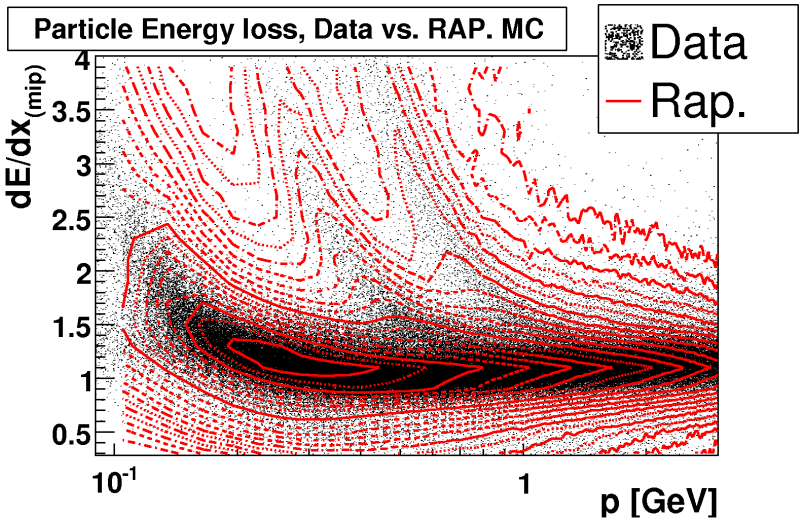


Figure 5.12: Energy loss for all selected tracks in D^* events for all of HERA II data. For the contour plot of RAPGAP the bin content was scaled by $z \rightarrow 0.1 \times \ln^2(1+z)$.

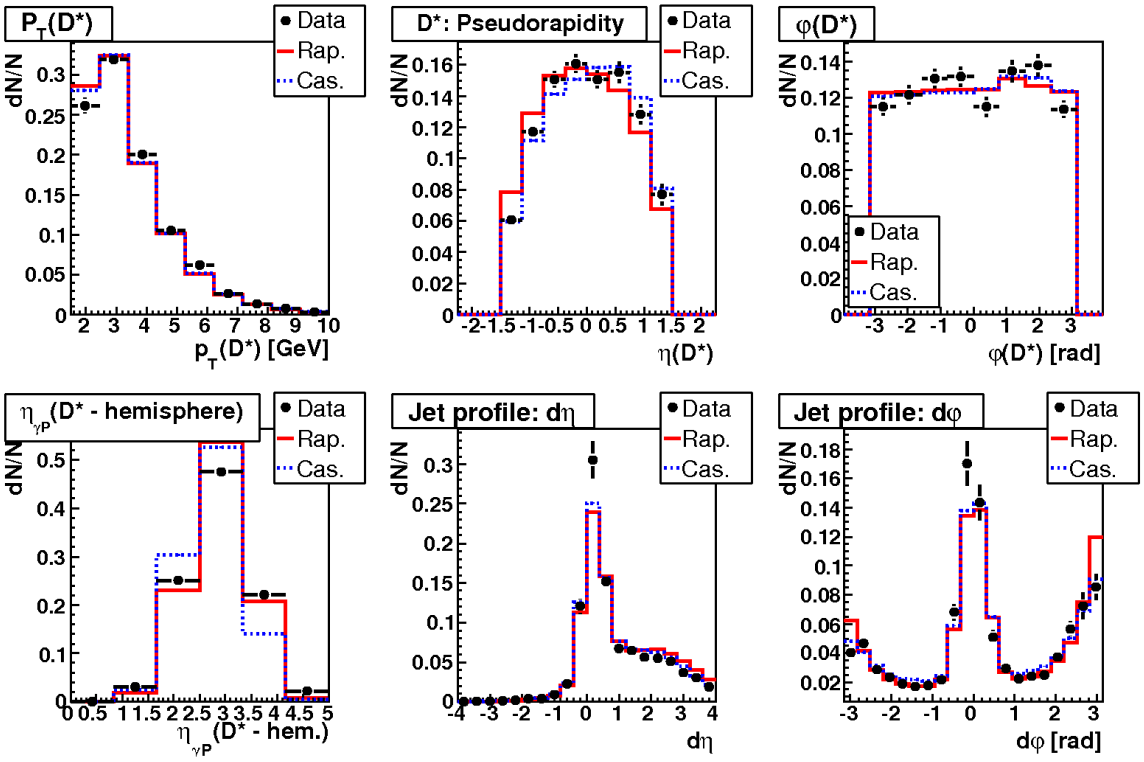


Figure 5.13: D^*- related variables for all of HERA II data.

rapidity distribution for the D^* hemisphere as reconstructed in the γP frame[2]. This observable is constructed from all particles in the hemisphere and is thus sensitive to a correct reconstruction of all of them. An imperfect simulation of some detector region may cause the MC predictions to differ from the data. The two MC models are sharing the same detector simulation, the difference between them might however arguably originate from resolved processes.

The last two distributions are related to the jet method and represent jet profiles in η and φ, i.e. the differences in η and φ between the jet axis and the full hadronic final state. The description of the jet profiles is acceptable.

[2]We refer here to the pseudorapidity with respect to the momentum of the whole D^* hemisphere.

Chapter 6

Fragmentation Measurement

In this chapter the hadron level corrected data for the three fragmentation observables are presented. These distributions represent the final outcome of the experimental measurement and are used to study charm quark fragmentation. The procedure of correcting the data to hadron level is based on methods described in chapter 3.4. Here we provide additional comments and clarifications concerning background subtraction and regularization parameters. In the second part of this chapter the evaluation of systematic errors is explained. The study of fragmentation models based on measured distributions is the subject of chapter 7.

6.1 Visible Range Definition and Previous Fragmentation Measurements

The distributions of the fragmentation observables z_{hem}, z_{jet} and $z_{\text{hem}}^{(\text{jet})}$ are measured in the visible range defined by

$$Q^2 > 5\,\text{GeV}^2,$$

$$0.05 < y < 0.6\,,$$

$$1.5\,\text{GeV} < p_T(D^*) < 15.0\,\text{GeV},$$

$$|\eta(D^*)| < 1.5\,,$$

and additionally for z_{jet} and $z_{\text{hem}}^{(\text{jet})}$

$$E_T(D^*\,\text{jet}) > 3.0\,\text{GeV}.$$

For all other constraints mentioned in chapter 4 the data are corrected using the Monte Carlo models.

This study of charm quark fragmentation is to be seen in the context of previously performed fragmentation measurements. Many of them studied charm fragmentation in e^+e^- annihilation (see references [63, 64, 65]) and provided interesting and important results. Our study can be most easily compared to measurements at HERA, one

by ZEUS in photoproduction and one by H1 in DIS. Detailed information about the preliminary ZEUS analysis [75] (especially MC settings) is not available. A comparison with the H1 HERA I analysis [66] is the most straightforward. Many techniques and methods used in the previous HERA I measurement were adopted, especially the definitions of our observables and some of the selection criteria. To emphasize new features of the present analysis and show its potential with respect to a better understanding of charm fragmentation, we list some of the major differences.

- New data with much more improved statistics ($36\,\mathrm{pb}^{-1}$ in the HERA I analysis vs. $354\,\mathrm{pb}^{-1}$ in this HERA II analysis).
- Different definition of the visible range, because of detector changes due to the HERA II collider and detector upgrade (electron energy, inelasticity and Q^2 limits are different).
- New detector components were studied and understood (BPC, new track trigger).
- Unfolding procedures. In addition to the bin-by-bin method used in [66], the SVD and the Bayesian approach are employed in this analysis.
- Some other minor differences in measurement and reconstruction methods, e.g. different signal extraction that respects better the signal shape (single Gaussian in [66] vs. double Gaussian function used here).

6.2 Unfolding Hadron Level Distributions

The unfolding methods having been described earlier, one still needs to provide more information about the procedure: explain background subtraction, bin limits, regularization parameters, treatment of QED radiative corrections and so on. These topics are covered in the following paragraphs.

6.2.1 Background Treatment, Migrations into Visible Range and Beauty Subtraction

When unfolding the data, background is an important issue which needs to be considered twice: firstly, when using the MC simulation to evaluate the response matrix and secondly, when measuring the actual distribution to be unfolded.

In the first case a problem could arise from background appearing at the MC reconstructed level used in the matrix determination. In the present analysis the error related to this effect is expected to be negligible. This is mainly because the signal MC simulation only was used to determine the response matrix and therefore the background reconstructed on detector level is very small in comparison with the signal (although existing). The main source of background is the combinatorial background which is of the order of 1% (figure 3.4). Its effect on the response matrix determination is however limited only to those events where a fake D^* is

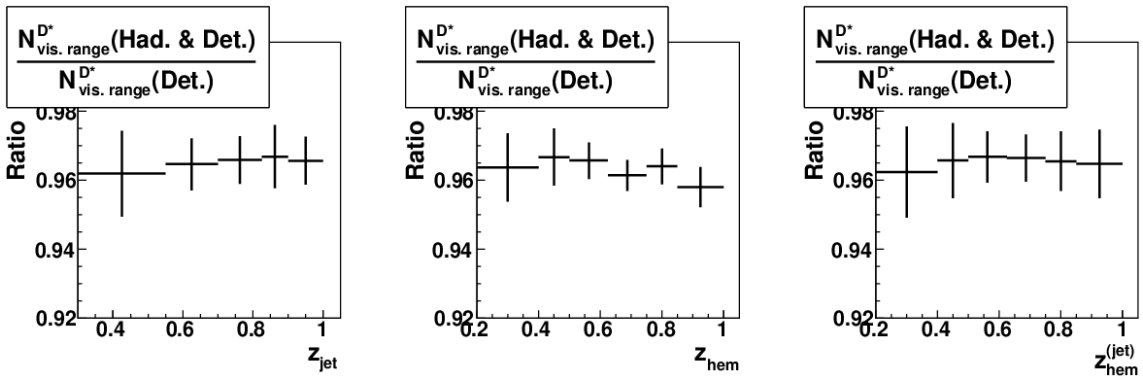


Figure 6.1: Ratio of the number of D^*s reconstructed and generated in the visible range to the total number of D^*s reconstructed in the visible range as predicted by RAPGAP MC simulation in bins of $z_{\rm hem}$, $z_{\rm jet}$ and $z_{\rm hem}^{\rm (jet)}$.

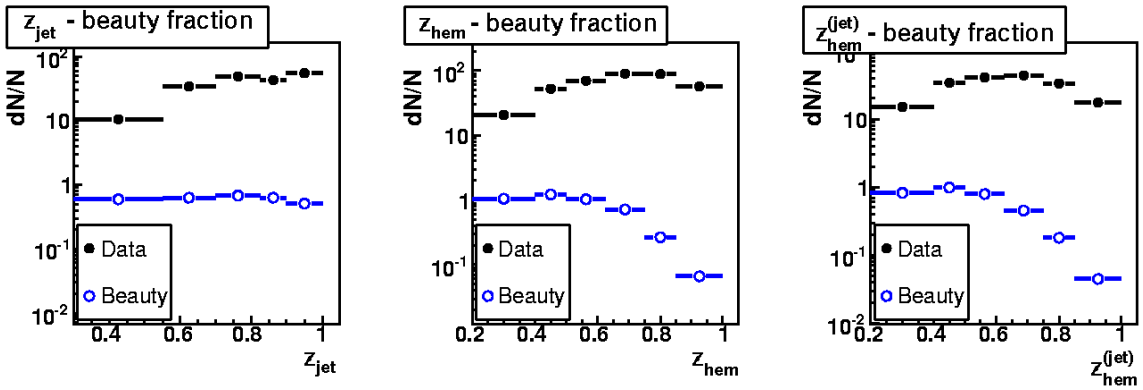


Figure 6.2: Measured data including D^* mesons from beauty decays in bins of $z_{\rm hem}$, $z_{\rm jet}$ and $z_{\rm hem}^{\rm (jet)}$ together with this beauty background as predicted by the RAPGAP MC model. The distributions are normalized to luminosity.

reconstructed at the detector level and at the same time a true D^* within the visible range is found at the hadron level so that the "fake" reconstructed $z_{\rm hem/jet}^{\rm Det.\,Lev.}$ is wrongly associated with $z_{\rm hem/jet}^{\rm Had.\,Lev.}$ on the hadron level. In addition, the response matrix is determined from events lying in the $\triangle M$ signal region only (a cut is applied) so that the combinatorial background situated far from the signal peak is ignored.

The second case which needs to be considered is related to background subtraction, a measured distribution incompletely corrected for background leads to wrong results. The reason for this is that one requirement for the different unfolding procedures from section 3.4 is a background-free data distribution. In the signal extraction procedure we fit the signal peak and thus the non-D^* background is filtered out. Possible remaining contamination can originate from true D^* particles which are still considered background in this analysis. There are two sources of such D^*s: particles that migrate from outside the hadron-level visible range into the visible range at detector level and decays of beauty mesons.

The magnitude of migration effects is estimated using the RAPGAP MC model.

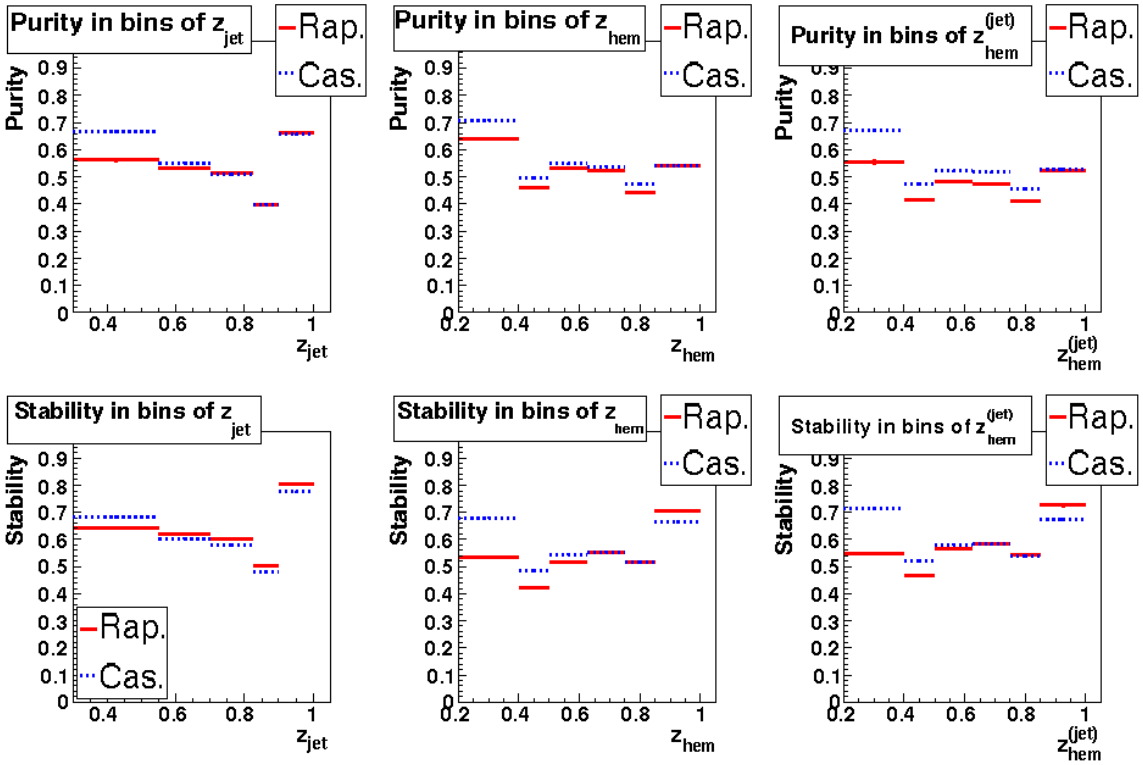


Figure 6.3: The purity and the stability for the $z_{\rm hem}$, $z_{\rm jet}$ and $z_{\rm hem}^{\rm (jet)}$ distributions as estimated with the RAPGAP and CASCADE MC simulations (for HERA II running conditions).

Approximately 4% of the reconstructed events were generated outside the visible range on hadron level as can be seen in figure 6.1. One observes that the relative[1] contributions due to migrations are almost constant, and therefore the shapes of the spectra are only slightly affected. The predicted background is subtracted from the measured distributions[2].

The same applies to the D^* contamination from beauty decays. This beauty-induced background as predicted by RAPGAP (see paragraph 5.1.2 for more details) is shown in figure 6.2 and is subtracted from data.

6.2.2 Purity and Stability

Purity and stability have been defined in paragraph 3.4.1. Unfortunately, these quantities do not improve with increasing statistics. They reflect detector effects which remain the same, independent of the amount of collected data. Purity and stability are related to migrations in bins between the hadron and the detector level distributions and one requires these migrations to be reasonably small for the

[1]Relative with respect to the charm signal.

[2]We do not associate a systematic error to migration-related background, since the influence on the shapes is negligible with respect to other systematic errors.

unfolding methods to work properly[3]. Therefore, the established values of purity and stability provide a constraint on the possible bin widths.

In this work the choice of the bin limits was driven by two ideas. Firstly, we require a binning such that purity and stability are never below 40%. Secondly, it is convenient to chose the bin limits such that an easy comparison with the previous analysis is possible. Thus, purity and stability with the binning of the HERA I analysis were studied and the results are shown in figure 6.3. One observes that both quantities are never smaller than 40%, although the bin limits in the high $z_{\mathrm{jet/hem}}$ region could be slightly redefined to make the distributions flatter. Since we prefer to have the possibility of an easy comparison with the previous measurement, the presented binning is kept. The actual bin limits are:

$$z_{\mathrm{hem}}\ \mathit{and}\ z_{\mathrm{hem}}^{(\mathrm{jet})}\ \mathit{distributions}:\ 0.2,\ 0.4,\ 0.5,\ 0.625,\ 0.75,\ 0.85,\ 1.0;$$

$$z_{\mathrm{jet}}\ \mathit{distribution}:\ 0.3,\ 0.55,\ 0.7,\ 0.825,\ 0.9,\ 1.0;$$

6.2.3 Unfolding Parameters

The choice of unfolding parameters was addressed already in paragraph 3.4. Here we provide more details about the regularization for the SVD and the Bayesian method used for extracting our distributions with the RAPGAP MC model. The bin-by-bin method does not involve regularization parameters and therefore does not need any further explanations.

Before commenting on the choice of parameters, it is pointed out that for both methods, SVD and Bayesian, a response matrix with an additional bin at the lower edge of the spectrum ($0.0 - 0.2$ for z_{hem} and $0.0 - 0.3$ for z_{jet}) is used. The added bin will not be used later in fits, because it contains large background contamination and a small signal. It is considered to take into account migrations from this bin to other bins in the z_{hem}, $z_{\mathrm{hem}}^{(\mathrm{jet})}$ and z_{jet} distributions.

