UNIVERSITÄT BONN
Physikalisches Institut

Measurement of beauty and charm cross sections in photoproduction using decays into electrons with ZEUS at HERA

by

Markus Jüngst

Photoproduction of heavy quarks in events with two jets and an electron associated with one of the jets has been studied with the ZEUS detector using the data recorded in the years 1996–2000 and 2006–2007, corresponding to an integrated luminosity of $\mathcal{L} = 120\,\mathrm{pb}^{-1}$ and $\mathcal{L} = 190\,\mathrm{pb}^{-1}$, respectively. The fractions of events containing b quarks, and also of events containing c quarks, were extracted from a likelihood fit using variables sensitive to electron identification as well as to semileptonic decays. For the second data-set the list of variables sensitive to the flavour could be extended using the additional information from the microvertexdetector. Total and differential cross sections for beauty and charm production were measured for both data-sets and compared with next-to-leading-order QCD calculations and Monte Carlo models.

Postal address:
Nußallee 12
D-53115 Bonn
Germany

BONN-IR-2010-03
Bonn University
February 2010
ISSN-0172-8741

UNIVERSITÄT BONN
Physikalisches Institut

Measurement of beauty and charm cross sections in photoproduction using decays into electrons with ZEUS at HERA

by
Markus Jüngst

Dieser Forschungsbericht wurde als Dissertation von der Mathematisch-Naturwissenschaftlichen Fakultät der Universität Bonn angenommen und ist auf dem Hochschulschriftenserver der ULB Bonn http://hss.ulb.uni-bonn.de/diss_online elektronisch publiziert.

Referent:	Prof. Ian C. Brock
Korreferent:	Dr. Jürgen Kroseberg

Angenommen am:	01.12.2009
Tag der Promotion:	04.02.2010

Contents

Chapter 1

Introduction

The first *elementary particle* was the *electron*, discovered in 1899 [1]. The properties were given by its mass and charge, which is one negative *elementary charge*, defined by the charge of the *proton*. The third property was defined by the *spin* of the electron which was introduced in the framework of quantum mechanics developed in the 1920*s*.

One of the most common ways to probe the structure of compound objects is to perform scattering experiments. The first of those experiments was performed by Rutherford in the year 1909, who shot α particles onto a thin gold foil. The Rutherford experiment contradicted the simple atom models, where e.g. J.J. Thomson described the electrons like plums in pudding. Rutherford described the result of the scattering experiment "as if you fired a 15-inch shell at a piece of paper tissue and it came back and hit you" [2]. As additional result to the atomic mass distribution he postulated a heavy neutral particle that would be necessary to bind the positively charged protons in the nuclei. This particle, called the *neutron*, was finally found in 1930.

Until 1935 the atomic picture was simple, with the proton and the neutron forming the atomic nucleus, and the electron cloud surrounding the nucleus, with the *photon* as the quantum for the electromagnetic radiation described by the *Quantum ElectroDynamics* (QED). At the end of the 1930s the *mesotron* and the *neutrino* entered the so-called particle zoo. The particle, which was needed for the exchange of nuclear forces, was later identified with the *pion*. The neutrino, which Wolfgang Pauli called "the story of this foolish child of my life crisis" [3], was postulated from the kinematics of radioactive decays. The neutrino was needed for reasons of spin and energy conservation, in order to explain a continuous energy spectrum of the emitted electron.

Until the middle of the twentieth century the theoretical predictions had the leading role in the progress of elementary particle physics. At this time several new particles were found and the advance become more initiated by new experimental results. Several models were developed to explain the connection between the long list of "elementary" particles. The model of *nuclear democracy* was found to be insufficient when the V particles, which were called *strange particles*, were found in 1947 [4]. Some years later, in 1949, Fermi and Yang speculated that the newly discovered particles could be compound objects from more elementary particles [5].

The theoretical models were mainly driven by Gell-Mann [6] and Zweig [7]. Gell-Mann also proposed the name *quarks*, inspired by James Joyce's *Finnegan Wake*, for these more fundamental particles. The model was developed into a consistent theory, *Quantum ChromoDynamics* (QCD), which is the field theory of quarks and *gluons* and their interaction which is the pendant to the QED for the strong interaction. The quarks, which have an additional degree of freedom called *colour*, are not observed as free particles, but are bound with an anti-quark into *mesons* or with two other quarks into *baryons*. The strong force, which is mediated by the gluons, linearly increases with the distance between two quarks and binds the quarks into these observable objects.

In the second half of the twentieth century, different quark and lepton types were discovered, completing the *Standard Model* (SM) [8, 9, 10]. This model consists of 12 spin-1/2 fermions (6 quarks and 6 leptons) which are summarised together with the interactions in Table 1.1, and 4 spin-1 gauge bosons describing three of the four known fundamental interactions. For each of the fermion exists an *antiparticle* with opposite charge. (A detailed description can be found e.g. in [11].)

	Generations			Q_{EM}	Interactions el. magn.	weak	strong
Quarks	$\begin{pmatrix} u \\ d \end{pmatrix}$	$\begin{pmatrix} c \\ s \end{pmatrix}$	$\begin{pmatrix} t \\ b \end{pmatrix}$	$+\frac{2}{3}$	✓	✓	✓
				$-\frac{1}{3}$	✓	✓	✓
Leptons	$\begin{pmatrix} \nu_e \\ e \end{pmatrix}$	$\begin{pmatrix} \nu_\mu \\ \mu \end{pmatrix}$	$\begin{pmatrix} \nu_\tau \\ \tau \end{pmatrix}$	0	–	✓	–
				-1	✓	✓	–

Table 1.1: *A list of the 6 quarks and leptons in the Standard Model with their corresponding electromagnetic charge and the interactions. The corresponding antiparticles (not shown) have the same masses and spins, but have opposite charges.*

The underlying symmetry for the SM is given by:

$$SU(3)_C \times SU(2)_{I_W} \times U(1)_Y, \tag{1.1}$$

where the first term corresponds to QCD with the colour charge, C, as the conserved quantum number. For the *electro-weak theory*, which is the unification of the QED and the weak interaction, quantum numbers are the *weak isospin*, I_W, and the *hypercharge*, Y. Four gauge bosons ($\gamma, W^{\pm}$ and Z) mediate the forces in this symmetry group. In contrast to the γ and the gluons the $W^{\pm}$ and the Z were found to have mass. An additional scalar field had to be introduced to rescue the gauge invariance via spontaneous symmetry breaking. A corresponding particle to this predicted *Higgs field*, a spin-0 Higgs boson, has not been observed yet.

To probe the inner structure of the proton whose dynamics are driven by the strong interaction, scattering experiments are required. Energies are needed that are sufficiently high to be sensitive to the dimensions of the protons. At the electron[1]-microscope HERA the scattering of electrons on the protons can be described as the interaction of a pointlike particle, which emits a virtual photon, providing the possibility to "look" into the proton. The advantage of using electrons is that the purely electromagnetic part of the interaction is well understood.

The production of the heavy quarks beauty (b) and charm (c) is a very interesting topic of high energy physics because it provides a testing ground for perturbative Quantum Chromodynamics (pQCD) [12]. For heavy quarks, produced with transverse momenta comparable or larger than the quark masses, theoretical calculations are expected to provide reliable predictions. At the ZEUS experiment, heavy quark cross sections have been measured using different approaches with complementary systematic uncertainties. Several analyses of charm used the reconstruction of $D^{*\pm}$ mesons, where for beauty and charm the identification of leptons from semileptonic decays (e.g. $D^0 \rightarrow \pi^- e^+ \nu_e$, $B^- \rightarrow D^0 e^- \bar{\nu}_e$) in events with jets was used.

The advantage of looking for the semileptonic decays is that the branching ratio is $\mathcal{B}(c(b) \rightarrow eX) \sim 0.1$ which is much larger than the more exclusive channel of e.g $\mathcal{B}(c \rightarrow D^* \rightarrow K\pi\pi) \sim 0.01$ [13]. In most of the analyses carried out so far semileptonic decay muons were used, as they are relatively easy to identify, whereas the sample of semileptonic decays into electrons suffer from substantial background from electrons that were produced in competing reactions, e.g. from photoconversions or Dalitz decays. The decay channels to electrons are interesting to complement the muon analyses in terms of the kinematic range and the systematic uncertainties.

[1] Hereafter unless explicitly stated both electrons and positrons are referred to as electrons.

In my diploma thesis I developed a likelihood method which allowed the identification of electrons from semileptonic decays, using the data-set from 1996-2000, corresponding to a luminosity of $120\,\mathrm{pb}^{-1}$ [14, 15]. The production of electrons from semileptonic decays (e_{SL}) was measured in events with at least two jets (jj) in photoproduction, $ep \to e\,bb\bar{b}\,X \to e\,jj\,e_{\mathrm{SL}}\,X'$, in the kinematic range $Q^2 < 1\,\mathrm{GeV}^2$ and $140\,\mathrm{GeV} < W_{\gamma p} < 280\,\mathrm{GeV}$, where Q^2 is the negative squared four-momentum exchanged between the electron and the proton and $W_{\gamma p}$ is the centre-of-mass energy of the γp system. During the first half of my PhD studies, this analysis, which was mainly focused on the beauty production, was finalised and published [16]. In the second half, I extended the analysis to the 2006-2007 ($\mathcal{L} = 195\,\mathrm{pb}^{-1}$) data-set where it was possible to benefit from detector improvements and newly available techniques of the decay identification. It was possible to reduce the uncertainties on both beauty and even more on the charm cross sections. The analysis of $ep \to e\,c\bar{c}\,X \to e\,jj\,e_{\mathrm{SL}}\,X'$, which was more of a by-product in the first analysis, is on a par with the beauty measurement.

The first half of my PhD time, I was living in Hamburg which gave me the possibility to work on the data analysis together closely with other members of the ZEUS collaboration, especially with the colleagues in the Heavy Flavour Physics Group. Beyond the analysis focused studies, I was also involved in the processes of data taking in a direct way, by taking shifts as well as by being responsible for the data quality validation of one detector component. When the HERA I analysis was finished, I returned to Bonn where I took more responsibility in the group by supervising and supporting other students. In this context, I am also closely involved in the analogue analysis in the kinematic region of deep inelastic scattering, for which the method of electron identification has been adopted. When the data taking was finished, I took over a more technical common task for the ZEUS collaboration, the ZeVis[2] coordination.

The outline of this thesis is as follows: After an introduction to heavy quark production in Chapter 2, the most important components of the ZEUS detector are summarised in Chapter 3. The reconstruction of kinematic variables and the selection procedure are described in Chapters 4 and 5. In the main part of the thesis, the method of signal identification (Chapter 6) and studies of the likelihood performance are shown (Chapter 7). After a short description of the fitting method in Chapter 8 systematic studies and checks are summarised in Chapter 9. In Chapter 10 total and differential cross sections are given and compared with leading order Monte Carlo and next-to-leading order QCD predictions. Finally, conclusions from the measurements are drawn in Chapter 11.

[2] ZeVis is the event display at ZEUS.

Chapter 2

Heavy Quark Production at HERA

This chapter briefly summarises the theoretical description of heavy flavour production at HERA. At the beginning a short definition of the kinematics in electron proton interaction is given. Model considerations motivate the way of implementation in event generators and the interpretation of the measured results comparing with theoretical calculations. At the end the decay channel used in this analysis is further specified.

2.1 Kinematics of Lepton Nucleon Interactions

In the SM the fundamental process in ep scattering is mediated either by the electromagnetic or the weak force through the exchange of either a photon in the first case, or a Z or a $W^{\pm}$ in the second case. The processes are further classified by the charge of the final state lepton. If a photon or a Z is exchanged, with a scattered electron or positron in the final state, the process is called *Neutral Current* (NC):

$$e^{\pm} + p \rightarrow e^{\pm} + X, \tag{2.1}$$

where X denotes the hadronic final state in the inelastic interaction. In the other type of process, *Charged Current* (CC), a $W^{\pm}$ is exchanged and the final state lepton is a neutrino (antineutrino):

$$e^{+}(e^{-}) + p \rightarrow \bar{\nu}(\nu) + X. \tag{2.2}$$

Both processes are displayed schematically in Figure 2.1, indicating the four-momenta of the initial-state and final-state particles.

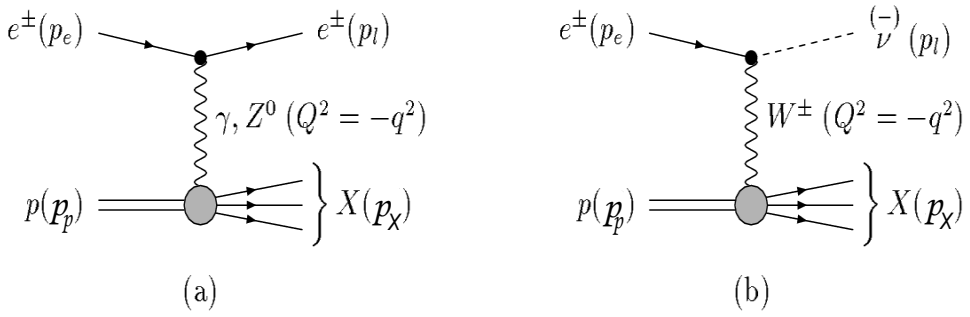


Figure 2.1: *Electron-proton scattering in Neutral Current (NC) (a) and Charged Current (CC) (b) processes. The hadronic final state is denoted by X. The picture, with slightly modified labels, has been taken from [17].*

In the following, the variables are listed that describe the kinematics of the NC process. They are calculated from the four-momenta of the initial-state and final-state particles. The centre-of-mass energy squared, s, is given by the energies of the two colliding particles:

$$s = (p_e + p_p)^2 = m_e^2 + m_p^2 + 2(E_e E_p - \vec{p_e} \cdot \vec{p_p}) \simeq 4E_e E_p\,; \tag{2.3}$$

The available resolution to probe the proton with the electron is directly connected to the photon virtuality, Q^2, which is the negative square of the four-momentum transfer,

$$Q^2 = -q^2 = -(p_e + p_l)^2\,. \tag{2.4}$$

The variable q determines the resolution as the quantum wavelength is given by $\lambda = \frac{h}{|q|}$. As not the total energy of the beam particles is transferred in the inelastic scattering of the hard subprocess, important variables to classify the event kinematics have to reflect these momentum fractions. For the proton, the momentum fraction carried by the parton (see Chapter 2.2.2) interacting with the lepton is given by the Bjorken-x scaling variable,

$$x = \frac{Q^2}{2p_p \cdot q}\,. \tag{2.5}$$

The inelasticity, y, is the fraction of the lepton momentum transferred to the exchanged boson, calculated in the proton rest frame:

$$y = \frac{p_p \cdot q}{p_p \cdot p_e}\,. \tag{2.6}$$

These variables can be combined into the effective centre-of-mass energy, W, of the $\gamma(Z)$-p system:

$$W^2 = (q + p_p)^2 \sim sy - Q^2\,. \tag{2.7}$$

There is an additional important relation between the variables describing NC processes. For fixed centre-of-mass energy, $\sqrt{s}$, the event kinematics is completely determined by the knowledge of two of these variables. Typically the (x,Q^2) or (x,y) pairs are chosen, while the third variable is fixed by:

$$Q^2 = s \cdot x \cdot y\,; \tag{2.8}$$

The kinematic coverage of the HERA experiment is depicted in Fig 2.2 using the kinematic plane of (x, Q^2).

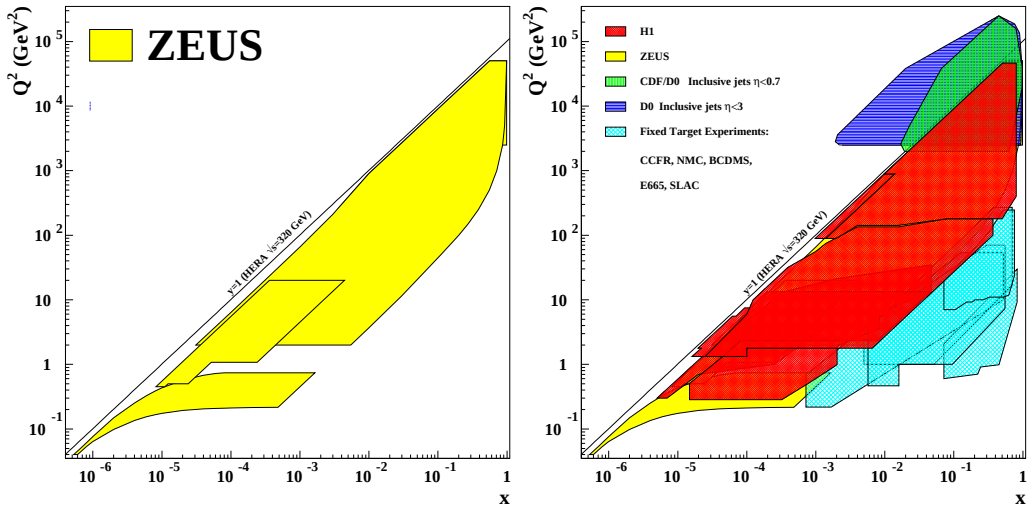


Figure 2.2: *(x, Q^2) plane covered by ZEUS (left) compared with H1, CDF and D0 and various fixed target experiments (right). The limiting line for $y = 1$ is given for the centre-of-mass energy of $\sqrt{s} = 320\,GeV$ [18].*

2.2 Model Considerations

2.2.1 The Quark-Parton Model

In the simple *parton model* [19], Feynman described the proton as a compound object of free point-like *partons*. In this picture, inelastic ep scattering can be seen as elastic scattering of the photon off one of these free partons, as depicted in Figure 2.3. Using the longitudinal momenta of the parton and the photon, ξp and q, the momentum fraction is given by:

$$\xi = \frac{Q^2}{2q \cdot p} = x. \tag{2.9}$$

This equation can be derived by using the momentum conservation and neglecting the proton and parton masses ($p'^2 = (\xi p+q)^2 = \xi^2 p^2+2\xi pq+q^2 \approx 2\xi pq-Q^2 \approx 0$).

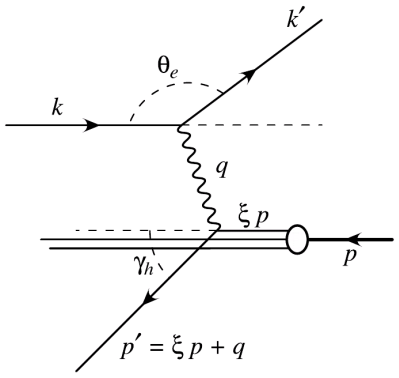


Figure 2.3: *Schematic representation of the deep inelastic scattering. The photon with momentum q interacts with the parton which carries a fraction, ξ, of the proton momentum p. The momentum of the outgoing parton is given by $p' = \xi p + q$.*

Therefore the Bjorken scaling variable, x, can be interpreted as the fraction of longitudinal proton momentum, ξ, carried by the parton in the scattering process. Bjorken predicted that, in the limit of high Q^2 and $\frac{pq}{M} \to \infty$, the structure functions of the proton, F_2 and F_1, which can be interpreted as the sum of parton densities, are *scale invariant* and depend only on one variable x:

$$F_2(x, Q^2) \;\; = F_2(x) \;\; = \sum_i e_i^2 x f_i(x), \tag{2.10}$$

$$F_1(x, Q^2) \;\; = F_1(x) \;\; = \frac{1}{2x} F_2(x), \tag{2.11}$$

where $f_i(x)$ is the parton distribution function and e_i denotes the charge of parton i. The second equation is known as the *Callan-Cross relation* [20], which is a consequence of the assumption to have pointlike charged particles with spin-1/2.

The parton model, with the identification of the partons as the quarks in the *Quark Model* [6], resulted in the naive *Quark-Parton Model* (QPM). A consequence of the QPM is that the proton momentum is a sum of the quark momenta. Measurements showed that only a part of the momentum is carried by the so-called *valence quarks*, and the other part is carried by the *sea quarks* and the *gluons*, which were discovered at DESY [21].

2.2.2 Quantum Chromodynamics

For the calculation and simulation of scattering processes, the production is split into two different parts according to the *factorisation theorem* [22, 12]. The hard process, which is the interaction of the partons, calculable in pQCD in terms of the strong coupling constant, α_s, and the formation of the hadrons in the final state, called hadronisation which cannot be calculated by pQCD. The cross section, including both the partonic process and the hadronisation is given by:

$$d\sigma(ep \to e'X) = \sum_{partons} \int_0^1 dx f_{i/p}(x, \mu_F^2) \cdot d\hat{\sigma}_i(\hat{s}, \alpha_s(\mu_R), \mu_R, \mu_F), \tag{2.12}$$

where $f_{i/p}(x, \mu_F^2)$ is the parton density for parton i in the proton, which depends on the scaling parameter μ_F, and $d\hat{\sigma}$ is the so-called reduced cross section for a given energy $\hat{s}$. The *factorisation scale*, μ_F, represents the border between soft and hard interactions. The cut-off variable, μ_R, is needed as a *renormalisation scale* to avoid divergences in the cross section calculations. Equation 2.12 has divergences for infinite momenta in the Feynman loop integrals (*ultraviolet divergence*), for gluon momenta close to zero (*infrared divergence*) and for radiations under vanishing angles (*collinear divergence*). The infrared divergences cancel out in the summation of real and virtual gluon graphs and the other two are finite after replacing the bare parameters by effective parameters, which depend on μ_R. In particular the strong coupling constant, α_s, transforms into a running coupling constant [23], which is given to first order by:

$$\alpha_s(\alpha_s^0, \mu_R) \quad = \quad \frac{12\pi}{(33 - 2n_f)\ \ln\left(\frac{\mu_R^2}{\Lambda_{QCD}^2}\right)}, \tag{2.13}$$

where n_f is the number of active flavours for a given scale. The value of the energy scale in QCD, Λ_{QCD}[1], represents the threshold at which α_s becomes large and pQCD is not valid anymore. The dependence of α_s on the renormalisation scale μ_R is shown in Fig. 2.4. The *asymptotic freedom* at large scales, and hence for small

[1] Typically, experimental values for Λ_{QCD} are ~200 MeV.

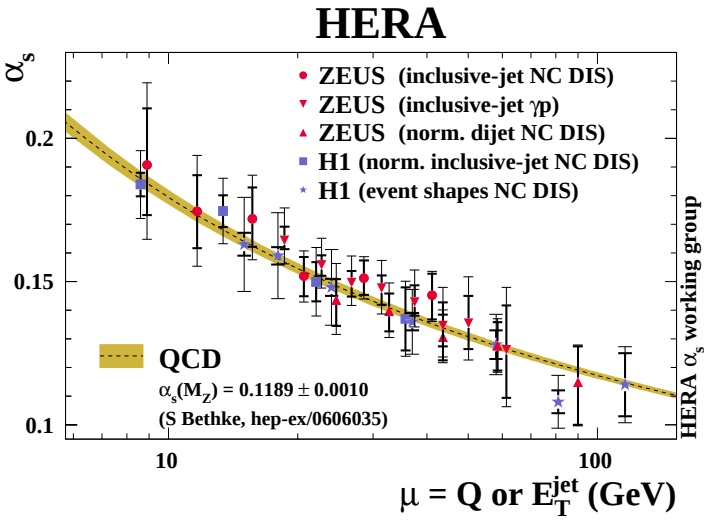


Figure 2.4: *Running of α_s with the photon virtuality or the jet energy as the scaling parameter μ. The data measured at HERA are compared with QCD predictions (shaded band) [24].*

distances, corresponds to the scattering off quasi-free partons. For long distances *colour confinement* explains the non-existence of free quarks. The heavy beauty and charm quark masses provide a hard scale resulting in an effective coupling constant of ∼0.2 and ∼0.4 respectively, making pQCD calculations applicable.

2.2.3 Evolution of Parton Densities

The parton density in Equation 2.12 is universal and depends only on the hadron type, i.e. for HERA on the proton. The experimental parton densities are determined for a given scale, μ_F, which can be provided by the heavy quark masses, the transverse energy or by large photon virtuality, Q^2. The method of extracting a complete set of quark ($q(x, Q^2)$) and gluon ($xg(x, Q^2)$) density functions requires a parametrisation which describes the x dependence at some cut-off value of $Q^2 = Q_0^2$. Here q denotes both the valence quarks and the sea-quarks which can be produced via g gluon-splitting. Typically, the parametrisation is chosen as follows [25]:

$$xf(x, Q_0^2) = A(1 - x)^\beta x^\alpha (1 + \epsilon\sqrt{x} + \gamma x), \tag{2.14}$$

with five free parameters[2]. In Figure 2.5 parton density functions are shown using the inclusive DIS cross sections ($Q^2 = 10\,\mathrm{GeV}^2$) from H1 and ZEUS [26].

[2]The fit has some additional constraint due to the valence quark and momentum sum rules.

The parton density fits show the contributions from the valence quarks, the sea quarks and the gluons. In the region of low x the proton dynamics are dominated by gluons and sea quarks. At $x \sim 0.1$, the valence quarks of the proton (uud) show the behaviour of the naive picture, where each of the three quarks contribute approximately to one third of the proton momentum.

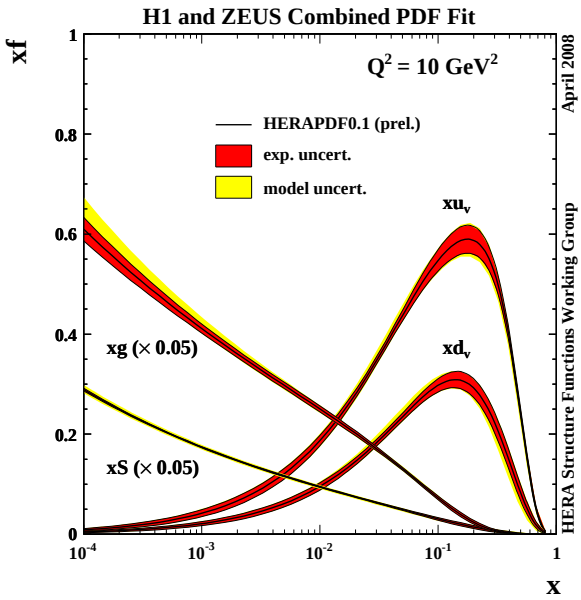


Figure 2.5: *Parton density functions of the valence quarks (u and d) [26], the gluons (g) and the sea (S) quark content of the proton. The curve shows the HERA I data, of combined H1 and ZEUS inclusive DIS cross sections as input for a next-to-leading order QCD parton distribution function fit. The shaded bands show the experimental uncertainty and the model uncertainties. Note that the contributions from the sea quarks and the gluons, which dominate the region of low x, are suppressed by a factor of 20.*

To calculate the densities with the PDFs from HERA measurements for other scale ranges, e.g. to be transformed to LHC energies, QCD *evolution equations* can be used. The *Dokshitzer-Gribov-Lipatov-Altarelli-Parisi* (DGLAP) equation [27, 28, 29, 30] describes the evolution of the quark momentum distribution in Q^2. In the extrapolation, the gluon emission, which is given in a parton ladder, is ordered in squared transverse momentum, k_T^2, larger than the cut-off scale ($Q_0^2 < k_{T,0}^2 < k_{T,1}^2 ... k_{T,n}^2 < Q^2$). Because terms of $log(1/x)$ are ignored in the summation, the DGLAP equation is valid especially at high x and Q^2. In contrast to this, in the *Balitzki-Fadin-Kuaev-Lipatov* (BFKL) equation [31, 32, 33], terms are expressed as powers of $(\alpha_s log(1/x))^n$. In this approach the gluon emissions are ordered by the angles of the emitted gluon with respect to the incoming proton.

The unification of the two methods would be a huge improvement in a consistent description of the heavy flavour production in both the *photoproduction* (PhP or γp) and the *deep-inelastic scattering* (DIS) regimes.

2.3 Photoproduction

The variable Q^2 is used to split the kinematic plane into two separate regions: the deep-inelastic and the photoproduction regimes, which are treated independently in both the measurement and reconstruction of events as well as in the way they are described by theoretical models. The DIS regime is defined by the requirement to have $Q^2 \gg m_p^2$, while PhP is typically defined by $Q^2 \leq 1\,\text{GeV}^2$. For small values of Q^2 the photon exchanged by the lepton and the proton is almost on-shell and can be treated as a quasi-real particle. Since the propagator for the weak interaction is proportional to $1/(M_Z^2 + Q^2)$ (where $M_{Z^0} = 91.2\,\text{GeV}$), its contribution is small and the ep interaction can be taken as a scattering of a real photon off the proton.

At high Q^2, the weak interaction becomes comparable in strength, e.g. the cross section depends less on the mass of the exchanged boson. In Figure 2.6, the measurement of electroweak unification for $Q^2 \sim M_W^2$ is shown.

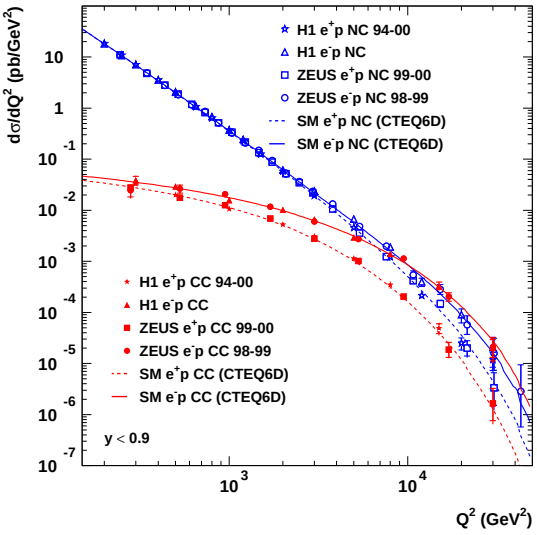


Figure 2.6: *Inclusive differential charged current (CC) and neutral current (NC) cross sections as a function of Q^2. For Q^2 smaller than M_W^2 the CC cross section is suppressed [34]. The measurement is compared with the predictions from the Standard Model calculation using CTEQ6 PDFs.*

2.3.1 Direct and resolved processes

The main contributions to photoproduction of heavy quarks stems from photon-gluon fusion (BGF) processes as shown in Figure 2.7. In photoproduction, the

photon can interact in two different ways. In the *direct* process (also called point-like component) the photon interacts directly in the hard subprocess. In the second case, the so-called *resolved* process, the photon has its own substructure, acting as a source of new partons. The substructure originates from complex intermediate states of the photon, like γ-fluctuations into $q\bar{q}$ pairs. In the direct BGF process, the photon couples to a quark-antiquark pair, emitted from a gluon of the proton. This process is characterised by two high energetic quark jets in the final state.

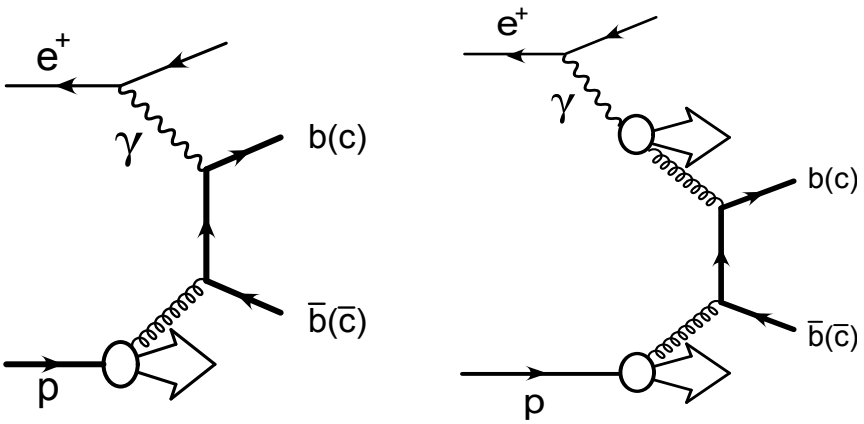


Figure 2.7: *Leading order Feynman diagram for direct (left) and resolved (right) PhP. In the case of direct PhP, here for boson-gluon fusion, the photon can be considered as a point-like object, where in the case of resolved processes, the photon acts as a source for new partons*

In the resolved processes (see Figure 2.7 right) only a fraction of the photon energy participates in the hard interaction. This fact is used to define the variable x_γ which correspond to the energy fraction of the photon contributing to the parton interaction:

$$x_\gamma = \frac{\sum_{q_{1,2}}(E_T^i \cdot e^{-\eta^i})}{2 \cdot y \cdot E_e}, \tag{2.15}$$

where E_T^i and η^i are the transverse energy and the pseudorapidity of the quarks. As in the direct process all the energy from the photon is going into the scattering process with the parton, the final state is expected to have higher transverse energy than the hadron-like process. The value of x_γ is equal to 1 for direct photons and less than 1 for resolved photons.

2.3.2 Next-to-Leading Order Processes

For the calculation of heavy quark production at *next-to-leading order* (NLO), two approaches are commonly used: in the *massless* calculation the zero-mass approximation is applied to the heavy quarks, even if their masses are much larger than Λ_{QCD}. For energies sufficiently high, the heavy quarks can be produced via gluon splitting and are treated as incoming partons. Below that threshold, only light quarks and gluons participate in the hard interaction. As the heavy quark masses are set to zero, it is possible to re-sum logarithms in p_T^2/m^2-terms to all orders using the perturbative fragmentation function technique. This allows reliable predictions for large transverse momenta $p_T \gg m$ to be obtained.