SVD

Within the SVD method (see section 3.4.2) the result of the matrix inversion is expressed as a sum of scalar coefficients times basis vectors with different frequencies. We denote with d_i these coefficients divided by their statistical errors with $i = 1$ referring to the least and $i = i_{max}$ to the most oscillating term . To achieve a certain level of smoothness in the solution one needs to suppress the highly oscillating components. The most straightforward way of doing this is to suppress those vectors which have coefficients for which $|d_i| \lesssim 1$, i.e. one looks for the first index k such that for $j > k$ one typically[4] has $|d_j| \lesssim 1$. One then chooses the value of the regularization parameter to be $\tau = s_k^2$, where s_k is the kth singular value (see reference [67]).

[3]Strictly speaking, this requirement is necessary only for the application of the bin-by-bin method. One however doubts about any unfolding method in case of huge migrations since in this case the matrix unfolding is usually close to an ill-defined problem.

[4]The values of $|d_i|$ are typically falling, however, depending on the concrete problem it might happen that a high-frequency vector has a $|d_j|$ significantly larger than one.

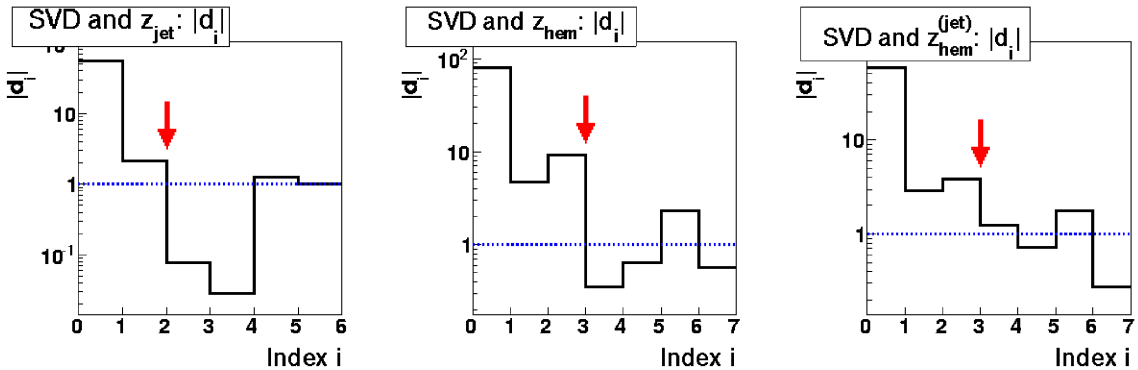


Figure 6.4: Coefficients $|d_i|$ as function of their index. The chosen values $k_{\mathrm{SVD}}(z_{\mathrm{hem}})$, $k_{\mathrm{SVD}}(z_{\mathrm{jet}})$ and $k_{\mathrm{SVD}}(z_{\mathrm{hem}}^{(\mathrm{jet})})$ are indicated by arrows.

This procedure was applied to the fragmentation observables and the corresponding response matrices. From figure 6.4 we obtain for the different observables:

$$k_{\mathrm{SVD}}(z_{\mathrm{jet}}) = 2,$$

$$k_{\mathrm{SVD}}(z_{\mathrm{hem}}) = 3,$$

$$k_{\mathrm{SVD}}(z_{\mathrm{hem}}^{(\mathrm{jet})}) = 3,$$

where we prefer to give the index position of the last significant vector rather than a hardly interpretable value of the parameter τ.

Bayesian Approach

A known feature of the Bayesian iterative approach (see section 3.4.3) is that the most important modifications in unfolded distributions occur in the very first steps, further steps lead only to slight changes in progressively unfolded spectra[5]. After a few of iterations the difference $\chi^2_k = \chi^2[d_{(k)}^{\text{Had. Level, Data}}, d_{(k+1)}^{\text{Had. Level, Data}}]$ between two consecutive distributions becomes small and the solution stabilizes. In our case we observe that for all three observables z_{jet}, z_{hem} and $z_{\mathrm{hem}}^{(\mathrm{jet})}$ the result is mainly driven by the first iteration, from the second one onwards the changes in distributions are tiny and very close to zero. Thus a fine-tuning of the number of steps does not make sense and we choose for the three observables the number of iterative steps to be 3, i.e.

$$k_{\mathrm{Bayes}}(z_{\mathrm{hem}}) = k_{\mathrm{Bayes}}(z_{\mathrm{jet}}) = k_{\mathrm{Bayes}}(z_{\mathrm{hem}}^{(\mathrm{jet})}) = 3.$$

6.2.4 Treatment of QED Radiative Corrections

The used MC models differ also with respect to the implementation of QED radiation (see section 1.4.1). The RAPGAP model is available with and without QED

[5]See for example page 496 of reference [68] where one reads "...one can realize that in most cases a good agreement is reached after a few iterations".

radiation, for CASCADE QED radiation is not implemented. This is the reason why the correction procedures are not the same for the two models. We study all presented unfolding methods (SVD, Bayes, bin-by-bin) with the RAPGAP simulation and apply only the bin-by-bin approach to the CASCADE model. In the case of RAPGAP, the different methods allow us to investigate the systematic error related to the unfolding procedure. The bin-by-bin approach applied to both, RAPGAP and CASCADE, enables us to establish the error related to the MC model.

RAPGAP

The unfolding procedure is rather straightforward for the RAPGAP model, where both options, with and without QED radiation, are available. First, the measured (detector level) distribution $d_i^{\mathrm{DL,\,Data}}$ is unfolded via one of the described methods (SVD, Bayes or bin-by-bin) to the radiative hadron level $d_i^{\mathrm{HL,\,Data,\,Rad}}$ using the response matrix from the radiative RAPGAP[6]. Next, the obtained distribution is corrected for radiative effects applying the bin-by-bin correction factors

$$c_i^{\mathrm{HL,\,RAP,\,QED}} = \frac{d_i^{\mathrm{HL,\,RAP,\,Non-Rad}}}{d_i^{\mathrm{HL,\,RAP,\,Rad}}} \,,$$

determined from radiative and non-radiative MC simulations. One thus has

$$d_i^{\mathrm{HL,\,Data,\,Non-Rad}} = c_i^{\mathrm{HL,\,RAP,\,QED}} \cdot d_i^{\mathrm{HL,\,Data,\,Rad}} \,.$$

The use of the bin-by-bin method in the last step is justified, because the difference between the two MC distributions does not involve detector effects or different physics, only QED radiation. Moreover, the coefficients $c_i^{\mathrm{HL,\,RAP,\,QED}}$ are precisely known due to the high statistics available in MC simulations at hadron level.

CASCADE

In CASCADE, QED radiative effects are not included, and therefore we estimate them from RAPGAP. Including the RAPGAP estimates of these effects into the SVD or Bayesian unfolding is complicated and thus we prefer to avoid it. Hence we correct the data with CASCADE using the bin-by-bin method only. For this purpose we add radiative corrections to both, detector and hadron level CASCADE distributions, as estimated from RAPGAP

$$d_i^{\mathrm{DL,\,CAS,\,Rad}} = c_i^{\mathrm{DL,\,RAP}} \cdot d_i^{\mathrm{DL,\,CAS,\,Non-Rad}} \,,\; c_i^{\mathrm{DL,\,RAP}} = \frac{d_i^{\mathrm{DL,\,RAP,\,Rad}}}{d_i^{\mathrm{DL,\,RAP,\,Non-Rad}}}$$

and

$$d_i^{\mathrm{HL,\,CAS,\,Rad}} = c_i^{\mathrm{HL,\,RAP}} \cdot d_i^{\mathrm{HL,\,CAS,\,Non-Rad}} \,,\; c_i^{\mathrm{HL,\,RAP}} = \frac{d_i^{\mathrm{HL,\,RAP,\,Rad}}}{d_i^{\mathrm{HL,\,RAP,\,Non-Rad}}} \,.$$

[6]Here d_i refers to number of entries in the ith bin, the same notation as in section 3.4 is used.

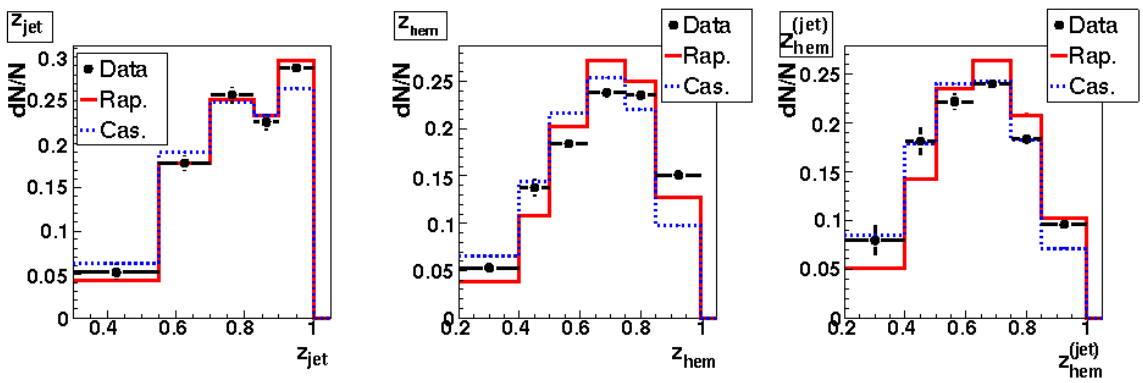


Figure 6.5: Area-normalized and beauty-subtracted detector level distributions for z_{jet}, z_{hem} and $z_{\text{hem}}^{(\text{jet})}$ together with charm MC predictions (all HERA II).

We define correction factors to correct the data from the detector level to the radiative hadron level (i.e. "unfolding")

$$c_i^{\text{DL}-\text{HL, CAS, Rad}} = \frac{d_i^{\text{HL, CAS, Rad}}}{d_i^{\text{DL, CAS, Rad}}}$$

and correction factors to correct for QED radiation on the hadron level

$$c_i^{\text{HL, CAS, QED}} = \frac{d_i^{\text{HL, CAS, Non}-\text{Rad}}}{d_i^{\text{HL, CAS, Rad}}} \, .$$

The hadron-level data distributions, corrected for QED radiative effects, can be written as

$$d_i^{\text{HL, Data, Non}-\text{Rad}} = c_i^{\text{HL, CAS, QED}} \cdot c_i^{\text{DL}-\text{HL, CAS, Rad}} \cdot d_i^{\text{DL, Data}} \, ,$$

or, after simplifications, as

$$d_i^{\text{HL, Data, Non}-\text{Rad}} = \frac{d_i^{\text{HL, CAS, Non}-\text{Rad}}}{d_i^{\text{DL, CAS, Non}-\text{Rad}}} \cdot \frac{d_i^{\text{DL, RAP, Non}-\text{Rad}}}{d_i^{\text{DL, RAP, Rad}}} \cdot d_i^{\text{DL, Data}} \, ,$$

where all ratios are calculated with the help of MC models at detector or hadron level and with or without QED radiative effects.

6.2.5 Hadron Level Unfolded Distributions

The information presented so far allows us to correct the measured data to hadron level and obtain normalized differential cross sections in z_{jet}, z_{hem} and $z_{\text{hem}}^{(\text{jet})}$. The measured spectra at detector level together with their predictions from RAPGAP and CASCADE are shown in figure 6.5. The description of the data by the MC models is not perfect, which is not completely unexpected, since the models have not been tuned with respect to the fragmentation parameters. Extraction of optimal

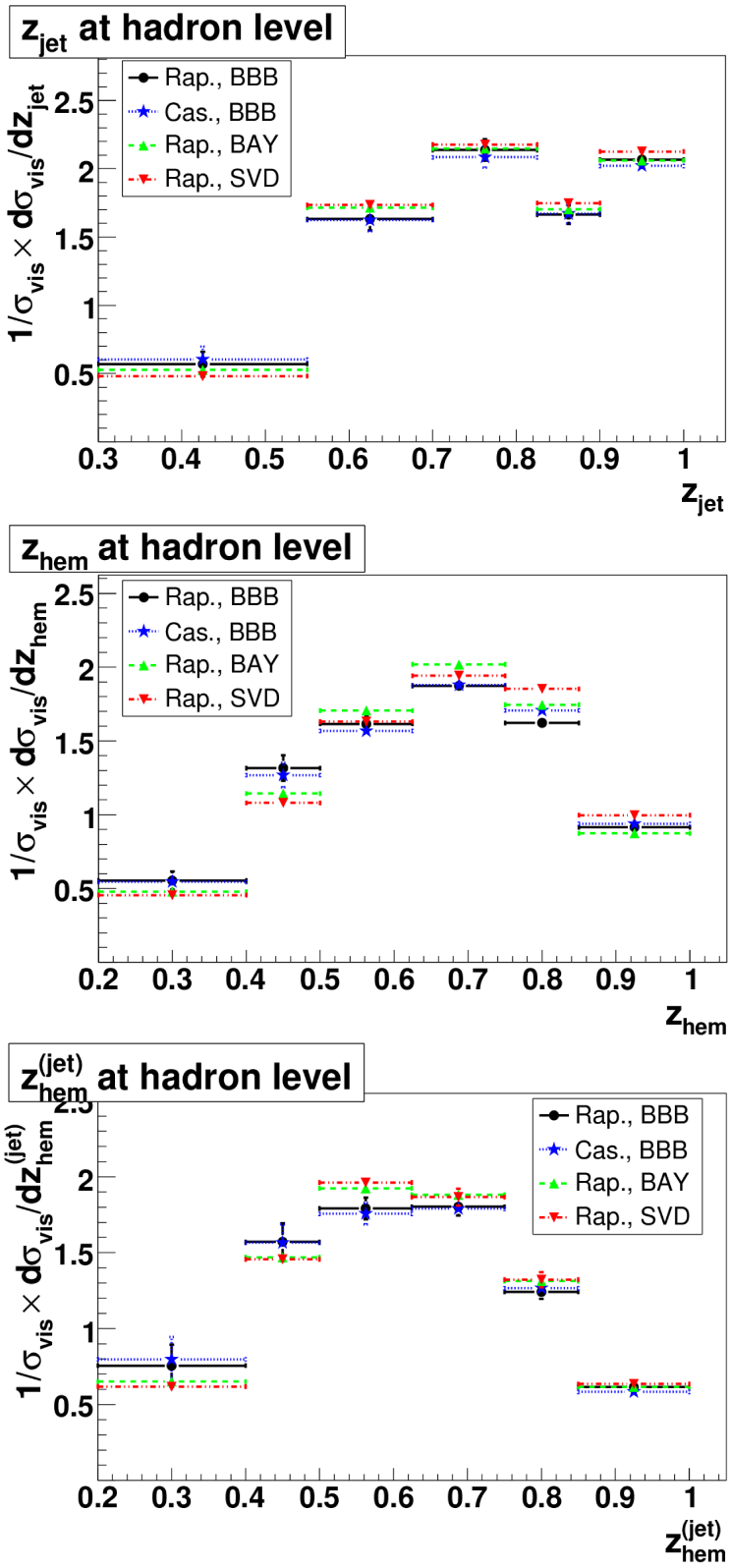


Figure 6.6: Area-normalized distributions for $z_{\rm jet}$, $z_{\rm hem}$ and $z_{\rm hem}^{\rm (jet)}$ corrected to hadron level (all HERA II) using different unfolding methods with statistical errors only. The meaning of the legend nomenclature is as follows: Rap. = RAPGAP, Cas. = CASCADE, BBB = bin-by-bin unfolding, BAY = unfolding based on Bayesian approach, SVD = SVD unfolding.

fragmentation parameters is what we are aiming for in this thesis. The spectra corrected to hadron level using different unfolding methods are presented in the figure 6.6. One observes that the results obtained with different methods agree well with each other in the case of $z_{\rm jet}$, the agreement is less good in the case of $z_{\rm hem}$ and $z_{\rm hem}^{\rm (jet)}$. The differences are not always covered by the statistical errors. For our final results we make use of the spectra (data points and their statistical errors) obtained via the SVD method. This method is expected (by construction - it takes into account migrations) to be better than the bin-by-bin approach, and its behavior with respect to the regularization seems to be more appropriate than in the case of the Bayesian procedure (the latter one having small sensitiveness to the value of the regularization parameter).

The last step towards the final results requires the evaluation of systematic errors.

6.3 Systematic Errors

Systematic errors result from imperfect knowledge of detector effects and inexact reconstruction and measurement methods. Most of them are studied using the RAPGAP MC model. A quantity under consideration may be varied in the MC simulation within its systematic uncertainty and the response matrix is evaluated. Using this response matrix the data are corrected to hadron level via the regularized SVD unfolding. The changes in the resulting distribution with respect to the nominal distribution are taken as systematic error due to the uncertainty of that quantity. Other systematic uncertainties are evaluated using the data. In the latter case one investigates the impact of using an alternative data treatment (for example a different signal extraction or a different unfolding method) on the fragmentation spectra.

We consider systematic errors related to uncertainties in:

- signal extraction,
- unfolding method,
- model dependence,
- trigger efficiency,
- electron energy,
- electron θ,
- energy loss dE/dx,
- hadronic energy scale,
- track momenta,
- beauty contribution to the D^* production and

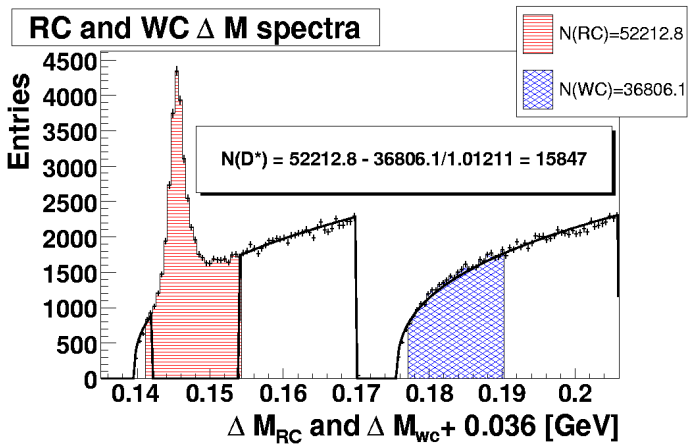


Figure 6.7: In an alternative approach, the D^* signal is estimated as difference between the hatched (red) and checkered (blue) area. The latter is divided by a normalization factor determined from the fit (black line).

- resolved photon component.

In the following, we briefly describe the evaluation of systematic errors related to each individual source. The numerical values for each bin in $z_{\rm jet}$, $z_{\rm hem}$ and $z_{\rm hem}^{\rm (jet)}$ will be given in summary tables at the end of this section.

6.3.1 Determination of Systematic Uncertainties

Signal Extraction

To evaluate the uncertainty in the number of D^* mesons due to our nominal signal extraction procedure, we consider an equally reasonable alternative method and compare the results. Since we do not want the statistical uncertainties to influence the outcome, we do not investigate the differences between the two methods in each bin of $z_{\rm jet/hem}$. Therefore we study the two approaches using the inclusive $\triangle M$ spectrum, where the statistical errors are small and hence their influence on the result of the fit.

The alternative procedure that we adopt can be summarized in the following steps:

- A simultaneous fit on the right-charge and wrong-charge background is performed, whereby the signal region in the right-charge spectrum is ignored in the fit. The fit function is

$$f'(x) = f_{\rm bg}(x_{\rm RC}) + k\, f_{\rm bg}(x_{\rm WC}),\ x = \Delta M$$

  where $f_{\rm bg}(x) = N_{\rm bg}(x - m_\pi)^\alpha$ and k is a multiplicative factor (a free parameter, together with $N_{\rm bg}$ and α) which is introduced in order to normalize the wrong-charge background to the right-charge one.

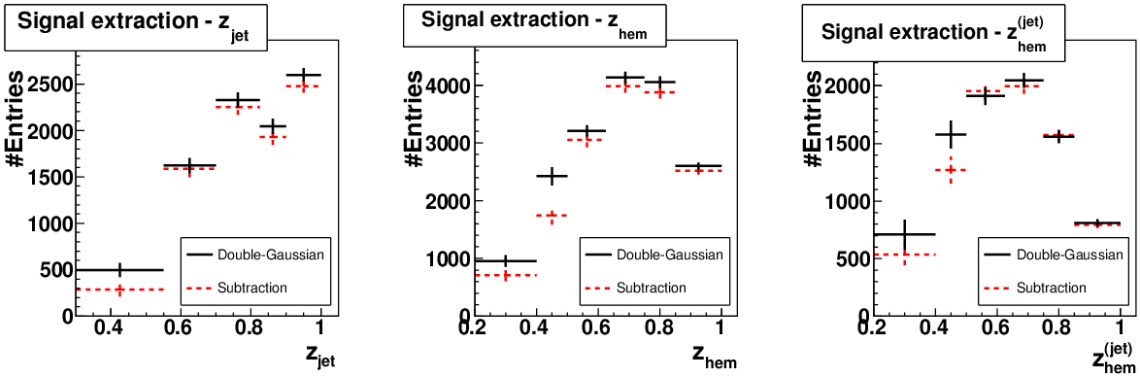


Figure 6.8: Comparison of two signal extraction methods.

- The (scaled) number of entries in the signal region of the wrong-charge spectrum is subtracted from the number of entries in the signal region of the right-charge spectrum and the result is considered as the extracted number of D^* mesons.

- The number from the previous step is compared to the number obtained using the double-Gaussian function and the relative error is calculated as

$$\sigma^{\text{sys.}}(N_{D^*}) = \frac{\left|N_{D^*}^{\text{Gauss}} - N_{D^*}^{\text{Sub.}}\right|}{N_{D^*}^{\text{Gauss}}}.$$

- These uncertainties are assigned to each bin of the respective observable distribution.

This procedure is illustrated in figure 6.7.

We also investigated the two signal extraction methods in bins of $z_{\text{jet/hem}}$ in order to locate possible correlations between them. If the relative difference between the two methods is constant in bins of an observable then the uncertainty extracted from the inclusive ΔM spectrum may be an over-estimate since we study the shape of the spectra, not their absolute normalization. The results are shown in figure 6.8 and one observes that a systematic shift is indeed present, although not in all bins. One however also sees significant differences at low $z_{\text{jet/hem}}$ values, bigger than what is expected from the inclusive estimate. These differences can be ascribed to a large extent to statistical fluctuations that have influenced the result of the fit or the subtraction. Thus, we consider the estimate coming from the inclusive data sample as more reliable, not over-estimating the uncertainty and we adopt it.

Unfolding Method

For the extraction of a systematic error related to the unfolding procedure we take a rather straightforward approach. We consider the three used methods (bin-by-bin, SVD and Bayesian approach) and for each bin in each of the $z_{\text{hem/jet}}$ distributions

we define the systematic error as

$$\sigma^{\text{sys.}}_{\text{Unfold.}}(d_i) = \frac{1}{2}[\max(d_i^{\text{BBB}},\, d_i^{\text{SVD}},\, d_i^{\text{BAY}}) - \min(d_i^{\text{BBB}},\, d_i^{\text{SVD}},\, d_i^{\text{BAY}})],$$

where d_i refers to the number of entries in a given bin. The error is scaled when the normalized cross sections $\frac{1}{\sigma_{\text{vis}}}\frac{d\sigma_{\text{vis}}}{dz_{\text{hem/jet}}}$ are presented.

Model Dependence

The use of two different models - RAPGAP and CASCADE - allows us to estimate the corresponding systematic uncertainty. Unfortunately, the unavailability of QED radiation in CASCADE makes the SVD and the Bayesian unfolding difficult to perform and therefore we do not use these unfolding methods here. Thus, to evaluate the uncertainty of our results related to the model, we rely on the bin-by-bin correction procedure as described in paragraph 6.2.4. On one hand this approach might be inconsistent to some extent with the SVD unfolding for our nominal results, on the other hand the possible bias is expected to be partially canceled, since we use the bin-by-bin method for both, RAPGAP and CASCADE. In this context, we are not interested in the unfolded distributions themselves, but only in variations between the obtained distributions using the two models. Thus we expect the bin-by-bin procedure to provide rather conservative estimates for the model uncertainty. One half of the difference in each bin between the (bin-by-bin) corrected distributions by RAPGAP and CASCADE is taken as model-related systematic uncertainty of the results.

Trigger Efficiencies

The uncertainties in the trigger inefficiencies are estimated from the plots presented in section 5.2. The uncertainty is calculated as luminosity-weighted average of (remaining[7]) trigger inefficiencies over the five running periods. An average global error is assigned to all bins of all distributions in case of the Spacal triggers. A bin-dependent error is applied in case of the tracking trigger inefficiencies. In the latter case the uncertainty reflects the remaining inefficiency that is still visible after the reweighting procedure. The trigger-related systematic uncertainties are small in comparison to other systematic uncertainties.

Electron Energy

The uncertainty of the energy of the scattered electron can be estimated from the double-ratio plot presented in figure 5.10. One observes that at low energies the deviation of the MC description reaches 4%, at higher energies it is of the order of 1%. Thus we apply to our detector simulation a linear energy-dependent variation, starting with $\pm 4\%$ for $E_{\text{elec.}} = 11.0\,\text{GeV}$ and going down to $\pm 1\%$ for $E_{\text{elec.}} = 27.6\,\text{GeV}$.

[7] In case of the track trigger.

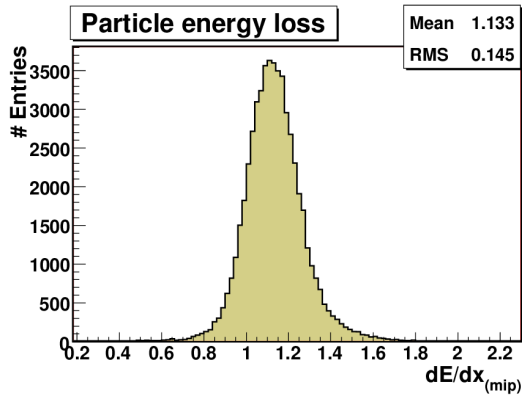


Figure 6.9: Energy-loss distribution for all selected tracks in events entering the ΔM distribution (data).

This variation is considered safe and conservative, since the deviation at lower energies also receives contribution from inaccuracies of the double-angle method (see section 5.4.1). Moreover, detailed studies ([74, 76]) were done for the HERA I period, the results of which have also some significance for HERA II, since the detector (Spacal) is the same apart from a reduced acceptance at small radii from the beam axis. These results support our estimates as being conservative.

Applying these variations and correcting the data with the modified MC simulation, we define the errors as the difference between the obtained distributions and the nominal ones obtained with the nominal MC model.

Electron Scattering Angle

The angular resolution of the BPC is approximately $\triangle\theta_{\text{elec.}} \approx 0.5$ mrad and that of the Spacal is $\triangle\theta_{\text{elec.}} \approx 1$ mrad. A large number of scattered electrons is however detected outside the BPC geometrical acceptance. An additional uncertainty of the order of $\triangle\theta_{\text{elec.}} \approx 0.8$ mrad due to misalignment is associated with the BPC measurement. Thus, preferring a conservative estimate, we vary the electron angle in the MC detector simulations by $\triangle\theta_{\text{elec.}} = \pm 1$ mrad. The procedure for extracting the resulting uncertainty is the same as in the previous cases.

Energy Loss

The uncertainty of the particle energy loss for the D^*- daughter particles is estimated in the following way. From the distribution[8] of dE/dx for all selected tracks in data events that enter the $\triangle M$ plot (figure 6.9) one estimates the dE/dx measurement resolution. The relative resolution is approx. 13% and variations (upwards and downwards) of this magnitude were applied to the MC dE/dx values.

We observe that in all bins of z_{jet}, z_{hem} and $z_{\text{hem}}^{(\text{jet})}$ distributions the extracted

[8]This distribution is just the projection on the vertical axis of the two-dimensional plot from figure 4.5.

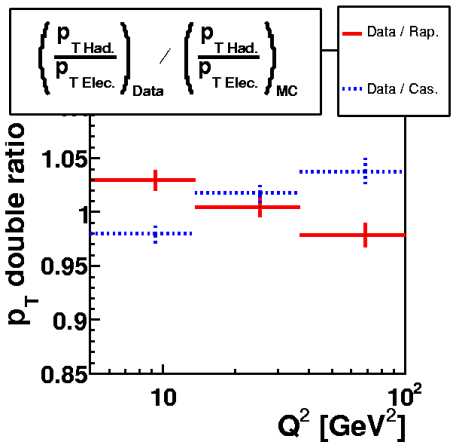


Figure 6.10: Double-ratio plot for transverse momentum p_T measured as function of Q^2 for hadronic final state and for the scattered electron in both, data and MC models.

relative error is negligible in comparison to other systematic errors, and thus it is neglected in summary tables.

Hadronic Energy Scale

The estimated uncertainty of the hadronic energy measurement is essentially based on the double-ratio plot in figure 6.10. It shows the comparison of data to MC simulations with respect to the p_T balance between the scattered electron and the hadronic final state. The displayed quantity $\frac{(P_T^{\text{Had.}}/P_T^{\text{Elec.}})_{\text{Data}}}{(P_T^{\text{Had.}}/P_T^{\text{Elec.}})_{\text{MC}}}$ is independent of the physics model[9] and thus reflects the measurement biases. The plot suggests an uncertainty of the order of 4%. This value can be considered as safe uncertainty estimate and is supported by HERA I analyses (the calorimeters remained the same in HERA II), see for example reference [77]. To account for the worse hadronic energy measurement in the Spacal compared to the Liquid Argon Calorimeter, we apply different shifts for hadrons measured in these two detectors. The calorimetric energy[10] of hadronic objects is then varied by $\pm$ 4% for the LAr Calorimeter and $\pm$ 7% for the Spacal. The systematic uncertainties associated with the hadronic energy scale were defined in a similar way to previous cases.

Track Momenta

The variation in tack momenta of the D^*- daughter particles was taken to be $\pm 0.5\,\%$ [78].

[9] In every model momentum conservation implies the balance in transverse momenta.

[10] In many cases the energy is taken from the track information. In this case the variation is not done.

Beauty Contribution to D^* meson production

The beauty fraction was raised by 100 % (doubled) with respect to the default value predicted by the RAPGAP model. The impact of this variation on the unfolded spectra defines the corresponding systematic uncertainty. A downward variation of the beauty contribution is not supported by existing studies in beauty physics and was therefore not done.

Resolved Photon Component

The last systematic uncertainty taken into account is the resolved photon component. It was so far neglected in our fragmentation study and in order to evaluate the impact of a possible resolved photon contribution, we investigate its effect by simulating resolved processes at the level as predicted by RAPGAP. However, a complication in unfolding arises, since QED radiative effects are not available when simulating resolved contribution with RAPGAP. To overcome this difficulty the bin-by-bin procedure was adopted. Although this does not seem to be fully consistent with the nominal procedure based on the SVD approach, we expect the systematic bias introduced by this approach to cancel in large parts, because we apply the bin-by-bin procedure to both, the nominal direct-only MC model as well as to the MC simulation including also a resolved photon component. The correction of QED radiative effects for the resolved contribution is done in a similar way as for CASCADE (see section 6.2.4). The systematic uncertainties due to the additional resolved contribution are defined as differences between the bin-by-bin corrected spectra using the direct-only MC events in one case and using a mixture of direct and resolved events in the other.

6.3.2 Summary of Statistical and Systematic Uncertainties

The three following summary tables contain the numerical values for the statistical and systematic uncertainties. They were obtained as described in the previous section. Systematic uncertainties which are correlated between different bins of the extracted distributions are visualized in the plots of appendix B.

z_{jet} interval	0.3 - 0.55	0.55 - 0.7	0.7 - 0.825	0.825 - 0.9	0.9 - 1.0
Statistical error	1.80 %	1.12 %	0.18 %	1.04 %	2.04 %
Signal extraction	4.49 %	4.49 %	4.49 %	4.49 %	4.49 %
Unfolding uncertainty	9.28 %	2.97 %	0.90 %	2.38 %	1.53 %
Model dependence	3.54 %	0.19 %	1.22 %	0.20 %	1.05 %
Spacal trigger efficiency	0.64 %	0.64 %	0.64 %	0.64 %	0.64 %
Track trigger efficiency	0.27 %	$<$ 0.01 %	$<$ 0.01 %	1.54 %	0.49 %
Electron energy [+ (4 - 1) %]	0.47 %	0.16 %	-0.04 %	-0.07 %	-0.38 %
Electron energy [- (4 - 1) %]	-0.66 %	-0.20 %	0.07 %	0.37 %	0.30 %
Electron θ [+1.0 mrad]	0.04 %	$<$ 0.01 %	-0.01 %	0.05 %	-0.05 %
Electron θ [-1.0 mrad]	-0.01 %	$>$ -0.01 %	-0.01 %	0.01 %	0.01 %
Had. energy scale [+4 $\%_{\text{Lar.}}$, +7 $\%_{\text{Spa.}}$]	-0.70 %	-0.42 %	-0.04 %	0.38 %	0.71 %
Had. energy scale [-4 $\%_{\text{Lar.}}$, -7 $\%_{\text{Spa.}}$]	0.65 %	0.41 %	0.03 %	-0.37 %	-0.70 %
Track momenta [+0.5 %]	0.41 %	0.24 %	0.03 %	-0.21 %	-0.43 %
Track momenta [-0.5 %]	-0.28 %	-0.17 %	$>$ 0.01 %	0.14 %	0.29 %
Beauty fraction [+100 %]	-0.42 %	-0.25 %	-0.02 %	0.22 %	0.43 %
Resolved contribution	2.85 %	0.53 %	-0.59 %	-0.88 %	-1.22 %

Table 6.1: Statistical and systematic errors for bins of z_{jet} distribution.

z_{hem} interval	0.2 - 0.4	0.4 - 0.5	0.5 - 0.625	0.625 - 0.75	0.75 - 0.85	0.85 - 1.0
Statistical error	4.03 %	2.63 %	1.58 %	2.04 %	2.32 %	3.02 %
Signal extraction	4.49 %	4.49 %	4.49 %	4.49 %	4.49 %	4.49 %
Unfolding uncertainty	10.83 %	10.93 %	2.78 %	3.75 %	6.21 %	6.19 %
Model dependence	0.85 %	2.21 %	1.42 %	0.15 %	2.31 %	1.17 %
Spacal trigger efficiency	0.64 %	0.64 %	0.64 %	0.64 %	0.64 %	0.64 %
Track trigger efficiency	0.60 %	1.09 %	0.02 %	1.05 %	0.12 %	0.79 %
Electron energy [+ (4 - 1) %]	2.41 %	1.23 %	0.50 %	-0.18 %	-0.89 %	-1.61 %
Electron energy [- (4 - 1) %]	-2.78 %	-1.69 %	-0.88 %	0.12 %	1.31 %	2.26 %
Electron θ [+1.0 mrad]	-0.13 %	-0.09 %	-0.07 %	-0.01 %	0.08 %	0.15 %
Electron θ [-1.0 mrad]	0.12 %	0.14 %	0.07 %	$>$ -0.01 %	-0.09 %	-0.15 %
Had. energy scale [+4 $\%_{\text{Lar.}}$, +7 $\%_{\text{Spa.}}$]	-3.29 %	-2.62 %	-1.40 %	0.36 %	1.88 %	2.84 %
Had. energy scale [-4 $\%_{\text{Lar.}}$, -7 $\%_{\text{Spa.}}$]	3.36 %	2.74 %	1.55 %	-0.35 %	-2.05 %	-2.97 %
Track momenta [+0.5 %]	0.39 %	0.40 %	0.26 %	-0.04 %	-0.30 %	-0.43 %
Track momenta [-0.5 %]	-0.34 %	-0.28 %	-0.23 %	$>$ -0.01 %	0.22 %	0.45 %
Beauty fraction [+100 %]	-2.53 %	-1.72 %	-0.55 %	0.56 %	1.09 %	1.24 %
Resolved contribution [switched ON]	2.11 %	0.38 %	-0.23 %	-1.01 %	-0.43 %	0.58 %

Table 6.2: Statistical and systematic errors for bins of z_{hem} distribution.