In the *fixed flavour numbering scheme* (FFNS), also called the *massive* calculation, the hadronic structure of the photon and the proton do not contain any heavy flavour contributions, and the heavy quarks do not contribute to the evolution of the running coupling constant. The heavy quark pair is produced in a hard scattering process as an interaction of the photon with a parton [35, 36]. At leading order, this is the boson-gluon fusion, whereas at NLO, additional processes have to be taken into account, providing three-parton final states:

$$\gamma g \rightarrow q\bar{q}g, \tag{2.16}$$

where q is a b or c quark in our consideration. The Feynman graphs for these corrections are shown in Figure 2.8. Additional terms come from virtual corrections (see Figure 2.9) which contribute to the interference with the leading order contributions.

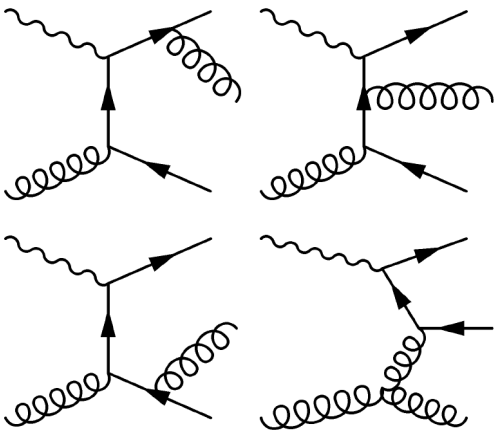


Figure 2.8: *Real NLO QCD contributions to heavy quark production with directly interacting photons. Courtesy of [37].*

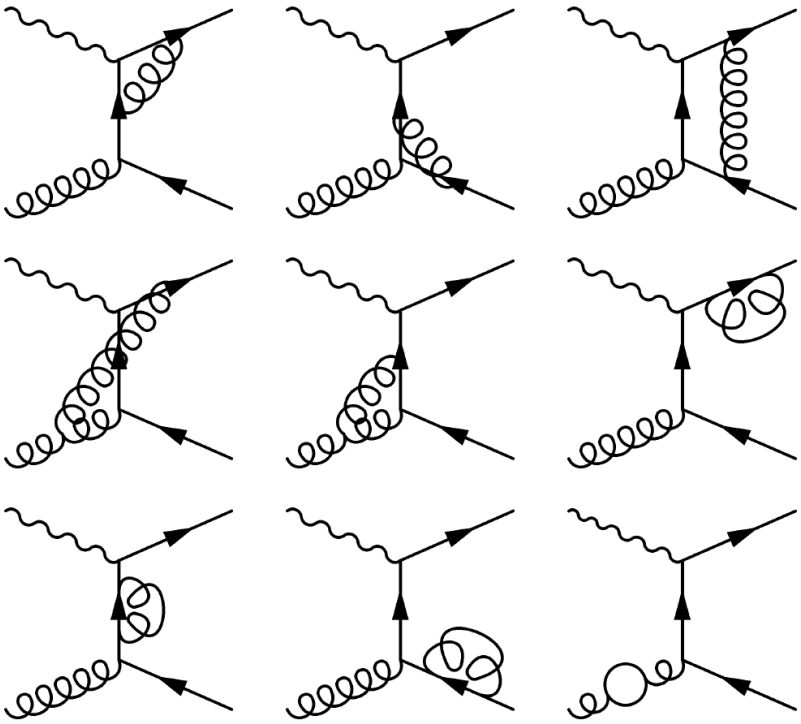


Figure 2.9: *Virtual NLO QCD contributions to heavy quark production with directly interacting photons. Courtesy of [37].*

In addition to *partonic* processes where the photon directly couples to the quarks, the *hadronic* component of the proton has to be added to calculate the total cross section. In the second case, the photon acts as a parton source, described by the photon-parton-density function, and only a component of the photon takes part in the hard process. The heavy quark mass acts as a cut-off for collinear singularities in the initial- and final-state radiation, and therefore defines the scale for the perturbative calculations.

2.4 Event Generators

Monte Carlo (MC) programs serve as essential tools in data analysis for various reasons. Specifically, they are used to check the detector performance and the response of different detector components, and also in the process of cross section unfolding for the extraction of efficiency and acceptance correction. With the aim of simulating the signature of γp events in the detector, the programmes are run through several steps. In the first step, different types of events are generated, describing the required physics processes as accurately as possible. In Figure 2.10, the different stages to generate stable physical objects out of the initial colliding partons are schematically summarised:

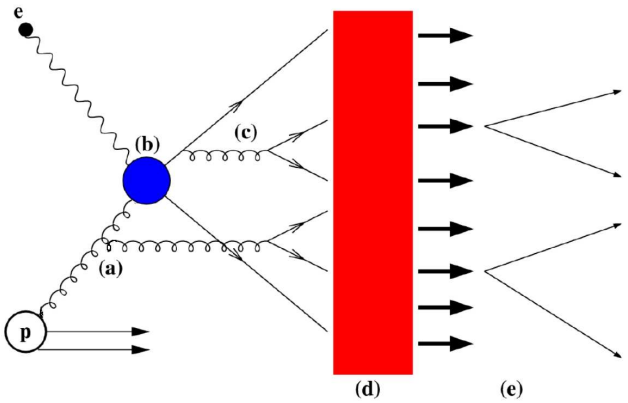


Figure 2.10: *Schematic view of the leading order plus parton shower simulation process. Initial (a) and final (c) state radiated gluons are simulated together with the partons from the hard subprocess (b). In the step of hadronisation/fragmentation (d) the partons transform into an intermediate step of unstable hadrons and decay to stable hadrons (e).*

- *Parton distributions: (PDFs)* Initial input for the calculation are the parton distributions of the colliding particles, which define the probability to find a specific parton with given longitudinal momentum fraction, x, and four-momentum transfer q^2. The photon flux from the incoming lepton is extracted using the Weizsaecker-Williams [38] spectrum.

- *Hard subprocess:* The main part of the event simulation is the parton interaction extracted from the proton and the photon or a photon constituent in resolved photon events. This part of the process can be calculated in fixed order perturbative expansion because it involves a hard scale. The processes, which are taken to leading order in the used Monte Carlo, fix the kinematics and the cross section for an event.

- *Parton shower:* The parton showering, which is not included in the LO calculations, uses QCD based models to describe higher order corrections to the event topology. Processes that contain charged and coloured objects are influenced by the emission of gluons and photons. Radiation of photons or gluons, as well as gluon splitting into quark pairs ($e \to e\gamma$, $q \to qg$, $q \to q\gamma$, $g \to q\bar{q}$), are included before and after the hard subprocess, denoted as *initial* and *final state radiation*.

- *Hadronisation:* Hadronic final states are only observed as colour-singlet states (confinement). The process where the coloured partons, bound by colour-strings, make a transition to colour-neutral objects, is called hadronisation. This process cannot be described in perturbation theory and therefore needs phenomenological models. The most common models are string and cluster models [39]. The Monte Carlo generator PYTHIA uses the string model where a colour string is modelled between the partons having an energy proportional to their distance in order to simulate the QCD-confinement. The quark and the antiquark move apart from each other and the strings break up when the energy is sufficient to produce new $q\bar{q}$ pairs. When the energy is not high enough to produce new quark pairs, hadrons are built. The output of the hadronisation is a bunch of particles and their 4-momenta, produced according to their measured spectra and relative branching ratios.

- *Particle decay:* Not all of the particles produced during the hadronisation process are stable. In the last step they decay according to their different decay chains into stable hadrons.

- *Beam remnant:* In ep interactions, the initial particles for the hard subprocess only carry a fraction of the energy of the initial beam particles. The rest of the energy is taken by the beam remnants. In contrast to the scattered lepton, the proton remnant as well as the resolved photon remnant carry colour and are so connected to the rest of the event. Therefore the fragmentation and reconstruction of the beam remnants are done in a coherent way.

The different steps described above are implemented in the Monte Carlo program PYTHIA6.3 [40], which uses leading order matrix elements, where the strong coupling constant is computed to first order in QCD using $\Lambda_{QCD} = 200\,\text{MeV}$. The CTEQ4L [41] parton distributions were used for the proton, while GRV-G LO [42] was used for the photon. The production of light quarks was simulated with leading-order matrix elements in the massless scheme using the same parton distributions as for the heavy flavour samples. The generated events were passed through a full simulation of the ZEUS detector, based on GEANT 3.13 [43]. Under

the HERA conditions, beauty-quark production is suppressed by approximately a factor of 200 with respect to charm-quark production. This is mainly caused by the larger mass of the beauty quarks (m_b = 4.75 GeV vs. m_c = 1.5 GeV). On the other hand, light quark production is one order of magnitude larger than charm quark production.

2.4.1 Excitation processes

In the massive scheme, only light flavours (u, d, s) are active in the partonic structure of the proton and the photon. Therefore the heavy flavours (b, c) have to be produced by gluon splitting, establishing the boson gluon fusion as the dominant process. The massive scheme contributions, where a heavy quark is directly coming from the partonic structure of the proton or the photon, so-called *excitation processes*, are not simulated and have to be treated separately. Figure 2.4.1 shows the two different kinds of excitation processes. They are simulated in the massless approach by splitting a gluon within the proton/photon into a $b\bar{b}$ or $c\bar{c}$ pair. These

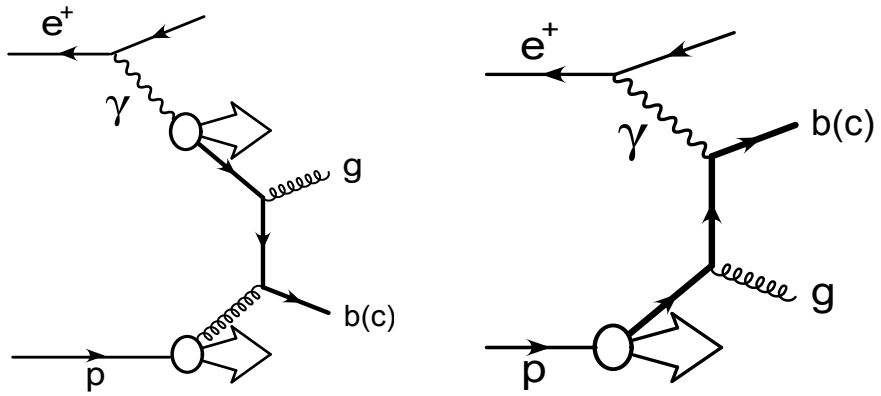


Figure 2.11: *LO Feynman diagrams for excitation in γ (left) and excitation in proton (right). The heavy quark is produced within the partonic structure of the proton or the photon. This kind of processes are not included in the massive approach where only light flavours are allowed in the partonic structure. Therefore they are simulated separately using the massless scheme.*

types of processes become important at high energies where terms of the form $\alpha_s log(p_T^2/m^2)$ become large. Therefore the massive calculation, which is used for the NLO QCD prediction, is expected to work at momenta around or larger than the heavy quark masses [44].

2.5 Semileptonic Decays of Heavy Hadrons

Heavy quark production is studied in this analysis using *semileptonic decays* of the heavy quarks into electrons. Figure 2.12 shows as an example the spectator model for a semileptonic B^-, where the beauty quark is transformed into a charm quark via a W^- emission. An electron and an antielectron neutrino are created in the W^- decay. This model is based on the assumption that the light quark in

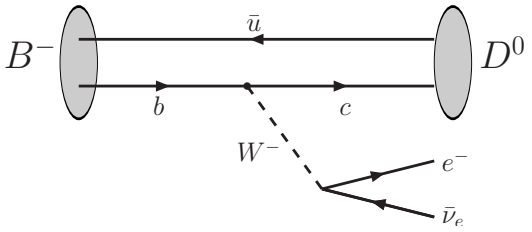


Figure 2.12: *Spectator model for the semileptonic b decay where a B^--meson transforms into a D^0 under an emission of a W^-. In this model the $\bar{u}$ does not take part in the interaction. In the channel under consideration the W^- emits an electron and an $\bar{\nu}_e$.*

the heavy hadron continues into the final state with a negligible influence on the other quark. This spectator model works better for beauty than for charm quarks, because the large mass suppresses the higher-order QCD effects.

The decay width is proportional to the square of the appropriate element of the *Cabbibo-Kobayashi-Maskawa* (CKM) matrix [45]. The transition from a beauty quark to a charm quark is given by $\Gamma \sim |V_{bc}|^2 m_b^2$.

Figure 2.13 shows, as another example, the spectator model for a semileptonic D^0 decay. As the charm quark mass is much closer to the light quark masses, corrections to the spectator model are needed. The improvements include gluon corrections, which carry the interaction between the light and heavy quark, corrections to the uncertainty of the quark mass, and the Fermi motion of the charm quark in the proton. Not only the production rates, but also properties as the electron momentum spectrum are different in beauty and charm semileptonic decays (see Section 6.3.2).

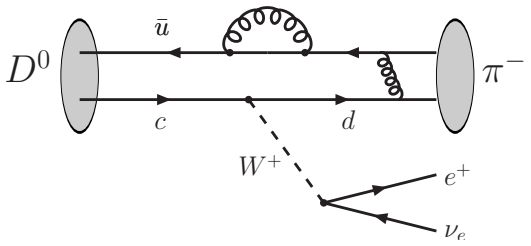


Figure 2.13: *Gluon corrections in the spectator model for the semileptonic c decay where a D^0-meson transforms into a π^- under an emission of a W^+. In the channel under consideration the W^+ emits a positron and an ν_e.*

2.6 HERA Measurements

Beauty and charm quark production cross sections have been measured over a wide rage of Q^2 by both the ZEUS [46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61] and the H1 [62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73] collaborations. The cross sections are well described in their dependence on kinematic variables by both leading order and next-to-leading order QCD predictions, while only NLO describes also the absolute values correctly. Both the deep inelastic scattering (DIS) at large Q^2 and photoproduction measurements at low Q^2 are reasonably well described by NLO QCD predictions. Most of the previous measurements of b-quark production used decay muons to tag semileptonic decays of the B hadrons. Electrons from b decays can be identified down to lower momenta, but their measurement is more difficult since background from other sources is substantially.

2.6.1 Beauty in Photoproduction

In Figure 2.14 the published cross sections for beauty in photoproduction at HERA are summarised as a function of the transverse momentum of the b quark, p_T^b. A first measurement of b-quark photoproduction from semileptonic decays to electrons (e^-) was presented in a previous publication [49]. It was based on e^+p collision data from the 1996–1997 running period corresponding to an integrated luminosity of $38\,\mathrm{pb}^{-1}$ (see open circles).

The published ZEUS result with electrons obtained for the HERA I data corresponding to $120\,\mathrm{pb}^{-1}$ [16] are shown by the black triangles having comparable or smaller uncertainties than other measurements. Comparing all results, covering a large momentum range, with the NLO QCD calculation, gives a consistent picture of beauty quark production for low Q^2.

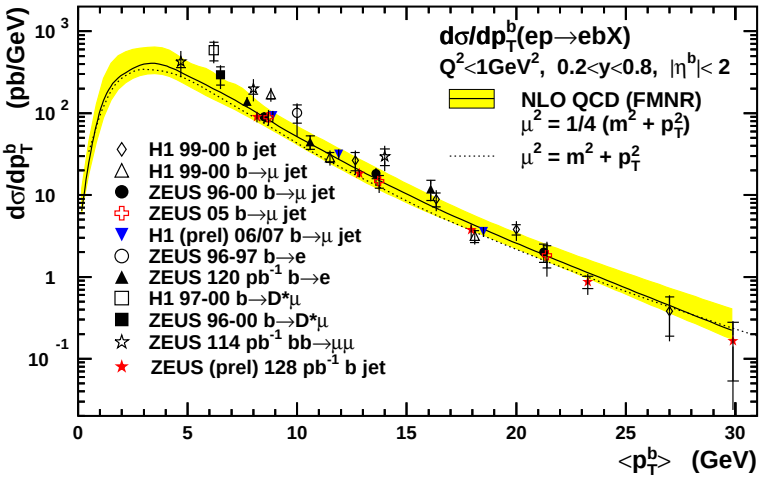


Figure 2.14: *Measurements of beauty in photoproduction at HERA. The results are given as a function of the transverse momentum of the b quark, p_T^b. The theoretical prediction has been extracted by the FMNR program. The variation of the scale parameters used in this analysis correspond to the dotted line, where variation for the error band used the scales as given in the Figure. The combination of the H1 and ZEUS measurements give a consistent picture of beauty production for low Q^2.*

Chapter 3

The ZEUS Detector at HERA

This analysis was performed using data collected with the ZEUS detector at the *Hadron-Electron-Ringanlage* (HERA) [74] in Hamburg. In this chapter the experimental facilities, which provide the environment for this analysis, will be presented, including the main detector components relevant for this thesis (a more detailed description can be found in [75]). At the end of the chapter the trigger and data acquisition systems will be described briefly.

3.1 The HERA Collider

HERA, the first electron (positron)-proton collider, was located at the *Deutsches-Elektronen-Synchrotron* (DESY) in Hamburg. The accelerator was constructed from 1984 to 1990 and, after one year of commissioning, started operation in 1992. From 1992 to 2007 it operated at beam energies of $E_p = 820 - 920\,\text{GeV}$ and $E_e = 26.5 - 27.5\,\text{GeV}$ yielding centre-of-mass energies, $\sqrt{s_{ep}} = \sqrt{4E_eE_p} = 296 - 318\,\text{GeV}$ and allowed the investigation of DIS and PhP processes at the highest energy scales accessed so far. The HERA ring is 6.3 km in circumference, around which four experiments were located, ZEUS, H1, HERMES and HERA B. The two colliding-beam experiments, ZEUS and H1, were located in the halls to the south and north, respectively. The HERMES experiment used polarised $e^{\pm}$ beams to study spin effects in lepton-nucleon interactions using a polarised nuclear target, whereas the HERA B experiment was designed to investigate B meson physics. Figure 3.1 shows the HERA ring system including the experimental halls in the straight sections of the ring and its pre-accelerator elements. The $e^{\pm}p$ collider had its first phase of operation, HERA I, from 1992 through 2000. The HERA collider was upgraded to increase the specific luminosity, as well as to provide longitudinally polarised lepton beams to the collider experiments [78]. The

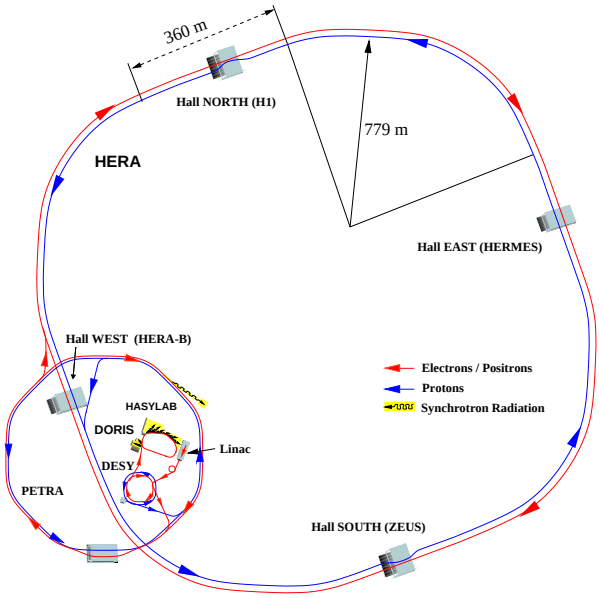


Figure 3.1: *Schematic view of the HERA collider [74], four experimental halls and the pre-accelerator ring with the injection system [76, 77].*

second phase of data taking, HERA II, began in 2003, after completion of the machine and detector upgrades. In the last three months of HERA operation in 2007, data with lowered proton beam energies of 460 GeV (MER) and 575 GeV (LER) was taken. In Table 3.1, the most important values for the HERAII running conditions are listed. Figure 3.2 shows that the HERA II integrated luminosity of $\sim 580\,\mathrm{pb}^{-1}$ greatly exceeds the $\sim 190\,\mathrm{pb}^{-1}$ delivered in the HERA I running period [79]. The integrated luminosities delivered for the different periods, split by lepton type and energy, are listed in Table 3.2.

3.2 The ZEUS Detector

The ZEUS detector was a multi-purpose detector covering almost the full solid angle. The detector comprised many components giving as much information as possible about the final-state particles from the ep interaction. Especially energies, directions and the nature of single particles as well as compound jet objects should be reconstructed with high precision. In order to accommodate for the particle boost due to the energy imbalance, the detector was more heavily instrumented

	Electrons	Protons
Maximum current (mA)	58	136
Maximum number of bunches	180	180
Maximum number of colliding bunches	174	
Specific luminosity ($10^{30}\,\mathrm{cm^2 s^{-1} mA^{-2}}$)	1.79	
Peak luminosity ($10^{31}\,\mathrm{cm^2 s^{-1}}$)	7.44	
Energy (GeV)	27.6	920
Centre-of-mass energy (GeV)	318	

Table 3.1: *HERA II design parameters.*

in the forward direction, which was defined by the proton beam direction. This direction defined the z–axis of the right–handed ZEUS coordinate system [80, 81], where the centre of the *Central Tracking Device* (CTD) defined the origin (Figure 3.3). In the following, only a brief overview of the ZEUS detector will be given, where the most important sub-components for this analysis are described. A detailed description of the detector is given in [82]. Figure 3.4 shows a schematic side–view of the ZEUS detector layout.

The innermost detector components surrounding the beam–pipe were the *Micro-Vertex Detector* (MVD) and the CTD, whose information were combined with the *Forward Detector* (FDET) to a tracking system covering a wide solid angle. Charged particle tracks could be reconstructed in the angular range of $7.5^\circ < \theta < 170^\circ$. For this analysis only the tracking system for the central region was important, so the MVD and the CTD is discussed more in detail in Sections 3.2.1 and 3.2.2. Outside the super-conducting solenoid, which provided a magnetic field of $\sim$1.43 T, were the *electromagnetic* (EMC), the *hadronic* (HAC) and the *Backing Calorimeters* (BAC). The function and the sub-components of the high resolution depleted-uranium scintillator calorimeter will be discussed in Chapter 3.2.3. Before and behind the BAC were muon chambers surrounding the iron yoke. Together with the central tracking system these chambers were used to identify the muons that passed the calorimeter system.

Important components for the event preselection were the VETO wall and the C5 detector. The C5 detector, which consisted of $2x2$ scintillator layers interleaved with layers of tungsten, was positioned 1.2 m[1] from the nominal interaction point in the electron flight direction in the cut-out region of the *Rear Tracking Detector* (RTD) [83], a package of three planar drift chambers placed in front of the RCAL. The VETO wall was used to shield the detector against particles from the

[1] Before the detector upgrade the C5detector was located at -3.14 m, behind the RCAL.

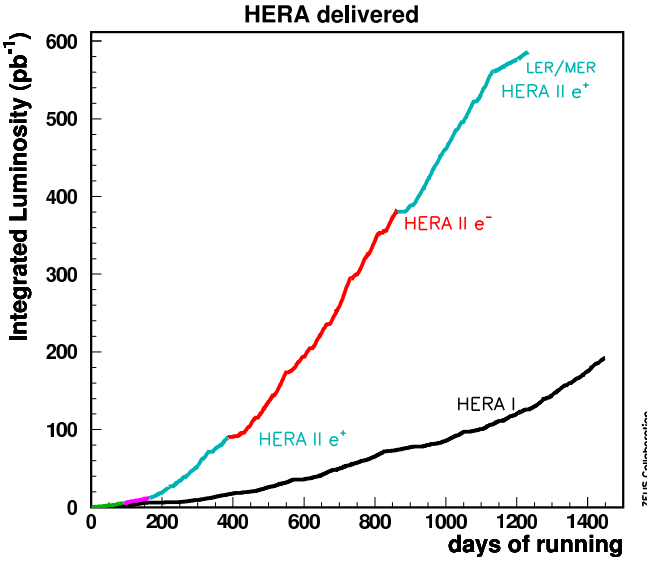


Figure 3.2: *Delivered integrated luminosity for each running period of HERA [79]. LER and MER describe the integrated luminosity for the lowered proton beam energies as described in the text.*

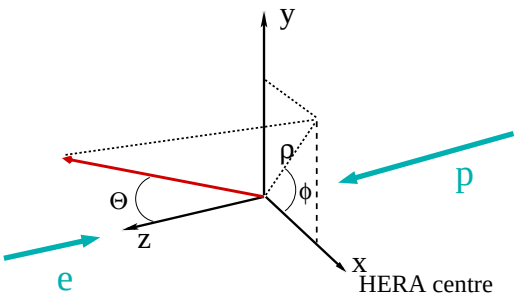


Figure 3.3: *ZEUS coordinate system. The z–axis was pointing in the direction of the proton beam while the x–axis was pointing towards the centre of the HERA ring, and the y–axis upwards. The azimuthal and the polar angle are denoted by ϕ and θ, respectively. Instead of θ the pseudo–rapidity $\eta = -\ln[\tan(\theta/2)]$ is frequently used.*

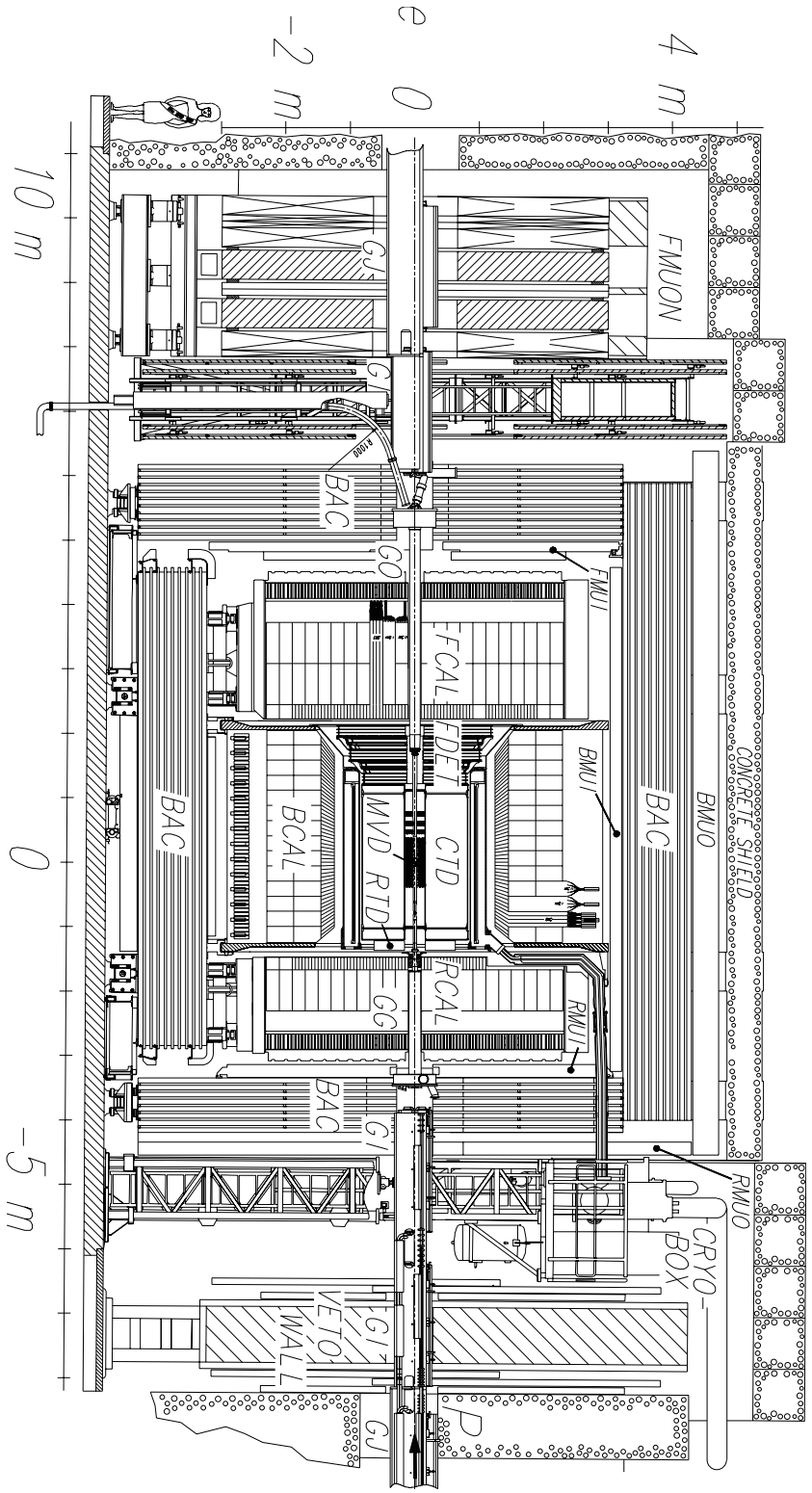


Figure 3.4: *Schematic side-view of the ZEUS detector showing the tracking system in the centre embedded into the calorimeter. In the outer region the ZEUS detector is covered by the muon chamber system.*

Years		Lepton	E_e/GeV	E_p/GeV	$\sqrt{s_{ep}}$/GeV	$\mathcal{L}^{del}$ /pb^{-1}
HERA-I	1992 – 1993	e^-	26.7	820	296	0.03
	1993 – 1994	e^-	27.5	820	300	2.2
	1994 – 1997	e^+	27.5	820	300	71
	1998 – 1999	e^-	27.5	920	318	25
	1999 – 2000	e^+	27.5	920	318	95
HERA-II	2003 – 2004	e^+	27.5	920	318	85
	2004 – 2005	e^-	27.5	920	318	205
	2006 – 2006	e^-	27.5	920	318	86
	2006 – 2007	e^+	27.5	920	318	180
	2007 – 2007	e^+	27.5	460	214	16
	2007 – 2007	e^+	27.5	575	251	9.4

Table 3.2: *HERA running conditions. In 1998 the energy of the proton beam was raised from* 820 *GeV to* 920 *GeV increasing the centre-of-mass energy,* $\sqrt{s_{ep}}$*, from* 300 *GeV to* 318 *GeV. Since 2003 HERA provided much higher specific luminosities. At the end of 2007 two periods with lowered proton beam energies were taken.*

proton beam halo and from beam-gas interactions. In addition there were photon and electron taggers for the luminosity measurement (Section 3.2.4) and detector stations for the observation of scattered protons in forward direction.

3.2.1 The Micro-Vertex Detector

Installed in 2001 during the break between HERA I and HERA II running periods, the MVD was the component of the detector closest to the interaction point. This component had been installed to allow the reconstruction of secondary vertex tracks, coming from the decay of long-lived particles with a lifetime of about 10^{-12} s, and to improve the global momentum resolution of the tracking system.

The MVD was designed to fit into the gap between the beam-pipe and the CTD, which had an inner radius of $\sim$20 cm. It was subdivided into the central (BMVD) and the forward (FMVD) parts in order to have a good matching with the existing detectors. Figure 3.5 shows a cross section of the BMVD. The sensors were primarily made from 320 µm thick n-doped silicon in which p-doped parallel strips were implanted on one side. They were arranged in three layers in the central region (600 detectors) and 4 wheels in the forward region with 112 silicon planes and resulted in at least 20 µm individual hit resolution for incident tracks, and a

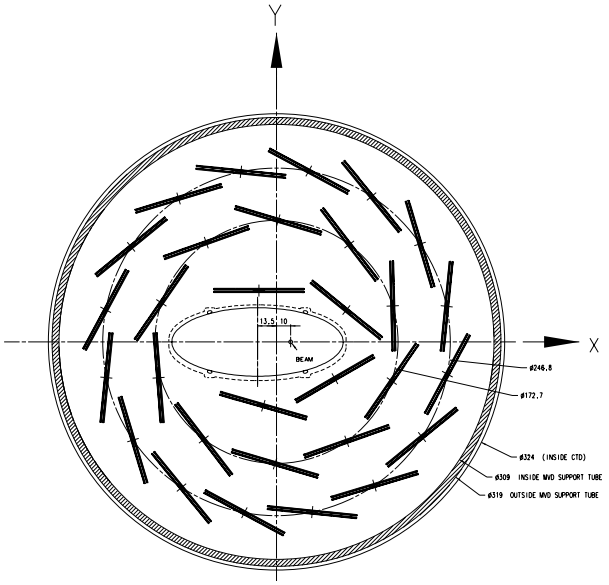


Figure 3.5: *Cross section of the BMVD, showing the planes of silicon sensors.*

two-track separation of 200 μm [84] with an angular coverage of $10^0 < \theta < 160^0$.