$z_{\mathrm{hem}}^{(\mathrm{jet})}$ interval	0.2 - 0.4	0.4 - 0.5	0.5 - 0.625	0.625 - 0.75	0.75 - 0.85	0.85 - 1.0
Statistical error	4.29 %	2.63 %	1.61 %	2.99 %	3.72 %	5.12 %
Signal extraction	4.49 %	4.49 %	4.49 %	4.49 %	4.49 %	4.49 %
Unfolding uncertainty	11.17 %	3.97 %	4.35 %	2.13 %	3.01 %	1.58 %
Model dependence	3.41 %	0.21 %	0.87 %	0.35 %	0.92 %	2.46 %
Spacal trigger efficiency	0.64 %	0.64 %	0.64 %	0.64 %	0.64 %	0.64 %
Track trigger efficiency	0.62 %	$<$ 0.01 %	0.06 %	0.32 %	0.94 %	0.68 %
Electron energy [+ (4 - 1) %]	0.67 %	0.13 %	-0.04 %	-0.17 %	-0.29 %	-0.13 %
Electron energy [- (4 - 1) %]	-0.55 %	0.17 %	0.28 %	0.06 %	-0.32 %	0.03 %
Electron θ [+1.0 mrad]	0.13 %	$>$ -0.01 %	-0.01 %	-0.04 %	-0.060 %	0.03 %
Electron θ [-1.0 mrad]	-0.01 %	0.03 %	0.05 %	0.01 %	-0.09 %	-0.04 %
Had. energy scale [+4 $\%_{\mathrm{Lar.}}$, +7 $\%_{\mathrm{Spa.}}$]	-2.73 %	-2.05 %	-0.87 %	0.77 %	2.41 %	3.63 %
Had. energy scale [-4 $\%_{\mathrm{Lar.}}$, -7 $\%_{\mathrm{Spa.}}$]	2.84 %	2.19 %	0.99 %	-0.94 %	-2.68 %	-3.50 %
Track momenta [+0.5 %]	0.79 %	0.67 %	0.28 %	-0.31 %	-0.78 %	-0.91 %
Track momenta [-0.5 %]	-0.86 %	-0.60 %	-0.29 %	0.32 %	0.74 %	0.94 %
Beauty fraction [+100 %]	-2.13 %	-1.42 %	-0.33 %	0.81 %	1.47 %	1.73 %
Resolved contribution [switched ON]	4.10 %	1.02 %	-0.07 %	-1.65 %	-1.68 %	-1.54 %

Table 6.3: Statistical and systematic errors for bins of $z_{\mathrm{hem}}^{(\mathrm{jet})}$ distribution.

Chapter 7

Results

7.1 Hadron Level Spectra with Statistical and Systematic Errors

The z_{hem}, z_{jet} and $z_{\text{hem}}^{(\text{jet})}$ distributions corrected to hadron level together with their statistical and systematic errors are one of the main results of this work, exploiting the four years long data taking period at the HERA II collider. The numbers of reconstructed D^* mesons for the whole HERA II running period are shown in table 7.1. The differential cross sections for the D^* meson production normalized to the visible total cross section within the $z_{\text{jet/hem}}$-range displayed in the respective subfigures are presented in figure 7.1 and the numerical values can be found in table 7.2. The visible range is defined by $Q^2 > 5\,\text{GeV}^2$, $0.05 < y < 0.6$, $1.5\,\text{GeV} < p_T(D^*) < 15.0\,\text{GeV}$, $|\eta(D^*)| < 1.5$, and additionally for z_{jet} and $z_{\text{hem}}^{(\text{jet})}$ by $E_T(D^*\,\text{jet}) > 3.0\,\text{GeV}$.

The different sources of systematic errors are considered as uncorrelated for each histogram bin. They are added in quadrature with the corresponding statistical uncertainty. If for a given error source two uncertainty estimates are available (one coming from an upward and the other from a downward variation of the quantity in consideration) then the average of their absolute values is taken. One should also be aware of the fact that correlations between different bins exist, but that they cannot be seen in figure 7.1 and in table 7.2. These correlations however are taken into account in the following section, where the corresponding covariance matrices are presented and fits of fragmentation functions using the unfolded distributions are performed.

7.2 Extraction of Fragmentation Parameters

The measured normalized D^* cross sections in z_{jet}, z_{hem} and $z_{\text{hem}}^{(\text{jet})}$ allow us to extract parameters for different parametrizations of the charm quark fragmentation function. This analysis is performed in the framework of the Lund String fragmentation model that is implemented in PYTHIA, which has been interfaced to both, RAPGAP and CASCADE. The parametrizations investigated are the ones by Pe-

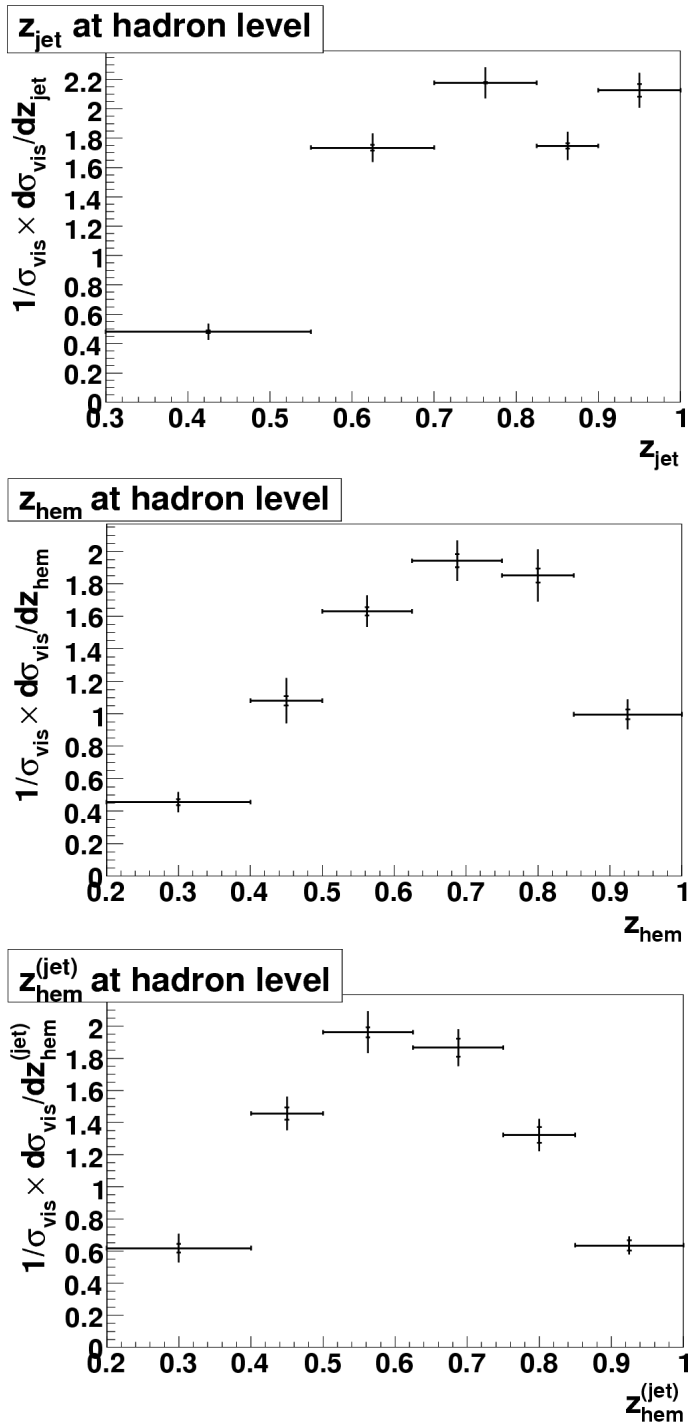


Figure 7.1: The differential cross sections for the D^* meson production normalized to the visible total cross section for $z_{\rm jet}$, $z_{\rm hem}$ and $z_{\rm hem}^{\rm (jet)}$. Total errors are shown, statistical uncertainties are denoted by short horizontal lines. The different behavior of $z_{\rm jet}$ distribution with respect to the two other distributions in the highest bin is caused by reconstructed jets containing only the D^* particle. For such jets $z_{\rm jet} = 1.0$.

z_{jet}	0.3 - 0.55	0.55 - 0.7	0.7 - 0.825	0.825 - 0.9	0.9 - 1.0
$\#D^*$	930	1819	2175	1146	1941

z_{hem}	0.2 - 0.4	0.4 - 0.5	0.5 - 0.625	0.625 - 0.75	0.75 - 0.85	0.85 - 1.0
$\#D^*$	1474	1869	3095	3981	3128	3010

$z_{\text{hem}}^{(\text{jet})}$	0.2 - 0.4	0.4 - 0.5	0.5 - 0.625	0.625 - 0.75	0.75 - 0.85	0.85 - 1.0
$\#D^*$	1076	1197	1813	1944	1183	924

Table 7.1: Number of reconstructed D^* mesons in bins of z_{jet}, z_{hem} and $z_{\text{hem}}^{(\text{jet})}$ for all of the HERA II data.

z_{jet}	0.3 - 0.55	0.55 - 0.7	0.7 - 0.825	0.825 - 0.9	0.9 - 1.0
$\frac{1}{\sigma_{\text{vis.}}}\frac{d\sigma_{\text{vis.}}}{dz_{\text{jet}}}$	0.480	1.735	2.177	1.748	2.127
Stat. err.	0.009	0.019	0.004	0.018	0.043
Syst. err.	0.054	0.095	0.105	0.095	0.101
Total err.	0.055	0.097	0.105	0.097	0.118

z_{hem}	0.2 - 0.4	0.4 - 0.5	0.5 - 0.625	0.625 - 0.75	0.75 - 0.85	0.85 - 1.0
$\frac{1}{\sigma_{\text{vis.}}}\frac{d\sigma_{\text{vis.}}}{dz_{\text{hem}}}$	0.455	1.080	1.631	1.942	1.852	0.997
Stat. err.	0.018	0.029	0.026	0.040	0.043	0.030
Syst. err.	0.059	0.136	0.094	0.119	0.156	0.086
Total err.	0.062	0.139	0.098	0.125	0.162	0.091

$z_{\text{hem}}^{(\text{jet})}$	0.2 - 0.4	0.4 - 0.5	0.5 - 0.625	0.625 - 0.75	0.75 - 0.85	0.85 - 1.0
$\frac{1}{\sigma_{\text{vis.}}}\frac{d\sigma_{\text{vis.}}}{dz_{\text{hem}}^{(\text{jet})}}$	0.618	1.456	1.962	1.866	1.322	0.635
Stat. err.	0.027	0.038	0.032	0.056	0.049	0.033
Syst. err.	0.085	0.097	0.126	0.101	0.087	0.044
Total err.	0.089	0.104	0.130	0.116	0.100	0.055

Table 7.2: The differential cross sections for the D^* meson production normalized to the visible total cross section and their uncertainties in bins of z_{jet}, z_{hem} and $z_{\text{hem}}^{(\text{jet})}$.

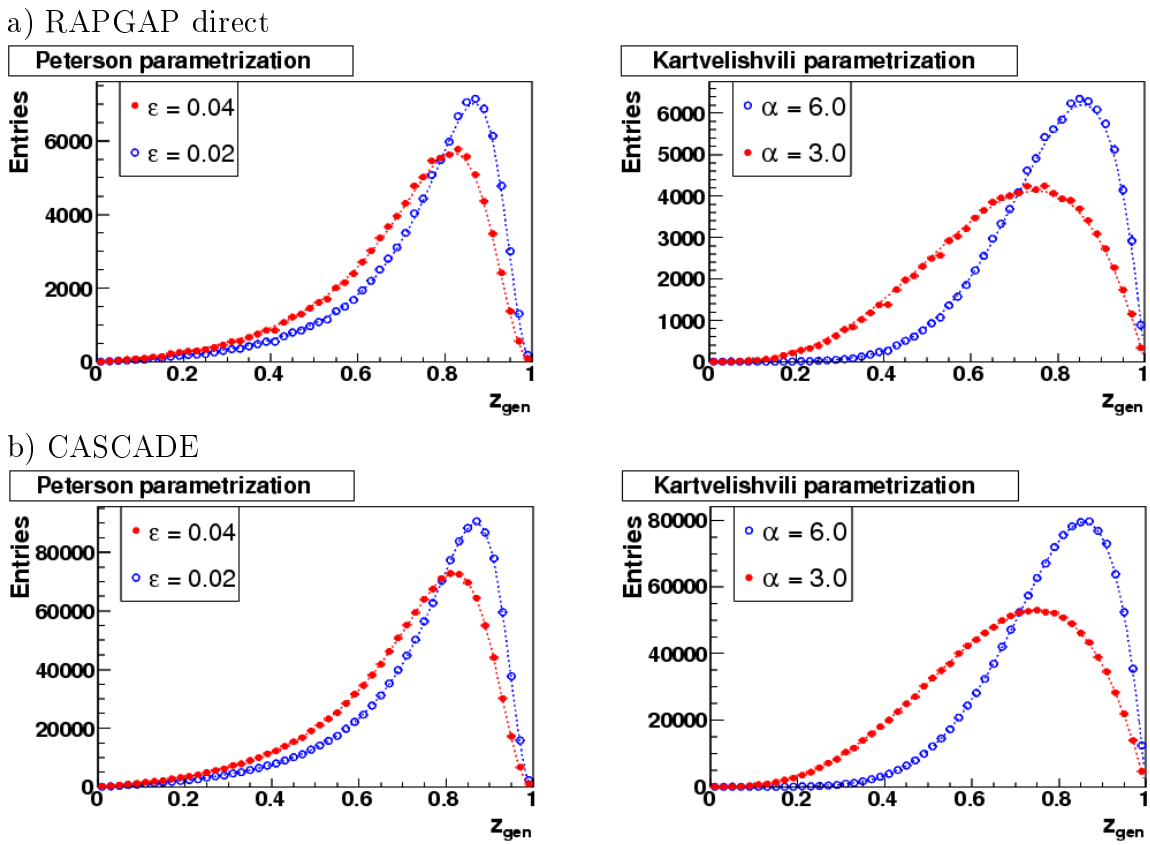


Figure 7.2: Cross-check of the reweighting procedure for events generated using RAPGAP direct and CASCADE. Events are produced using the Peterson function with $\varepsilon = 0.04$ and $Q^2 > 1$ GeV and are reweighted to other parametrizations ($\varepsilon = 0.02$ for Peterson and $\alpha = 6.0$, 3.0 for Kartvelishvili). Graphs of the Peterson and Kartvelishvili functions themselves are also shown (dotted lines); they match well with the reweighted distributions (an appropriate normalization was chosen).

terson and Kartvelishvili. They depend on a single parameter. The extraction of fragmentation parameters is based on calculating χ^2 from the measured data and the MC distributions, each MC distribution corresponding to a different fragmentation parameter. First, the production of MC distributions with different fragmentation parameters is explained, next the χ^2 method and the obtained results are presented.

7.2.1 Reweighting of Monte Carlo Models in z_{gen}

In order to extract fragmentation parameters one needs to have a large number of hadron-level MC distributions, with negligible statistical error, corresponding to different fragmentation functions in order that the minimization procedure using the χ^2 method works properly and the optimal parameter as well as its error can be determined reliably. A straightforward way of fulfilling this requirement would be to run the MC simulation many times, each time with a different fragmentation parameter. Even though no detector simulation is needed for our purpose, this

procedure still requires a large amount of computing time. This is the reason why a different approach was chosen. Let us denote the relevant MC quantity used in the string breaking as z_{gen}. The MC programs were modified such that the relevant values of z_{gen} were written into an output bank (gki bank). Then just one high statistics MC run was made with the Peterson parametrization and the parameter $\varepsilon = 0.04$. The z_{jet}, z_{hem} and $z_{\text{hem}}^{(\text{jet})}$ distributions (at hadron level, with all visible range requirements) for different fragmentation functions were then obtained by reweighting the generated MC events with respect to z_{gen}. This procedure was constructed such that, after the reweighting, the z_{gen} distribution follows the analytical shape of the Peterson function with the chosen value for the parameter ε or the Kartvelishvili function with a chosen α. This procedure has several nice features: it is quick, flexible and, in addition, it enables us to cross-check our understanding of the implementation of the Lund String fragmentation in MC programs and observe that the initial z_{gen} distribution indeed follows the Peterson function with $\varepsilon = 0.04$.

The reweighting procedure was tested. For that purpose events were generated with the Peterson fragmentation function and $\varepsilon = 0.04$. We required $Q^2 > 1\,\text{GeV}^2$, which was needed because of the steeply rising cross section at low Q^2; the absence of this cut would lead to a large number of events lying outside the visible range. This requirement might have an impact on the shape of the z_{gen} distribution and can, in principle, interfere with the reweighting procedure. However, the results presented in figure 7.2 indicate a negligible influence. The figure also illustrates that the initial z_{gen} distributions (for RAPGAP and CASCADE) indeed follow well the Peterson fragmentation function with $\varepsilon = 0.04$ (a small bias can be explained by the Q^2 cut) and that after reweighting the z_{gen} distributions respect the functions they were reweighted to.

When extracting fragmentation parameters using RAPGAP, resolved events were taken into account too, since these events have slightly different z_{jet}, z_{hem} and $z_{\text{hem}}^{(\text{jet})}$ spectra. The same reweighting procedure was applied, and direct and resolved events were combined with respect to their cross sections as predicted by the MC model. This combination was then used in the χ^2 method.

7.2.2 χ^2 Method and Extracted Fragmentation Parameters

Method of χ^2 Evaluation

The optimal fragmentation parameters are determined via a χ^2 comparison between the generated MC distributions and the measured data. Since the unfolding introduces correlations of statistical uncertainties among different histogram bins, and since some of the systematic errors are also correlated, instead of the simple χ^2 formula an expression with the full covariance matrix is used:

$$\chi^2(A) = u^T(A) C^{-1} u(A)\,,$$

where A denotes the used fragmentation parameter ($A = \varepsilon, \alpha$), u a vector with the number of bins as length $u_i = z_i^{\text{Data}} - z_i^{\text{MC}}(A)$ and C the covariance matrix related

to the statistical and systematic errors. C is the sum of three matrices:

$$C = X^{\text{stat.}}_{\text{corr.}} + Y^{\text{sys.}}_{\text{corr.}} + Z^{\text{sys.}}_{\text{uncorr.}},$$

where $X^{\text{stat.}}_{\text{corr.}}$, $Y^{\text{sys.}}_{\text{corr.}}$ and $Z^{\text{sys.}}_{\text{uncorr.}}$ are the covariance matrices corresponding to the correlated statistical errors, the correlated systematic and the uncorrelated systematic errors, respectively.

The matrices $X^{\text{stat.}}_{\text{corr.}}$ for each of the three observables z_{jet}, z_{hem} and $z^{(\text{jet})}_{\text{hem}}$ are obtained following the prescription given in section 6 of reference [67] :

- z_{jet} :

$$X^{\text{stat.}}_{\text{corr.}} = \begin{pmatrix} 1135 & 1500 & 164.6 & -678.4 & -2164 \\ 1500 & 1992 & 237 & -896.9 & -2889 \\ 164.6 & 237 & 58.91 & -99.16 & -367.5 \\ -678.4 & -896.9 & -99.16 & 405.9 & 1294 \\ -2164 & -2889 & -367.5 & 1294 & 4209 \end{pmatrix},$$

- z_{hem} :

$$X^{\text{stat.}}_{\text{corr.}} = \begin{pmatrix} 23380 & 13940 & -522.2 & -21490 & -15230 & -8303 \\ 13940 & 9890 & 3647 & -12390 & -11810 & -7905 \\ -522.2 & 3647 & 11660 & 4136 & -7806 & -10310 \\ -21490 & -12390 & 4136 & 25590 & 11350 & 346 \\ -15230 & -11810 & -7806 & 11350 & 15830 & 12600 \\ -8303 & -7905 & -10310 & 346 & 12600 & 16090 \end{pmatrix},$$

- $z^{(\text{jet})}_{\text{hem}}$:

$$X^{\text{stat.}}_{\text{corr.}} = \begin{pmatrix} 10810 & 5916 & -1988 & -9981 & -5644 & -2713 \\ 5916 & 3848 & 255.9 & -5540 & -3973 & -2366 \\ -1988 & 255.9 & 3980 & 2421 & -1380 & -2421 \\ -9981 & -5540 & 2421 & 10780 & 4773 & 828.1 \\ -5644 & -3973 & -1380 & 4773 & 4636 & 3335 \\ -2713 & -2366 & -2421 & 828.1 & 3335 & 4128 \end{pmatrix}.$$

These matrices need to be further modified for two reasons. Firstly, the numerical value of a given matrix element reflects the number of events, but when comparing the area-normalized distributions of the data with those of the MC models, we need to scale the matrices so that the diagonal elements correspond to the squares of the established statistical errors. Secondly, the distributions obtained via unfolding have been further modified[1], together with their statistical errors, and consequently an inconsistency has been introduced between the diagonal elements of the presented matrices and the final statistical errors. Therefore, it is not possible to scale the matrices by a single scalar factor to match the diagonal elements and the statistical errors, but one needs to define bin-dependent scaling factors $s_i^2 = X_{ii}/\sigma_i^2$ such that

[1]The bin-by-bin radiative corrections were applied.

a given matrix element is scaled by $X_{ij} \longrightarrow X_{ij}/s_i s_j$. This means that the diagonal matrix elements are by construction set to the correct values. To make sure, that this procedure does not introduce a bias, the ratios s_i/s_j were calculated for all scaling factors. It was found that they are always close to one, meaning that the applied procedure is not very different from one using a scalar normalization factor.

The covariance matrix $Y^{\text{sys.}}_{\text{corr.}}$ of the correlated systematic errors of a given observable is defined as the sum of matrices containing the different systematic error sources, $Y^{\text{sys.}}_{\text{corr.}} = \sum_k Y^k$ (see tables 6.1, 6.2 and 6.3). We considered that a correlation between different bins exists for the uncertainties related to the electron energy and azimuthal angle θ, the hadronic energy scale and the track momenta, and the beauty and resolved contributions. The plots in appendix B show a simple type of correlation - the bins on opposite sides of the histograms are anticorrelated with an approximately linear dependence. Thus, we define[2] the covariance matrix elements to be $Y^k_{ij} = \vartheta_{ij}\sigma_i^{\text{sys. corr.}}\sigma_j^{\text{sys. corr.}}$, where ϑ_{ij} gives the sign (± 1) of the product $\sigma_i^{\text{up}} \cdot \sigma_j^{\text{up}}$, where σ_i^{up} and σ_j^{up} correspond to the uncertainties of bin i and j estimated by the up variation of the respective quantity.

The diagonal elements of the covariance matrix for uncorrelated systematic errors of a given type is defined as $Z^k_{ii} = (\sigma_i^{\text{sys. uncorr.}})^2$, the off-diagonal elements are set to zero.

Fits and Fragmentation Parameters

The method described above allows us to determine the dependence of χ^2 on the fragmentation parameters, the resulting plots are presented in appendix C. The obtained points are fitted by a parabola in the region near the χ^2 minimum, and the minimum of the parabola defines the best fit value for the extracted parameter. The error on this parameter is determined by a variation of the parameter leading to a rise of χ^2 by 1. The comparison of the data and the MC predictions with near-to-optimal parameters[3] is shown in figure 7.3, and the numerical results are presented in table 7.3.

7.3 Results from e^+e^- Experiments[4]

Fragmentation universality is an important and still opened issue that requires further studies. In this section we focus on the results of three e^+e^- experiments, BELLE [65], CLEO [64] and ALEPH [63] and extract the optimal Peterson fragmentation parameter. These three experiments studied charm fragmentation into D^* mesons, BELLE and CLEO at energies close to $b\bar{b}$ production threshold ($\approx 10.52 - 10.58$ GeV) and ALEPH at energies corresponding to the Z resonance (91.2 GeV).

[2] If for a given bin two estimates of the systematic error $\sigma_{i,1}^{\text{sys. corr.}}$ and $\sigma_{i,2}^{\text{sys. corr.}}$ are available (variation up and down), we use $\sigma_i^{\text{sys. corr.}} = \frac{1}{2}\left(\left|\sigma_{i,1}^{\text{sys. corr.}}\right| + \left|\sigma_{i,2}^{\text{sys. corr.}}\right|\right)$.

[3] The parameters used to produce the shown MC distributions correspond to the MC settings that lead to the points lying next to the parabola minimum, see appendix C.

[4] The work presented in this section was done with the help of Thomas Lübbert, a DESY summer student in 2007 whom Dr. Grindhammer and I were supervising, see [79].

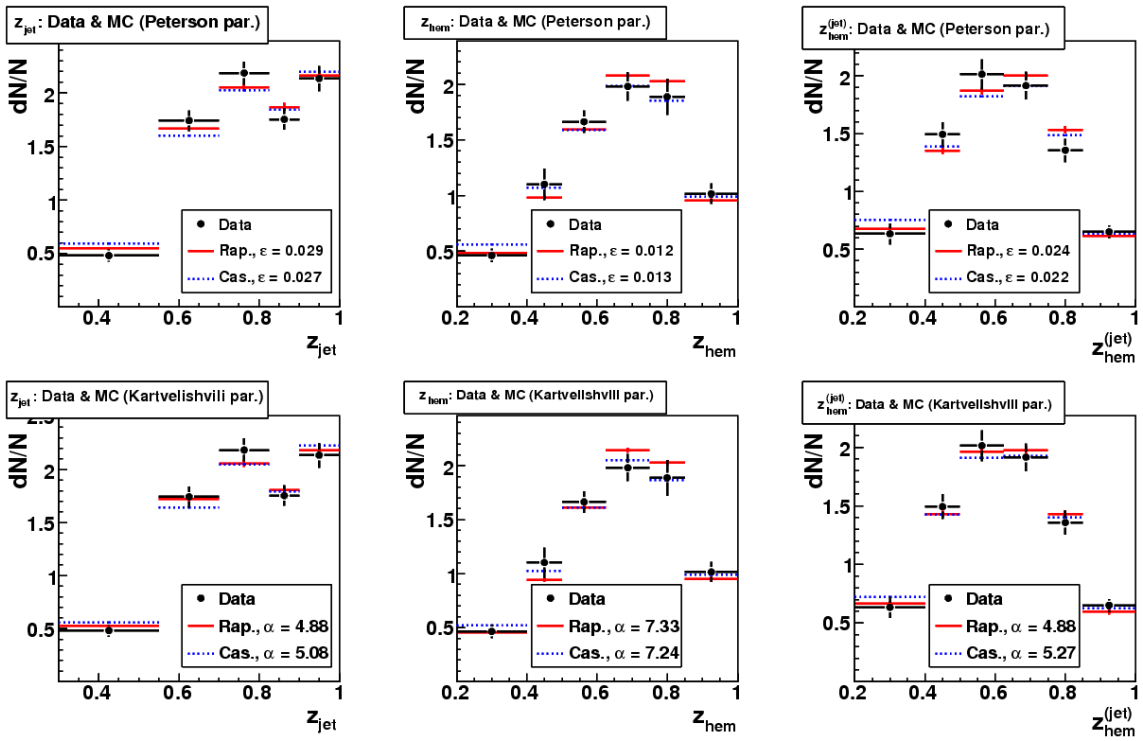


Figure 7.3: Comparison of data and MC distributions, the MC predictions were produced with near-to-optimal parameters.

z_{jet}	Parameter (α, ε)	$\chi^2/n.d.f.$
Peterson's ε	Rap.: $\varepsilon = 0.0285 \pm 0.0028$	$5.28/3 = 1.76$
	Cas.: $\varepsilon = 0.0273 \pm 0.0027$	$10.43/3 = 3.48$
Kartvelishvili's α	Rap.: $\alpha = 4.88 \pm 0.23$	$2.71/3 = 0.90$
	Cas.: $\alpha = 5.08 \pm 0.24$	$5.89/3 = 1.97$

z_{hem}	Parameter (α, ε)	$\chi^2/n.d.f.$
Peterson's ε	Rap.: $\varepsilon = 0.0116 \pm 0.0031$	$3.35/4 = 0.84$
	Cas.: $\varepsilon = 0.0129 \pm 0.0030$	$3.65/4 = 0.91$
Kartvelishvili's α	Rap.: $\alpha = 7.33 \pm 0.91$	$4.49/4 = 1.12$
	Cas.: $\alpha = 7.24 \pm 0.81$	$2.16/4 = 0.54$

$z_{\text{hem}}^{(\text{jet})}$	Parameter (α, ε)	$\chi^2/n.d.f.$
Peterson's ε	Rap.: $\varepsilon = 0.0241 \pm 0.0040$	$8.92/4 = 2.23$
	Cas.: $\varepsilon = 0.0223 \pm 0.0036$	$8.48/4 = 2.12$
Kartvelishvili's α	Rap.: $\alpha = 4.88 \pm 0.45$	$3.76/4 = 0.94$
	Cas.: $\alpha = 5.27 \pm 0.47$	$3.48/4 = 0.87$

Table 7.3: Extracted fragmentation parameters together with the $\chi^2/n.d.f.$ value of the fit.

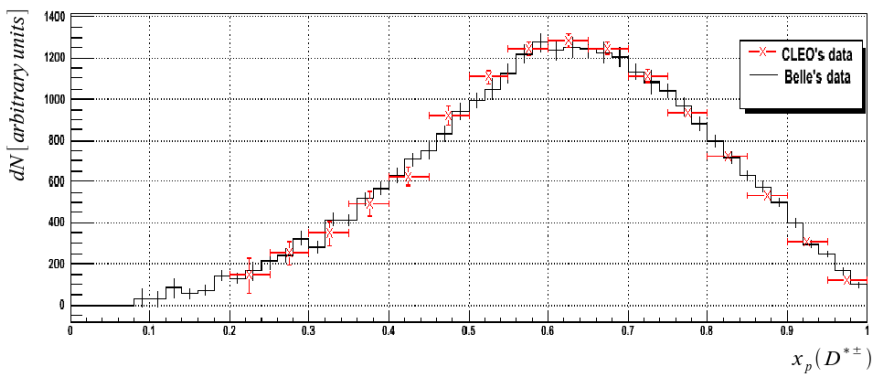


Figure 7.4: Area-normalized D^* spectra as a function of x_p as measured by the CLEO and BELLE collaborations. The data points agree within the given errors.

Some of these collaborations extracted fragmentation parameters for the Peterson and/or the Kartvelishvili function, we however will not compare our results to their numbers here. The reason is our lack of detailed knowledge about the MC models used for their parameter extraction. It is known that the results depend on whether higher charm resonances are taken into account, on the particles masses used as well as on various other steering parameters. Therefore, we rely only on their published data and run the PYTHIA MC model in e^+e^- mode with the Peterson fragmentation function and settings identical to what we used in the *ep* study. The optimal Peterson parameter ε is then determined by the χ^2 minimum found by comparing the published data and generated MC distributions.

CLEO

The CLEO measurement is in many aspects very similar to the BELLE one, both used the same observable x_p (reduced momentum, see section 3.3) and both were running at similar CMS energies. However, the BELLE measurement has better statistics and thus is preferable. In order to check the compatibility of the two measurements we normalized and overlaid their published spectra in figure 7.4. One observes that the two distributions are in good agreement within their errors. Due to the bigger errors we omit the CELO data and base our study on fragmentation universality only on the BELLE and ALEPH measurements, which are briefly described in the following.

BELLE

The published BELLE data allow for an easy analysis, since they do not contain any $b\bar{b}$ contamination and are corrected for all detector effects. We consider the measured x_p spectrum determined from the charged $D^* \to D^0\pi$ decay channel, where for $x_P < 0.5$ only continuum data ($\sqrt{s} = 10.52\,\text{GeV}$, no decays of B mesons) were used and for $x_p > 0.5$ a weighted average of continuum and "on resonance" data ($\sqrt{s} = 10.58\,\text{GeV}$) was used. The beauty component in the "on resonance" data is for $x_p > 0.5$ strongly suppressed. The data are not corrected for QED radiation

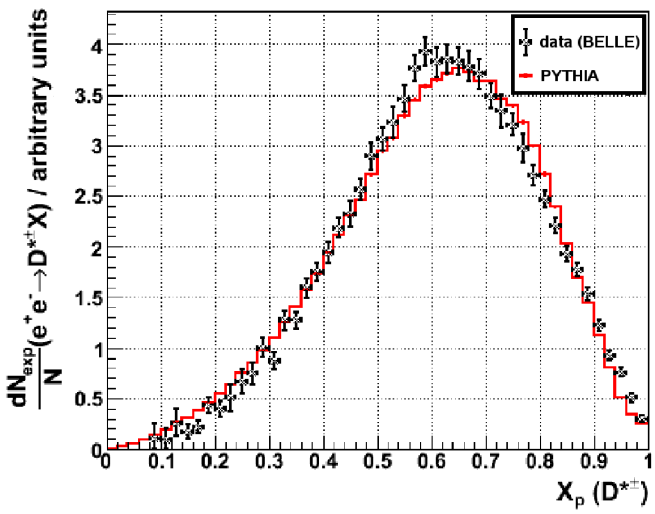


Figure 7.5: Comparison of data measured by BELLE and the prediction of the PYTHIA model with the Peterson fragmentation parameter $\varepsilon = 0.032$.

effects, and therefore we use a radiative MC simulation for the χ^2 comparison. The comparison of the BELLE data with the MC prediction based on a near-to-optimal parameter value of $\varepsilon = 0.032$ is shown in figure 7.5. In order to fit the fragmentation parameter, values of χ^2 are calculated[5] with respect to the measured data for MC predictions with different ε and a parabola is fitted to the χ^2 points close to the observed χ^2 minimum. The total measurement errors (statistical and systematic) are taken into account and are considered as uncorrelated; the information about correlations being not available. The ε corresponding to the minimum is considered to be the optimal fragmentation parameter, and the spread in ε leading to the variation of χ^2 by 1 is considered as $\pm 1\sigma$ error. The obtained results are

$$\varepsilon = 0.0316 \pm 0.0006$$

and

$$\chi^2_{\text{min.}}/n.d.f. = 148.95/45 = 3.31.$$

The important statistics accumulated by BELLE and the small statistical errors resulting from it reflect in both, the small uncertainty of the extracted parameter and a rather bad value of $\chi^2/n.d.f.$. The latter suggests that the Peterson parametrization and the MC model are not able to provide an adequate description of these precise data.

The BELLE collaboration using their MC model obtained $\varepsilon = 0.054$ for the Peterson parameter; this result and the value $\chi^2_{\text{min.}}/n.d.f. = 55.6$ suggest that the higher excited charm states were not included in their MC simulation, in contrast to our study.

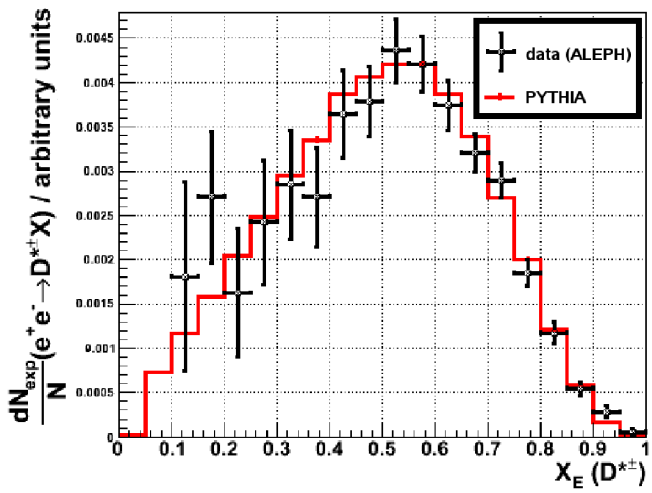


Figure 7.6: Comparison of data measured by ALEPH and the prediction of the PYTHIA model with the Peterson fragmentation parameter $\varepsilon = 0.042$.

ALEPH

The ALEPH measurement differs significantly from the BELLE one in several points. Unlike BELLE, the ALEPH collaboration uses the fractional energy x_E as the observable quantity (see section 3.3), the accumulated statistics is much smaller and the higher CMS energy entails more background. An important fraction of D^* mesons comes from decays of beauty hadrons and a non-negligible fraction originates from charm quarks that are created in the splitting of a perturbative gluon into a $c\bar{c}$ pair. Although the last case involves genuine charm quark fragmentation into D^* mesons, these processes are regarded as background, since the observable x_E is constructed to reflect the expectation coming from the lowest-order $c\bar{c}$ production diagram (see figure 3.5) and the splitting of a perturbative gluon involves higher order diagrams.

The ALEPH data were analyzed in a similar way to the BELLE data. We used PYTHIA in the $e^+e^- \to c\bar{c} \to D^*X$ mode. We are able to reproduce the shape of the x_E spectra for both, D^*s coming from direct $c\bar{c}$ production and D^*s from gluon splitting. The relative normalization of these distributions is different from the one observed in ALEPH data, but this can be explained by the fact that our simulation does not include the production of light flavors with gluon radiation, i.e. other gluon splitting processes such as $e^+e^- \to q\bar{q}g \to q\bar{q}c\bar{c} \to D^*X$, which are present in the data. So, comparing the ALEPH and PYTHIA x_E-spectra for the direct $c\bar{c}$ production, we determined the optimal Peterson fragmentation parameter using the χ^2 method for non-empty bins

$$\varepsilon = 0.042 \pm 0.003$$

and

$$\chi^2_{\text{min.}}/n.d.f. = 13.26/17 = 0.78.$$

[5]In the χ^2 calculation only the bins with non-zero content are taken into account.