3.2.2 The Central Detector

The CTD [85, 86] was one of the most important tracking components used for the track reconstruction and particle identification. It consisted of a cylindrical drift chamber with an outer radius of 85 cm and an overall length of 240 cm. It provided a large angular coverage of $15^\circ < \theta < 164^\circ$. The track position and the energy loss were measured in nine concentric super-layers, each with eight layers of sense wires (see Figure 3.6). In total, $\sim 4\,600$ sense wires with positive voltage and $\sim 20\,000$ field wires with opposite potential were arranged parallel to the z-axis. In the odd numbered (axial) super-layers the wires were installed parallel to the z-axis, whereas the ones in the even numbered (stereo) super-layers were rotated by a small ($\pm 5^0$) angle. By this arrangement it was possible to measure both the $r\phi$ as well as the z coordinates accurately. Charged particles ionised the gas of the chamber releasing atomic electrons, which drifted to the positively charged signal wires; in contrast, positively charged ions moved more slowly to the field wires. The positions of the tracks were reconstructed with an accuracy of 190 μm in (r, ϕ) and 2 mm in z. For fast reconstruction used by the trigger, the inner three superlayers were instrumented by a z-by-timing system which was able to reconstruct the position along the wire with a resolution of 4 cm by measuring the difference in time of the pulse arriving at each end of the wire. The CTD was filled with a gas mixture of 83% Ar, 5% C_2H_6 and 12% CO_2 close to at-

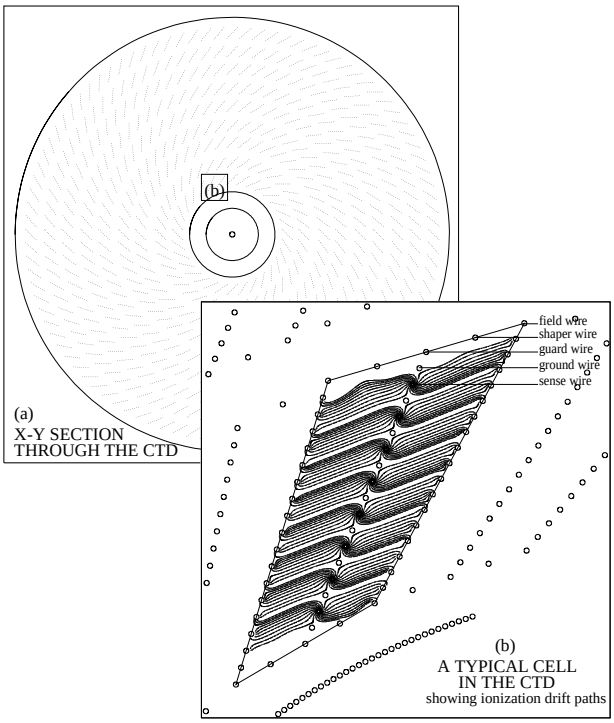


Figure 3.6: *Cut through the CTD in the xy–plane. In (a) the nine concentric superlayers surrounding the beam–pipe are shown. In (b) a single cell consisting of sense and field forming wires and the ionisation drift paths are depicted.*

mospheric pressure. As the charged tracks were curved within the magnetic field of the surrounding solenoid, the measured radius of the reconstructed tracks can be transformed into a momentum measurement. With the position measurement and the track reconstruction (see [87,88]), the resolution in transverse momentum using only the CTD information was [89]:

$$\frac{\sigma_{p_T}}{p_T[\text{GeV}]} = 0.0058 p_T \oplus 0.0065 \oplus \frac{0.0014}{p_T[\text{GeV}]}\,, \tag{3.1}$$

where the first term corresponds to the resolution of the hit positions, the second term to smearing from multiple scattering within the CTD and the last term to multiple scattering before the CTD. The track reconstruction combining the different tracking components and the achieved resolutions are described in Chapter 4.1. The CTD was also able to measure the energy loss of the particles by ionisation of the chamber gas, which was used for particle identification. As the energy loss is one of the key issues for this analysis, the procedure of using this measurement, –including corrections and simulation– are described separately in Chapter 4.2.

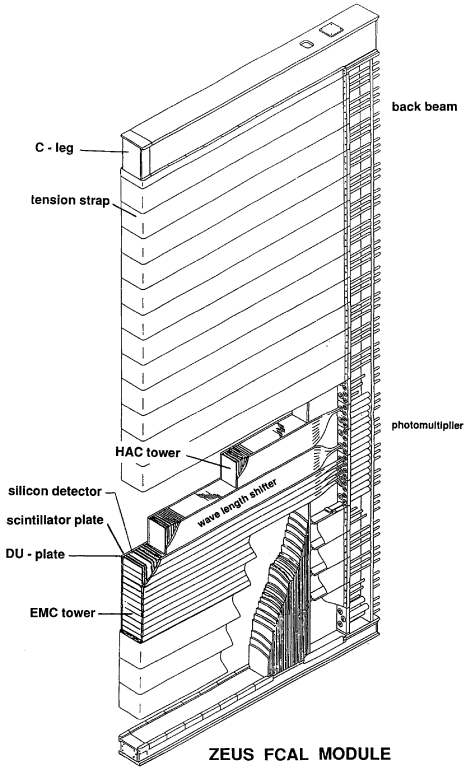


Figure 3.7: *Cut–away view of a FCAL module [75]. The particles enter the module from the left and traverse the sandwich structure of the towers made of uranium and plastic scintillators. The wavelength shifters mounted on the right collected (mainly blue) light emitted by the scintillators, converted it and guided it to the photomultiplier converting it to green light on the way.*

3.2.3 The Uranium-Scintillator Calorimeter

The ZEUS calorimeter (CAL) [90, 91, 92, 93] was a high resolution compensating calorimeter, which was built to measure the energies of the final-state particles in the event. For the interpretation of the energy deposits, the most important aims of the calorimeter design were the reconstruction of compound jet objects and a good energy and angular resolution, as well as hadron-electron separation for both isolated electrons and electrons in the jets. In order to achieve compensation and the best possible energy resolution, the CAL consisted of alternating plates of depleted uranium and plastic scintillators. The uranium, acting as absorber material with a thickness of one radiation length (3.3 mm), provided equal response of electrons and hadrons in the scintillator (2.6 mm). The different modules of the CAL were read out optically via wavelength shifter bars, light guides and photomultipliers. Figure 3.7 shows one of these modules, which were segmented in depth into an electromagnetic and one or two hadronic sections (EMC, HAC1 and HAC2) (see Figure 3.8). The typical tower size was 5 cm $\times$ 20 cm in the EMC and 20 cm $\times$ 20 cm in the HAC. The CAL was subdivided into three sections for forward (FCAL), barrel (BCAL) and rear (RCAL) regions with 7, 5 and 4 absorption lengths, respectively (see Table 3.3), covering almost the full angular region.

Section	Polar angle	Pseudo-rapidity
FCAL	$2.2° < \theta < 39.9°$	$1.0 < \eta < 4.0$
BCAL	$36.7° < \theta < 129.1°$	$-0.7 < \eta < 1.1$
RCAL	$128.1° < \theta < 176.5°$	$-3.5 < \eta < -0.7$

Table 3.3: *Angular acceptance of the CAL*

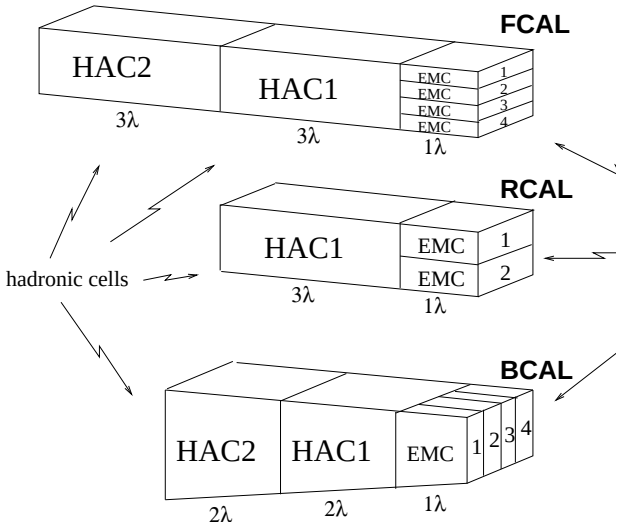


Figure 3.8: *Segmentation of the CAL towers, with λ denoting the interaction lengths. Each section consisted of an electromagnetic (EMC) and a hadronic (HAC) part, where in the FCAL and BCAL the latter were subdivided into HAC1 and HAC2.*

A parametrisation of the shower development [94] showed that for a 10 GeV hadron, 95 % of the transverse energy was contained in a cylinder with radius of approximately 20 cm. As electrons and photons interact mostly electromagnetically (QED processes like Compton-scattering, bremsstrahlung or pair–production), the showers are smaller. The size of the shower can be described by the Molière radius[2] which was typically $R_M = 2$ cm at the ZEUS energies. Muons lose only part of their energy within the CAL as they behave like minimum ionising particles. The different shower types which can be used for particle identification are shown schematically in Figure 3.9. The ZEUS calorimeter was compensating, i.e. electrons and hadrons of equal energies yield the same pulse height within 3 % for momenta above 3 GeV.

[2] $\sim$ 95 % of the energy is contained within twice the Molière radius

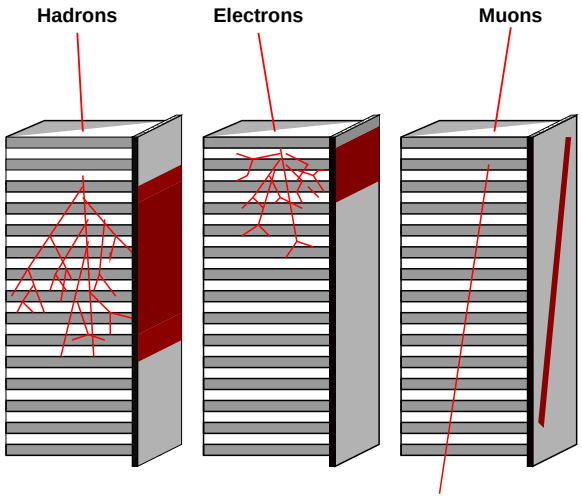


Figure 3.9: *Different types of shower shapes in the calorimeter. The uranium–scintillator structure is drawn for three towers with the wavelength shifters on the right side. The penetrating hadrons produce a large shower. The electron shower is small. The light signal of minimum ionising muons is equally spread over the depth of the tower.*

In test beam measurements the energy resolution has been determined to [90]:

$$\frac{\sigma}{E} = \frac{18\,\%}{\sqrt{E[\mathrm{GeV}]}} \oplus 0.2\,\% \qquad \text{for electrons and} \tag{3.2}$$

$$\frac{\sigma}{E} = \frac{35\,\%}{\sqrt{E[\mathrm{GeV}]}} \oplus 0.3\,\% \qquad \text{for hadrons.} \tag{3.3}$$

The backing calorimeter (BAC), surrounding the CAL, used the iron plates of the magnet yoke as absorber material, in order to measure the energy of late showering particles. The material of the solenoid had a thickness of 0.9 radiation lengths so that electrons, photons and hadrons had minimal energy loss.

3.2.4 Luminosity Monitor

At ZEUS the luminosity was monitored by the measurement of the bremsstrahlung process, $ep \to e'\gamma p$, for which the cross section could be calculated accurately as the integrated cross section is given by the Bethe–Heitler formula [95]. The total cross section of this QED process is about 326 mb for photon energies $0.1 < E_\gamma < 26.7$ GeV [96]. The measurement was based on the detection of the photon which were emitted at very small angles. The bremsstrahlung process was measured with high acceptance, while background photons originating from PhP and

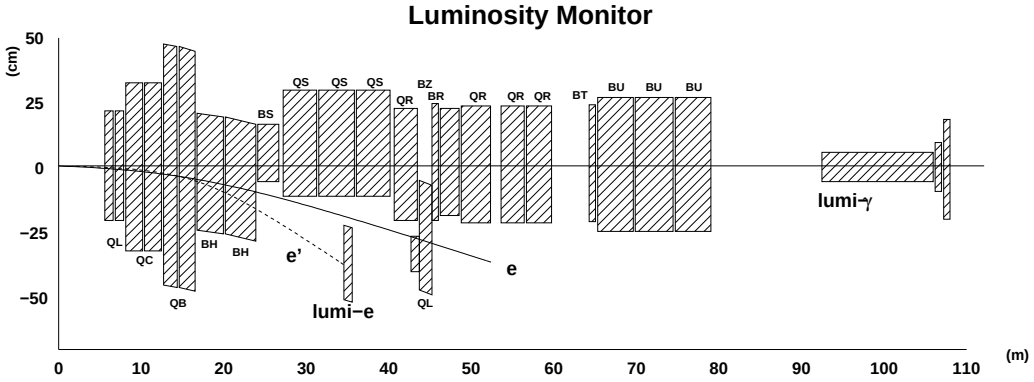


Figure 3.10: *The ZEUS luminosity monitoring system in HERA I. The hatched blocks indicate bending (B) and quadrupole (Q) magnets. The detectors at* 35 m *and* 44 m *measured electrons that were scattered under very small angles and were used for taggin events with a corresponding centre of mass energy within the range* $70 < \sqrt{s_{\gamma p}} < 120\,GeV$.

DIS events were detected by measuring the scattered electron or the initial state radiative photons. The general layout of the detection scheme (used in the HERA I period) is shown in Figure 3.10. The photons from the bremsstrahlung process propagated inside the proton pipe and could leave it at 82 m through a copper-beryllium window. The luminosity monitor (LUMI) consisted of a 22 radiation lengths deep photon calorimeter at a distance z of 104 to 107 m and an electron detector near the electron beam at a distance of $z = 35, 44$ m. The energy resolution under experimental conditions was $\sigma(E)/E = 26\,\%/\sqrt{E}$, with E measured in GeV. [97]

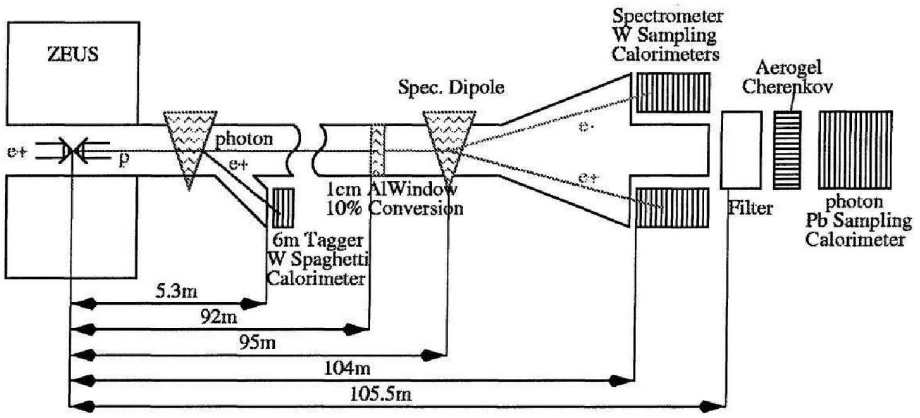


Figure 3.11: *The ZEUS luminosity monitoring system in HERA II.*

Due to the luminosity upgrade in 2001 it was necessary to improve the luminosity system to account for the increased synchrotron radiation, the larger number of overlayed bremsstrahlung events (pile-up) and the different settings for polarised electrons and positrons. In the HERA II data taking period the luminosity was measured by two independent monitors, the *Photon Calorimeter* (PCAL) and the *Spectrometer* (SPEC). The schematic view of these detector system is shown in Figure 3.11. As the uncertainty of the luminosity measurement was larger for the SPEC (3.5 %) than for the PCAL (2.6 %), the final cross section calculation was done based on the result from the PCAL [98].

3.2.5 Background Rejection

As the aim of the ZEUS detector was to detect particles arising from ep interactions, it was of crucial importance to get rid of particles from other sources which could mimic such interactions. One source was beam-gas interactions, where the beam-particles interacted with residual gas in the beam-pipe or directly with the beam-pipe material. Another important background was induced by halo muons or muons from cosmic rays passing the detector. Halo muons mostly originated from decays of charged pions created in proton-beam interactions with accelerator components. The ZEUS detector was shielded against cosmic muons as it was located ~ 25 m underground and covered by ~ 1 m of concrete, so that only a low rate of such muons was observed.

The main purpose of the iron VETO wall at $z = -7.5$ m was to protect the detector from proton beam halos directly as a passive absorber and in addition by providing timing information from scintillation counters on both sides of the wall. The background from non-ep interactions was further reduced at trigger level by using the timing information of the VETO wall, the C5 counters and the *Small-angle Rear Tracking Detector* (SRTD), which was used for detection of small-angle scattered electrons as well as for beam-gas rejection. The SRTD was located around the beam-pipe between the RTD and the CAL.

3.2.6 Trigger and Data Acquisition

HERA bunches crossed every 96 ns giving a bunch crossing rate of approximately 10 MHz. The maximum rate of events stored on tape should not exceed $3 - 5$ Hz. As the majority of the interactions was coming from beam-gas interactions, it was the purpose of the trigger system to decide which ep interactions are of interest. For these events the detector response which is stored in a pipeline system was passed to be written on tape. At ZEUS this system was realised in three levels and

operated in a pipeline mode in order to avoid dead time [99] and to increase the time to perform more detailed calculations. Finally, typical output trigger rates were around 1 kHz for the *First Level Trigger* (FLT), 100 Hz for the *Second Level Trigger* (SLT) and 10 Hz for the *Third Level Trigger* (TLT) corresponding to a rejection factor of 10^5 with a large acceptance of interesting physics events.

First Level Trigger (FLT)

The main focus of the FLTs was to provide a fast decision within 2.5 μs storing their data in pipelines, synchronised with the HERA 10 MHz clock. The output of the local FLTs was collected in the *Global First Level Trigger* (GFLT), where variables of the main detector were used to make an "OR" decision based on 64 trigger slots. The data processing at the GLFT was performed on hardware level using programmable look-up memory tables. Important and fast reconstructible information were the total energy (E_T) and missing transverse energy (E_T^{mis}) provided by the CAL, the number of reconstructed tracks in the CTD, primary and secondary vertex positions as well as the number of muons identified by the muon chambers. After 4.4 μs the GFLT made the decision, and if one of the slots was accepted, the decision was transferred to all components. After this selection the rate was reduced to approximately 1 kHz with a typical dead-time below 0.1 %.

Second Level Trigger (SLT)

More time was available at the SLT (about 8 ms), which allowed a more complex analysis of the data. If the GFLT accepted the event, the data of all components was read-out and stored in the SLT buffer, so that all information from the GFLT was available in the SLT, where it could be combined with CAL timing information, reconstructed electron candidates, hadronic clusters and muon candidates. The CAL timing information could be used to reject events from non-ep interactions, because the timing was shifted for beam-related background and showed differences in the upper and the lower part of the CAL when the response was originating from cosmics. Here the trigger was software-based on a transputer network. The algorithms used were more complex and included physics filters for particular event topologies. For the HERA II a new *Global Tracking Trigger* (GTT) has been developed in order to take advantage of the additional tracking information of the MVD. This algorithm, which was running on a PC farm, combined the MVD hits with the track segment from the CTD, which were then used to determine a more precise primary vertex, kinematic variables and reconstruction of meson candidates. Like the GFLT, the *Global Second Level Trigger* (GSLT) combined the trigger information and produced sub-trigger bits. If one of

the 32 bits was fulfilled, the GSLT accepted the event and passed the information to the EventBuilder (EVB).

Third Level Trigger (TLT)

The main task of the EVB was to collect the data, build a data structure, consistent with the ADAMO [100] database tables, and distribute the events over the processor nodes of the TLT. The ADAMO format was the data structure used by all subsequent levels. The implementation of the TLT was based on the off-line reconstruction software running on a farm of Linux-PC machines. Typically ~ 0.3 s after the bunch crossing the TLT passes the decision with an event rate of ~ 5 Hz. As for the GFLT and in the GSLT, the TLT output is a set of sub-trigger bits. The TLT bits have been specified by the physics analysis groups and were modified several times. At this level they were based on quantities like the energy and angle of the scattered electron, jet energies or the invariant mass of the final state. If the event passed one or more of the physics filters they were transferred to the DESY Computer Centre and recorded on tape.

Event Reconstruction

As mentioned above, events accepted by the TLT were stored on the mass storage system. For the reconstruction of the collected data the *ZEUS Event Physics Reconstruction* (ZEPHYR) program was used. In a first step the data had to be corrected for calibration effects and for detector imperfections like dead or noisy channels, e.g. of the MVD or CTD. Based on this, the specific reconstruction packages could build up tracks from the CTD and MVD, clusters from the CAL or muons from the muon chamber information. At the end of the ZEPHYR reconstruction process physical events like charmed meson or J/Ψ production were reconstructed giving input for analysis-specific physics filters. Stronger cuts, based on the corrected values, defined the data which was stored on *Data Summary Tapes* (DST).

Monte Carlo Simulation

Most of the physics analyses at ZEUS are based on the comparison of the event distributions with model predictions. This demands a detailed Monte Carlo simulation of the experiment, including the response and the trigger acceptance. To simulate this, the output of the Monte Carlo generators has to be passed to the

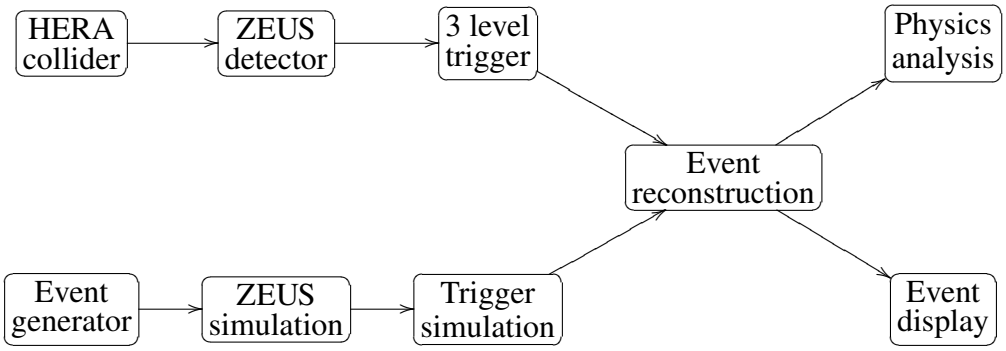


Figure 3.12: *Overview of the data and simulation chains.*

ZEUS detector simulation program MOZART[3]. This program is based on the CERN GEANT 3.21 package [101], and includes a detailed description of all detector components and material. The trigger chain simulation is done by a dedicated program called ZGANA. Finally simulated event samples are reconstructed using ZEPHYR and stored in the ADAMO format consistent with the one used for real data. With this procedure the individual user can analyse and visualise data and simulated event samples in the same way. An overview of the data and simulation chain is presented in Figure 3.12.

[3]Monte Carlo for ZEUS Analysis Reconstruction and Trigger

Chapter 4

Track Finding and Event Reconstruction

In this chapter a brief overview of the track finding and event reconstruction are presented. This includes physical objects, like tracks and jets, as well as kinematic quantities that are needed for the event selection and analysis. In addition a short summary of the specific energy loss measurement in the CTD, dE/dx, is given.

4.1 Track Reconstruction

The track finding and reconstruction is performed in various steps at ZEUS. In a homogeneous magnetic field charged particles traversing the MVD and the CTD follow a helix curve around the z-axis. The pattern recognition program VCTRAK [102] starts at the outermost superlayer of the CTD where track segments are found. These segments are extended to tracks by extrapolating inwards through the axial superlayers. They are fitted to a circle in the xy-plane. When MVD hits are added to the track, the parameters are updated and the track is extrapolated towards a virtual point on the beam line. By adding the z-by-timing information, i.e. the hits in the stereo CTD superlayers and from MVD-hit information, a linear fit in the sz-plane (with s as the path length along the path's arc) determines the five parameters of the track helix. The five parameters are $(\phi_H, Q/R, Q \cdot D_h, Z_H, cot\Theta)$, where ϕ_H is the azimuthal angle in the point-of-closest-approach Q denotes the charge of the track, given by the curvature, and R the radius of the helix. D_H and Z_H are the distances in the xy-plane with respect to the origin and the z-position of the point-of-closest-approach in $r\phi$ and Θ denotes the polar angle. In Figure 4.1 the helix parameters are shown schematically in the xy-plane. The fit method

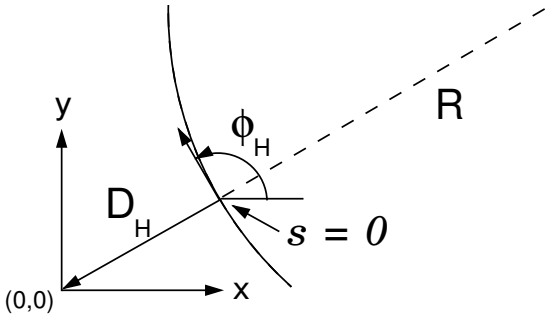


Figure 4.1: *Helix parameter in the xy-plane. R is the helix radius in the xy-plane, s the point with the smallest distance, D_H, to the origin and ϕ_H, the inclination of the tangent at this point.*

to determine the track parameters from the hits using a Kalman filter algorithm is described in [103].

4.2 dE/dx Measurement

As charged particles interacted with the gas within the CTD volume, the CTD hits could provide not only information about the position of the particles passing through the CTD. After the interaction the electrons and the ions were accelerated by the electric field in opposite directions so that they could not recombine. Passing the field the signal was amplified ($\sim O(10^4)$) close to the wire before it reached the sense wire. The integrated pulse size gives a measure of the energy loss of the particle. The dE/dx of a particle in the gas is given by the Bethe-Bloch formula [104].

$$\frac{dE}{dx} = \frac{4\pi N e^4}{m_e c^2 \beta^2}\left(\ln\frac{2m_e c^2\beta^2\gamma^2}{I} - \beta^2\right), \tag{4.1}$$

where z denotes the charge of the travelling particle, N the number of density of electrons in the matter traversed and I the mean excitation energy of the atom in matter. Equation 4.1 cannot be used directly for in this application, as it is only valid for particles much heavier than the scattered particles and does not take into account statistical fluctuations due to secondary tracks at high energies. In contrast to the analytical approach using the Landau formalism [105] a statistical formulation using the *photo-absorption-ionisation model* by Allison and Cobb [106] was implemented to simulate the $\beta\gamma$ dependence of the dE/dx.

It is calculated by the VCTRAK package using a truncated mean. This takes all the hits on a track in order of pulse size, discarding bottom 10 % and top 30 %, and then takes the mean of the remaining pulse sizes to give the dE/dx measurement. The corrected dE/dx measurement was normalised such that the corrected value is 1.0 for a minimum ionising particle (mip). A detailed description of all corrected effects can be found in [107]. Different samples of identified particles were used to

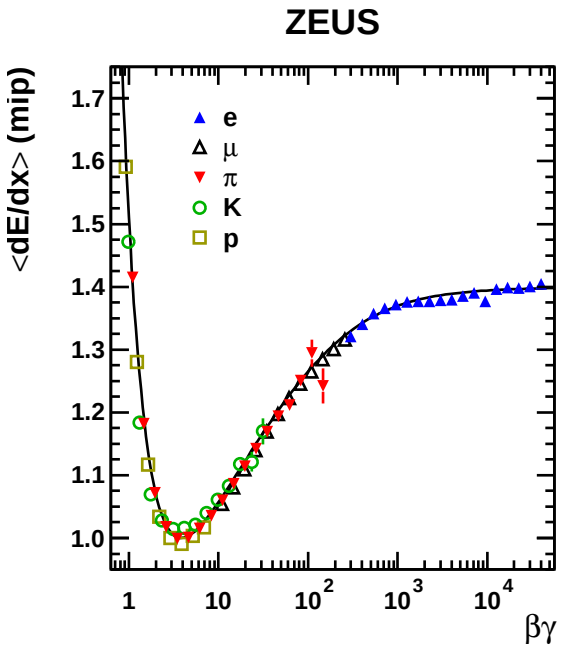


Figure 4.2: *The mean dE/dx measured in the CTD, $< dE/dx >$, as a function of $\beta\gamma$ for different samples of identified particles as denoted in the figure. The curve shows a physically motivated parametrisation of the $< dE/dx >$ dependence on $\beta\gamma$ [16]*

calibrate and validate the dE/dx measurement. The samples used for calibration were:

- $e^{\pm}$ from photon conversions, J/ψ decays and DIS electrons;
- $\pi^{\pm}$ from K^0 decays with $0.4\,\text{GeV} < p < 1\,\text{GeV}$, where p is the measured track momentum.

The samples used for validation were:

- $\pi^{\pm}$ from K^0 outside the momentum range used for the calibration sample, as well as $\pi^{\pm}$ from ρ^0, Λ and D^* decays;
- $K^{\pm}$ from ϕ^0 and D^* decays;
- $p, \bar{p}$ from Λ decays;
- cosmic $\mu^{\pm}$.

Typical sample purities were above 99% for the calibration samples and well above 95% for the validation samples [107]. After several corrections the measured dE/dx only depends on the particle boost $\beta\gamma = \frac{p}{m}$ (see Figure 4.2). All

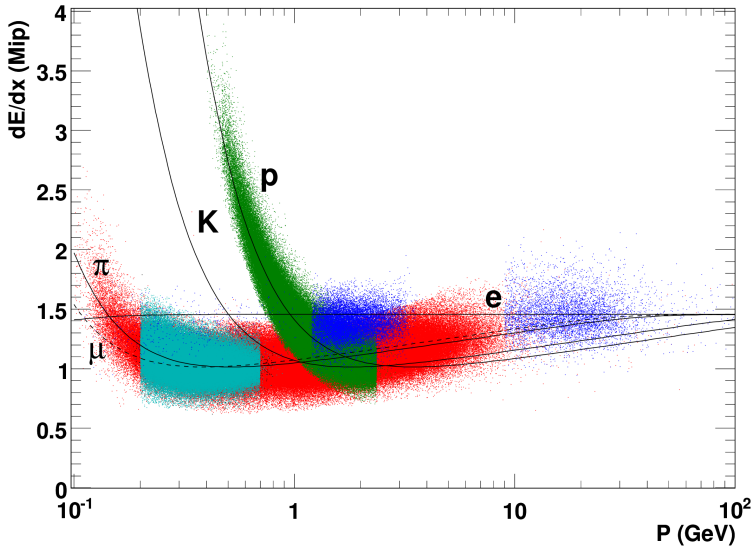


Figure 4.3: *The dE/dx measured in the CTD, as a function of the momentum, p, for the different validation samples. The curve show a physically motivated parametrisation of the $< dE/dx >$ dependence on $\beta\gamma$ unfolded to a momentum dependence using the corresponding particle masses.*

particle types are well described using a single physically motivated parametrisation of the mean energy loss as a function of $\beta\gamma$ with five free parameters, p_i, following Blum and Rolandi [108]:

$$\frac{dE}{dx} = \frac{p_1}{\beta^{p_4}}\left(p_2 - \beta^{p_4} - \ln\left(p_3 + \frac{1}{(\beta\gamma)^{p_5}}\right)\right). \tag{4.2}$$

This equation, which based on the photo-absorption-ionisation model, described the measured ionisation curves over a wide kinematic range.

As the CTD also provided a precise momentum measurement, this can be used to distinguish particles of different masses. A detailed description of this identification follows in Section 6.4.1. Figure 4.3 shows the measured values of dE/dx as a function of the particle momentum for the different validation samples defined above. The particle curves were obtained from the parametrisation as a function of $\beta\gamma$ using the appropriate particle masses.

This parametrisation determined from the data samples have directly been used to simulate the dE/dx in the Monte Carlo samples because the simulation of the energy loss obtained by the GEANT program were found to be not sufficient. Using the knowledge of the $\beta\gamma$ the mean dE/dx can be extracted from Equation 4.2. To simulate the dE/dx distribution in the Monte Carlo in addition to the the

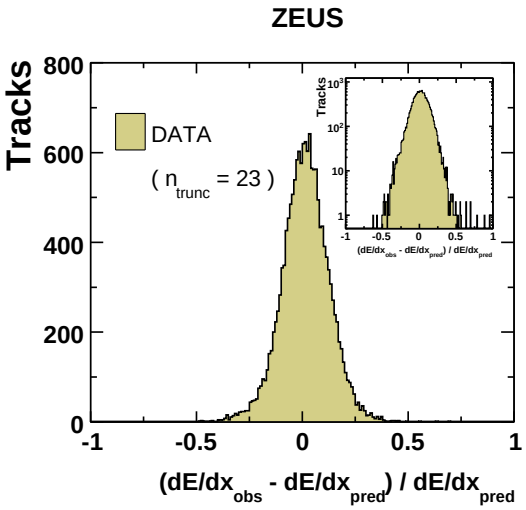


Figure 4.4: *Distribution of the relative difference between the observed (dE/dx_{obs}) and predicted (dE/dx_{pred}) specific energy loss for the track sample with $n_{trunc} = 23$. The inset shows the same distribution with a logarithmic scale for the HERA I data [16].*

mean dE/dx also the residua have to be described correctly. These residua depend on the number of CTD hits, $\sim \frac{1}{\sqrt{N_{trunc}}}$. Figure 4.4 shows the distribution for n_{trunc} equal to 23 which is a typical value for the samples under study. The random procedure to produce dE/dx values which follow these distributions can be found in [109]. The mean dE/dx and the residua have been redetermined for the HERA II data samples [110].

4.3 HERA I Primary Vertex Reconstruction

Once the tracks have been found and fitted, the primary vertex, as well as secondary vertices originating from the decay of long lived particles or the interaction of particles with the material of the detector, can be reconstructed. The primary vertex (position of the ep interaction) is fitted for each event using a global vertex fit [111] followed by a deterministic annealing filter [112]. Tracks that are consistent with originating from the primary vertex are refitted with the primary vertex as an additional step.

4.4 MVD Information

Since the end of 2002 a new version of VCTRAK including the MVD information was available. This allows improved studies of decays containing heavy quarks such as charm or beauty, having long lifetimes ($c\tau \geq 100\,\mu$m) . The direct impact of the improved tracking was [113]:

- a significant improvement in the precision of the track helix parameter determination;
- a better separation power of two nearby tracks down to 200 μm;
- a global increase in track finding efficiency from ~93.5% to ~97%;
- and an improvement in the determination of primary as well as secondary vertex positions.

To extract lifetime information from the precise tracking and secondary vertices, firstly the best possible knowledge of the primary interaction point is required.

4.4.1 Beam-spot

In cases where the spread of a variable is smaller than its single reconstruction resolution, the accuracy of that variable can be improved by replacing the value by its average. This is the case for the vertex position in the xy-plane while the z position still has to be recalculated for each event, as its width is mainly influenced by the proton bunch length. The average of the xy-spot calculated as the mean vertex position for a set of ep events is called the beam-spot. Typical values for the sigmas of a Gaussian fit to the distributions are $\sim 180\,\mu\text{m}$ for x and $\sim 170\,\mu\text{m}$ for y with a precision of $O(1\,\mu\text{m})$ for the mean. Detailed studies how to extract the values can be found in [113]. As the beam and detector axis are not parallel but have a tilt of the order of a few mrad the beamspot has to be corrected for the z position of the primary vertex event-by-event. This correction is taken into account to extract the impact parameter, i.e. the point of closest approach of a helix with the beamspot as a reference point.