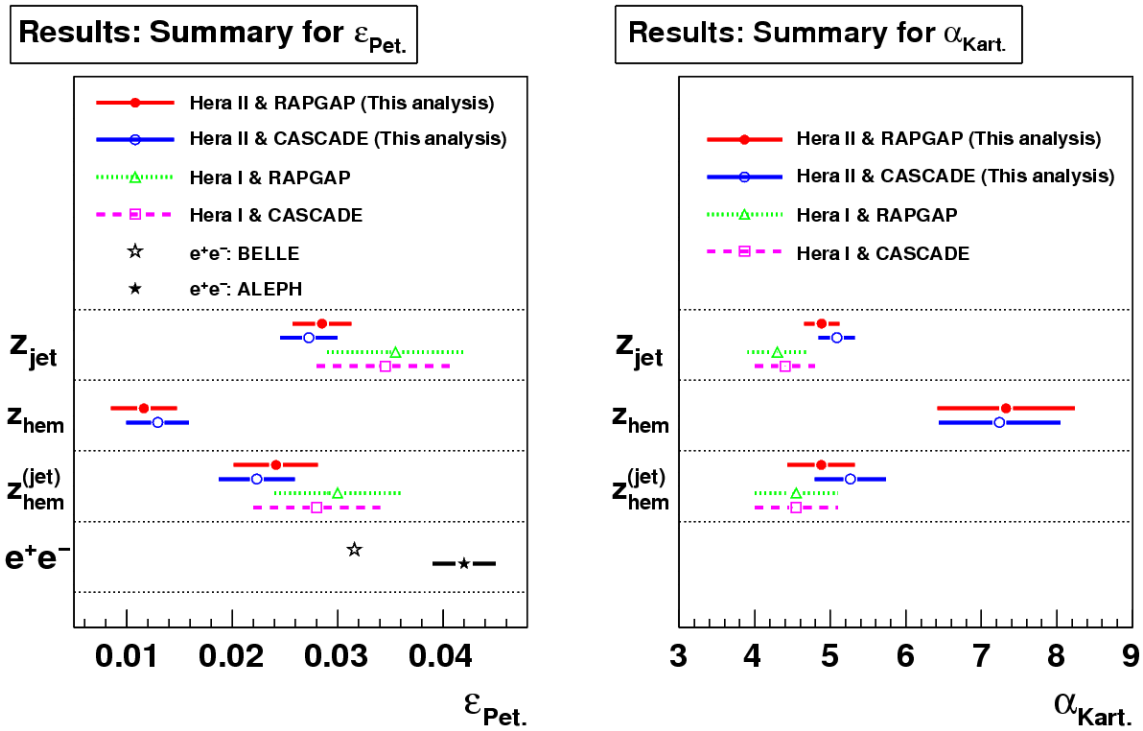


Figure 7.7: Summary of the results for the Peterson and the Kartvelishvili fragmentation parameter. Shown are the parameters extracted in this analysis from the HERA II data, the parameters from the HERA I data and from the data published by BELLE and ALEPH.

The data and the MC distribution produced with the fitted parameter, are shown in figure 7.6.

The ALEPH measurement lead to a better value of $\chi^2/n.d.f.$ than the BELLE results. In addition, one observes that the BELLE and the ALEPH values for the fragmentation parameter are not consistent ($\sim 3.5\,\sigma$) suggesting that even within e^+e^- experiments charm fragmentation is, at least for the Lund String model and the Peterson parametrization, not completely understood.

7.4 Comparison of Results and Conclusions

7.4.1 Observations

The results from this work together with the most recent results from the HERA I analysis [80] are summarized in figure 7.7.

Concerning the Peterson parameter, one observes good agreement for the values extracted using the z_{jet} and $z_{\text{hem}}^{(\text{jet})}$ method for both, the HERA I and HERA II running periods. These results are roughly consistent with the result obtained from the BELLE data, a deviation of $2.58\,\sigma$ is observed for the result obtained by the $z_{\text{hem}}^{(\text{jet})}$ method used in combination with CASCADE in HERA II. The parameters deviating

significantly from those extracted via $z_{\rm jet}$ and $z_{\rm hem}^{\rm (jet)}$ in HERA II are the ones using $z_{\rm hem}$ and the one from the ALEPH data. The results based on $z_{\rm hem}$ suggest a harder fragmentation, the result extracted from ALEPH data a softer one.

For the extracted Kartvelishvili fragmentation parameter one observes a similar behavior. The parameters from $z_{\rm jet}$ and $z_{\rm hem}^{\rm (jet)}$ for HERA I and HERA II agree reasonably well with each other, the ones from $z_{\rm hem}$ indicate a preference for a harder fragmentation.

For both parametrizations, results from HERA I exploiting the $z_{\rm hem}$ observable and using an unfolding with a response matrix are not available. However, results based on the bin-by-bin procedure exist [66] as well as results obtained with matrix unfolding and based on the $z_{\rm hem}$ observable reconstructed for events with no D^* jet only [80]. These results confirm the discrepancy between parameters from the $z_{\rm hem}$ and $z_{\rm jet}$ methods.

The observed features of the presented results can be summarized as follows

- The results based on the $z_{\rm jet}$ and $z_{\rm hem}^{\rm (jet)}$ observables are consistent and both are inconsistent with the results based on $z_{\rm hem}$. For both fragmentation parameters the $z_{\rm hem}$ method suggests a harder fragmentation function.

- The results obtained with two different MC models (RAPGAP and CASCADE) are consistent.

- The results from the HERA I analysis are consistent with the HERA II results obtained in this analysis, the HERA II results having smaller total errors (mainly driven by smaller statistical errors).

- In case of the Peterson fragmentation parameter, the results from the two e^+e^-experiments differ from each other by$\sim 3.5\sigma$, a discrepancy which has been seen also by other authors [81]. The result extracted from the BELLE measurement, which is based on larger statistics and less background (no beauty decays, no gluon splitting) is roughly consistent with our $z_{\rm jet}$ and $z_{\rm hem}^{\rm (jet)}$ results.

- The $\chi^2/n.d.f$ is acceptable for most HERA II results, it tends to be somewhat smaller for the Kartvelishvili parametrization than for the Peterson parametrization. The $\chi^2/n.d.f$ is rather large for the BELLE fit and suggests that the Lund String model with the Peterson function is not able to provide an adequate description of the precise data of BELLE.

7.4.2 Interpretation of Results

The most important observations with respect to possible interpretations are the independence (or small dependence) of the obtained results on the model of parton evolution (DGLAP vs. CCFM) and also on the running period and analysis (HERA I vs. HERA II, analyzed independently). The existing discrepancy between the results from the $z_{\rm hem}$ and $z_{\rm hem}^{\rm (jet)}$ methods (the latter being consistent with the $z_{\rm jet}$ result) is presumably coming from a different event selection (events with jet only vs. all

events) and indicates that the observed difference in parameters extracted via $z_{\rm hem}$ and $z_{\rm jet}$ does not originate in the different definitions of the observables but rather in the physics of the events. The $z_{\rm hem}^{\rm (jet)}$ and $z_{\rm jet}$-based results correspond to events significantly above the charm production threshold where enough energy is available for a jet with $E_T > 3\,\mathrm{GeV}$. One further observes that the HERA "jet" results are (in most cases) consistent with the value extracted from the BELLE data, the BELLE experiment having a CMS energy comparable with the average CMS energy of the photon-gluon system in BGF events entering the $z_{\rm hem}^{\rm (jet)}$ and $z_{\rm jet}$ distributions[6].

Taking these considerations into account, the most plausible interpretation of the obtained results is the inadequacy of existing models[7] in describing charm fragmentation over the whole range of production energies. Using the Lund String model with the Peterson or Kartvelishvili parametrization of the fragmentation function, the data suggest harder fragmentation near the production threshold and softer fragmentation for higher energies. The result derived from the ALEPH data fits well into this picture: the ALEPH data correspond to the highest CMS energy, and they lead to the softest fragmentation function. One should also notice that the HERA II data show a slight preference for the Kartvelishvili parametrization.

The discrepancy between the values derived from the ALEPH and BELLE data is of the same order as the discrepancy between parameters extracted via $z_{\rm hem}$ and via $z_{\rm hem}^{\rm (jet)}$ and $z_{\rm jet}$. Thus the question of fragmentation universality is obscured by inconsistencies within data having the same particles in the initial state and cannot be addressed easily. Taking "our" BELLE result at face value and comparing it with "our" ALEPH result, one would conclude that within the PYTHIA model we do not find universality. The agreement between the $z_{\rm hem}^{\rm (jet)}$ and $z_{\rm jet}$-based results and the parameter extracted from the BELLE data suggests that, at least for the corresponding charm production energies and for the used models, the assumption of fragmentation universality is valid. However, it is difficult to analyze the universality issue for a model that seems not to describe the data of same types at different charm production energies.

The possible causes of the inadequacy of the models to describe the range in $\sqrt{\hat{s}_{c\bar{c}}}$ might be clarified by further theoretical work; investigating of more appropriate parametrizations of the fragmentation function, analysis of different fragmentation models or study the validity of the factorization theorem for charm production in deep-inelastic ep collisions.

[6]The charm production energy is $\sqrt{\langle \hat{s}_{c\bar{c}} \rangle} \sim 10\,\mathrm{GeV}$ for events with jet and $\sqrt{\langle \hat{s}_{c\bar{c}} \rangle} \sim 8\,\mathrm{GeV}$ for all events (with and without a jet).

[7]By model we understand the Lund string model together with the Peterson or Kartvelishvili fragmentation function.

Chapter 8

Summary and Outlook

In this work the fragmentation of the charm quark into $D^{*\pm}$ mesons in deep-inelastic electron-proton collisions was studied. The data were taken by the H1 detector during the years 2004 - 2007 (HERA II running period), exploiting the beams of the HERA collider. We defined three observables quantities (z_{jet}, z_{hem} and $z_{\text{hem}}^{(\text{jet})}$) which are sensitive to the fragmentation process and measured the normalized differential $D^{*\pm}$ production cross sections for these observables in the visible range defined by the phase space requirements $Q^2 > 5\,\text{GeV}^2$, $0.05 < y < 0.6$, by the cuts on the D^* meson $1.5\,\text{GeV} < p_T(D^*) < 15.0\,\text{GeV}$, $|\eta(D^*)| < 1.5$, and additionally for z_{jet} and $z_{\text{hem}}^{(\text{jet})}$ by $E_T(D^*\,\text{jet}) > 3.0\,\text{GeV}$. The RAPGAP and CASCADE Monte Carlo models were used in combination with a method of regularized unfolding to correct the measured data for detector effects. The measured cross sections, together with their statistical and systematic errors, are visualized in figure 7.1 and the numerical values can be found in table 7.2.

The Lund String fragmentation model was used to extract the optimal fragmentation parameters for the Peterson and Kartvelishvili parametrizations of the fragmentation function. For that purpose two Monte Carlo generators (RAPGAP and CASCADE) with different parton evolutions were used, both generators being interfaced with PYTHIA 6.2, where the Lund String fragmentation is implemented. The corrected data were compared to predictions of the models using different values for the fragmentation parameter and, using a χ^2 method with a full covariance matrix, we extracted optimal fragmentation parameters for the three observables. In total 12 values for fragmentation parameter were extracted (for 2 models, 2 parametrizations and 3 observables). The results are summarized in table 7.3.

In order to check the universality of the charm fragmentation function with the model under our control we analyzed the data published by the BELLE and ALEPH collaborations. Using the PYTHIA 6.2 program in the e^+e^- mode, but otherwise with parameter setting for the analysis of our *ep* data, we fitted the fragmentation parameter for the Peterson parametrization.

All results are summarized in figure 7.7. Their comparison suggests that charm fragmentation is not fully understood yet, neither when comparing the e^+e^- results with *ep* results, nor when comparing results for different observables or experiments based on data with the same initial state. In the case of electron-proton deep-inelastic

collisions, the z_{hem} observable suggest a harder fragmentation than what is found when studying the z_{jet} and $z_{\mathrm{hem}}^{(\mathrm{jet})}$ distributions, the two latter being consistent with each other. In the case of e^+e^- annihilation, the fragmentation parameters extracted from BELLE and ALEPH data are not consistent; this discrepancy has been observed also by other authors [81]. The ALEPH data prefer a softer fragmentation than what is obtained from the BELLE data.

The fact that the results are consistent for different parton evolution models (DGLAP in RAPGAP, CCFM in CASCADE) and also for different running periods (HERA II analyzed here, HERA I analyzed in [80]) suggests that the discrepancies in results might originate from the description of the fragmentation itself. It seems that the Lund String model with the Peterson or Kartvelishvili parametrization of the fragmentation function is unable to describe consistently both, the physics near the charm production threshold (sensitiveness of the z_{hem} observable) and processes away from the threshold (z_{jet} and $z_{\mathrm{hem}}^{(\mathrm{jet})}$ - requiring a jet with $E_T > 3$ GeV), since a dependence of the extracted fragmentation parameter on the charm production energy is observed. The result deduced from the ALEPH data, the data with the highest CMS energy we have investigated, fits well into this picture and predicts the softest fragmentation function.

It is difficult to unambiguously address the question of fragmentation universality when inconsistencies are observed within the data having the same particles in the initial state. However, if we restrict our attention to the BELLE result only and the results obtained via the z_{jet} and $z_{\mathrm{hem}}^{(\mathrm{jet})}$ methods, where the charm production energies are comparable (~ 10 GeV), then we observe agreement. This suggests that in this domain fragmentation universality is, in the context of the tested model, observed.

The present knowledge of charm fragmentation and its universality can be improved. Concerning the H1 data, some room for further improvements and fine-tuning remains especially with respect to systematic errors, but since the data taking at HERA is over, the presented results are expected to be final or near-to-final. The latter is true only on condition that further efforts will be invested into the study of charm fragmentation in H1. Possible future steps could be the usage of the FTT simulation in the Monte Carlo programs, the use of improved track reconstruction methods and detector calibrations. One could also produce a fully inclusive Monte Carlo sample in order to improve the understanding of the D^* signal extraction. And exploit new Monte Carlo generators with NLO matrix elements and parton showers that should become available soon.

From the theory point of view, further efforts could be made in the investigation of the validity of the factorization theorem, fragmentation universality and in the study of parametrizations of fragmentation functions or, eventually, in the study of new fragmentation models.

Appendix A

Trigger Efficiencies in $z_{\text{hem}}^{(\text{jet})}$ and Additional Control Plots

In this appendix the track trigger efficiencies as a function of $z_{\text{hem}}^{(\text{jet})}$ as well as additional control plots are presented which have not been show previously. The control plots are related to the event variables, detector quantities and D^* observables. Like in section 5.4, the distributions are area-normalized, and the data are compared to the two MC models - RAPGAP and CASCADE. The data shown correspond to all of HERA II data.

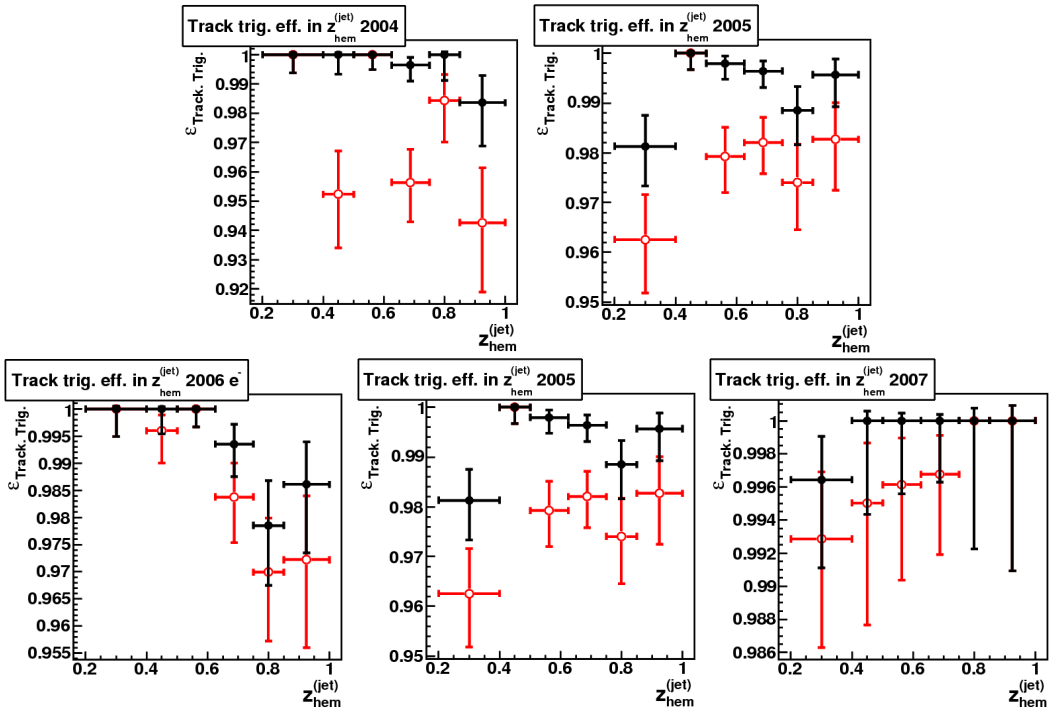


Figure A.1: Track trigger efficiencies in $z_{\text{hem}}^{(\text{jet})}$ for different run periods. The open circles (red) correspond to the status before the reweighting, the full markers (black) to the situation after the reweighting was applied.