4.4.2 HERA II Primary Vertex Reconstruction

For the HERA II data set the primary vertex could be determined with higher precision. Origins for the higher quality were the improved tracking and the beamspot as an additional ingredient for the fit. The beamspot, which was calculated using the primary vertex position in the $r\phi$-plane determined by taking the mean position in a few thousend events, had a positive effect especially for events with only a few good tracks or heavy flavour events, where the tracks are expected to have a larger displacement on average.

4.4.3 Impact Parameter

The *impact parameter*, IP, is a track-based variable, which can be used to separate tracks that are consistent with originating from the primary interaction point, from tracks which originate from a long-lived particle, and have some displacement. Of special interest is the signed IP, where the sign is defined by the angle of the displacement vector with respect to the axis of the associated jet. This is illustrated in Figure 4.5.

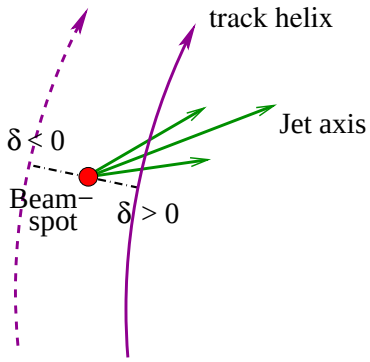


Figure 4.5: *Schematic view of the impact parameter. The minimal distance between the track helix and the beamspot is called impact parameter. The sign of the impact parameter is defined by the relative direction to the jet axis.*

Due to the long hadron lifetime the tracks from heavy flavour interactions have some displacement in the direction of the hadron and therefore in the direction of the jet. This results in a positive tail in the impact parameter distribution. To have a better control of resolution effects and therefore, whether a displacement is consistent with no displacement, the significance of the IP is calculated:

$$\delta_{IP} = \frac{IP}{\sigma_{IP}}, \tag{4.3}$$

where σ_{IP} is the uncertainty on the impact parameter calculated from the beamspot and the helix uncertainties. Several methods to use the impact parameter based variable in the signal extraction are described in Chapter 6.

4.4.4 Decay Length

Another approach to use lifetime information provides the measured *decay length*, DL, which, in contrast to the IP, is a jet-based (see Chapter 4.6) variable. To calculate secondary vertices, which are candidates for the heavy flavour decay, the good tracks in a jet are combined and a vertex is fitted. The good tracks are defined by a minimal momentum of $p_T > 0.5\,\text{GeV}$ and some minimal number of

hits ($N_{MVD} \geq 4$ and $N_{SL} \geq 2$) The decay length significance, δ_{DL}, can also be calculated:

$$\delta_{DL} = \frac{DL}{\sigma_{DL}}, \tag{4.4}$$

where σ_{DL} is the error of the decay length calculated from the beam-spot and the covariance matrix of the vertex position. The sign of the DL is calculated from the angle between the decay length direction and the jet axis, where the DL is negative if the angle in the xy-plane is larger than 90°. To avoid discontinuities at angles close to 90° and to include the expectation that the displacement should be in the hadron/jet direction the DL is projected onto the jet axis. Figure 4.6 shows a schematic view of the decay length and how the projection is done. As more in-

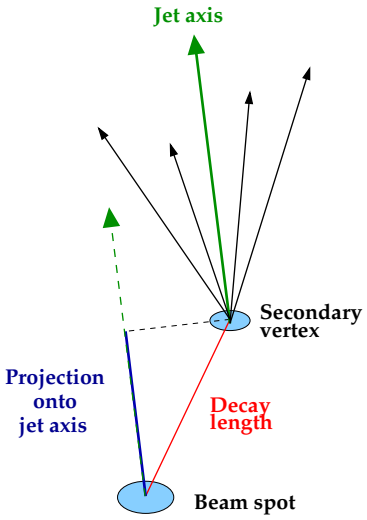


Figure 4.6: *Schematic view of the decay length calculation. The distance between the beamspot and the refitted secondary vertex is calculated in xy. An additional option is to use the projected decay length which is the component in the direction of the associated jet. Courtesy of [114].*

formation is combined in one variable, the DL has a larger separation power than the IP. On the other hand, there are more systematic effects affecting the calculation. In addition larger samples are needed, as the efficiency for reconstructing the secondary vertices is much lower than for the IP, which is defined for all tracks.

Another variable that could be included to distinguish between light flavour and heavy flavour contributions is the invariant mass of a reconstructed secondary vertex. The invariant mass is reconstructed using the 4-momenta of the tracks that enter the secondary vertex fit. For the heavy flavour identification one could either use the shape of the decay length significance distribution or split the samples into different mass regions in order to obtain beauty, charm and light flavour enriched regions. Figure 4.7 shows the distribution of the decay length significance for the whole mass region and for a subsample including high masses only ($2 < m_{vtx} < 7.5\,\text{GeV}$). The filled histograms shows the Monte Carlo distribu-

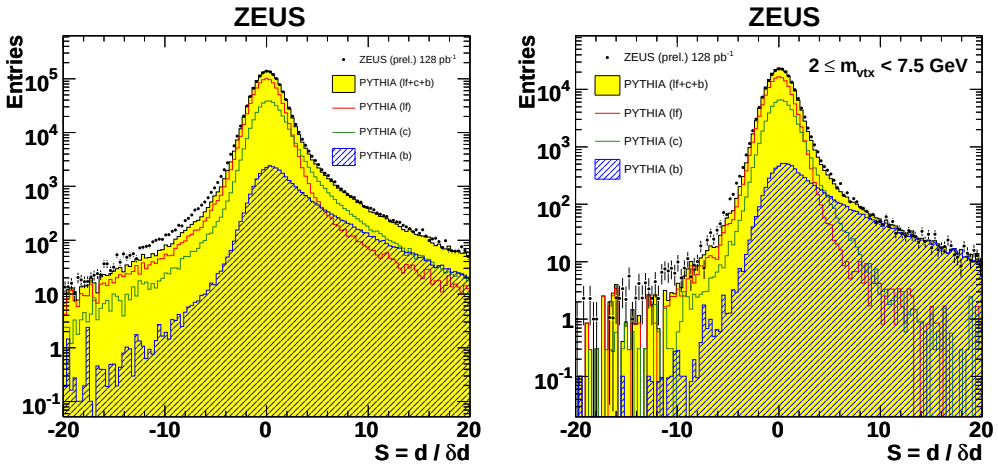


Figure 4.7: *Distribution of the decay length significance for all data (left) and for the beauty enriched sample (right). The sample has been enriched for heavy flavour [115].*

tion, split in the contributions from beauty, charm and light flavour events, which describes the data reasonable well. The non-vanishing decay length of the light flavour contribution originates from resolution effects, hence its distribution is symmetrically distributed around zero. In contrast to this, the heavy flavour contributions have larger tails due to the longer lifetime. The right distribution shows that, using the decay length significance, it is possible to extract a beauty sample with high purity.

4.5 Hadronic Final State Reconstruction

The information in the hadronic final state consists of charged as well as neutral particle components. The energy of both were measured by the CAL whereas for charged particles the momentum was measured by the tracking detectors. In the following the algorithm used to combine information in *Energy Flow Objects* (EFOs) is described.

The energy in the CAL was clustered into two-dimensional objects, called *islands*, associating energy deposits to single particle objects or to bunches of particles. The cells in the EMC as well as in the HAC1 and HAC2 were clustered layer by layer into islands, where the centre of the islands was defined by the weighted mean position of the contributing cells. To take the exponential decline of the shower energy distribution into account, the weight was determined with

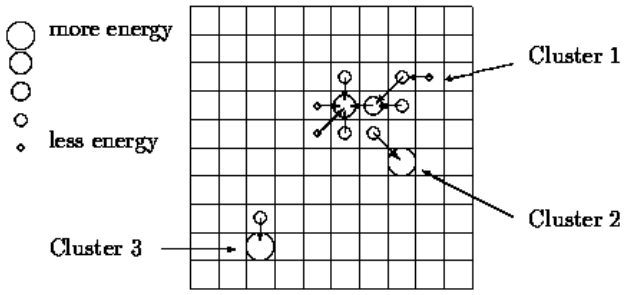


Figure 4.8: *Schematic view of the island clustering algorithm in the ZEUS calorimeter.*

the logarithm of the cell energy. Figure 4.8 shows a schematic view of the island clustering algorithm. In a second step the islands of different cell layers were combined into three-dimensional objects, called *cone islands*. To benefit from the precise momentum measurement of the tracking and the energy measurement of the energy depositions in the CAL, the tracks are extrapolated to the CAL surface and matched to the cone islands. In case of charged particles, the EFOs can consist of either single tracks, cone islands, or combined objects. For the latter, the energy is either determined from the CAL energy or, if the relative uncertainty of the momentum measurement is smaller than the energy measurement in the CAL, from the momentum under the assumption of a pion mass. For neutral particles only cone islands are seen. Figure 4.9 illustrates the different types of energy flow objects. A detailed description of the EFO reconstruction is given in [116].

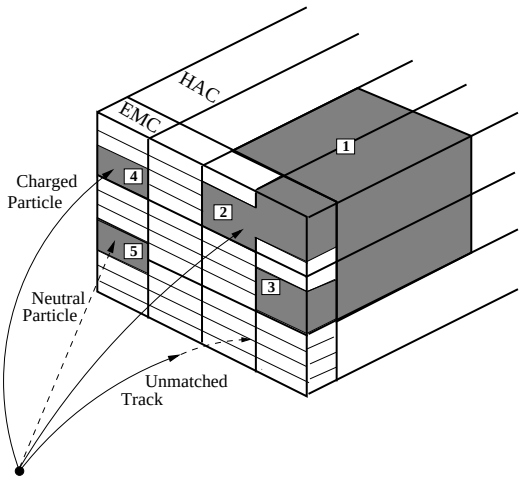


Figure 4.9: *Reconstruction of energy flow objects: EMC cell islands 2 and 3 are joined with HAC cell island 1 to form a cone island. In the next step the cone islands are matched to tracks (4). Good tracks which are not associated with any calorimeter object, i.e. unmatched tracks, are counted as charged energy. Calorimeter objects not associated with any track are counted as neutral energy (5). Courtesy of [116].*

The reconstructed EFOs are the input for the jet finding algorithm and the kinematic variable reconstruction as well as for the electron candidates as described in Chapter 5.3.2.

4.6 Jet Reconstruction

High p_T events are typically characterised by abstract objects, called jets, which reflect the energy and direction of the originating partons, fragmented into a bunch of neutral and charged particles. In order to compare experimental measurements with theoretical predictions, these jets have to be defined at *Detector Level* (DL) and *Hadron Level* (HL) as well as at *Parton Level* (PL).

As the number of jets and their properties are used for trigger purposes, a fast reconstruction algorithm is also needed. For this purpose the cone algorithm was used. This algorithm maximises the transverse energy, summed over the cells in a cone of radius R[1] ($R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$), where $\Delta\eta$ and $\Delta\phi$ are the distances in pseudorapidity and transverse momentum between the energy deposit and the jet axis. This purely geometrical algorithm was used for trigger preselection. Problems occur in case of overlapping jets, where the energy in the common region has to be split between two jets, or both jets are merged into one object.

Important requirements for the used jet finding algorithms are collinear and infrared safety. This means that the result is not affected if a massless parton is split into two massless partons or if an additional infinitely soft parton is added. A more advanced algorithm, which fulfils these requirements, is the commonly used k_t algorithm [117]. For this algorithm the important variable is $k_t = min(E_i^2, E_j^2)(1 - \cos\Theta_{i,j})$, where E_i, E_j are the particle energies and $\Theta_{i,j}$ denotes the opening angle between all possible cluster pairs. If the value of the pair with the minimal k_t is below some cut-off parameter, $(y \cdot E_T^2)$, where E_T is the total transverse energy of the event, the pair is combined into a single object. The proton direction acts in this procedure as an additional object with infinite energy, to associate clusters in the forward region with the proton-remnant. These clusters are not considered further in the jet-finding procedure. In an iterative procedure the pairs are merged into pseudo-particles until no pair with k_t below the cut-off is found. The algorithm allows the merging procedure to clearly separate jets from the proton-remnant, which has relatively little transverse energy. This avoids the contamination from the proton-remnant in the jets near the forward beam-pipe. Also, the k_t algorithm avoids the ambiguities related to the overlapping and merging of jets.

This algorithm has been used to define jets on the three different physical levels, the jet algorithms are initialised with different input particles. For the detector level, the jet finder is running over corrected EFOs to take the energy impact of charged and neutral particles to be as close as possible to the hadron level, at which the cross sections are defined. For Monte Carlo events the jets at hadron level are defined by the true hadronic final state. Therefore the algorithm runs

[1] typically the maximum radius is set to $R = 1$

over all stable[2] final state particles except neutrinos.

The theoretical calculations in pQCD are given on parton level. Therefore jet objects also have to be defined at parton level. At parton level the jet finder runs over all partons before the fragmentation process. The input particles for each level are assigned with a reconstructed four momentum $q^i = (p^i_x, p^i_y, p^i_z, E^i)$ and define the kinematics of the jets as:

$$q^{jet} = \sum_i q^i, \quad E^{jet}_T = E^{jet} \cdot p^{jet}_T / p^{jet}, \tag{4.5}$$

where the sum runs over the particles described above. E^{jet}, p^{jet} and p^{jet}_T are the energy, momentum and transverse momentum of the jet, respectively.

4.7 Reconstruction of Kinematic Variables

As shown in Section 2.1 two of the variables x, y and Q^2 are necessary to describe the event kinematics. There are several methods to reconstruct these variables. In DIS where the 4-momentum of the scattered electron can be used to reconstruct the event properties, methods like the *Electron Method* and the *Double Angle Method* can be used [118]. In events where the scattered electron cannot be detected, as in photoproduction where the electron remains in the beam-pipe, the *Jaquet-Blondel Method* [119] can be used. In this method the kinematic variables are defined by summing over the four-momenta of the EFO objects.

$$Q^2 = \frac{(\sum_i p^i_x)^2 + (\sum_i p^i_y)^2}{1 - y_{jb}} \tag{4.6}$$

$$x_{jb} = \frac{(\sum_i p^i_x)^2 + (\sum_i p^i_y)^2}{s y_{jb}(1 - y_{jb})} \tag{4.7}$$

$$y_{jb} = \frac{\sum_i (E^i - p^i_z)}{2E_e} \tag{4.8}$$

4.8 Event Display

Figure 4.10 show different views at the top of the figure and zoom factors for an example event in the ZEUS event display, ZeVis. Some of the kinematic variables

[2] All particles with a lifetime of $\tau > 0.01$ ns are defined as stable

calculated from the hadronic system are listed. In the display the jet objects are shown as black arrows, indicating the direction and length of the arrow proportional to the energy. The EFOs are shown as green arrows built from the track objects and CAL entries. The lower plots show the zoomed region of the MVD, including two secondary vertices with significance larger than three.

With these reconstructed variables it is now possible to define the event and track selections which define the accessible kinematic region at detector level.

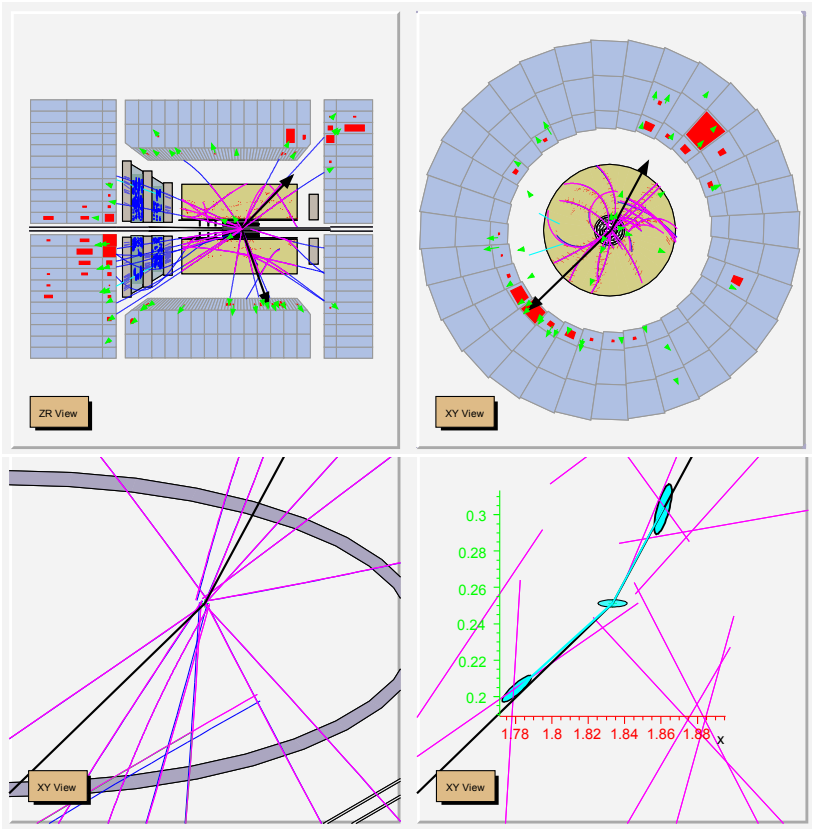


Figure 4.10: *Reconstructed objects illustrated for an example event in the ZEUS event display ZeVis. The whole event including the inner detector and the calorimeter is shown in zr-view (upper left) and in xy-view (upper right) where the two reconstructed jets, the reconstructed tracks and the calorimeter energy are shown. On the lower row the xy-view is shown zoomed into the MVD region (lower left) and zoomed to µm-range (lower right). The ellipses show the beamspot and the secondary vertices with their uncertainties, which are used to calculate the decay length significance.*

Chapter 5

Event Selection

In this chapter the selection of dijet photoproduction events and of candidates for electrons from semileptonic decays in dijet γp events is summarised. The selection should enrich the heavy flavour processes:

$$e^{\pm} p \rightarrow b\bar{b}(c\bar{c})e'^{\pm}X \rightarrow e^{\pm}_{SL} + jj + e'^{\pm}X, \tag{5.1}$$

where $e^{\pm}_{SL}$ is the electron originating from a semileptonic decay of a b- or c-hadron. The scattered electron $e'^{\pm}$ remains undetected as it is scattered under too low an angle. The aim of the event selection is to accept as many events of interest as possible and to reject background events. A good background removal is important for the final signal extraction, but even more important to be able to describe the accepted data by the simulated Monte Carlo set. Therefore it has to be checked how well the used variables are reproduced by the Monte Carlo. The preselection cuts defined below have been applied to the following data and Monte Carlo sets.

5.1 Data and Monte Carlo Sets

For the HERA I analysis the full data-set from 1996-2000 has been used (see Table 3.2). This data-set corresponds to an integrated luminosity of $\sim 120\,\mathrm{pb}^{-1}$ taken by the ZEUS detector. For the extension to the HERA II period the data of the years 2006-2007[1] has been used. In Table 5.1 the used Monte Carlo sets are summarised. The samples were split according to the production process which are described in detail in Chapter 2.4. For the HERA I analysis the beauty contribution has been removed from the inclusive dijet γp Monte Carlo and was simulated

[1] At the time when this analysis was performed, this was the data-set made with the latest corrections and software version.

separately using the massive calculations. For HERA II the charm contributions have been treated analogously to the beauty Monte Carlo and had to be removed from the inclusive sample as well. The electron spectrum in the charm sample was corrected to the measured spectrum from the CLEO collaboration [120]. In Figure 5.1 the momentum spectrum of leptons in the centre-of-mass-system of the charmed hadron is given for the PYTHIA Monte Carlo sample, which is compared with the measured spectra of D^0 and D^+ decays. The spectrum extracted from the PYTHIA Monte Carlo has a good agreement with the measurements from CLEO.

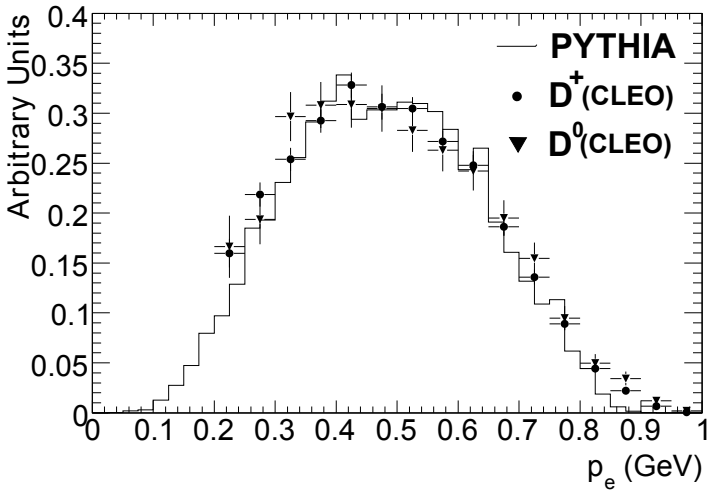


Figure 5.1: *Momentum spectrum of electrons originating from semileptonic decays in PYTHIA, measured in the rest frame of the charm hadron. The spectrum is compared with the spectra for D^0 and D^+ decays measured by the CLEO collaboration [120].*

5.2 Online Selection

As described in Section 3.2.6 the ZEUS trigger system was based on a three level trigger for fast data selection. For the preselection an inclusive dijet γp trigger (HFL05) was used. The trigger corresponded to the following cuts:

$$E_T^{jet_{1,2}} > 4.5\,\text{GeV}, |\eta^{jet_{1,2}}| < 2.5, \tag{5.2}$$

$$(E - p_z) < 100\,\text{GeV}, \tag{5.3}$$

	$\mathcal{L}(\text{pb}^{-1})$					σ_{318}(pb)
Process	$96/97e^+$	$98/99e^-$	$99/00e^+$	$06e^-$	$06/07e^+$	
b in dir.	250	27	380	490	1500	4100
b in res.	260	99	360	500	1600	710
b in exc. γ	240	72	360	540	1400	1300
b in exc. p	290	85	420	550	1800	410
c in dir.	-	-	-	270	560	150000
c in res.	-	-	-	270	660	12000
c in exc. γ	-	-	-	270	570	260000
c in exc. p	-	-	-	280	570	56000
light flavour + (c) in dir.	44	18	73	69	140	660000
light flavour + (c) in res.	45	17	65	69	140	7500000

Table 5.1: *Integrated luminosities of the used Monte Carlo samples. The corresponding cross sections are given for a centre-of-mass energy of $\sqrt{s} = 318\,GeV$. For the HERA I samples, the charm processes are simulated together with the light flavour processes in an inclusive sample, with the beauty processes subtracted The direct (dir.), resolved (res.) and excitation (exc.) processes were simulated separately.*

$$\frac{p_z}{E} < 0.95, \tag{5.4}$$

where E and p_z denote the energy and z-momentum of the hadronic system. Additionally a reconstructed ep vertex was required to reduce beam-gas and cosmic events already at this stage.

5.3 Offline Selection

The kinematic cuts on the reconstructed variables (detector level) were the same as the truth selection. Offline selection criteria are much less critical than the trigger selection, because events that do not pass the trigger are lost and are hard to correct for. The selection at the stage of private data set production or at the final analysis of the data sample can be tuned and tested for inefficiency by varying

the cuts to study systematic effects. In the following, the selection at first for the dijet γp events and secondly for the e_{sl} candidates are summarised.

5.3.1 Event Selection

The following selection was applied on the well reconstructed[2] events:

- *γp events:* the kinematic region of this analysis was restricted to $Q^2 < 1\,\mathrm{GeV}^2$ and $0.2 < y < 0.8$. On the reconstructed variables this phase space was selected by cutting at $0.2 < y_{jb} < 0.8$. The inelasticity was restricted to reduce remaining DIS events in the region of high y_{jb} and beam-gas events which were predominant at low y_{jb}. Secondly, as the scattered electron was undetectable for γp events, a veto on electrons found by the Sinistra finder[3] with energy of more than 5 GeV was applied:

$$\mathcal{P}^{sira} > 0.9, E_e > 5\,\mathrm{GeV}, y_{el} < 0.9, \tag{5.5}$$

  where $\mathcal{P}^{sira}$ is the probability which is given by the Sinistra finder.

- *Dijet events:* Dijet events were defined by requiring at least two high energetic ($E_T^{jet^{1(2)}} > 7(6)\,\mathrm{GeV}$) jets in the central part of the detector ($|\eta^{jet^{1,2}}| < 2.5$). In contrast to the trigger level, the k_t cluster algorithm was used here to reconstruct the jets.

- *Cleaning cuts:* The protons and electron beams collided at the nominal interaction point at $z = 0\,\mathrm{cm}$. For HERA II (HERA I) only events were accepted that fulfil $|Z_{vtx}| < 30(50)\,\mathrm{cm}$. The cut reduced background from beam-gas events, collisions with the collimators, satellite bunch interactions and events with a mismeasured vertex.

5.3.2 Electron Selection

In this section the preselection cuts on the electron candidates are summarised. The measurement was performed for $p_T^e > 0.9\,\mathrm{GeV}$. To be able to use dE/dx for the particle identification, it was necessary to restrict the electron candidate selection to the central region ($|\eta^e| < 1.5$). The likelihood method used for particle identification is described in detail in Chapter 6.

[2]For a good event, which is characterised technically by the flag EVTAKE the main detector components like CTD and, for HERA II, MVD must have worked properly.

[3]Electron finder to detect and reconstruct the scattered electron [121, 122].

- *EFO object:* Electrons are charged particles with a clear energy deposit in the electromagnetic calorimeter, so they are expected to have a well reconstructed track which can be matched via the EFO algorithm to an island in the calorimeter (”1:1” EFOs). This unique connection is also needed for the particle identification. In Figure 5.2 the distribution of the EFO types is shown for all EFOs and for the subsample of true electrons[4]. The signal-to-

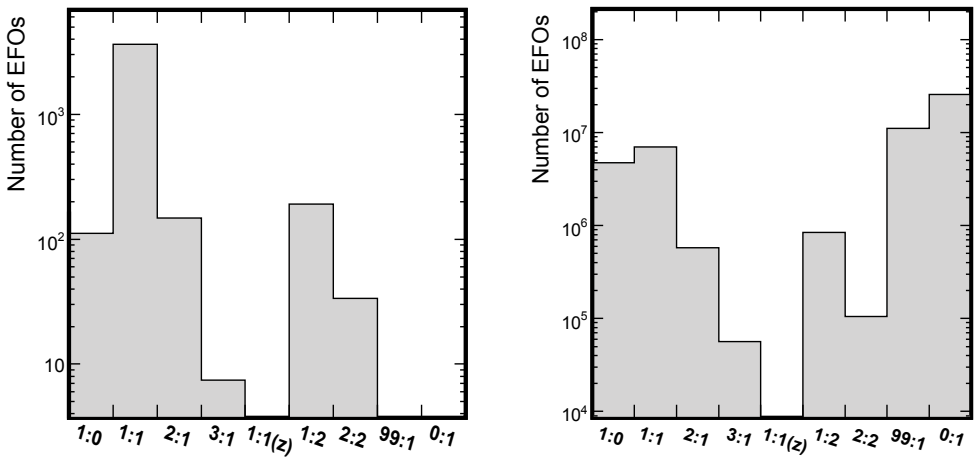


Figure 5.2: *EFO types in the Monte Carlo sample before the ”1:1” EFO requirement for the true electrons (left) and for all candidates (right). The dominant part of the true electrons are reconstructed as ”1:1” so that only a small signal fraction is lost by this cut. The notations ”99:1” or ”1:1(z)” are technical description for the EFOs were the association to the tracks, or CAL information, respectively, did not work properly. The Figure is taken from [14].*

background ratio was significantly improved by this cut without losing too many signal electrons.

- *Energy deposit:* One important variable to distinguish electrons from hadrons, using the different shower shapes in the calorimeter, was the fraction of EFO energy taken from the calorimeter information, E^{CAL}, which was deposited in the electromagnetic part, E^{EMC}. For the HERA I analysis the variable E^{EMC}/E^{CAL} was taken as an input variable for the likelihood hypothesis (see Chapter 6), while for the HERA II analysis this has been replaced by a hard cut at $E^{EMC}/E^{CAL} > 0.95$ and at $E^{CAL} > 0.5\,\text{GeV}$ (see Section 7.3), motivated by studies showing that the gain of the additional information does not compensate for the possible systematic uncertainty.

[4]Note that for (CAL 99:1) and (CAL 0:1) it is not possible to make a matching to the Monte Carlo truth information.

- *Track quality:* For the HERA II data-set, where it was desirable to exploit the lifetime information, additional requirements on the track quality were necessary. To reduce long-living K^0 and Λ particles and particles coming from secondary interactions with detector material, the tracks must have an impact parameter with $IP < 1\,\text{cm}$. To have a well-reconstructed impact parameter at least four MVD hits were required.

- *dE/dx measurement:* In contrast to the HERA I analysis where a minimal dE/dx of 1.1 mip had been required to suppress insufficiently described fake electrons, this cut was skipped for the HERA II analysis and was replaced by a more sophisticated cut on the later described dE/dx likelihood (see Section 6.4.1) at $-2\ln T_e^{dE/dx} < 3$ was applied.

- *Photoconversions:* To reduce electrons originating from photoconversions, candidates found by the conversion finder were removed. The way how the conversion finder was used and improved during the analysis is described in Section 5.4.

- *Jet association:* For the calculation of some variables it was necessary to associate the track to a good jet ($E_T^{jet} > 6\,\text{GeV}, |\eta^{jet}| < 2.5$) in the $\eta\phi$-plane with $\Delta R[e, jet] = \sqrt{\Delta\eta^2 + \Delta\phi^2} < 1.5$. For the HERA II analysis the cut was tightened to $\Delta R[e, jet] < 1$. The effect at the detector level was negligible but avoided ambiguities in the extrapolation of the jets to hadron or detector level. Studies on the jet association can be found in [14].

- *Geometry cuts:* At the transition between FCAL and BCAL or RCAL there were larger gaps between the CAL blocks (supercracks). If cells from different calorimeter regions were combined in the clustering process, the energy and the position cannot be reconstructed properly (see Figure 5.3). The energy reconstruction and the connection of the cell entries to the tracks via the EFO objects is very important for the electron reconstruction, so the problematic region had to be cut out. The HERA I cut the EFO depth, $d < (\frac{(\theta-90)^2}{500} + 18)\,\text{cm}$, was replaced by a cut based on the detector-geometry. The radius of the EFO position was restricted to $r_{cell} < 136\,\text{cm}$ in the central region and $r_{cell} < 175\,\text{cm}$ in the other regions. Additionally the supercracks were cut at $-160\,\text{cm} < z_{cell} < -110\,\text{cm}$ and $178\,\text{cm} < z_{cell} < 232\,\text{cm}$. On the upper side of the BCAL, a small region for the so-called chimney was not well instrumented. This region at $z < -110\,\text{cm}$ with $y_{cell} > 80\,\text{cm}$ and $|x_{cell}| < 12\,\text{cm}$ was removed from the reconstruction. The combination of these cuts is illustrated by the shaded area in Figure 5.4 – on the left for the chimney cut and for the depth and the supercracks on the right.

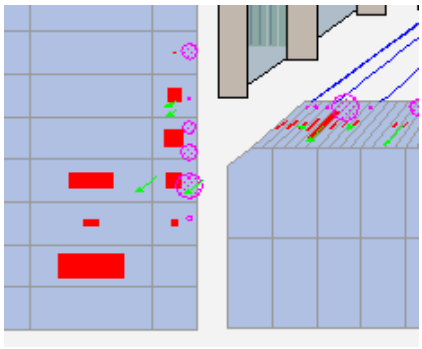


Figure 5.3: *Example event with track going into the η region where the energy are deposited around the supercrack. If the energy of a track is deposited in two regions of the CAL, the clustering process cannot work properly. Both the energy as well as the position of the EFO objects were affected by this.*

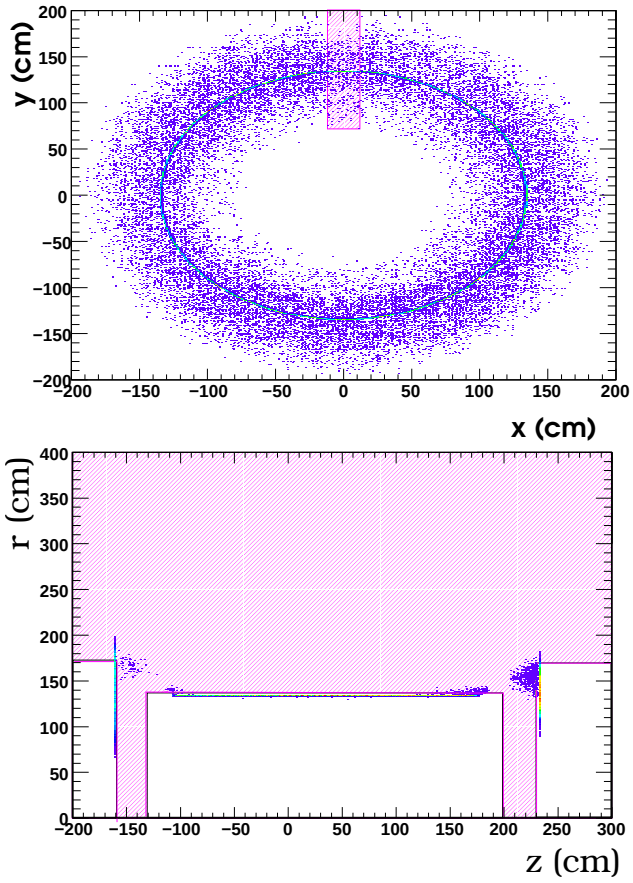


Figure 5.4: *Distribution of EFO position of the electron candidates in xy (top) and zr (bottom). The region removed by the geometry cuts are indicated by the dashed areas.*

5.4 Conversion Finder

Not only fake electrons like pions, protons or kaons contribute to the background, but also electrons from sources other than semileptonic decays. An important source of electrons were photons originating from π^0 decays, which produce electron pairs via photon conversion ($\gamma \rightarrow e^+e^-$). These conversions took place via electromagnetic interactions with the detector material, e.g. the beam-pipe. It was important to have a reliable suppression of these electrons, to avoid conversion vertices reconstructed positions as secondary vertices from long lived particles. In order to identify and suppress these electrons, a conversion finder was used, and extended to take the improved tracking of the HERA II data into account. This programme used the opening angle between the tracks of the electron-positron pair and the reconstructed invariant mass of the photon (see Figure 5.5). Both the opening angle as well as the invariant mass of the two opposite charged tracks should be distributed around zero, while the widths depend on the resolution of the track reconstruction.

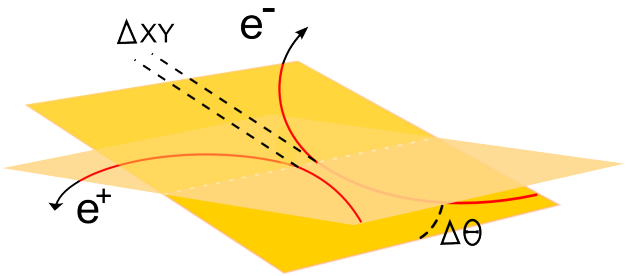


Figure 5.5: *Schematic view of the conversion planes. The angle $\Delta\theta$ denotes the opening angle between the two track planes and Δxy is the minimal distance between the two tracks in the xy plane. From the four momenta of the two tracks one can calculate the invariant mass of the photon candidate.*

Using the resolution, the opening angle, $\Delta\theta$, and the minimal distance in the xy-plane, Δxy, a combined dimensionless distance parameter, D, could be defined:

$$D = \sqrt{\left(\frac{\Delta xy}{\sigma_{xy}}\right)^2 + \left(\frac{\Delta\theta}{\sigma_\theta}\right)^2} < 15$$

For the HERA I analysis the following values were used [123]:

$$\sigma_{xy}^{data} = (0.08 \pm 0.01)\,\text{cm}, \tag{5.6}$$
$$\sigma_{xy}^{MC} = (0.11 \pm 0.02)\,\text{cm}, \tag{5.7}$$
$$\sigma_{\theta}^{data} = (0.017 \pm 0.0001)\,\text{rad and} \tag{5.8}$$
$$\sigma_{\theta}^{MC} = (0.017 \pm 0.0002)\,\text{rad}. \tag{5.9}$$

As mentioned above, the second important quantity distinguishing electrons from photoconversions were the invariant mass of the photon. For electron pairs, each with $P_e \geq 250\,\text{MeV}$, the invariant mass must be below $m_x < 250\,\text{MeV}$[5]. In addition, the z-difference of the two tracks at the point of closest approach must be consistent with zero.

5.4.1 HERA II Extension

For the HERA II data the routines had to be extended to take into account the combined tracking and to benefit from the improved tracking resolution. The precision of the conversion parameters, reflected in the resolution of the input variables, depended on the track quality. The additional MVD information yielded to more track classes available than in the HERA I data. Due to the larger statistics which was used to determine the resolutions it was possible to split the sample into subsamples, yielding a better separation power of the conversion candidates. The main variables affecting the resolutions were the number of hit superlayers in the CTD (N_{CTD}), the number of reconstructed MVD hits (N_{MVD}), and the radial position of the conversion vertex. The five selected classes for the tracks were *CTD* ($N_{CTD} \geq 3 \,\&\, N_{MVD} < 4$), *MVD* ($N_{CTD} < 4 \,\&\, N_{MVD} \geq 4$) and *CTD+MVD* ($N_{CTD} \geq 3 \,\&\, N_{MVD} \geq 4$), where the two classes containing MVD hits were further split by their vertex position to be inside or outside the first MVD layer.

Figure 5.6 shows the resolution of the opening angle for *CTD* and for *MVD* tracks reflecting the improved tracking of the micro-vertex detector. To estimate the remaining background in the conversion finder and the effect on the resolution, the conversion finder had been applied for same-charged candidates as well. As the background had no preference for the charge combination, the same charge distribution could be used to make a background subtraction. The figures show that both the opposite-charged as well as the same-charged candidates are well described by the Monte Carlo.

[5]This cut was used for rejecting conversions. To select conversions with high purity harder cuts were used.

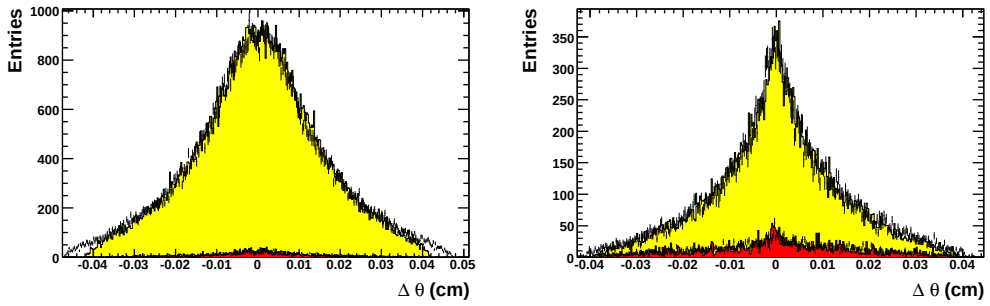


Figure 5.6: *Comparison of the opening angle distributions for CTD tracks (left) and MVD tracks (right). The histograms show the Monte Carlo distributions for the opposite-charge combination (yellow) and same-charge combination (red). The distributions agree quite well with the two corresponding data distributions shown as points.*

For each of the samples, the width of the distribution has been determined. To subtract the background contribution, the distributions for same-charge candidates had been fitted by a Gaussian function and the subtracted distributions were used to determine the resolutions. Figure 5.7 shows the subtraction procedure on the example of the $\Delta\Theta$ distribution for *MVD* tracks.

	$\Delta\Theta$ (rad)		Δxy (cm)	
	DATA	MC	DATA	MC
MVD + CTD (out)	0.019	0.017	0.060	0.050
MVD + CTD (in)	0.018	0.017	0.051	0.045
MVD (out)	0.012	0.014	0.060	0.060
MVD (in)	0.013	0.019	0.066	0.061
CTD	0.021	0.019	0.070	0.066

Table 5.2: *Resolutions of $\Delta\Theta$ and Δxy for the five different track qualities defined in the text. The resolutions have been determined separately for data and MC [124].*

Table 5.2 shows the resolutions for the five subsamples, calculated for data and Monte Carlo. The best resolution in $\Delta\Theta$ could be achieved for the *MVD* tracks where the precision was significantly better than for *CTD* tracks and also better than the values determined for HERA I. For the resolution in xy, the differences were less pronounced but have similar trends. Most of the classes do not show significant differences between data and Monte Carlo, where e.g. differences in the $\Delta\Theta$ resolution in the inner MVD region are still not understood.

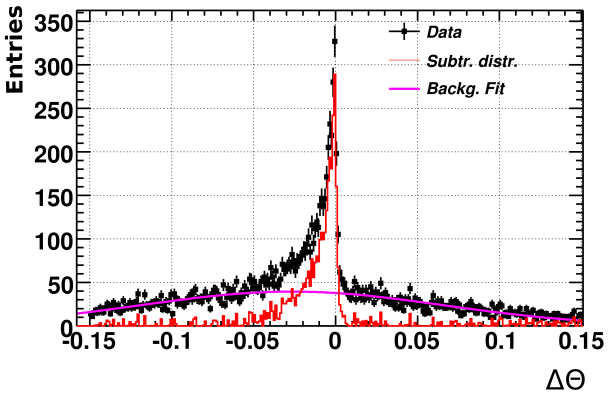


Figure 5.7: *Example plot for the determination of the* $\Delta\Theta$ *resolution for MVD tracks. The opposite-charged background has been fitted separately before fitting the background subtracted-distribution. The fits have been performed for data and for MC separately.*

Truth Information

As a cross-check, the background contribution was extracted by using the truth information in addition to the opposite-charge method. Each track had been matched to a Monte Carlo particle, and by using the truth decay information, it was checked whether the electrons originated from a photon. The amount and shapes for the truth signal were found to be consistent with the subtracted distributions. In addition the truth information was used to identify all true photoconversions in the Monte Carlo events. Figure 5.8 shows the position of the conversion vertex using the Monte Carlo truth information. The distribution of the vertices reflects the detector material.

Reconstructed Vertex Position

The description of the detector geometry and its material could be checked by monitoring the position of the reconstructed vertex position of the electrons originating from photoconversions. The distribution of the vertices was reconstructed both in data and Monte Carlo, and comparing them, a reasonable agreement was found. Regions with discrepancies were identified with cooling pipes and cables from the laser alignment system that had not initially been implemented into MOZART, but were included in the finally used Monte Carlo set. In Figure 5.9, the reconstructed vertex position of the photoconversions are shown in the $r\phi$-plane (left) and in the xy-plane (right). The figure has been extracted from the

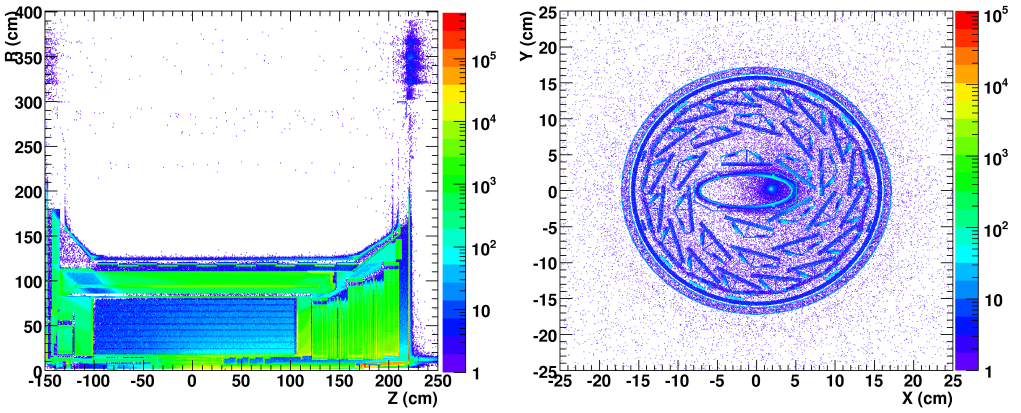


Figure 5.8: *Distribution of the conversion vertex using Monte Carlo truth information for the rz-plane (left) and for xy-plane in the central region (right). The position of the vertices reflects the distribution of the dead material in the detector. The colour scale represents the number of entries per bins.*

light flavour Monte Carlo sample. Due to the resolution, the detector structure was not as clearly visible as in Figure 5.8. In addition the range that can be scanned is limited by the detector acceptance.

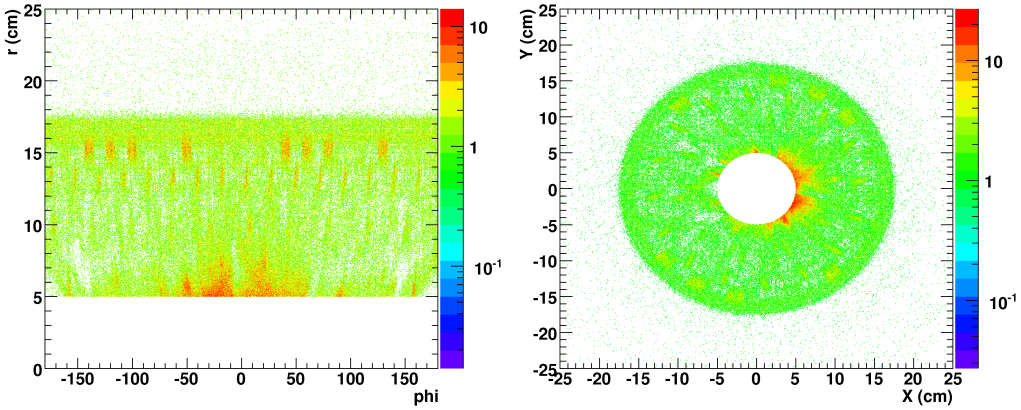


Figure 5.9: *Distribution of the conversion vertex position in data using the reconstructed vertex position for the rϕ-plane (left) and for xy-plane (right).*

Improvements

Figure 5.10 shows the improvement of the conversion finder using the new tracking and the re-determined resolution values. On these figures, where the quality

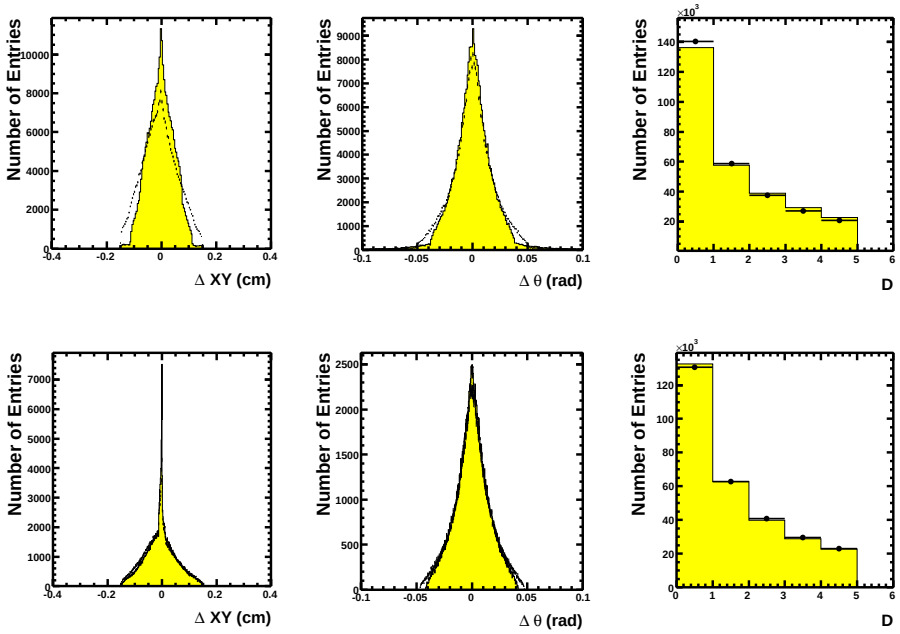


Figure 5.10: *The upper row shows the minimal distance, Δxy, the opening angle, $\Delta\Theta$, and the quality parameter, D, for data (points) and Monte Carlo (yellow) using the old resolution parameters. The same variables are shown on the lower row using the new resolution parameters. The resolution has been improved significantly and the Monte Carlo simulation describes the data much better with the new resolution parameters.*

variables are shown using the old (upper row) and the new (lower row) resolution values, one can see that the Monte Carlo description as well as the tracking was significantly improved. In Figure 5.11, some control plots for the conversion candidates are shown. The upper two figures show the quality parameter and the invariant mass, which both entered the conversion finding algorithm, while the lower two figures show the momentum and η distribution of the conversion candidate tracks. The remaining non-photoconversion contribution could be estimated using the same charge combination (red). All distributions show reasonable agreement between data and Monte Carlo. The discrepancies in the invariant mass had only small effects on the finding efficiency, as only a small fraction had values

above the cut value. Compared to the previous version of the conversion finder all distributions are better described.

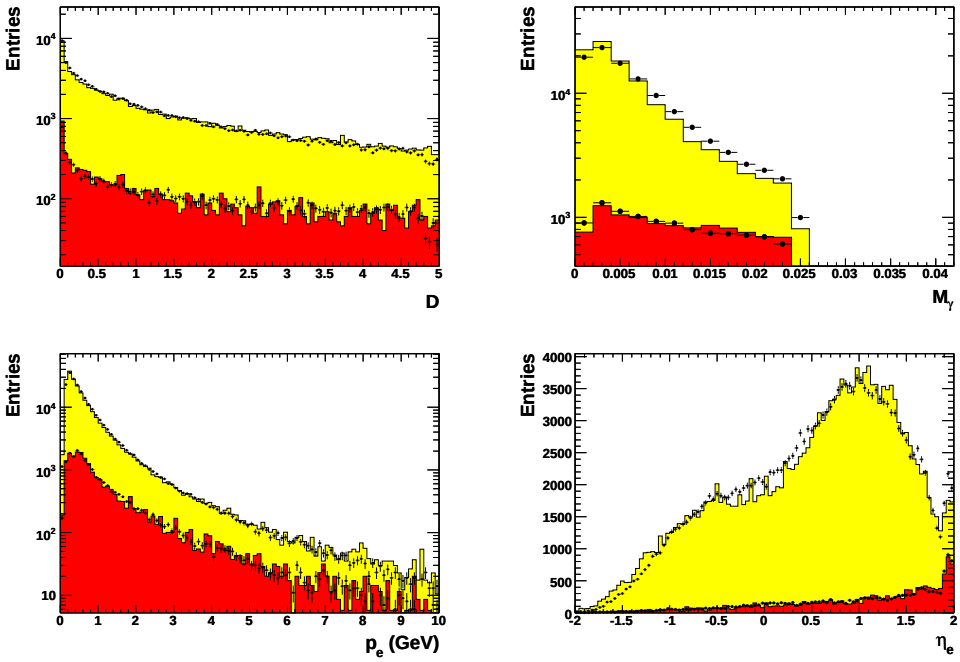


Figure 5.11: *Some control plots for the photoconversion finder. The figures show the distribution for the quality parameter, D, (upper left), the invariant mass, M_γ, (upper right), the electron momenta, p_e, (lower left), and the electron pseudorapidity, η_e, (lower right). The yellow histograms show the right charge combinations while the red contribution comes from wrong charge combinations. The points show the corresponding distributions for data.*

To estimate the impact of the improved conversion finder on the analysis, it was studied how many truth electrons remained in the sample after cutting on the candidates found by the conversion finder. Figure 5.12 shows the distribution of the electron likelihood hypothesis,$-2\ln T_e$, (defined in Chapter 6). The new conversion finder improved the conversion background suppression by more than a factor of two, with a negligible signal loss.

The updated conversion finder can detect $\sim 75\,\%$ of the conversions in the data sample. After additional track selection (see Section 5.3.2) the remaining fraction was further increased. In the final candidate sample $\sim$ 2% of the conversions remain.

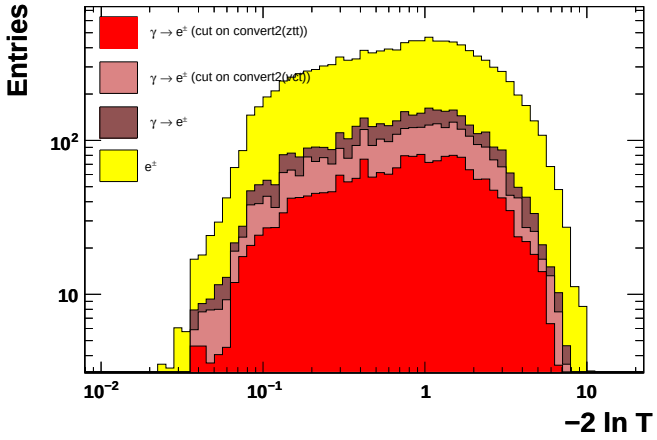


Figure 5.12: *Electron likelihood hypothesis (see Chapter 6) for all true electrons (yellow). The dark brown distribution shows the subsample of electrons originating from photoconversions. After the cut on the conversion finder based on the HERA I tuning (vct) the light brown contribution remains, which can further be suppressed by using the HERA II tune (ztt). The remaining electrons from photoconversions are shown in red.*

5.5 Control Distributions

After the selection of the electron candidates, the data quality has been checked with several control plots. All plots were done after the event and particle selection, which are summarised in Tables 5.3 and 5.4.

Figure 5.13 shows some example plots for the event variables, and in Figure 5.14, a selection of track control plots is shown. The figures show a reasonable description, but there are some contributions which are not well simulated. Possible origins for the deficits in the y distribution are remaining events from deep inelastic scattering in the data set or background from beamgas events. A reason for the bad description of the jet energy is the known bad description of the jet multiplicity. These regions had to be improved further, by cutting on variables which were used for the signal extraction, or restrict the fit region to well simulated areas. If this was not possible, one has to show that the discrepancies are either really coming from background and do not influence the signal extraction, or calculate their systematic effects.

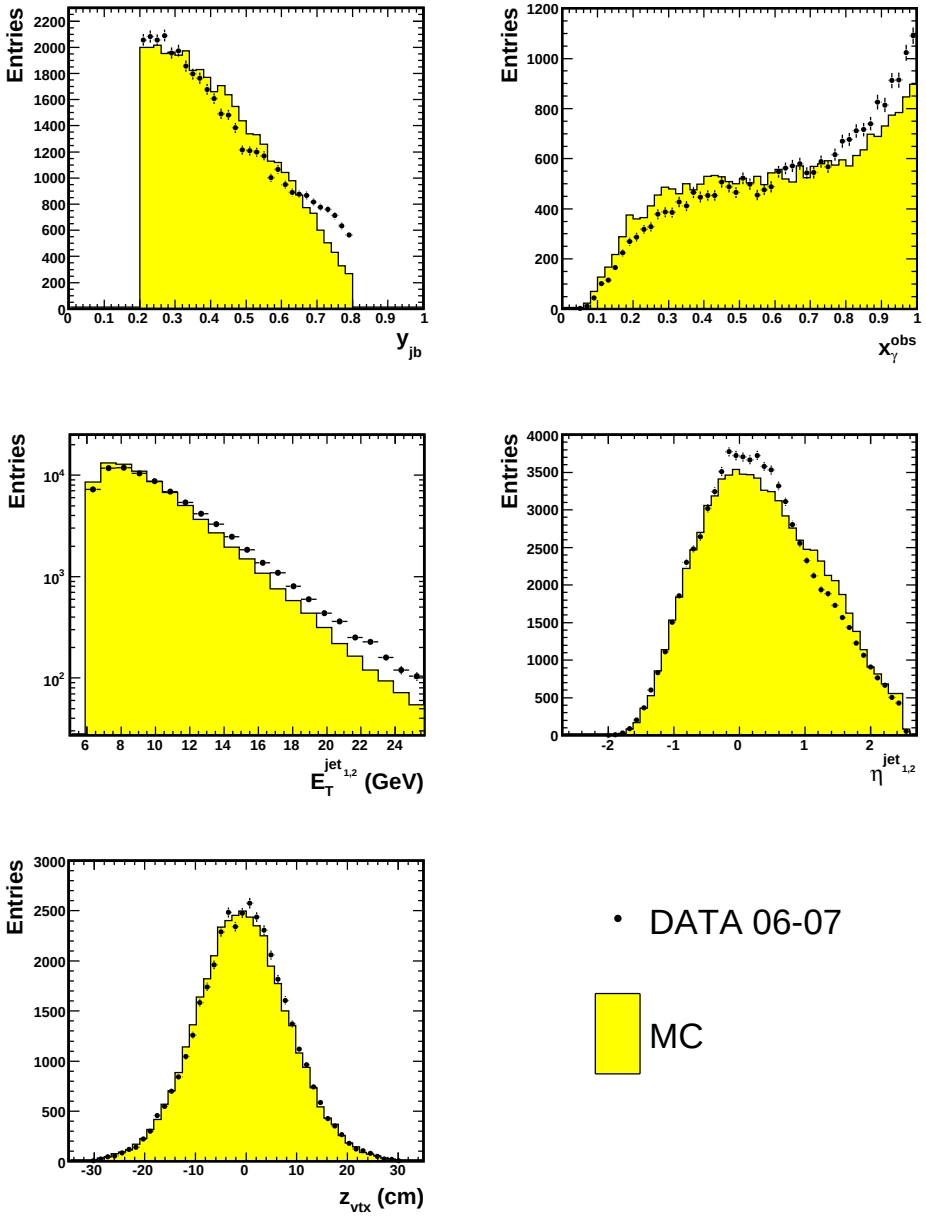


Figure 5.13: *Control plots showing how the Monte Carlo describes the data after preselection. The figures show from top left to bottom right y_{jb}, x_γ, $E_T^{jet_{1,2}}$, $\eta^{jet_{1,2}}$ and z_{vtx}.*

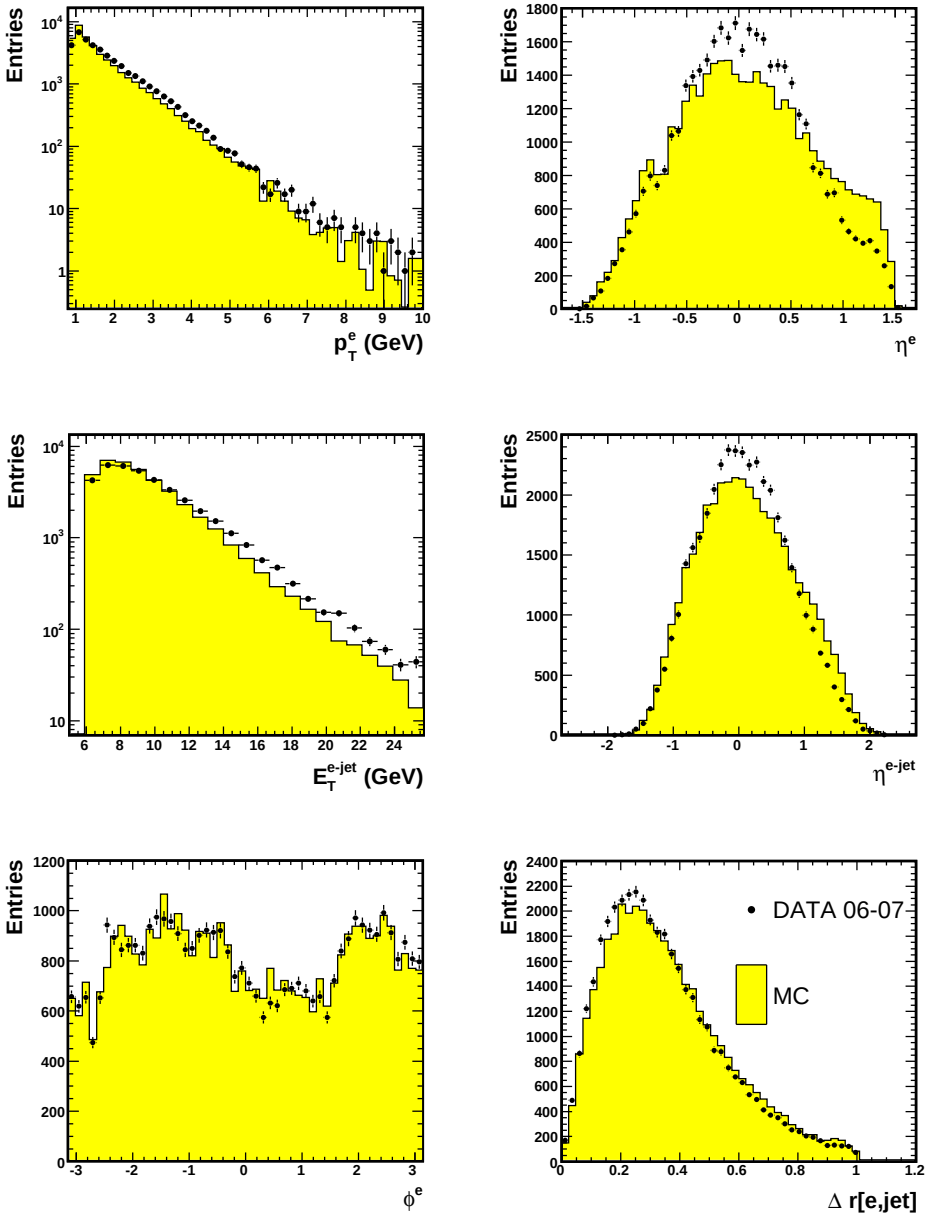


Figure 5.14: *Control plots showing how the Monte Carlo describes the data after preselection. The figures show from top left to bottom right p_T^e, η^e, E_T^{e-jet}, η^{e-jet} and $\Delta r[e, jet]$.*

	HERA I	HERA II
Photoproduction	$0.2 < y_{jb} < 0.8$	$0.2 < y_{jb} < 0.8$
	no DIS e'	no DIS e'
Dijet	$E_T^{jet1(2)} > 7(6)\,\text{GeV}$	$E_T^{jet1(2)} > 7(6)\,\text{GeV}$
	$\lvert\eta^{jet1,2}\rvert < 2.5$	$\lvert\eta^{jet1,2}\rvert < 2.5$
Primary vertex	$\lvert Z_{vtx}\rvert < 50\,\text{cm}$	$\lvert Z_{vtx}\rvert < 30\,\text{cm}$
Data quality	EVTAKE	EVTAKE & MVD

Table 5.3: *Cuts applied for the event selection. The difference between the HERA I and HERA II analysis is a tighter cut for the primary vertex position and the requirement of a working MVD for the latter one.*

	HERA I	HERA II
Kinematic range	$p_T^e < 0.9\,\text{GeV}$	$p_T^e < 0.9\,\text{GeV}$
	$\lvert\eta^e\rvert < 1.5$	$\lvert\eta^e\rvert < 1.5$
EFOs	1:1	1:1
EMC fraction	$E^{EMC}/E^{CAL} < 0.9\,\text{GeV}$	$E^{EMC}/E^{CAL} < 0.95\,\text{GeV}$
	-	$E^{CAL} > 0.5\,\text{GeV}$
Displacement	-	$\lvert IP\rvert < 1\,\text{cm}$
$\gamma \to e^+e^-$	Conversion finder HERA I	Conversion finder HERA II
dE/dx	$dE/dx < 1.1\,\text{mip}$	$-2\ln T_e^{dE/dx} < 3$
Jet association	$\Delta r[e, jet] < 1.5$	$\Delta r[e, jet] < 1$
Cell depth	$d < (\frac{(\theta-90)^2}{500} + 18)\,\text{cm}$	(see section 5.3.2)

Table 5.4: *Cuts applied for the electron candidate selection. Several cuts were implemented or specified for the HERA II data-set.*

Chapter 6

Signal Extraction

In this chapter the method used for signal extraction is presented. To differentiate the relatively small signal from the large light flavour background a likelihood method has been developed that has been maximised in the separation power. While powerfull, limitations of the method include imperfections in the simulation, which result in systematic effects in the fitting method. Therefore careful checks of the variable description in both the signal as well as in the background region is necessary before combining the available variables to one global discrimination variable.

6.1 Likelihood Function

The aim of the likelihood function is to combine the information of several input variables into one discriminating variable without losing too much information. Several variables have been studied and a selection has been made in order to get a large discrimination power under controllable systematic uncertainties. The likelihood function for hypothesis i is given by $\mathcal{L}_i = \alpha \cdot \prod_j \mathcal{P}_i^j$, where α is the particle abundance and $\mathcal{P}_i^j$ is the probability density function for variable j and for particle under study i. In this analysis the main particle types under investigation are $e^{\pm}, \pi^{\pm}, K^{\pm}$ and $p/\bar{p}$. It was found that after the cuts the muon contribution was negligible and therefore they are not listed here. In addition the electrons can further be sub-classified into electrons from semileptonic b decays, e_b, from semileptonic c decays, e_c, and from other contributions, e_o. Since the hard subprocess is of interest to compare with theoretical calculations, processes where a charmed hadron is produced in a $b\bar{b}$-event, called cascade decays ($b \to c \to e$), are counted as beauty signal. Studies also showed that the distributions of the discriminating variables and, therefore also of the likelihood, are very similar to

the other beauty events, which justifies this treatment. Similar arguments are valid for events where a τ is produced semileptonically and then decays into an electron ($b \to \tau \to e$), or for electrons from J/Ψ. In the latter case the event finding efficiency is worse, as the neutrino signature is different. In contrast to the HERA I analysis, where these two processes were treated as background, they were included in the signal for the HERA II analysis increasing the cross section by a few percent. The relative contributions of all these decays are listed in Table 6.1. How likely it is to identify a particle depends on the relative abundance of this

Decay	PYTHIA6.2	PDG2008	PDG/PYTHIA
$b \to e^- X$ (direct)	0.105	0.1086 ± 0.0035	1.03 ± 0.03
$b \to c \to e^+ X$	0.090	0.0802 ± 0.0019	0.89 ± 0.01
$b \to \bar{c} \to e^- X$	0.014	0.016 ± 0.005	1.14 ± 0.36
$b \to J/\Psi \to eX^*$	0.0014	0.00138 ± 0.00013	0.99 ± 0.10
$b \to \tau^\pm \to eX$	0.0072	0.0043 ± 0.0004	0.60 ± 0.06
Total indirect	0.113	0.102 ± 0.005	0.90 ± 0.04
Total (direct + indirect)	0.217	0.210 ± 0.006	0.97 ± 0.03

Table 6.1: *Branching ratios for B decays into electrons from PYTHIAand from the Particle Data Book for the LEP B hadron mix (*$0.400B^+ + 0.400B^0 + 0.114B_s + 0.086\Lambda_b$*) [13].*

particle type, and for a given set of discriminating variables, on the probability to get this variable combination. This probability is given by the product of the probability density functions of each variable. The usage of likelihood ratios for the particle identification makes the method independent of the normalisation of the probability density functions.