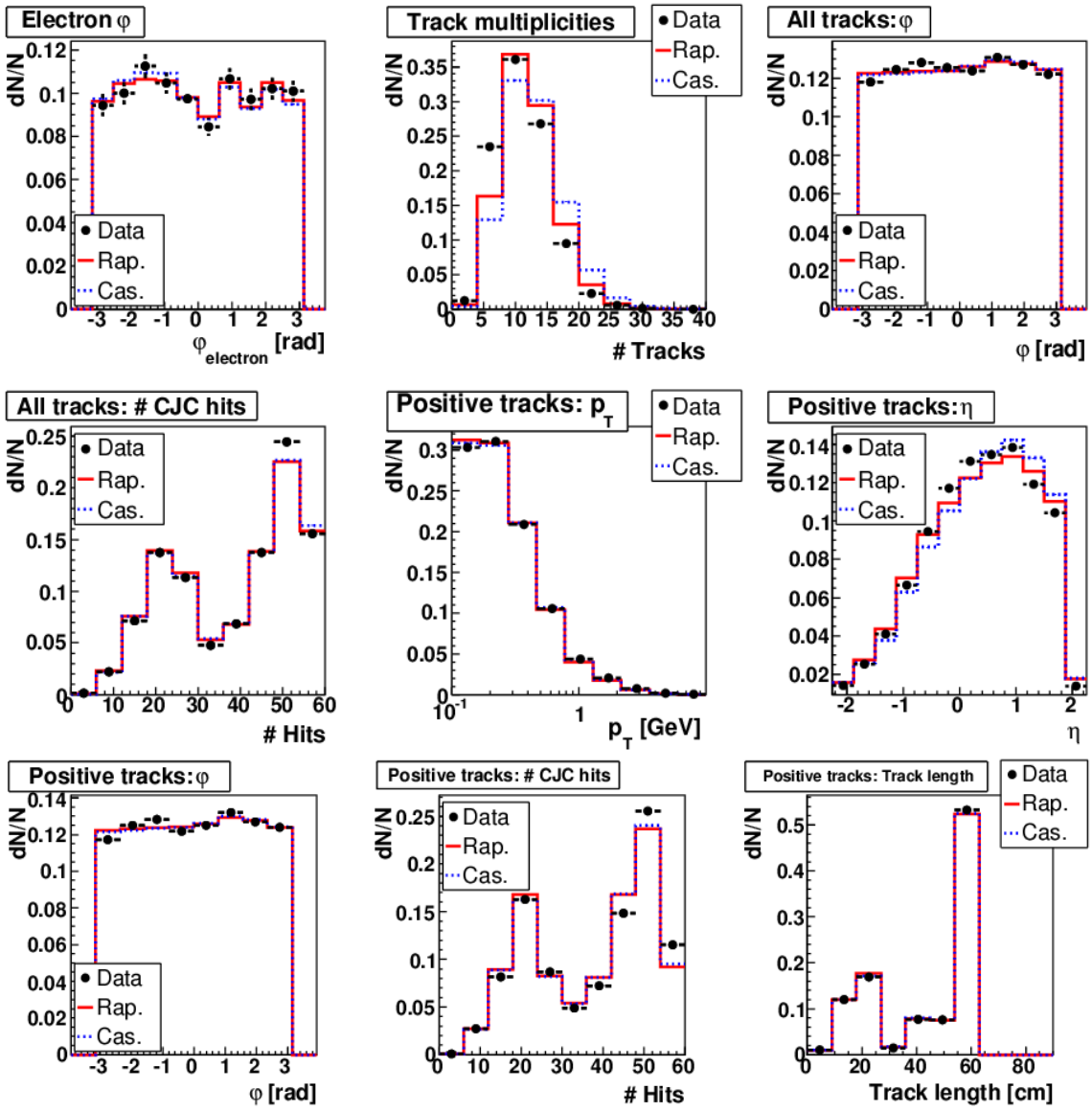


Figure A.2: Two event quantities (electron azimuthal angle, track multiplicity) and some track-related quantities for all "selected" tracks and for tracks with positive electric charge.

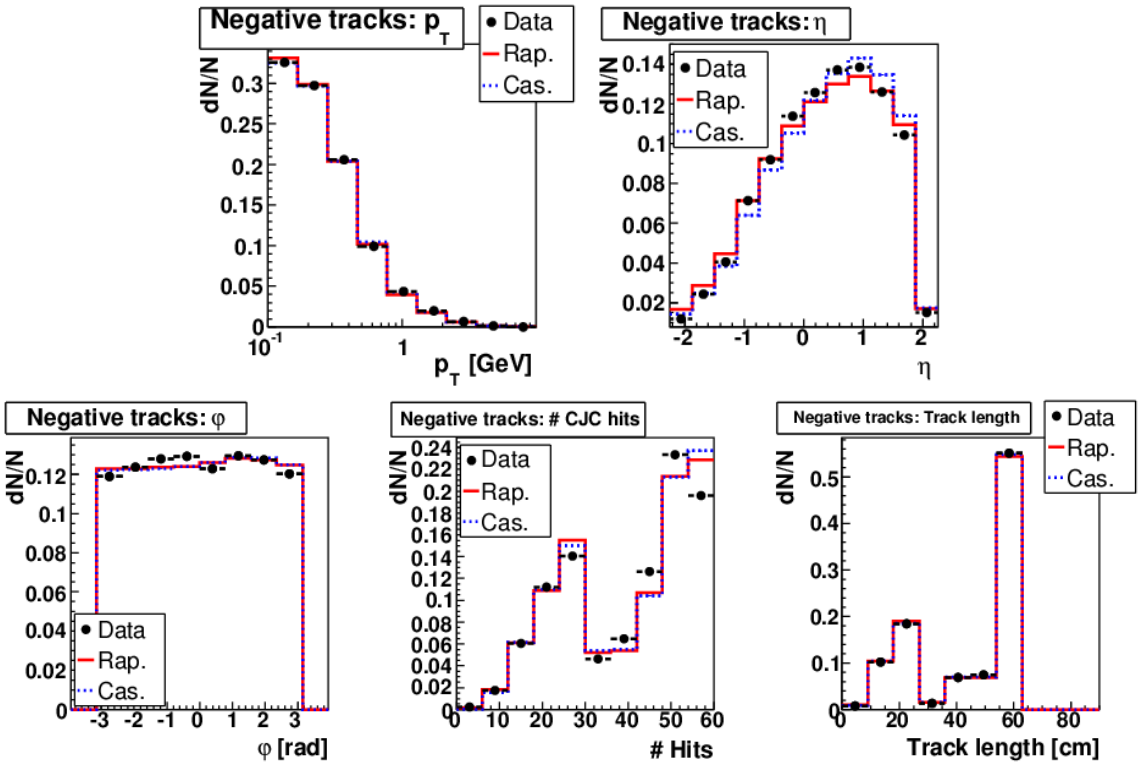


Figure A.3: Track-related quantities for tracks with negative electric charge.

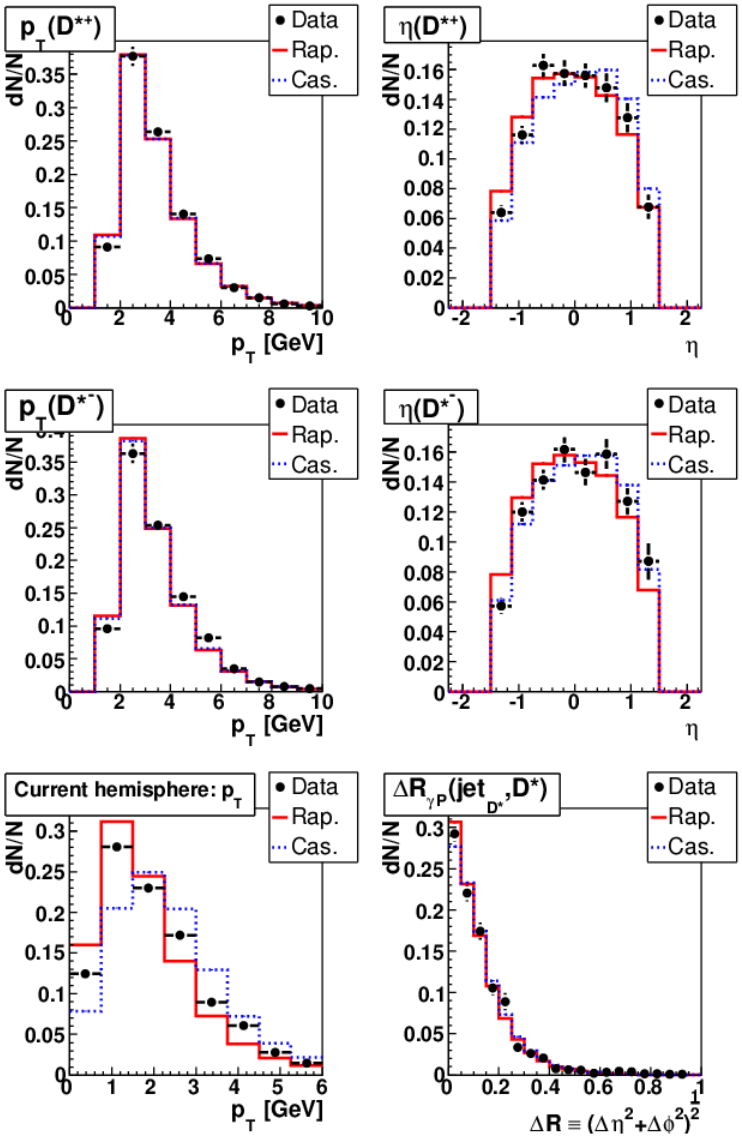


Figure A.4: D^*-related quantities: transverse momentum and pseudorapidity of negative and positive D^* particles, transverse momentum of the current hemisphere and $\triangle R$ of the D^* jet.

Appendix B

Plots of Systematic Errors

In addition to the numbers given in tables 6.1, 6.2 and 6.3, we provide here plots of correlated systematic errors. The vertical scales of the plots are fixed so that the uncertainties can be directly compared. The "Relative difference" refers to the relative difference with respect to the "default" distributions that have been obtained via unfolding based on a MC model without systematic shifts.

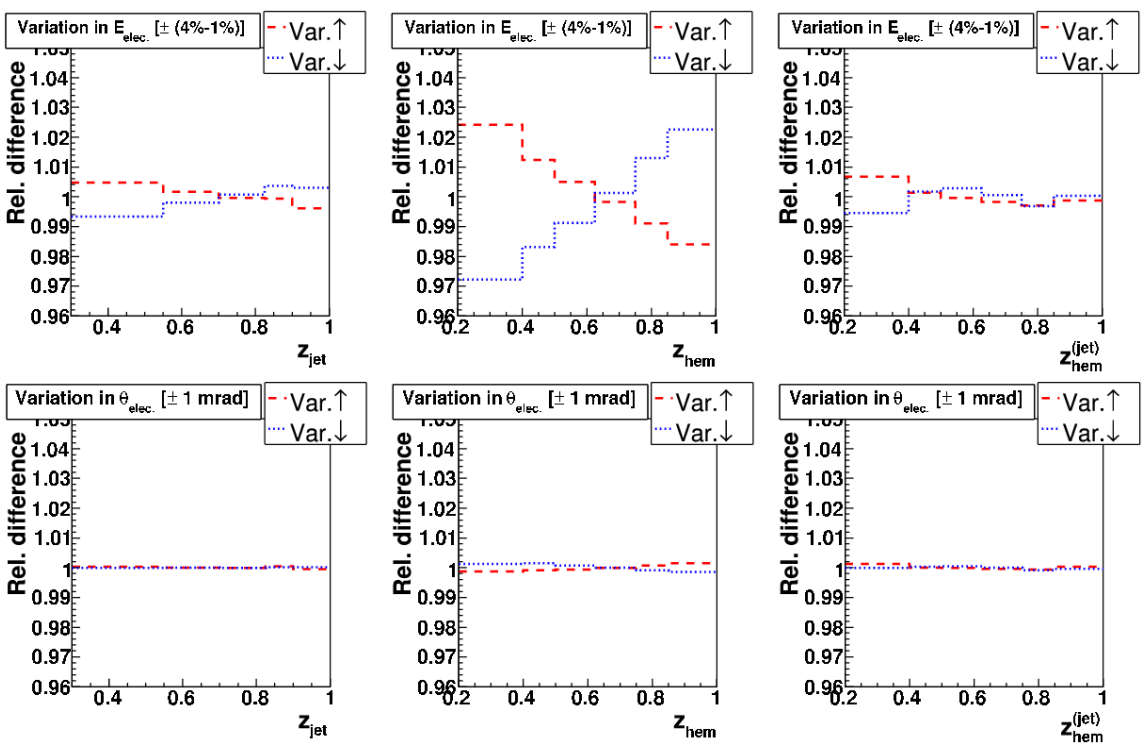


Figure B.1: Systematic errors related to the electron energy measurement and the measurement of the electron polar angle.

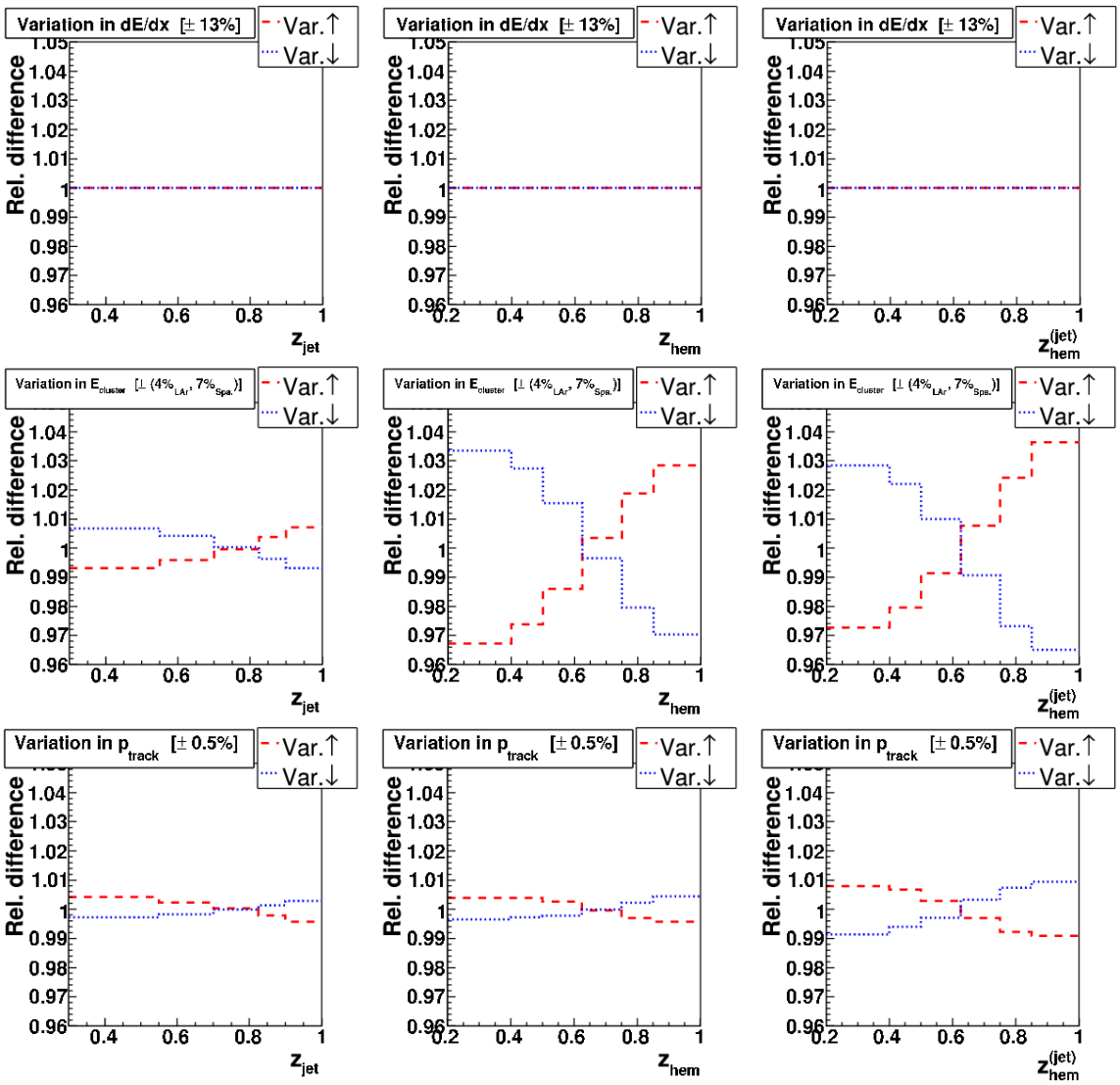


Figure B.2: Systematic errors related to the measurement of particle energy loss, energy of calorimetric clusters and track momenta.

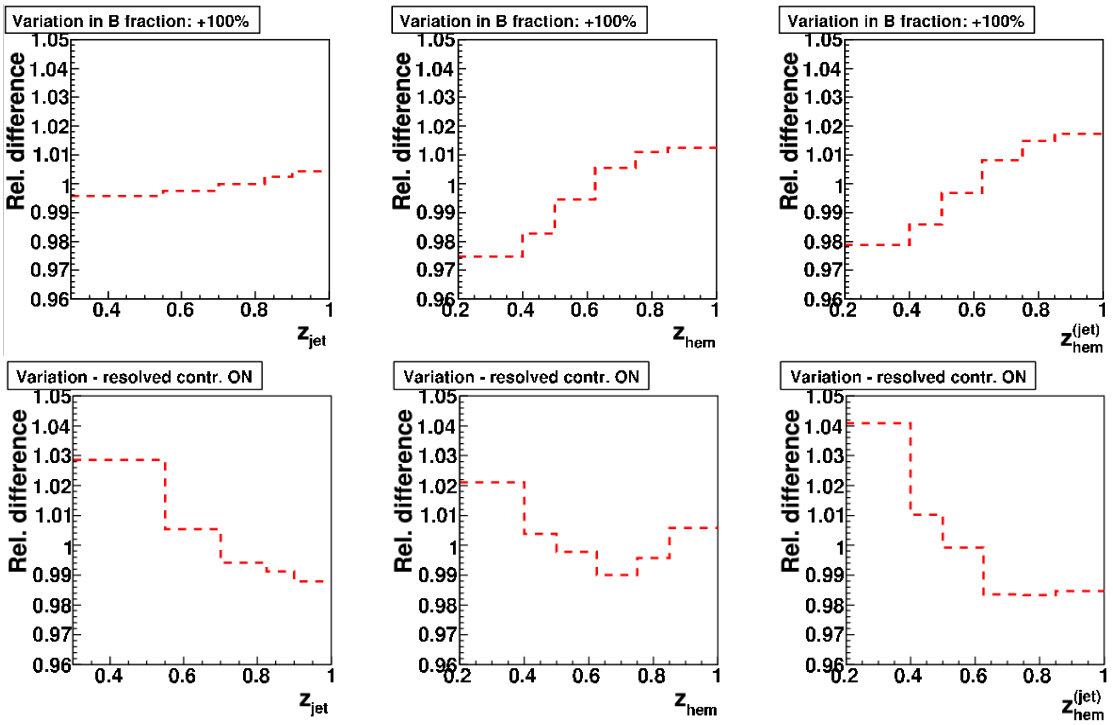


Figure B.3: Systematic errors related to the beauty component and to the contribution of resolved events.

Appendix C

χ^2 Fits and Parameter Extraction

The following figures show the (ε, χ^2) and (α, χ^2) plots used in the parameter extraction together with the resulting fits and the extracted parameter values. A parabola with three free parameters was chosen as fitting function; the fit determines the χ^2 minimum and the symmetric uncertainties. The fit was performed only to the five points with the lowest χ^2, fitting a bigger range in many cases leads to a biased determination of the χ^2 minimum.