6.2 Abundances

The first contribution to the likelihood function is the particle abundance, α. The prior abundance was taken from the Monte Carlo simulation and checked with the posterior probability which were determined during the measurement. As the relative abundances vary over the kinematic region of interest, the interpretation of the likelihood as a finding probability can be improved by determination of the abundances including kinematic dependencies. Therefore the particle abundances for electrons, pions, protons and kaons have been extracted in a 2-dimensional grid of η and p_T of the tracks. It was not possible to find an analytical function which describes the shapes of the abundances for all particle types, therefore the

abundance has been determined in ηp_T-bins. A small binning is desirable as large steps in the relative abundances would result in binning effects in the likelihood distributions[1]. To avoid statistical fluctuations bins with low statistics have been combined afterwards. Figure 6.1 shows the absolute particle abundance for the four given particle types where the entries have been scaled with the bin size.

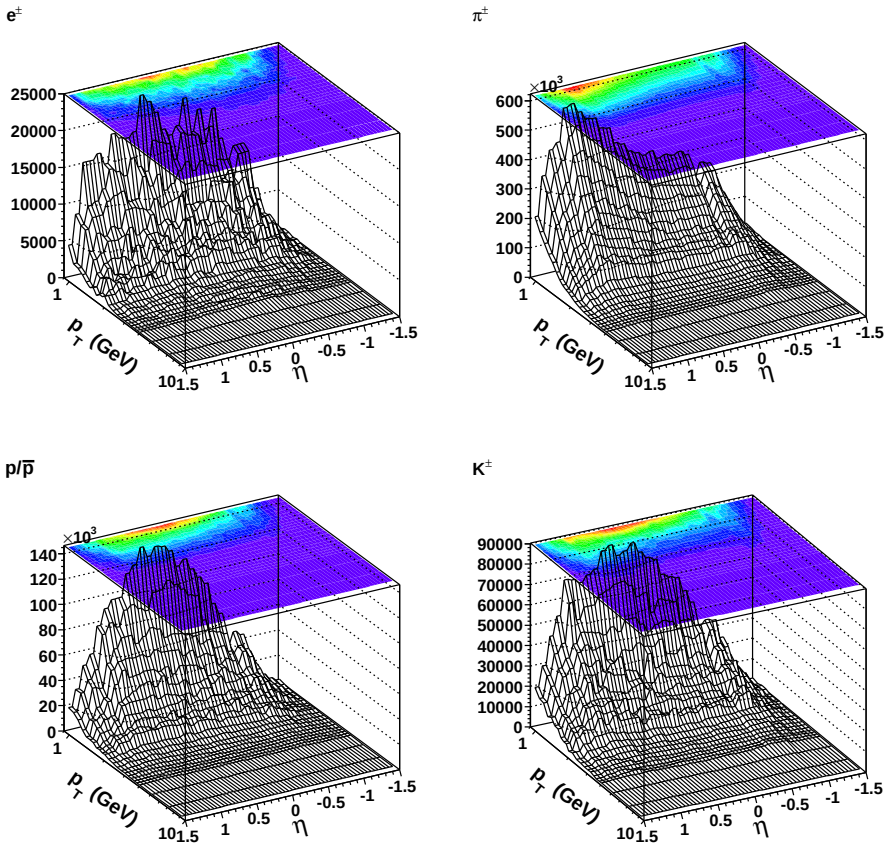


Figure 6.1: *Particle abundances for $e, \pi, p/\bar{p}$ and K binned in η and p_T.*

The relative abundances for e_b, e_c and e_o are binned in p_T, because no clear η-dependence was found. Figure 6.2 shows the abundance for the three types. For low momenta the dominant sources are the electrons from photoconversions and from Dalitz decays, because they have a much softer momentum spectrum than the electrons originating from a semileptonic decay of a heavy flavour hadron. Due to the larger quark mass the momentum spectrum of the e_b is higher than e_c making it easier to identify electrons with large momenta. It was studied that for a selection

[1] Which is from first principle no problem if the Monte Carlo describes the data perfectly.

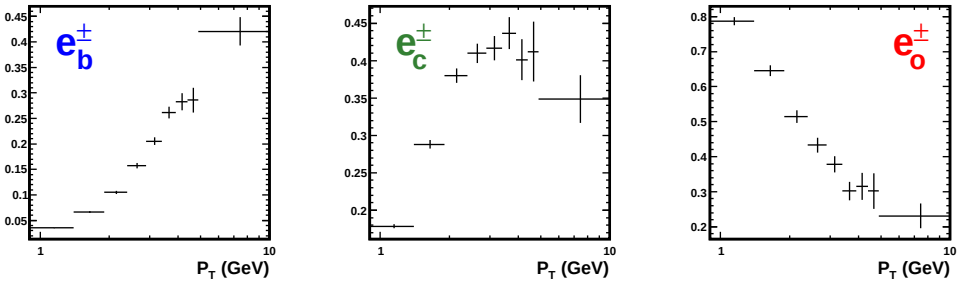


Figure 6.2: *Decay abundances for electrons from semileptonic decays from b quarks, from c quarks, and from other sources as a function of p_T.*

of DIS events the scattered electrons have a significant effect on the spectrum of the background electrons, resulting in a harder momentum spectrum. If the contribution of electron and positron running is not balanced, the abundances and PDFs have to be extracted separately for positively and negatively charged tracks. For this analysis it has been verified, that the DIS background is small and does not have a significant effect on the cross section determination.

6.3 Probability Density Functions

The second ingredient for the likelihood function are the probability density functions, PDFs. They reflect the relative probability to find a particle at a given value of a selected quantity. These PDFs are extracted from the Monte Carlo simulation and have to be checked with the data afterwards. Hence it is necessary to have a reasonable description of the signal as well as of the background region. Some more detailed studies will be given in Chapter 7. The product of several PDFs allows the information from different variables to be combined into one single quantity.

6.3.1 Electron Identification

For the electron identification the following variables have been found to have good separation power and were used as input variables for the likelihood:

- dE/dx, the specific energy loss per unit distance;
- E^{CAL}/p^{trk}, the ratio of energy deposited in the calorimeter to the track momentum measured in the CTD;

- E^{EMC}/E^{CAL}, the fraction of energy deposited in the electromagnetic part of the calorimeter;
- d^{cell}, the depth of the energy deposit in the calorimeter.

Energy Loss due to Ionisation

Combined with the track momentum, the dE/dx information can be used to identify particles, as the average energy loss per distance depends on the momentum and the particle mass (see Figure 4.3). The information gained from the calibration and test samples have been unfolded into relative abundances to identify a particle of a given class. The output of this probability is implemented in the common ZEUS software so that the user can pick up the required information and combine it with the analysis-specific PDFs and the variables useful for his application.

Figure 6.3 shows the output of the four particle types for the sample after preselection. All four distributions are well described, so that it can be used as an input for the likelihood test function. The influence of the remaining discrepancies, especially in the electron and kaon cases, was studied by varying the PDF in data and Monte Carlo. The distribution has also been evaluated as a function of the kinematic variables p_T^e and η^e at different selections. The influence of these uncertainties on the cross section determination is described in Chapter 9.

Calorimeter Energy over Track Momentum

The ZEUS calorimeter was optimised for the accurate determination of jet energies in the region of $10 < E^{jet} < 100\,\text{GeV}$. For lower energies there are deviations in the compensation, i.e. the energy response differs for hadrons and electrons at the same energy. As also the resolutions in the CAL differ for hadrons and electrons at the same energy, the energy distributions vary for the different particle types. In the low momenta region the momentum resolution of the CTD is better than the CAL energy resolution. By taking the ratio of energy and momentum measurements, E^{CAL}/p^{trk}, the scale dependence of the relative uncertainty is reduced and hence this variable can be used for particle identification.

Originally E^{CAL}/p^{trk} was only intended to be used for the identification of antiprotons [125]. Protons and antiprotons behave differently when they interact with the detector material. Owing for the annihilation energy, the distribution of E^{CAL}/p^{trk} is shifted to larger values for the antiprotons. For relatively light particles the E^{CAL}/p^{trk} ratio should be close to one. The energy in the calorimeter

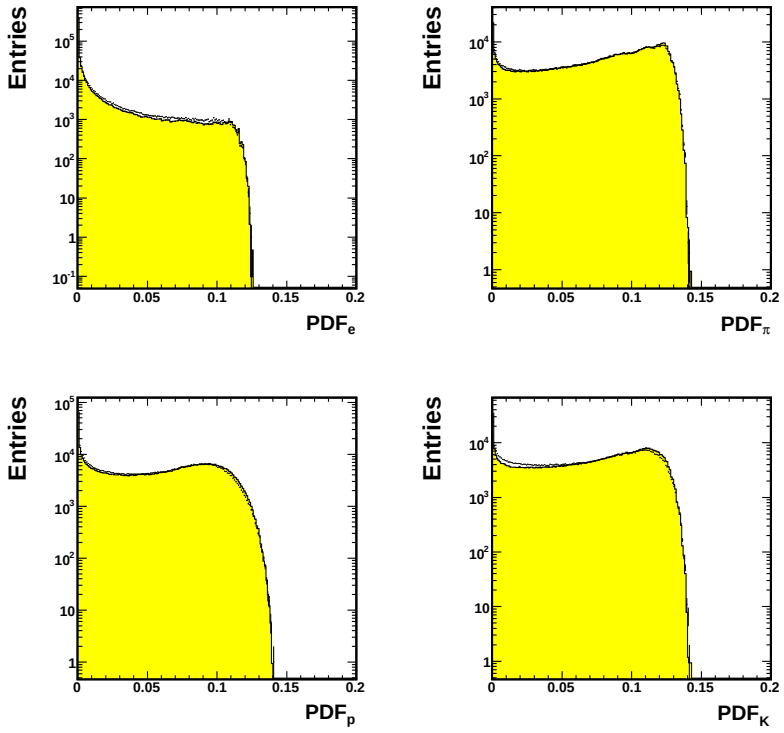


Figure 6.3: *Distribution of a given particle and its kinematic quantities to be an electron, pion, kaon or proton. The yellow histogram is the Monte Carlo distribution while the points represent the data. To extract the relative probability a hypothesis will be normalised by the sum of all hypotheses.*

is now increased by $\sim 1\,\text{GeV}$ due to the recombination with the detector material. This effect is visible in Figure 6.4, where the maximum is shifted to ~ 2 for antiprotons.

Different behaviour of the detector response for protons and antiprotons observed has been observed at ZEUS, but the separation used for the particle identification has its origin in the mass dependent shift of the central value [126]. The values of E^{CAL}/p^{trk} moves to lower values for the heavier particles, while the lighter particles are shifted to high values.

In Figure 6.13(a) the normalised distributions are shown for electrons, pions, kaons, protons and antiprotons. For the first three cases the samples of particles and antiparticles are combined, as no significant difference was observed. E^{CAL}/p^{trk} was the only variable which had to be treated in this way, as all other

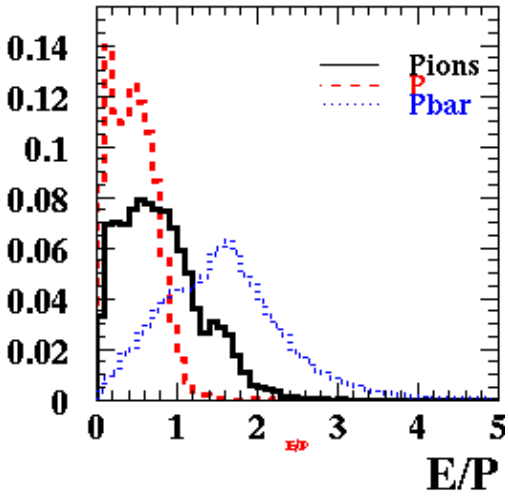


Figure 6.4: *Monte Carlo distribution of E^{CAL}/p^{trk} for pion, protons and antiprotons. The antiprotons are shifted to higher values due to the annihilation process in the calorimeter. The result is taken from [125], where a sample of 70000 single track events with a flat momentum distribution in* $[0.75, 2.5]$ *was used to investigate the different detector responses of protons and antiprotons with respect to the pion background.*

variables were consistent for positively and negatively charged particles. With the higher precision of the HERA II data low values of E^{CAL}/p^{trk} showed some imperfections for tracks having small energies. These candidates did not affect the final result, but to improve the background description candidates with $E^{CAL} < 0.5\,\text{GeV}$ were not considered to have a reasonable description. The quality of the variable description was tested using calibration samples of the dE/dx simulation and using particle enriched sample as described in the next Chapter. Figure 6.5 shows a comparison between data and Monte Carlo for particle samples used in a DIS analysis [127]. The distributions are shown for pion, proton and electron samples, where the left figures are for the negatively charged and the right figures for positively charged tracks. In the case of electrons and pions no clear charge dependence is observed, but for protons and antiprotons the distributions differ.

Fraction of Energy in the el. Calorimeter

For the HERA I analysis the fraction of calorimeter energy deposited in the electromagnetic part, E^{EMC}/E^{CAL}, was used as an ingredient for the likelihood whereas in the later developments the PDF was replaced by a hard cut and by d^{cell}, a partly correlated variable. As the shower signature differs between electrons, muons and hadrons (See Chapter 3.2.3), variables sensitive to the shower shape can be used for the electron identification. By construction design the fraction of the energy deposited in the innermost layer of the calorimeter should be close to one for electrons. Figure 6.6 shows the shape of this variable for electrons, muons and hadrons. Most of the electrons have an EMC-fraction close to one, where the hadrons deposit a significant fraction of their energy in the hadronic calorimeter,

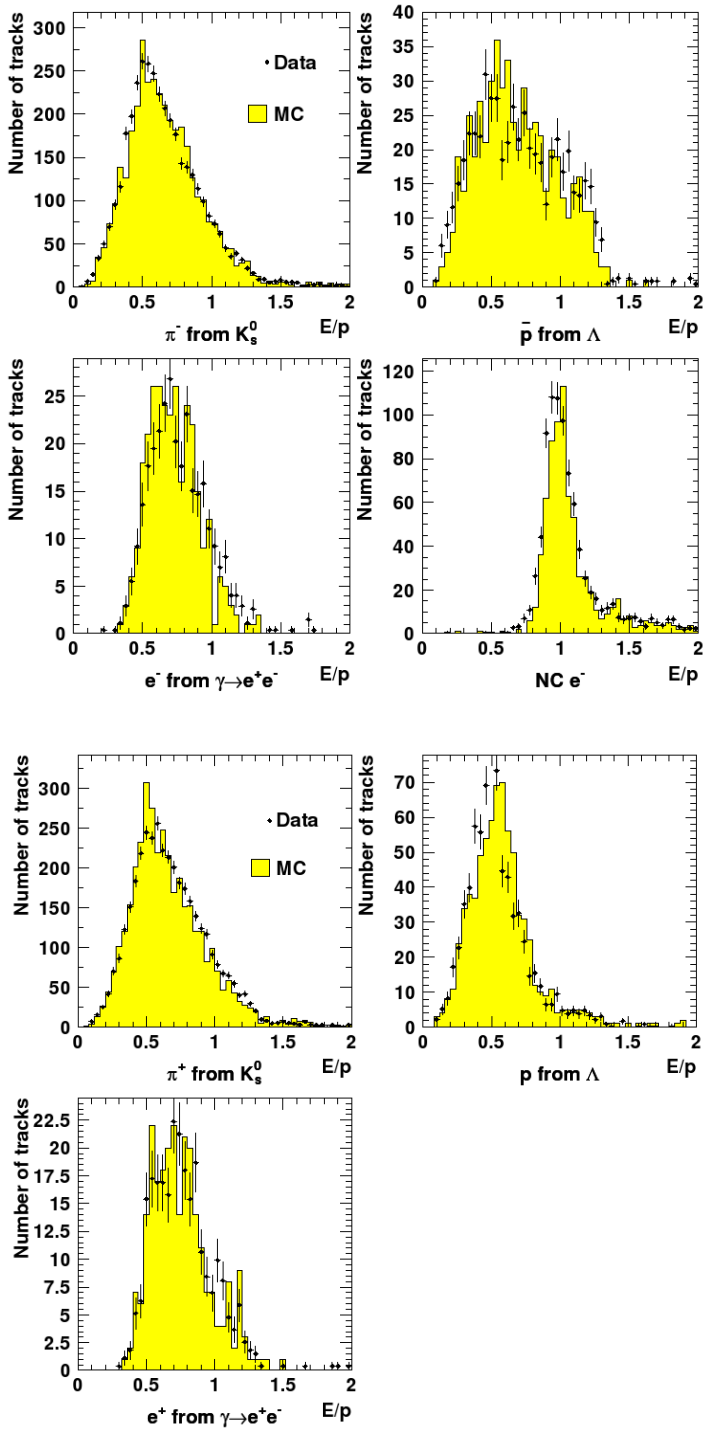


Figure 6.5: *Comparison of data and Monte Carlo for E^{CAL}/p^{trk} distributions for $\pi^{\pm}$ from K^0_s, $p\bar{p}$ from λ, $e^{\pm}$ from γ conversions, and NC e^- [127]. The upper four figures show the samples using only the negatively charged tracks and the lower three figures the samples using the positively charged tracks. For protons and antiprotons a different behaviour is observed.*

which is reflected in lower values of E^{EMC}/E^{CAL}. A better separation is achieved for the muons in the sample, whose energies are split between the calorimeter and the muon system.

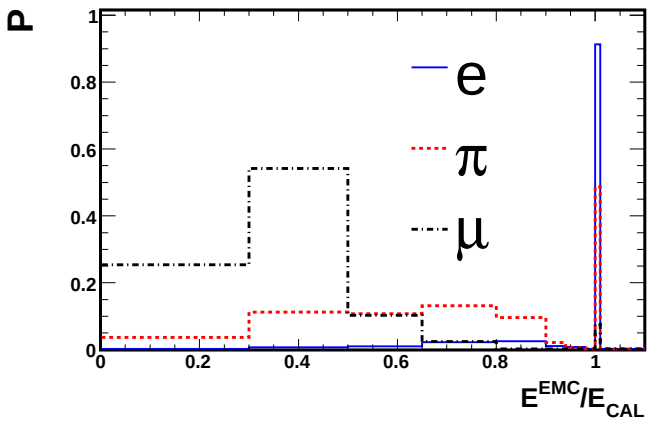


Figure 6.6: *Normalised distribution of E^{EMC}/E^{CAL} for electrons (blue), pions (red) and muons (black). Most of the electrons have an EMC-fraction close to one, where the hadrons deposit a significant fraction of their energy in the hadronic calorimeter. The muons, which are minimal ionising particles, can pass the calorimeter, and deposit part of their energy in the muon chambers, giving a mean value of E^{EMC}/E^{CAL} around 0.4.*

Related variables are the opening angle of the shower, denoted by r_{cell}, the radius of the cone containing the cells associated to the EFO object, the number of cells clustered together, n_{cell}, or the cell depth, d^{cell}, the centre-of-gravity for the summed cells in the CAL. The drop of E^{EMC}/E^{CAL} as an input variable for the likelihood in the HERA II analysis mainly had two reasons. Firstly the geometry cuts were improved reducing the separation power of the EMC-fraction. Secondly the observed deficits in the description of E^{CAL}/p^{trk} have been found to be correlated to the EMC-fraction. Even if this not fully understood background was reduced by cleaning cuts, cross-correlation of systematic effects between likelihood variables should be avoided.

Depth of Energy Deposit

In the HERA I analysis it was already studied whether the position of the EFO, which is the centre-of-gravity in the calorimeter, can be used to separate electrons and hadrons further [14]. The depth of the energy deposit in the CAL was studied separately for the different calorimeter regions, split into transverse and parallel

components relative to the surface at the impact point. The description of this variable has been improved by additional cleaning cuts, and especially by tighter geometry cuts. The distribution of d^{cell} is shown in Figure 6.13(a). Even if the separation power is small compared to the dE/dx it has been used as it is expected to have a small systematic uncertainty. The variable is quite stable against uncertainties in the Monte Carlo description as the distribution is very similar for all background particles, so that a good description in electron and background enriched samples indicates the robustness against changing contributions of the background.

6.3.2 Decay Identification

For the decay identification the following three variables were found to have good separation power and are used as input variables for the likelihood:

- $p_{\perp}^{rel}$, the transverse energy of the electron candidate with respect to the associated jet;
- $\Delta\phi$, the azimuthal angle between the electron candidate and the missing transverse momentum vector;
- δ_{IP}, the signed impact parameter significance of the electron candidate with respect to the beamspot.

Separation from the Jet

Due to the larger mass of the B hadron the decay products originating from it have more phase-space accessible, and therefore the spectrum of electrons from semileptonic b decays have a harder spectrum than particles from light flavour decays. The spectrum is measured as the transverse momentum of the electron candidate relative to the associated jet, p_T^{rel}. In this context the jet axis represents the direction of the original mother particle in which reference system the transverse momentum should be reconstructed. The transverse momentum is then calculated using the the following equation:

$$p_T^{rel} = \frac{|\vec{p}^{jet} \times \vec{p}^{e}|}{|\vec{p}^{jet}|}, \tag{6.1}$$

where $\vec{p}^{jet}$ and $\vec{p}^{e}$ are the momentum vectors of the jet and the electron candidate. Schematically the reconstruction of this variable is shown in Figure 6.7. This variable, which is the most important variable for the heavy flavour separation

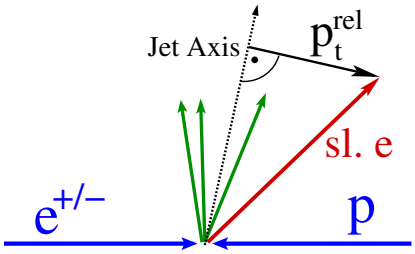


Figure 6.7: *Schematic view of the reconstruction of* p_T^{rel}

has already been used in previous heavy flavour analyses using both muons and electrons as semileptonic lepton particles (e.g. [128, 113]). The different spectra of e_b, e_c and e_o are shown in Figure 6.12(a). In contrast to the good separation of the beauty signal due to the harder spectrum, the variable is not able to distinguish between charm and light flavour. The fact that the spectrum of the charm signal is even lighter than for the light flavour contribution is understood as a consequence of the higher track multiplicities in the charm jets. This multiplicity shifts the p_T^{rel} to lower values.

Reconstruction of the Neutrino

To benefit from the knowledge that the events of interest contain an undetected neutrino the missing energy distribution was studied. More information can be drawn from a combination of the direction and the size of the missing energy vector. As the z-component of the energy sum is not well measured and the Lorentz boost of the hadronic four momentum is not known, the missing transverse momentum vector has been determined in the xy-plane to:

$$p_T^{miss} = -\sum \vec{p}_T, \tag{6.2}$$

where the sum runs over all EFO objects. This vector represents the direction of the escaped neutrino from the semileptonic decay. Previous studies have shown that there is a good correlation between the direction of the neutrino and p_T^{miss} even if more soft neutrinos are present in the result [14]. To use the decay signature of the neutrino the opening angle between p_T^{miss} and the electron momentum, $\Delta\phi$ was used. By using the electron as a reference system it is a complementary variable to the p_T^{rel} which is highly correlated to the angle between the electron and the jet direction. Figure 6.8 shows schematically the reconstruction of $\Delta\phi$. The normalised distribution for e_b, e_c and e_o is shown in 6.12(b). Due to its sensitivity to the presence of a neutrino originating from a semileptonic decay, $\Delta\phi$ has a similar response for beauty and charm, but a different behaviour for light flavour

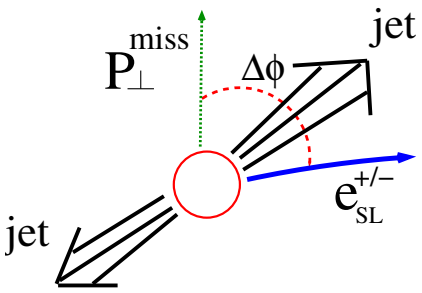


Figure 6.8: *Schematic sketch of $\Delta\phi$, the difference of the azimuthal angle between the electron and the neutrino, represented by the p_T^{miss}-vector.*

events. The asymmetry in the light flavour contribution is caused by the track multiplicity. The charged track multiplicity is not balanced and as p_T^{miss} points on average to the side with less tracks, more candidates are found having a large angle to p_T^{miss}. Combining p_T^{rel} and p_T^{miss} the likelihood is sensitive to both the beauty and the charm contributions.

Using Lifetime Information

For the HERA II analysis the improved tracking using the MVD information gives the possibility to determine the track displacement with respect to the primary interaction point with much higher precision. This analysis benefits from the lifetime information by using the signed impact parameter significance, δ_{IP}, as defined in Equation 4.3. The impact parameter for light flavour events, except for particles originating from K_L or Λ decays, is only treated by resolution effects, and should therefore be consistent with zero. As the resolution effects are not shifting the IP to any favoured direction, the IP distribution should be symmetric for the light flavour contribution. The normalised distribution of δ_{IP} is shown in Figure 6.9. The light flavour background is symmetrically distributed around zero, where both heavy flavour contributions have positive tails. Due to the longer lifetime of B hadrons the separation works better for e_b than for e_c. This variable, which is only accessible for the HERA II analysis, helps to improve the overall separation, even if the separation is not as good as for p_T^{rel} or $\Delta\phi$.

Corrections

Especially for the HERA I analysis deviations in the p_T^{rel} and $\Delta\phi$ distributions have been observed. Studies showed, that the discrepancy is visible in the background but not in the signal enriched region. Figure 6.10 shows the ratio of data to Monte

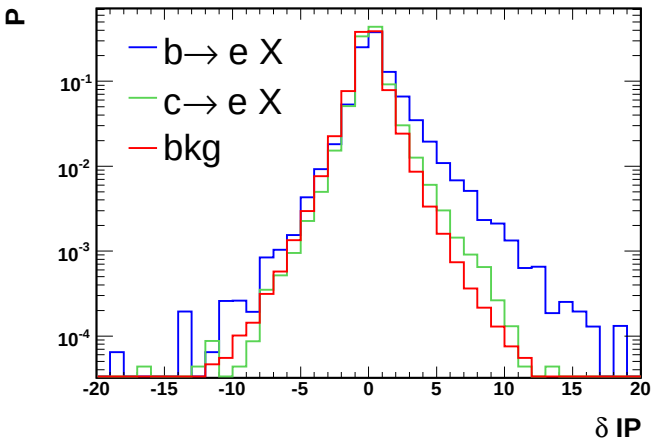


Figure 6.9: *Probability density functions for δIP given for e_b (blue), e_c (green) and e_o (red). The points show the binned distribution extracted from the Monte Carlo. The background values are distributed symmetrically around zero, where the heavy flavour contributions have positive tails. The separation from the light flavour background is better for semileptonic decays from beauty.*

Carlo as a function of p_T^{rel} and $\Delta\phi$ in a background dominated sample. The pure background sample to calculate these corrections has been extracted by cutting on the test function T_e, which is described at the end of this Chapter. The fractions of e_b and e_c have been found to be negligible after this selection. With these factors the charm and the light quark p_T^{rel} distributions were corrected for. For low values of p_T^{rel} the correction is very small, at $p_T^{rel} = 1.5\,\text{GeV}$, where the purity of the b contribution is largest, it is of the order of $\sim 15\,\%$ and increased up to factors of $\sim 2-3$ for very high values of p_T^{rel}. The region of physical interest for this analysis is between $0.5 < p_T^{rel} < 2$ (see Chapter 7). In the HERA II data set the discrepancy at high values of p_T^{rel} is reduced significantly. The correction was found to be of the order of $\sim 5\,\%$ and raised to $\sim 20\,\%$. In case of the variable $\Delta\phi$ the correction was of the order of $\sim 5\,\%$. The correction has also been applied, but the effect was much smaller than for the p_T^{rel} correction. The deficit in the description of p_T^{rel} and the necessity to correct for it was one of the main systematic uncertainty sources for the HERA I analysis. For the HERA II analysis the contribution to the total systematic is substantially reduced due to the the better description of this variable. In the case of p_T^{rel} the correction has also been calculated for each bin of the variables for the final differential cross section. As an example the deviation in the Monte Carlo relative to data in bins of p_T^e is shown in Figure 6.11 For low values of p_T^{rel}, where the deviation is small, no dependence on p_T^e was observed, in contrast to the region of higher values, where the discrepancies can fluctuate and

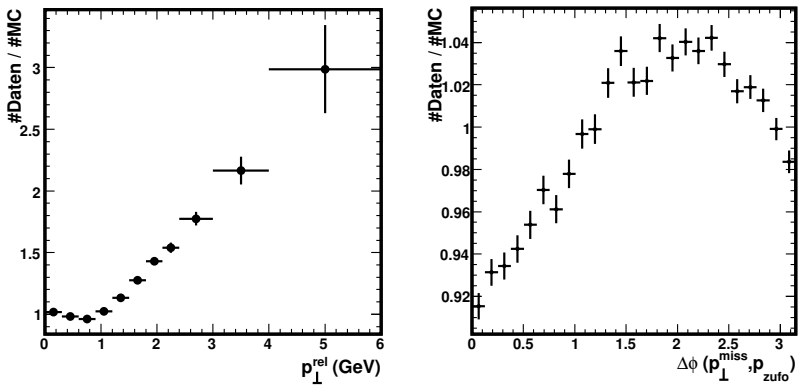


Figure 6.10: *Ratio of data to Monte Carlo distribution calculated from a background enriched sample to correct the PDF of p_T^{rel} and $\Delta\phi$ [14].*

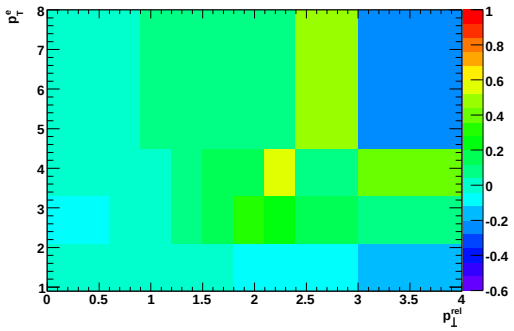


Figure 6.11: *Deviation of the Monte Carlo distribution relative to data in bins of p_T^{rel} and p_T^e. The deviation as a function of p_T^e had to be checked, in order to test the systematic effect of the correction depending on the bins of the final differential cross sections.*

should be taken into account in the calculation of the systematic uncertainties.

6.3.3 Variable Description

Like in the treatment of abundances, where a fine binning with small fluctuations was aimed for, the shape of the variables used in the likelihood should be well described. A rough binning has the advantage of high precision for the average over a bin, but can cause problems in case of migrations at the bin edges. In addition, large steps of likelihood probabilities at the bin edges should be avoided, as this would drive to separations in likelihood responses of candidates, nearby in phase space, but with values around these edges. To avoid such effects, in the first step the distributions were extracted using a binning which is wide enough to avoid

statistical fluctuations. In the second step a smooth distribution was achieved, which describes the shape of the distribution. These distributions were found by fit-functions or different smoothing procedures.

Fit with Function

Wherever it is possible to define an analytical function to fit the PDF distributions this function was used to define the PDFs. Such a function with reasonable number of fit parameters and good description of the points was found for p_T^{rel} and $\Delta\phi$. Figure 6.12(a) shows the binned distribution for beauty, charm and background particles extracted from the Monte Carlo including the line which represents a convolution of a Gaussian and a Landau function fitted to the points. The fit curve describes the points very well and has also been checked with a finer binning in region of high statistics. For $\Delta\phi$ a polynomial of fourth order has been used as a fit function. The fit curve and the binned distributions are shown in Figure 6.12(b).

In this case the curves describe the points very well again. It was found that the influence on the cross section by replacing the binned histogram with the smooth curve is negligible. The stability is not reflected in a better description, but there are less steps in the control plots.

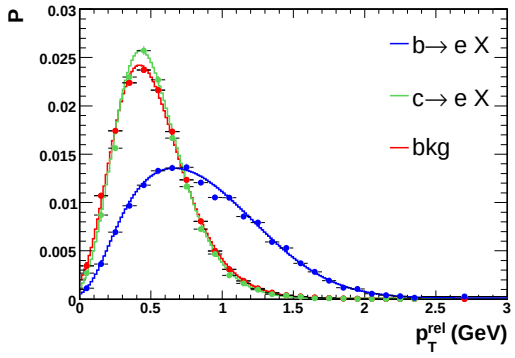
Smoothing

For the variables E^{CAL}/p^{trk} and d^{cell} no explicit function could be found to describe all distributions. In most of the cases it was possible to define a function which describes the distribution for one particle type, but not one function for all of them. The smoothing routines implemented in the software package ROOT [129] were not sufficient; therefore a combination of a B-Spline and a Bézier curve was finally used [130, 131]. The B-Spline is defined by:

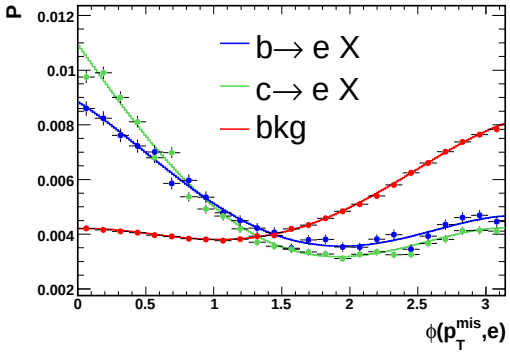
$$P'(t) = \sum_{i=0}^{n} N(i,k,t) \cdot P_i, \tag{6.3}$$

where P_i are the start values for bin i of the histogram with n bins, and $N(i,k,t)$ is defined by:

$$\begin{aligned}
N(i,1,t) &= 1 \qquad \text{if}(\tau[i] \le t \le \tau[i+1]) \\
N(i,1,t) &= 0 \qquad \text{else} \\
N(i,k,t) &= \frac{t-\tau[i]}{\tau[i+k-1]-\tau[i]} \cdot N(i+1,k-1,t) \\
&\quad + \frac{\tau[i+k]-t}{\tau[i+k]-\tau[i+1]} \cdot N(i,k-1,t),
\end{aligned} \tag{6.4}$$



(a) p_T^{rel} PDF



(b) $\Delta\phi$ PDF

Figure 6.12: *Probability density functions for p_T^{rel} (a) and $\Delta\phi$ (b) given for e_b (blue), e_c (green) and e_o (red). For $\Delta\phi$ the zero is suppressed on the y-axis. The points show the binned distributions extracted from Monte Carlo. The lines show the fitted distributions which were finally used to calculate the probabilities.*

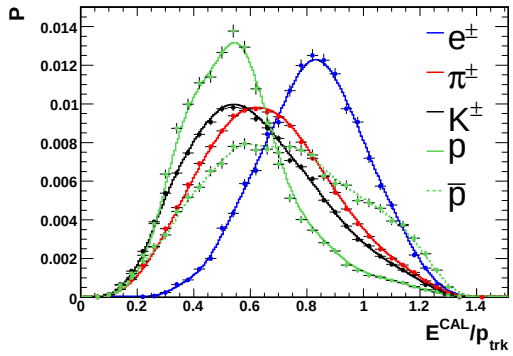
where $\tau[i]$ are the knot points spread over the range under consideration. In the first step intermediate knot points were calculated for the PDFs. The values $P'(t)$ were extracted using a spline up to order $k = 3$. It was checked that higher orders did not improve the description of the input histogram any more. In the second step these points were smeared using a Bézier curve which is defined by:

$$P''(s) = \sum_{i=0}^{n'} B(i, n', s) \cdot P'_i, \tag{6.5}$$

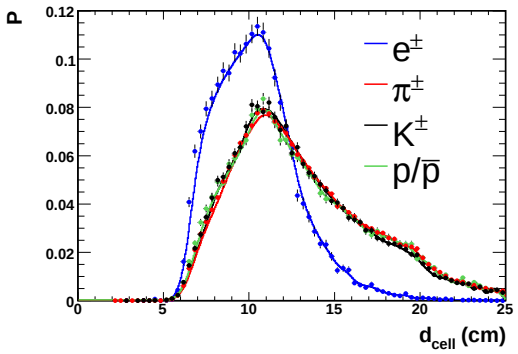
with the coefficients:

$$B(i, n', s) = \binom{n'}{i} s^i (1 - s)^{n'-1}, \tag{6.6}$$

where P'_i are the n' input points $P'(t)$ from the B-Spline. In Figure 6.13 the smooth curves, $P''(s)$, are compared with the points, which represent the binned Monte Carlo distributions, P_i, for the different particle sorts.



(a) PDF for E^{CAL}/p^{trk}



(b) PDF for d^{cell}

Figure 6.13: *Probability density functions for E^{CAL}/p^{trk} (a) and d^{cell} (b) given for e_b (blue), e_c (green) and e_o (red). The points show the binned distributions extracted from the Monte Carlo. The lines show the smoothed distributions which were finally used to calculate the probabilities.*

For both E^{CAL}/p^{trk} (Figure 6.13(a)) and d^{cell} (Figure 6.13(b)) the points are well reproduced by the smooth curves.

6.4 Test Function

Combining these PDFs with the relative abundances the likelihood function can be computed for various hypotheses. In this consideration for the hypothesis of the different particle sorts (e, π, K, p) and in the case of the electrons, for the different decay types. In the case of a hypothesis for electrons from a specific decay type, the abundance is a product of the electron and the corresponding decay abundance. In order to interpret the likelihood as a probability and to have a direct

comparison of the goodness of the candidates, the likelihood ratio (test function T) is used, where the hypothesis is divided by the sum of all different hypotheses. The following three test functions are of main interest:

- T_e, the test function for being an electron,
- T_b, the test function for being an electron from a semileptonic b decay and
- T_c, the test function for being an electron from a semileptonic c decay.

The first one is given by:

$$T_e = \frac{\mathcal{L}_e}{\mathcal{L}_e + \mathcal{L}_\pi + \mathcal{L}_\pi + \mathcal{L}_K + \mathcal{L}_p} \tag{6.7}$$

with

$$\mathcal{L}_e = \alpha_e(\eta, p_T) \cdot \mathcal{P}_e(dE/dx) \cdot \mathcal{P}_e(E^{CAL}/p^{trk}) \cdot \mathcal{P}_e(d^{cell}) \tag{6.8}$$

and by adding the decay abundance and the three decay variables T_b (and analogously T_c) is given by:

$$T_b = \frac{\mathcal{L}_b}{\mathcal{L}_b + \mathcal{L}_c + \mathcal{L}_o + \mathcal{L}_\pi + \mathcal{L}_K + \mathcal{L}_p} \tag{6.9}$$

with

$$\mathcal{L}_b = \mathcal{L}_e \cdot \alpha_b(p_T) \cdot \mathcal{P}_b(p_T^{rel}) \cdot \mathcal{P}_b(\Delta\phi) \cdot \mathcal{P}_b(\delta_{IP}) \tag{6.10}$$

The beauty hypothesis was used to extract the beauty and charm fractions and the electron and charm hypotheses were used to extract signal enriched samples at different selection stages to control the signal and background description.

6.4.1 dE/dx Likelihood

The most important contribution to the electron identification is the dE/dx PDF. To study the separation power of this variable, one can define a test function with dE/dx as the only input variable. Figure 6.14 shows the performance of the likelihood, using only the dE/dx probability density functions (see Figure 6.3) as input for the likelihood ratio:

$$T^e_{dE/dx} = \frac{\mathcal{P}_e(dE/dx)}{\mathcal{P}_\pi(dE/dx) + \mathcal{P}_e(dE/dx) + \mathcal{P}_p(dE/dx) + \mathcal{P}_K(dE/dx)} \tag{6.11}$$

The left figure shows the distribution of the $-2\ln T^e_{dE/dx}$ for the electron candidates defined in the event selection (dropping the cut of $-2\ln T^e_{dE/dx} < 3$) split into the four relevant particle types. The signal efficiency and background suppression can be calculated from the normalised distribution of the true electrons and the fake electrons in Figure 6.14 (left). By cutting at $-2\ln T^e_{dE/dx} < 2$ the efficiency for the electron selection is $\sim$98% with a background suppression of $\sim$62%.

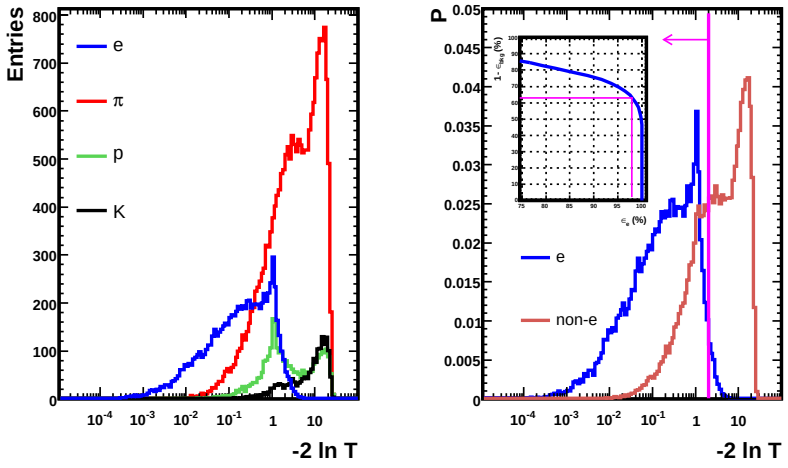


Figure 6.14: *dE/dx likelihood hypothesis test for e, p, K, π (left) and the normalised distribution for e and non-e. The line indicates the cut at $T^{e}_{dE/dx} < 2$, where the signal efficiency and background rejection is shown in the inlay plot.*

6.5 Summary

Combining all ingredients, the likelihood provides one single powerful variable to separate the heavy flavour events from the light flavour contribution. The distribution of this variable was used to determine the fractions of electrons originating from semileptonic decays. In Figure 6.15 the likelihood test function is shown for the beauty (upper) and the charm (lower) hypotheses. Both distributions, T_b and T_c, separate the corresponding signal from the background.

For the final cross section determination only the beauty hypothesis has been used. The charm test function was used to extract charm enriched samples, which are very important to study systematic effects from imperfections in the signal description. In the following chapter studies of the likelihood performance including control plots are presented. The studies are important to understand the influence of a single variable to the likelihood and to be able to estimate the systematic uncertainty due to the likelihood methods. Systematic effects can be caused by bad description of input variables to the likelihood or, more subtle, by effects on the likelihood by differences in the correlation between the variables. The unfolding of the cross sections is described in Chapter 10. During the optimisation process of the signal determination the input of the likelihood was varied using alternative variables for the decay identification. Some examples of these variables are listed in the next chapter.

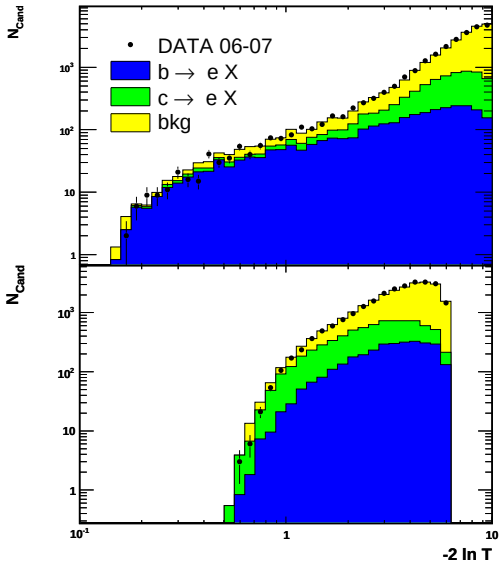


Figure 6.15: *Test function response for the beauty hypothesis, T_b, (upper Figure) and charm hypothesis, T_c (lower Figure). The histograms show the Monte Carlo split into their contributions from e_b (blue), e_c (green) and light flavour background (yellow). The distribution is compared to data (points). For the final cross section determination only the beauty hypothesis has been used, where the charm hypothesis is a very useful tool to extract charm enriched samples for systematic studies.*

Chapter 7

Likelihood Performance

In this chapter some studies to understand the performance of the likelihood and the influence of the different variables are discussed, with emphasis on the separation power of the dE/dx measurement. Control plots will give an impression of the signal and background description by the Monte Carlo. At the end some alternative variables, tested for the signal extraction, are presented.

7.1 dE/dx Hypothesis

As explained in Chapter 6 the variable dE/dx is not only able to distinguish between electrons and background particles, but can also be used for other particle hypotheses. Figure 7.1 shows the likelihood test function using only the dE/dx as input variable for the electron hypothesis, $T^e_{dE/dx}$, as well as for pion, kaon and proton hypothesis for the different particle types. For the following studies the candidates were selected with the cuts listed in Tables 5.3 and 5.4 without $-2\ln T^e_{dE/dx} < 3$. The upper left figure shows the separation for the electron hypothesis, which is able to distinguish between electrons and pions or kaons. The proton distribution has two peaks, one in the pure background region and one close to the electron peak, resulting in a separation limited to the subsample of the right peak. The origin of the double-peak structure are the crossing points of the bands in the $(dE/dx, p)$-plane. In the following the influences on the kinematic distributions of the remaining particles are described. The upper right and lower left figures show that it is also possible to make pion and proton enriched sample. Due to the overlaying structure of the dE/dx bands the possibility to obtain a sample with reasonable kaon purity is limited.

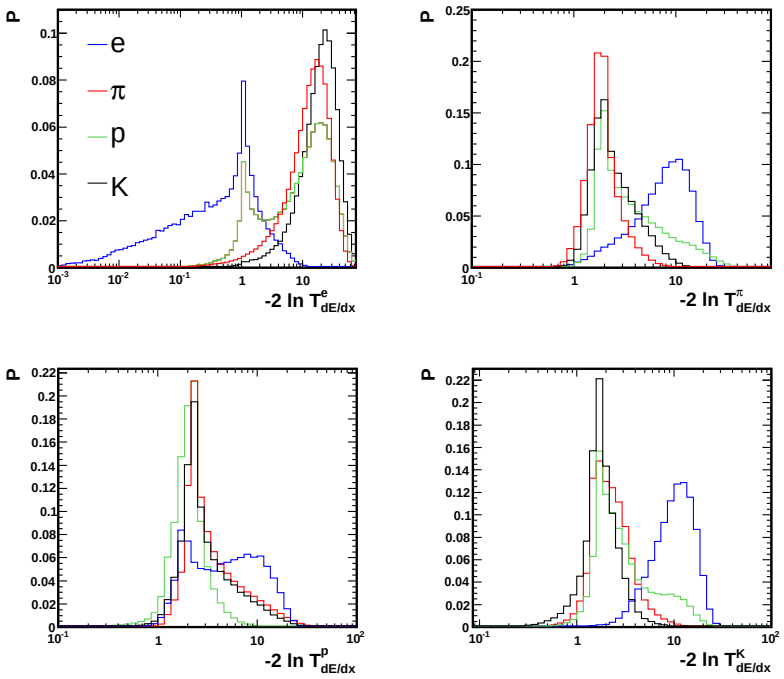


Figure 7.1: *Normalised distributions of the likelihood test function for $e, \pi, p/\bar{p}$ and K using the electron (upper left), pion (upper right), proton (lower left) and kaon (lower right) hypotheses.*

7.1.1 Particle Samples

Figure 7.2 shows a scatter plot of dE/dx vs. momentum for all electron candidates. At this selection stage the band structure for protons with high dE/dx at low momenta is already reduced by the preselection cuts.

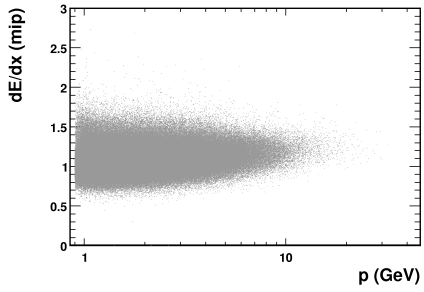


Figure 7.2: *Scatter plot of mean energy loss, dE/dx, vs. track momentum, p, for all electron candidates.*

By cutting on $-2 \ln T^{x}_{dE/dx} < 3(1.5)$ a medium (highly) enriched sample, for particle sort x, can be obtained. In Figure 7.3 these samples are shown for an electron,

pion, proton and kaon enriched region – for the lower cut on the left side and the harder cut on the right side.

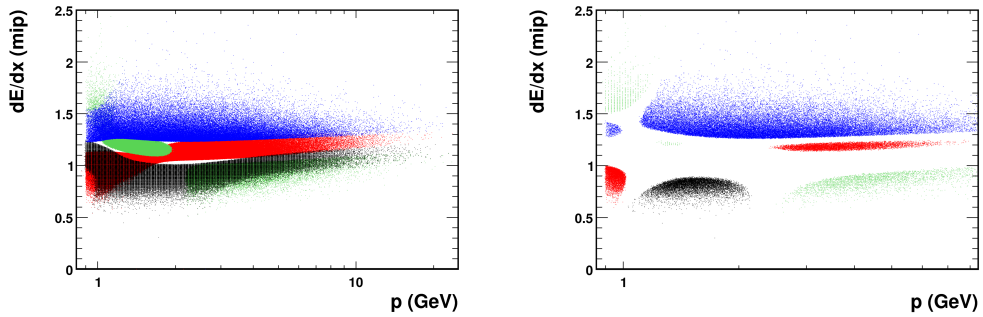
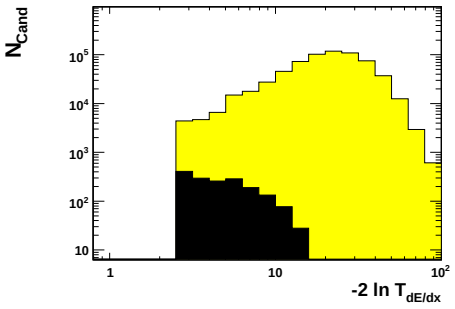


Figure 7.3: *Scatter plot of mean energy loss, dE/dx, vs. track momentum p. The medium (left) and highly (right) enriched regions were extracted by cutting on the different likelihood hypotheses as described in the text. The samples shown are the electron (blue), pion (red), proton (green) and kaon (black) samples.*

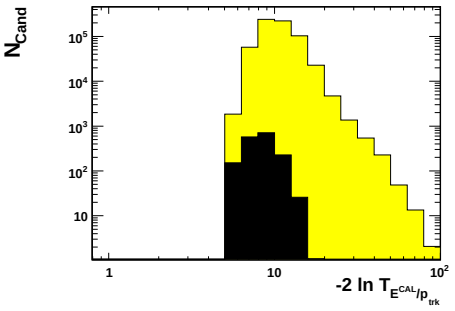
For the medium enriched case the samples still have overlapping regions, while they are clearly separated in the highly enriched sample. The likelihood value depends on the separation of the bands, where crossing-points lead to indistinguishable regions. The likelihood cuts cause disconnected subparts of the particle bands. For example the proton candidates are split into three regions: the low momentum and high dE/dx region, the high momentum and low dE/dx region and the region in the gap of the pion and the electron band in the intermediate region. The right figure also shows that it is possible to identify electrons down to 0.9 GeV. Additionally it shows that it is complicated to identify electrons in the region around $\sim$ 1.1 GeV, and that a hard cut on the electron likelihood would depopulate this region.

7.2 Separation Power

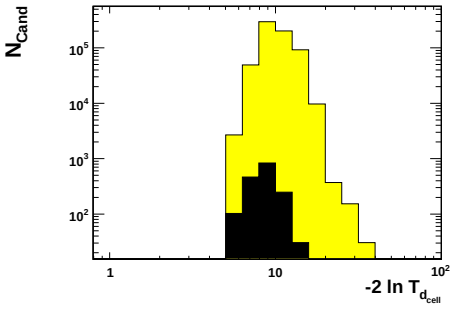
To compare the separation power of the other variables with the one of the dE/dx test function, the test function has been computed after replacing the dE/dx PDFs by the PDFs of the other variables and combining probabilities with the abundances as described in the previous chapter. In Figure 7.4 the hypothesis test function is shown for the six different input variables. The likelihood distribution is shown for the whole Monte Carlo sample (yellow) and for the beauty signal (black). Figure 7.4(a) illustrates the importance of the dE/dx for the electron identification.



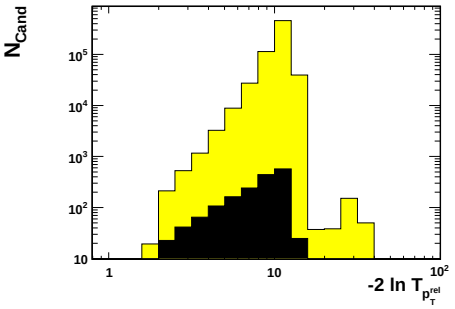
(a) Test function using dE/dx



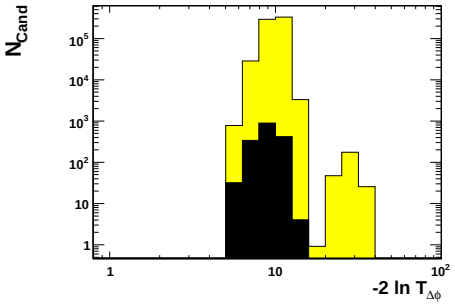
(b) Test function using E^{CAL}/p_{trk}



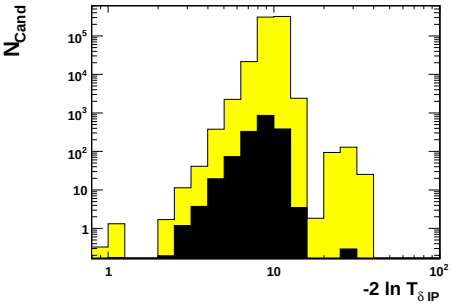
(c) Test function using d^{cell}



(d) Test function using p_T^{rel}



(e) Test function using $\Delta\phi$



(f) Test function using δ_{IP}

Figure 7.4: *Test function shown for the complete Monte Carlo sample (yellow) and for the beauty signal (black). The test functions have been computed by using the abundance plus a signal additional input variable.*

The large light flavour background is shifted towards values above ~ 10. With decreasing separation power also E^{CAL}/p^{trk} (7.4(b)) and d^{cell} (7.4(c)) help to separate the electron and non-electron contributions. The three decay variables do not distinguish between electrons and background particles, but enrich the heavy flavour contribution. Especially for the p_T^{rel} variable (see Figure 7.4(d)) the shape of the likelihood distributions differ. The range of the test function and the separation is lower for $\Delta\phi$ (7.4(e)) and δ_{IP} (7.4(f)), but both variables contribute to the final separation of e_b, e_c and the light flavour background. The study of the correlations between the variables and their influence on the systematic uncertainty was important for the selection of the likelihood variables.

7.3 Correlations

As mentioned in Chapter 6 it is complicated to quantify the total systematic uncertainty for correlated variable, while not much additional separation power can be obtained from highly correlated variables. In Figure 7.5 the correlation matrix for some variables under study is shown. The variables used for the HERA I analysis are highlighted by the magenta box. By this studies of the correlations

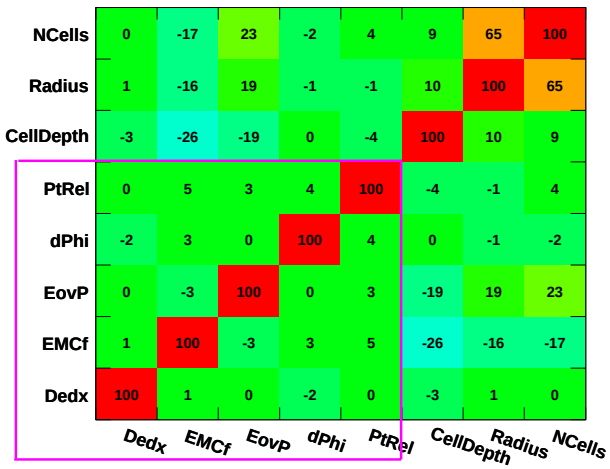


Figure 7.5: *Correlation matrix for possible variables in the likelihood. The variables used for the HERA I analysis are marked by the magenta box. The other three variables are highly correlated and are also connected with the EMC-fraction. The correlation studies were performed using the software package TMVA [132]*

the cut at $E^{CAL} > 0.5\,\text{GeV}$ (see Section 5.3.2) was driven to stabilise the selection and reduce the imperfection in the E^{CAL}/p^{trk} simulation. As a consequence of

dropping the EMC fraction from the likelihood for the HERA II analysis, it was studied which of the three additionally shown correlated variables could improve the electron separation. The cell radius, r^{cell}, was found to be not well simulated and to be too sensitive to detector geometry. The distribution is a superposition of several step-functions, where the step size and width depends on the calorimeter block size for a given η region. Also the number of hit cells, N^{cell}, was excluded because this integer variable showed some migrations which are not simulated by the Monte Carlo. Therefore d^{cell} was finally chosen to improve the electron identification.

7.4 Control Plots

The test function with different hypotheses uses the following variables: dE/dx, E^{CAL}/p^{trk}, d^{cell}, p_T^{rel}, $\Delta\phi$ and δ_{IP}. With these tools of different hypotheses, it was possible to check the quality of the data description. Only well described input variables can be used to have same likelihood response for data and Monte Carlo. In the following some examples of checks for the background and signal description are shown. The distributions of all candidates quantifies the quality of the background description as the selection at this stage is clearly background-dominated. Additionally the distributions were checked for signal enriched regions, where cuts on the likelihood hypothesis have been used to enrich the samples with beauty or charm events. In the following some control plots are shown. The background enriched plots correspond to all candidates and the signal enriched plots to candidates with $-2\ln T_b < 1.5$ for beauty and $-2\ln T_c < 5$ for charm. Additional plots can be found in Appendix A. Control plots for the HERA I analysis can be found in [16]. The cuts for the selection have been varied to scan various levels of purities and identify regions of imperfections to decide whether they influence the results.

7.4.1 Likelihood Variables

The first variables to check are the likelihood input variables. To be sensitive to the signal to background separation power, the distributions must be reasonably well described for both the signal as well as for the background contributions. Differences in the shapes for different selection stages show the regions of sensitivity as well as the influence of the variable to the likelihood. In Figure 7.6 p_T^{rel}, which is the main variable for the beauty separation, is shown as an example for the likelihood variables. The distribution is reasonably well described for all three selections. Uncertainties due to remaining discrepancies were studied by a

variation of the Monte Carlo distribution as described in Chapter 9. The region of interest for the charm-enriched plots is clearly separated from the region for the beauty-enriched plot. This reflects the probability to identify the different sources at a given value of p_T^{rel} using its PDF (see Figure 6.12(a)).

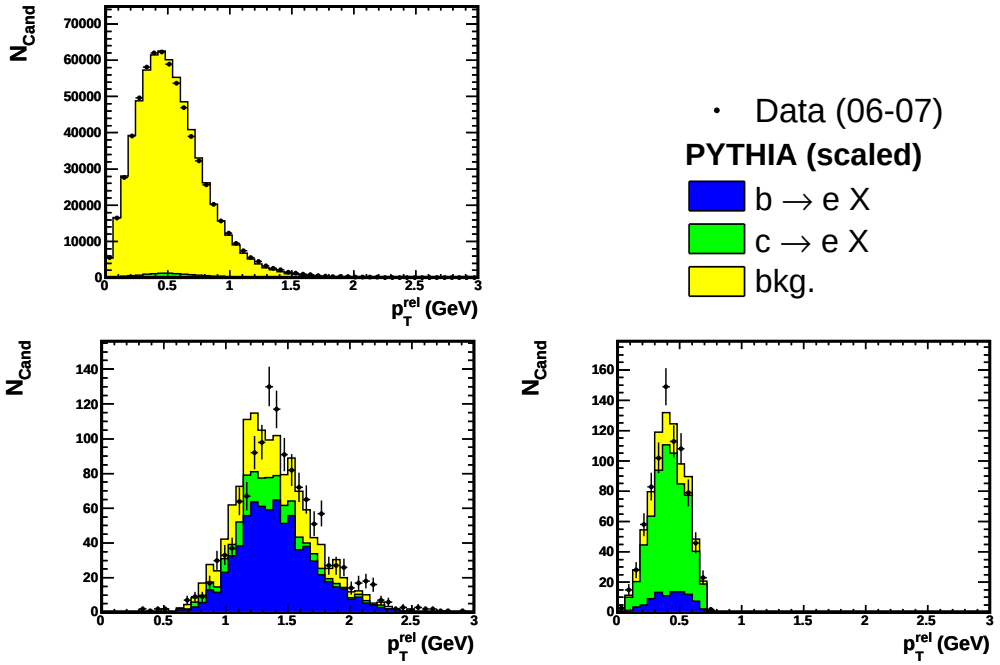


Figure 7.6: *Distribution of the likelihood variable p_T^{rel}. The three histograms show the distribution for all candidates which is background dominated (upper left), for beauty enriched (lower left) and charm enriched (lower right). The Monte Carlo is subdivided into the beauty signal (blue), the charm signal (green) and the background contributions (yellow). The three contributions were scaled by the fit result (see Chapter 10).*

A similar effect can be seen for the variable δ_{IP} which is shown in Figure 7.7. The three figures show the distribution for a background dominated region and for beauty and charm enriched regions, where the negative side of the distribution is highly suppressed after the likelihood cuts. The variable is well described at all three selection stages. Small discrepancies were found when studying the control plots in bins of the kinematic variables of the track. The effect is most pronounced for tracks going to the forward or backward regions in the upper half of the detector. Figure 7.8 shows δ_{IP} in bins of p_T^e and η^e, where a shift is visible in the lowest left and right figure. The effect of this shift on the final cross section determination was not very large, nevertheless it would be an important step in the further increase in precision of the HERA II results to find a correction procedure.

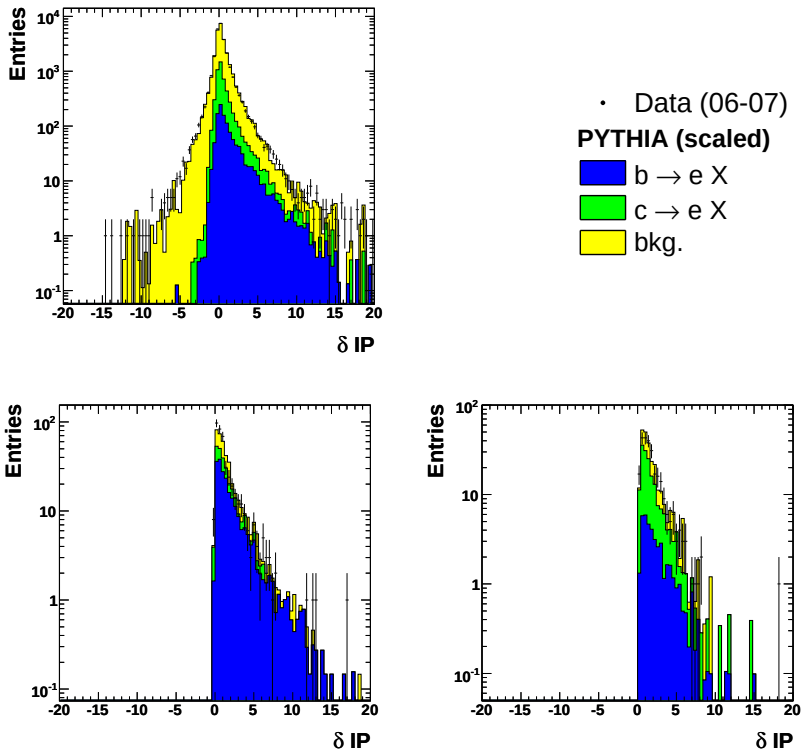


Figure 7.7: *Distribution of δ_{IP} for background (upper left), beauty (lower left) and charm (lower right) enriched samples. The sum of the Monte Carlo contributions from beauty (blue), charm (green) and background (yellow), scaled with the factors as described in Chapter 10, describe the data distribution (points) in all three selections.*

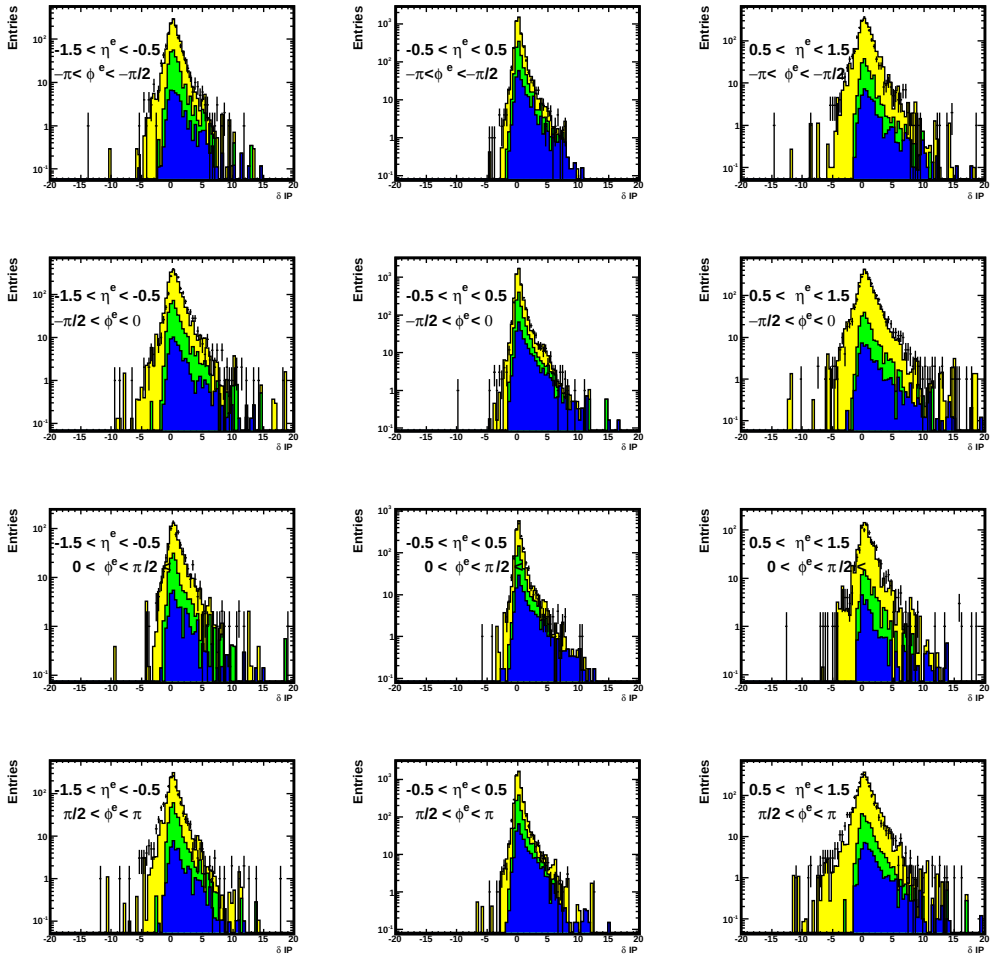


Figure 7.8: *Distribution of δ_{IP} in different bins of p_T^e and η^e (as denoted in the figures) for the background enriched sample. Other details as in caption of Figure 7.7. Except for the last row, where a closer view showed a small shift for the forward and backward tracks, no major discrepancies are observed.*

7.4.2 Kinematic Variables

A more indirect influence of the selection is visible in the kinematic track variables. These do not enter directly into the likelihood but are of large interest as they define the kinematic phase space under study. In Figure 7.9 the transverse momentum of the electron candidate, p_T^e, is shown for the three different selections. The signal electrons clearly have a harder spectrum than the background particles. As the momentum enters the likelihood only indirectly via the relative abundance and via the dE/dx PDF, the separation is not that clear as in the case of p_T^{rel}. If the momentum would directly enter the likelihood the region of low momenta would be depopulated making it impossible to measure the production rates at these energies.