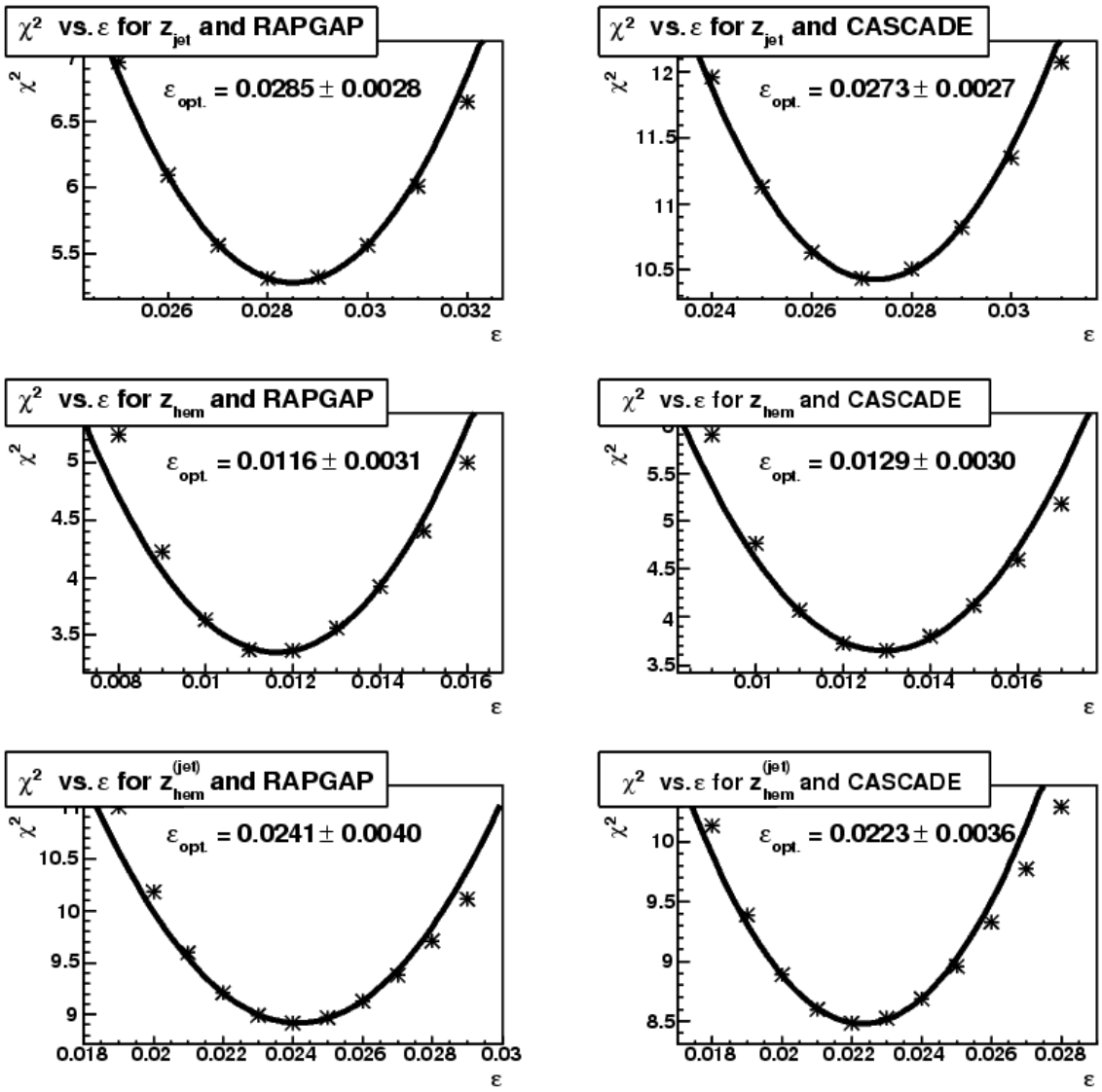


Figure C.1: χ^2 plots and corresponding parabola fits used in the extraction of the Peterson fragmentation parameters for different MC models and observables.

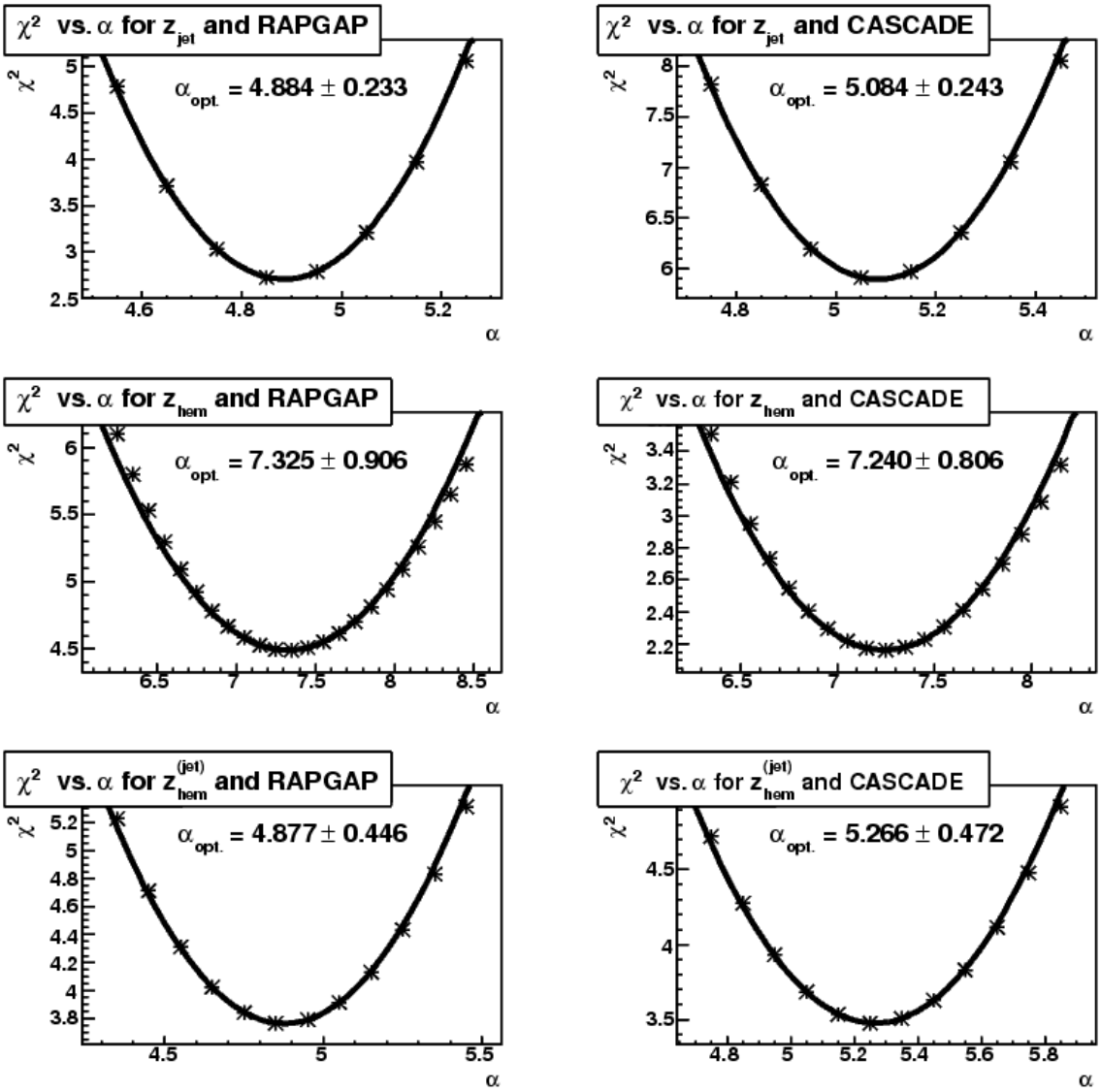


Figure C.2: χ^2 plots and corresponding parabola fits used in the extraction of the Kartvelishvili fragmentation parameters for different MC models and observables.

Bibliography

[1] M. K. Gaillard, P. D. Grannis, F. J. Sciulli, Rev. Mod. Phys. (1999) 71.

[2] S. Eidelman et al., Phys. Lett. B529 (2004) 1.

[3] W. Hollik, J. Phys. G: Nucl. Part. Phys. 23 (1997) 1503-1537.

[4] G. Altarelli, Annu. Rev. Nucl. Part. Sci. 1989. 39: 357-406.

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

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

[7] V. N. Gribov, L. N. Lipatov, Sov. J. Nucl. Phys. 15 (1972) 438 and 675.

[8] L. N. Lipatov, Sov. J. Nucl. Phys. 20 (1975) 94.

[9] G. Altarelli, G. Parisi, Nucl. Phys. B126 (1977) 298.

[10] Y. L. Dokshitzer, Sov. Phys. JETP 46 (1977) 641.

[11] E. A. Kuraev, L. N. Lipatov, V. S. Fadin, Sov. Phys. JETP 45 (1977) 199.

[12] I. I. Balitsky, L. N. Lipatov, Sov. J. Nucl. Phys. 28 (1978) 822.

[13] M. Ciafaloni, Nucl. Phys. B296 (1988) 49.

[14] S. Catani, F. Fiorani, G. Marchesini, Phys. Lett. B234 (1990) 339.

[15] S. Catani, F. Fiorani, G. Marchesini, Nucl. Phys. B336 (1990) 18.

[16] G. Marchesini, Nucl. Phys. B445 (1995) 45.

[17] J. Baines et al., Summary report of the heavy-flavour working group of the HERA-LHC Workshop, hep-ph/0601164.

[18] B. Mele and P. Nason, Nucl. Phys. B361 (1991) 626.

[19] J. D. Bjorken, Phys. Rev. D17 (1978) 171.

[20] M. Suzuki, Phys. Lett. B71 (1977) 139.

[21] C. Peterson, D. Schlatter, I. Schmitt and P. M. Zerwas, Phys. Rev. D27 (1983) 105.

[22] V. G. Kartvelishvili, A. K. Likhoded and V. A. Petrov, Phys. Lett. B78 (1978) 615.

[23] J. Kuti and V. Weisskopf, Phys. Rev. D4 (1971) 3418.

[24] R. P. Feynman and R. D. Field, Nucl. Phys. B136 (1978) 1.

[25] X. Artru and G. Mennessier, Nucl. Phys. B70, (1974) 93.

[26] B. Andersson, G. Gustafson, G. Ingelman, T. Sjöstrand, Phys. Reports 97 (1983) 31.

[27] G. Marchesini et al., Comp. Phys. Comm. 67 (1992) 465.

[28] G. Corcella et al., JHEP 0101, 010 (2001).

[29] D. Amati and G. Veneziano, Phys. Lett. B83, (1979) 87.

[30] R. Brun et al., CERN-DD/EE-84-1 (1997).

[31] H. Jung, Comput. Phys. Comm. 86 (1995) 147.

[32] G. Ingelman, A. Edin, J. Rathsman, Comp. Phys. Comm. 101 (1997) 108.

[33] T. Sjöstrand et al., Comp. Phys. Comm. 135 (2001) 238.

[34] A. Kwiatkowski, H. Spiesberger and H.-J. Möhring, Comput. Phys. Commun. 69 (1992) 155.

[35] H. L. Lai et al., Eur. Phys. J. C 12 (2000) 375.

[36] G. Schuler, T. Sjöstrand, Phys. Lett. B 376 (1996) 193.

[37] H. Jung, Comp. Phys. Comm. 143 (2002) 100.

[38] J. Wagner, Charm and Beauty production at HERA with D^*-Muon Events, Ph.D. thesis, Universität Hamburg (2004).

[39] H. Jung, Unintegrated Parton Density Functions in CCFM, in XII International Workshop on Deep Inelastic Scattering (DIS 2004), eds. D. Bruncko, J. Ferencei, and P. Stríženec, Vol. 1 (2004) 299, hep-ph/0411287.

[40] I. Abt et al. (H1 Collaboration), Nucl. Instrum. Meth. A 386 (1997) 310.

[41] I. Abt et al. (H1 Collaboration), Nucl. Instrum. Meth. A 386 (1997) 348.

[42] M. Cuje et al., H1 high luminosity upgrade 2000 - CIP and Level 1 Vertex Trigger, DESY-PRC 98/2 and H1 internal note H1-01/98-535 (1998).

[43] J. Becker et al., Nucl. Instrum. Meth. A 586 (2008) 190.

[44] D. Pitzl et al., Nucl. Instrum. Meth. A 454 (2000) 334, hep-ex/0002044.

[45] B. List et al., Nucl. Instrum. Meth. A 549 (2005) 33.

[46] B. Andrieu et al. (H1 Calorimeter Group), Nucl. Instrum. Meth. A 336, (1993) 460.

[47] M. zur Nedden, B. Reisert and T. Schörner, H1 Liquid Argon Trigger: Overview, Simulation and Performance, H1 internal note H1-04/01-592 (2001).

[48] R.-D. Appuhn et al. (H1 Spacal Group), Nucl. Instrum. Meth. A 386 (1997) 397.

[49] T. Nicholls et al. (H1 Spacal Group), Nucl. Instrum. Meth. A 374 (1996) 149.

[50] R.-D. Appuhn et al. (H1 Spacal Group), Nucl. Instrum. Meth. A 382, (1996) 395.

[51] M. O. Bönig, Messung des D^*-Meson-Produktionsquerschnitts in tiefinelastischer Streuung mit dem H1-Experiment, Ph.D. thesis, Universität Dortmund (2007).

[52] B. List, Tracking at H1, presentation on DESY IT Seminar, 30.10.2006.

[53] C. Kleinwort, private communication.

[54] J. Kretzschmar, A Precision Measurement of the Proton Structure Function F2 with the H1 Experiment, Ph.D. thesis, Humboldt-Universität zu Berlin (2007).

[55] A. A. Glazov, Measurement of the Proton Structure Function F2 and FL with the H1 Detector at HERA. Ph.D. thesis, Humboldt-Universität zu Berlin (1998).

[56] M. Peez, B. Portheault and E. Sauvant, H1 internal note, H1-01/05-616 (2005).

[57] QHQTRK manual, (Heavy Flavor Working Group track selection code by Lee West).

[58] J. Cao and Z. Zhang, H1 internal note, H1-12/99-580 (1999).

[59] S. Catani, Y. L. Dokshitzer, M. H. Seymour and B. R. Webber, Nucl. Phys. B406 (1993) 187.

[60] S. D. Ellis, D. E. Soper, Phys. Rev. D48 (1993) 3160.

[61] M. Cacciari and G.P. Salam, Phys. Rev. D4 (1971) 3418.

[62] G. J. Feldman et al., Phys. Rev. Lett. 38 (1977) 1313.

[63] R. Barate et al., Eur. Phys. J. C 16 (2000) 597.

[64] M. Artuso et al., Phys. Rev. D 70, 112001 (2004).

[65] R. Seuster et al., Phys. Rev. D 73, 032002 (2006).

[66] Z. Rúriková, Measurement of Charm Fragmentation in DIS at HERA, Ph.D. thesis, Universität Hamburg (2005), MPP-2006-10.

[67] A. Höcker and V. Kartvelishvili, Nucl. Instrum. Meth. A 372 (1996) 469, hep-ph/9509307.

[68] G. D'Agostini, Nucl. Instrum. Meth. A 362, (1995) 487.

[69] S. Bentvelsen, J. Engelen and P. Kooijman, Reconstruction of (x, Q^2) and extraction of structure functions in neutral current scattering at HERA, DESY HERA Workshop 1991:0023-42.

[70] F. Moreau (representing the H1 Collaboration), Nucl. Phys. B (Proc. Suppl.) 61B (1988) 132.

[71] J. Riedlberger, The H1 Trigger with Emphasis on Tracking Triggers, H1 internal note , H1-01/95-419 (1995).

[72] A. Nikiforov, Measurements of the Neutral Current $e^{\pm}p$ Cross Sections Using Longitudinally Polarised Lepton Beams at HERA II, Ph.D. thesis, Ludwig-Maximilians-Universität München (2007).

[73] S. Schmidt, Messung charminduzierter Zweijetereignisse in tief inelastischer ep-Streuung mit dem H1-detektor bei HERA, Ph.D. thesis, Technischen Universität München (2004), MPP-2004-140.

[74] O.Behrendt, A.Vargas, S.Glazov, M.Klein, V.Lenderman, A.Petrukhin (representing H1 Collaboration), Measurement of the Inclusive ep Scattering Cross Section at low Q2 and x at HERA, Paper draft to be submitted to EPJ C or PRD (2008).

[75] S. Fang (ZEUS Collaboration), Charm Fragmentation Function and Fragmentation Fractions of Charm Mesons at ZEUS, in XV International Workshop on Deep Inelastic Scattering (DIS 2007), eds. G. Grindhammer and K. Sachs, Vol. 2 (2007) 817.

[76] Alexandre Glazov, private communication.

[77] R. Kogler, S. Osman, Systematic Error Determination for HFS in Analyses Using Hadroo II, talk given on H1 HaQ Meeting on 05/02/2008, DESY (Hamburg), Germany.

[78] Z. Rúriková, private communication.

[79] T. Lübbert, Study of charm fragmentation in e^+e^- annihilation and ep scattering, Summer Student Report, DESY, 2007.

[80] F. D. Aaron et al. (H1 Collaboration), Study of Charm Fragmentation into D^* Mesons in Deep-Inelastic Scattering at HERA (2008), To be published in Eur. Phys. J. C.

[81] M. Cacciari, P. Nason and C. Oleari, JHEP 0604:006 (2006), hep-ph/0510032v1.

Acknowledgements

The completion of this work would not be possible without help of many people. I am especially grateful to

- Dr. Günter Grindhammer for being my supervisor, for discussions, his time, help, suggestions and kindness.
- Zuzana Rúriková for her help, fruitful discussions, innumerable advices and patience.
- Juraj Braciník for bringing me to experimental physics and for discussions.
- Adil Aktas for help with programming and computers.
- My colleagues and fellows from the H1 collaboration for discussions and help; specifically those from the heavy flavour group.
- My former teachers and professors in Slovakia for having taught me the physics and for their support; specifically to RNDr. Stanislav Dubnička DrSc., doc. RNDr. Martin Mojžiš PhD, doc. RNDr. Vladimír Černý CSc. and Prof. RNDr. Ján Pišút DrSc..
- My parents, my brother and all my relatives for their support.
- All of my friends from Hamburg and Slovakia for relax I had with them during this period.
- All people that helped me, supported me and made my work possible.

Thank you!