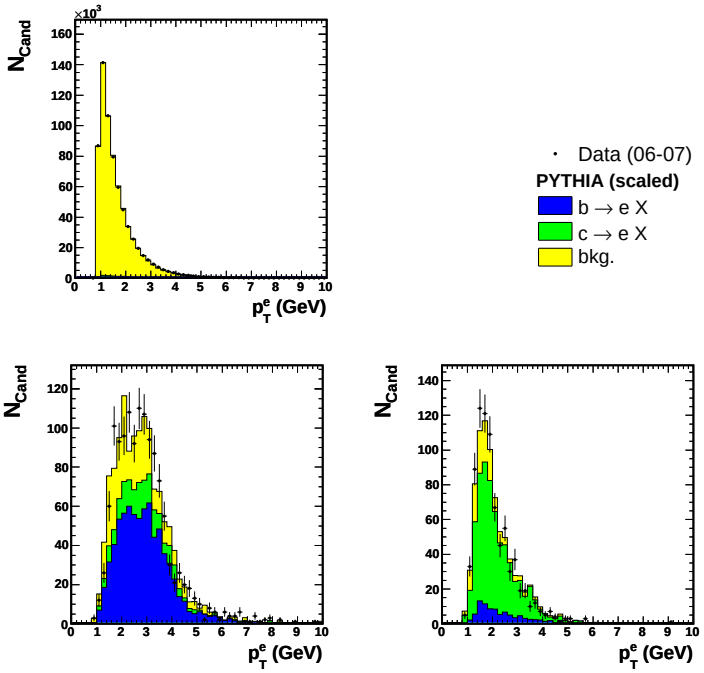


Figure 7.9: *Distribution of the kinematic variable p_T^e. The three histograms show the distribution for all candidates which is background dominated (upper left), for beauty enriched (lower left) and charm enriched (lower right). The Monte Carlo is subdivided into the beauty signal (blue), the charm signal (green) and the background contributions (yellow). The three contributions were scaled by the fit result (see Chapter 10).*

7.5 Alternative Variables

In the previous chapter some alternative variables for the electron identification were sketched briefly. During the HERA II analysis the studies were focused on the improvement of the dE/dx description and on different variables using the lifetime information. The variables not used in the likelihood hypothesis are still useful for systematic studies providing independent techniques for the signal enrichment. For another analysis [133] the variable $\Delta\phi$ was replaced by $p_T^{miss,\|e}$, the component of the missing transverse momentum parallel to the lepton direction. Figure 7.10 shows the control plot and the normalised distributions of $p_T^{miss,\|e}$ on a logarithmic (left) and linear (right) scale. The data distribution is well described

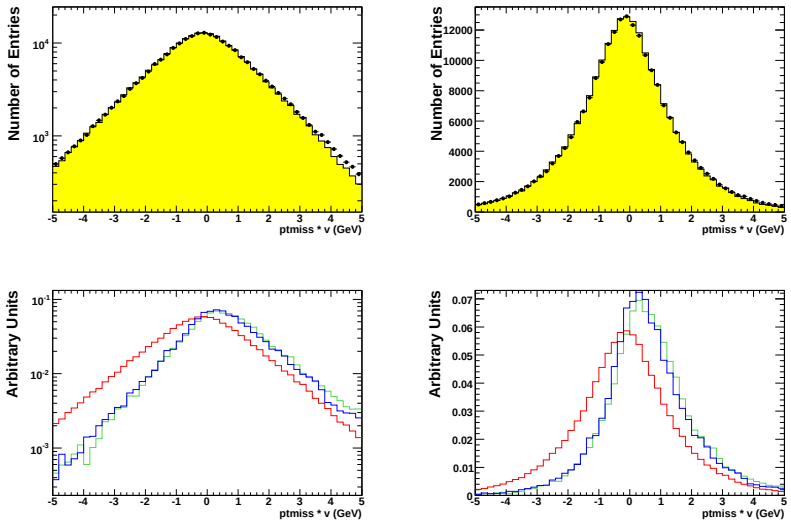


Figure 7.10: *Distribution of the transverse momentum component parallel to the electron direction, $p_T^{miss,\|e}$. The upper row shows the control plot for all candidate in Monte Carlo (yellow) and data (points), where the lower row shows the PDF for e_b (blue), e_c (green) and bkg. (red). To compare the description in both the maximum as well as in the tails, the plots are shown on linear and logarithmic scale.*

by the Monte Carlo. Similar to $\Delta\phi$, this variable is able to distinguish between the heavy and the light flavour contributions. The separation power was found to be a little bit better, but in photoproduction events the systematic uncertainty due to this variable is larger than for $\Delta\phi$ so that this variable was only used as a systematic cross check.

An alternative parametrisation of the impact parameter is the impact parameter weight, W_{IP}, which is defined as following:

$$W_{IP} = \int_{-\inf}^{\delta_{IP}} f(x)dx, \tag{7.1}$$

where $f(x)$ is the normalised fit function of the symmetric background distribution of the impact parameter significance, and δ_{IP} is the measured impact parameter significance of the electron candidate. The distribution of W_{IP}, which is shown in Figure 7.11, is reasonably well described for both signal and the background enriched regions.

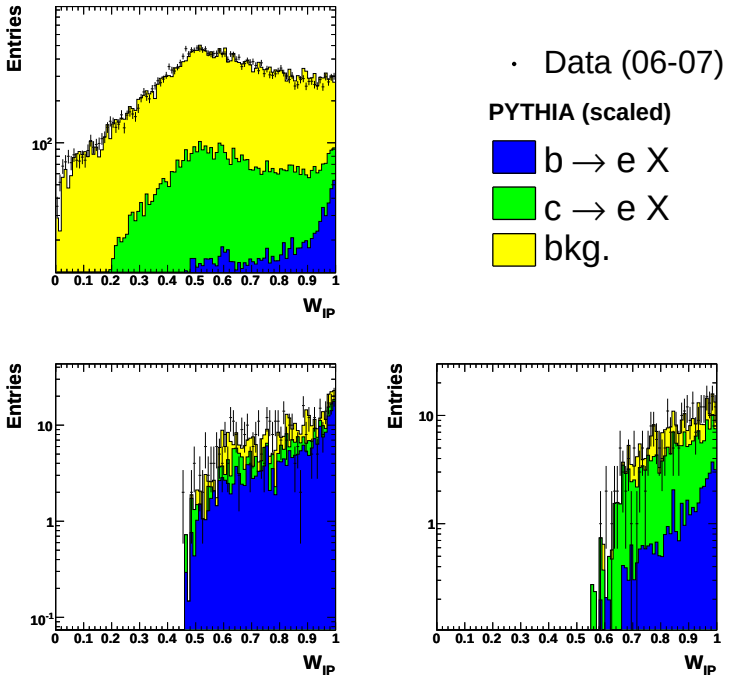


Figure 7.11: *Distribution of the lifetime variable W_{IP}. The three histograms show the distribution for all candidates which is background dominated (upper left), for beauty enriched (lower left) and charm enriched (lower right). The Monte Carlo is subdivided into the beauty signal (blue), the charm signal (green) and the background contributions (yellow). The three contributions were scaled by the fit result (see Chapter 10).*

This parametrisation has the advantage that the separation power is directly visible, as the weight is connected to the probability of being a signal particle. Due to the integration procedure the light flavour contribution is not symmetric anymore which is necessary for the subtraction methods used in other analyses [115], and which provides a good probe for the background reduction efficiency.

To perform tests on signal enriched samples the light flavour contribution has been subtracted statistically by assigning all events with negative decay length with a negative weight. Due to the symmetric distribution of the light flavour background the contribution from the light flavour should cancel in variables which are not correlated to the determination of the decay length. For the likelihood function only the ratio of the PDFs matters. Therefore this approach did not give any additional separation power.

Another variable which has a very high potential for the signal extraction is the decay length significance, δ_{DL}. The samples for the beauty, charm and light flavour enriched region are shown in Figure 7.12. Also this variable is reasonably well

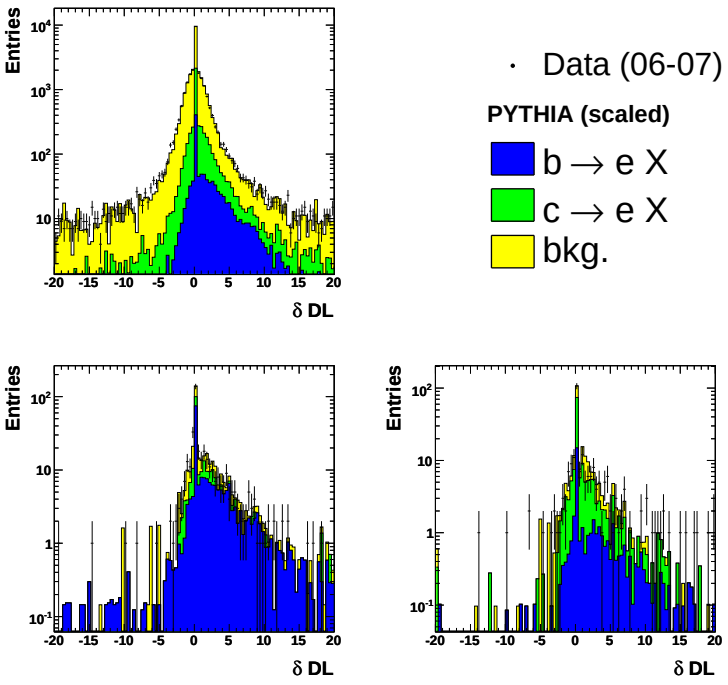


Figure 7.12: *Distribution of the lifetime variable δ_{DL}. The three histograms show the distribution for all candidates which is background dominated (upper left), for beauty enriched (lower left) and charm enriched (lower right). The Monte Carlo is subdivided into the beauty signal (blue), the charm signal (green) and the background contributions (yellow). The three contributions were scaled by the fit result (see Chapter 10). The zero bin corresponds to candidates where no secondary vertex could be assigned.*

described and the separation power is higher then for δ_{IP}, it did not enter the likelihood at the end. The reason is the low selection efficiency for the secondary vertex requirement. The zero bin shows the number of candidates where no secondary vertex could be assigned, and therefore δ_{DL} could not be calculated. From

all selected candidates $\sim 70\,\%$ do not have a proper secondary vertex and would be lost. For the charm enriched sample the fraction is quite similar, whereas for the beauty enriched candidates only $\sim 30\,\%$ of the candidates are lost. This reflects the higher efficiency to reconstruct a secondary vertex in beauty events. The uncertainty on the final cross section was higher using the subsample with the decay length information. Even by adding the additional separation power of m_{vtx} a more precise result was achieved when using the larger sample.

Chapter 8

Cross Section Determination

In this chapter the fitting method and the calculation of the cross sections from the Monte Carlo fractions will be discussed. The fitting method is used for both the determination of the central values for the cross sections as well as for the calculation of the systematic uncertainty which is described in Chapter 9.

8.1 Fitting

The determination of the cross section from the likelihood distributions of data and the three different classes in Monte Carlo is based on the fact that the data consists of the same three classes of particles as treated in the three Monte Carlo sources (see Equation 8.3). The aim of the fit are the fractions f_x in data (see Equation 8.2).

$$\begin{aligned} N^{data} &= N_b^{data} + N_c^{data} + N_o^{data} && (8.1)\\ &= N^{data} * f_b + N^{data} * f_c + N^{data} * f_o && (8.2)\\ &\simeq k_b * N_b^{MC} + k_c * N_c^{MC} + k_o * N_o^{MC}, && (8.3) \end{aligned}$$

where $k_{b/c}$ are the factors with which the Monte Carlo templates have to be multiplied so that the Monte Carlo distributions sum up to the data distribution. The fractions f_b, f_c and f_o are varied within the fit, so that in each bin the sum of the scaled Monte Carlo entries gets close to the sum in data (Equation 8.1). Since the statistics can get low in some bins a *binned maximum likelihood fit* has been used. Also the statistical uncertainty of the Monte Carlo samples is not negligible in all bins, therefore this uncertainty has to be taken into account in the fit. A method, where the shape of the Monte Carlo samples is not completely fixed, but is allowed to vary within its uncertainty, has been developed by Barlow et al. [134] and is implemented in the ROOT TFractionFitter [135, 136].

8.1.1 HERA I

In Figure 8.1 the fit of the Monte Carlo fractions for the HERA I analysis is shown using the TFractionFitter.

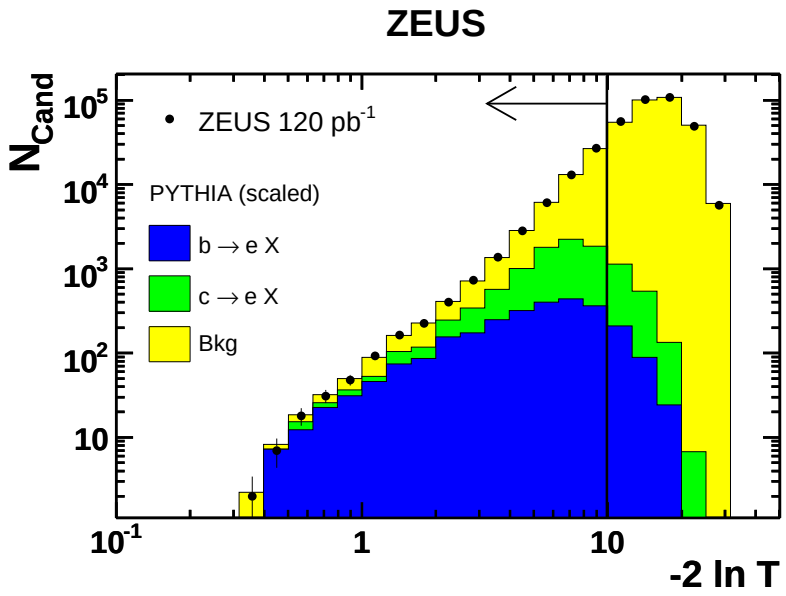


Figure 8.1: *The likelihood distribution for electron candidates, N_{Cand}, in data compared to the Monte Carlo expectation after the fit using the TFractionFitter. The arrow indicates the region included in the fit ($-2\ln T < 10$). The shaded areas show the fitted contributions from b quarks, c quarks and background as denoted in the figure.*

The fractions were determined with a $\chi^2/d.o.f = 13/12$ to be:

$$\begin{aligned} f_b &= (4.8 \pm 0.4)\,\% \\ f_o &= (8.8 \pm 1.0)\,\% \\ f_c &= (86.4 \pm 0.9)\,\%. \end{aligned}$$

These results were found to be consistend with an alternative fit method which was implemented for the HERA II analyses (see Section 8.1.2). The correlation matrix for the three parameters is given in Table 8.1.

	$b \to eX$	$c \to eX$	Bkg.
$b \to eX$	1.00	0.60	-0.21
$c \to eX$	0.60	1.00	-0.75
Bkg.	-0.21	-0.75	1.00

Table 8.1: *Correlation matrix for the three fit parameters using the HERA I data-set.*

8.1.2 HERA II

For the HERA II analysis this method was also used; but to check the uncertainty on the fractions more directly during the fit procedure a new method used in another analysis [133, 137] was adapted. In this new approach the statistical uncertainty of the Monte Carlo distributions were implemented by simulating a large set of Monte Carlo templates with a Gaussian distribution around the central value. By assuming Poisson uncertainties the likelihood is defined by a convolution of two Poisson functions for the variation of the Monte Carlo and the Monte Carlo sum to the data. The distribution of fractions for the Monte Carlo sets, directly gives a handle on the localisation and the width of the minima. In Figure 8.2 the fit for the total visible cross section in the HERA II analysis is shown using the new fit procedure. For the HERA II analysis the beauty and charm fractions are a bit higher:

$$\begin{aligned} f_b &= (5.8 \pm 0.2)\,\% \\ f_o &= (13.9 \pm 0.4)\,\% \\ f_c &= (80.3 \pm 0.5)\,\%. \end{aligned}$$

The absolute values of the fractions cannot be compared directly, as the number of reconstructed data and the efficiency depends on the selection criteria and the fit ranges. The positive effect of higher signal fractions are the purities, improve the stability of the fit. The uncertainties of the signal fractions are significantly reduced, were the main influence is coming from the improved separation power and not from the higher luminosity. Due to the increased statistics the distributions were split into more bins, while the worse description is reflected in $\chi^2/d.o.f = 69/42$. In comparison to the HERA I analysis the entries in the correlation matrix (see Table 8.2) are reduced which has a positive effect on the error propagation. By fixing one of the parameters one can extract a $n - \sigma$ contour in an iterative fit procedure. The $1 - \sigma(68\,\%)$ and $2 - \sigma(95\,\%)$ areas for two free parameters are shown in Figure 8.3.

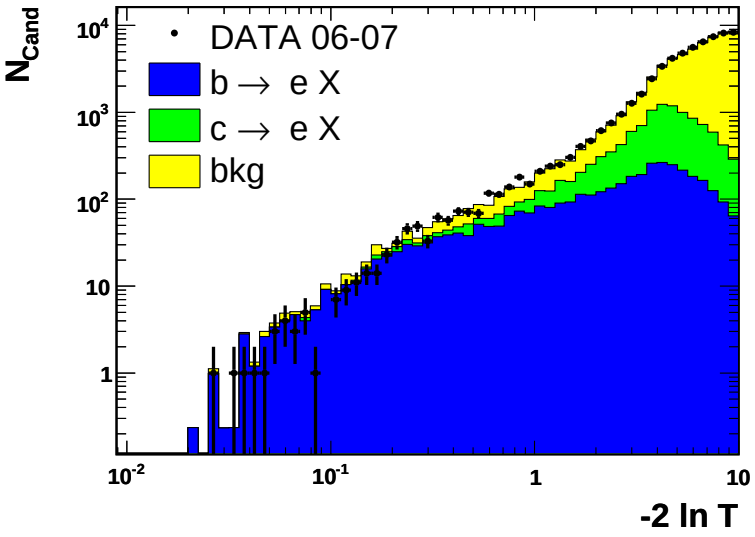


Figure 8.2: *The likelihood distribution for electron candidates, N_{Cand}, in data compared to the Monte Carlo expectation after the fit using the fit described in the text. The shaded areas show the fitted contributions from b quarks, c quarks and background as denoted in the figure.*

A different fit method using a Bayesian approach was also used to extract the Monte Carlo fractions. Therefore a fraction fit was implemented in the existing software package BAT [139]. The results were found to be consistent, but due to time-consuming calculations this fit program was only used as a cross-check. As the BAT program provides not only the most probable fit result including the $1-\sigma$ error but also a probability distribution of the allowed parameter space, it is a very useful tool for the error interpretation. From the probability distributions it has been checked that there are no additional minima which are separated from the found minimum, and that treating the errors as symmetric around the central value is a reasonable approach.

8.2 Acceptance Corrections

For a given number of signal particles, N, and a given integrated luminosity, $\mathcal{L}$, the cross section is given by $\sigma = \frac{N}{\mathcal{L}}$. To determine a cross section, the number of reconstructed signal candidates, which is the signal fraction times the number of

	$b \to eX$	$c \to eX$	Bkg.
$b \to eX$	1.00	-0.35	-0.04
$c \to eX$	-0.35	1.00	-0.62
Bkg.	-0.04	-0.62	1.00

Table 8.2: *Correlation matrix for the three fit parameter using the HERA II data-set.*

reconstructed candidates, N^{data}, has to be corrected for acceptance affects, and in the case of differential cross sections, normalised by the bin width, Δv_k:

$$\frac{d\sigma_b}{dv_k} = \frac{N^{data} \cdot f_b^{data}(v_k)}{\mathcal{A}^b_{v_k} \cdot \mathcal{L} \cdot \Delta v_k}, \tag{8.4}$$

where $A^b_{v_k}$ is the acceptance to find a candidate in bin k of variable v. For the total cross section this formula simplifies to:

$$\sigma_b = \frac{N^{data} \cdot f_b^{data}}{\mathcal{A}^b \cdot \mathcal{L}}. \tag{8.5}$$

The acceptance including detector and software effects was determined using the Monte Carlo simulation and is defined as the ratio of the number of reconstructed particles, N_b^{obs}, to the number of generated particles, N_b^{had}:

$$\mathcal{A} = \frac{N_b^{obs}}{N_b^{had}}, \tag{8.6}$$

The final cross sections are given at hadron level, therefore the number of generated particles are also given at this stage. Here the same kinematic cuts were applied using the generator level and the jet cuts were applied on the hadron level jets (see Chapter 4.6), where the jet finding algorithm has been applied to all final state particles with a lifetime of $\tau > 0.1$ ns. The acceptances and purities for the differential cross sections in the HERA II analysis are shown in Figures B.1 and B.2. The purity, $\mathcal{P}$, is defined as:

$$\mathcal{P} = \frac{N_b^{obs,true}}{N_b^{obs}}, \tag{8.7}$$

where $N_b^{obs,true}$ is the number of candidates reconstructed and selected in the same kinematic bin on hadron and detector level.

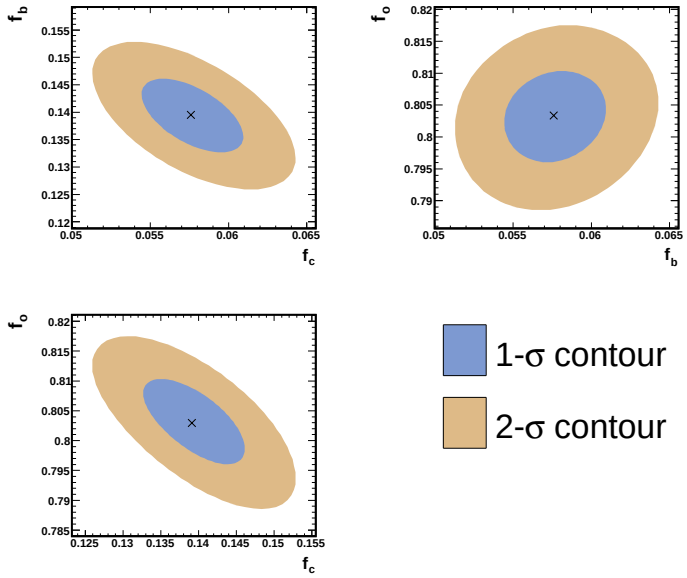


Figure 8.3: *$1-\sigma$ and $2-\sigma$ contour plot. The three plots show the correlation between the three fit variables beauty, charm and background. To calculate a contour line, one variable has been fixed, and by migrating in the 2D space of the other two variables the values were determined, which are $1(2)-\sigma$ away from the central value. (The correlation ellipses were calculated using Minuit [138]*

).

Chapter 9

Systematic Studies

The sources of uncertainty on the cross section which do not have their origin in the statistical fluctuations are called systematic effects. Limitations in the physics and detector simulation can bias the measurement. To estimate these uncertainties parameters like analysis cuts, likelihood parameters but also background sources have to be varied within their uncertainty. The analysis has been performed after each variation redoing the fit to the likelihood distribution to study the relative deviation, $\Delta\sigma/\sigma$ of the cross section bin-by-bin to:

$$\frac{\Delta\sigma}{\sigma} = \frac{\sigma_{syst} - \sigma_{nominal}}{\sigma_{nominal}}. \tag{9.1}$$

By comparing the simultaneous variation of two parameters with the separate variation, possible correlations were checked. Quadratic summation of uncertainties originating from correlated variables would cause overestimation of the systematic uncertainties. Variations that have the same effect on the selection should be avoided[1]. In the following the systematic variations are listed.

9.1 Uncertainties

In this section only the sources of systematics having a significant contribution to the overall uncertainty are listed. For the HERA I analysis the uncertainties were calculated for the total visible cross section and taken as a constant deviation over all bins. This procedure was used as the statistical uncertainty was larger so that no significant fluctuation could be determined. Due to the increased statistics and improved separation power in the HERA II analysis it was possible to calculate

[1] Information about the calculation of systematic uncertainties can be found in [140, 141].

the systematic uncertainties for each bin. A list of the systematic uncertainties classified in three different classes can be found in Table 9.1.

9.1.1 Energy Scale

The CAL energy was shifted according to Equation 3.2 and 3.3 by ±3 % [142]. As the variation was lower than the statistical uncertainty the variation has been made over a wider range and the deviation has been fitted as a function of the variation (see Figure 9.1). From the linear fit the uncertainty at a variation of ±3 % has been

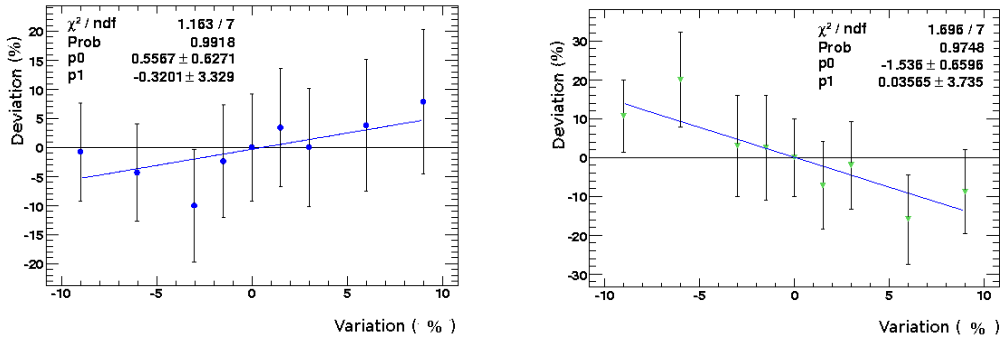


Figure 9.1: *Dependence of the systematic deviation of the beauty (left) and charm (right) cross section from the variation of the energy scale. The energy was shifted in a range of −9 % to +9 % to unfold the statistical and the systematic uncertainty. The deviation was fitted with a linear dependence to determine an uncertainty at* ±3 %.

determined to ±2 % for beauty and ±5 % for charm. For the HERA II analysis this procedure was repeated for each bin. The uncertainty on the total cross section was found to be lower for the beauty and of the same size for the charm cross section. The largest uncertainty is in the first energy bin of the jet, as the energy scale directly affects the dijet selection and migration of the jets in the energy bins.

9.1.2 Electron Background

As the background electrons have larger probability in the likelihood hypothesis than the fake electron contribution, the relative contributions of true electrons in the background sample influences directly the signal fraction. Due to the shape of the distribution, the variation mainly influences the charm fraction. The possible contributions of the true electrons from photoconversions and Dalitz decays are shown in Figure 9.2

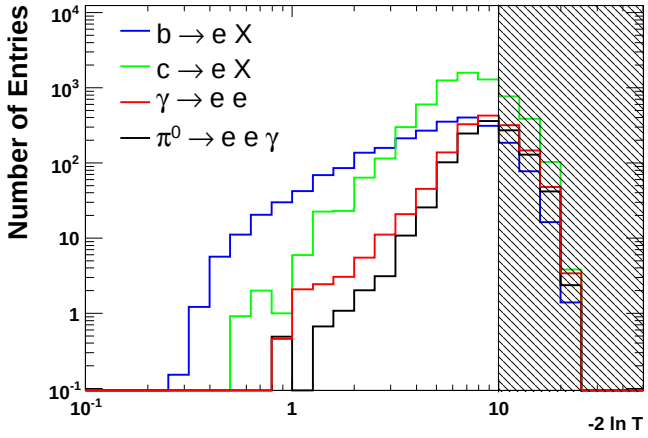


Figure 9.2: *Likelihood test function for beauty (blue) and charm (green) signal as well as for electron from photoconversions (red) and Dalitz decays (black). Uncertainties on the number of background electrons mainly influence the charm fraction as the shape of e_o is similar as for e_c.*

- *Photoconversions*: The upper limit on efficiency uncertainty of the conversion finder has been determined to be 25 %. Therefore the normalisation of this contribution has been varied by this amount. The deviation in the cross section results was up to ±1 % for beauty and ±4 % for charm.

- *Dalitz decays*: In an analogous way the contributions from Dalitz decays have been varied by 20 %. The calculated uncertainties were ±1 % for beauty and ±2.5 % for charm.

- *DIS electrons*: The analysis selection has been applied on a RAPGAP DIS Monte Carlo sample to study the DIS rejection. The number of DIS events surviving the selection has been found to be small. As additional check, the analysis has been performed separately for the negative and positive charged candidates in electron and positron running to exclude DIS electrons in half of the samples. The cross sections have been found to be consistent. Therefore no additional systematic uncertainty has been assigned to the DIS background.

9.1.3 x_γ Reweighting

As shown in Chapter 5 the description of the x_γ distribution is not very good. The origins of this could be imperfections in the description of the detector response as this variable is highly sensitive to a shift in the jet energy and reconstruction of their directions. Physical reasons for differences could be a wrong relative contribution of the direct/resolved processes in the Monte Carlo samples. As a consistency check the contributions of the different Monte Carlo processes were reweighted. Figure 9.3 shows the x_γ distribution for data compared with the best fit of the direct and resolved contributions. The relative contribution of the

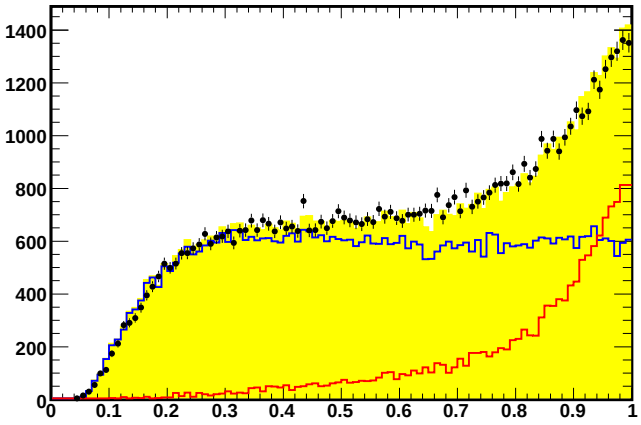


Figure 9.3: *x_γ distribution with reweighted Monte Carlo contributions from direct (red) and resolved (blue) processes. The Monte Carlo sum (yellow) is compared with the data distribution shown as points.*

direct/resolved processes were varied in the range of 25:75 up to 35:65. This variation yields to fluctuations in the cross sections of up to $\sim 3\,\%$. As a final systematic uncertainty the whole x_γ distribution was reweighted in the Monte Carlo. The systematic effect of the x_γ reweighting was $\sim 2\,\%$ for both the beauty and charm cross sections in the HERA I results. As the description of x_γ was improved for the HERA II analysis the effect on the cross section was lower. Therefore this uncertainty was not added to the other uncertainties.

9.1.4 Likelihood Variables

More complicated than variation of cuts or background contributions is the effect of the method using a likelihood fit. To estimate the uncertainty due to the likelihood all ingredients have been varied within a reasonable range. In the following only the effects having a measurable influence on the likelihood are listed. Some more variations and studies are explained in the consistency checks.

dE/dx

The variables where the main separation power was observed are dE/dx for the electron identification and the three variables for the decay identification. Figure 9.4 shows the distribution of the electron and the pion dE/dx probability for data and Monte Carlo. The distributions have been smeared by $\pm 10\,\%$. The limits of uncertainties are be reflected by the red and blue curve.

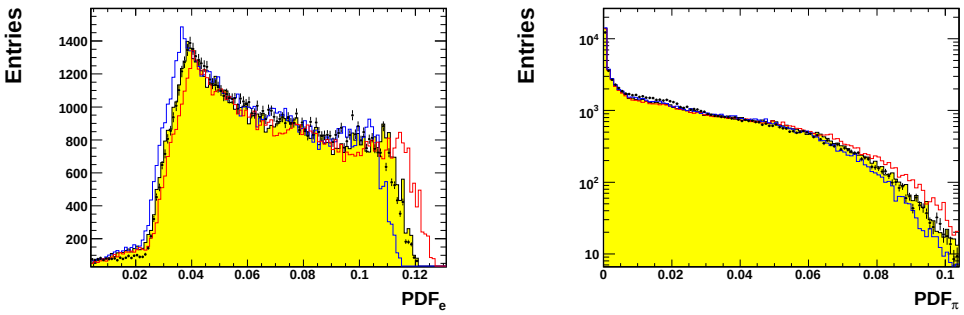


Figure 9.4: *Distribution of the electron probability, PDF_e (left), and pion probability. PDF_π (right). The uncertainty due to discrepancies between data (points) and Monte Carlo (yellow) are studied by smearing the Monte Carlo distributions. The variations are shown by the red and the blue histograms.*

The larger uncertainty was found to come from the electron probability. The uncertainties varied between $\sim 1\,\%$ and $\sim 4\,\%$ for both beauty and charm. For the HERAI analysis this procedure was not possible, as the probability distribution was not available in this way. Therefore the simulation of the Monte Carlo dE/dx was directly varied by varying the width and mean of these dE/dx resolutions. The extracted uncertainties were found to be $^{+1}_{-5}\,\%$ for beauty $^{+10}_{-3}\,\%$ for charm. The stability on the dE/dx variable was significantly improved using the improved dE/dx description.

$\mathbf{p_T^{rel}}$, $\Delta\phi$ and $\delta_{\mathbf{IP}}$

The effect of the decay variables has been studied in a similar way as the dE/dx uncertainty. The data were compared with the Monte Carlo distributions and the PDFs were varied within their uncertainties. Figure 9.5 shows the distributions for p_T^{rel}, $\Delta\phi$ and δ_{IP} including the distributions for the variation.

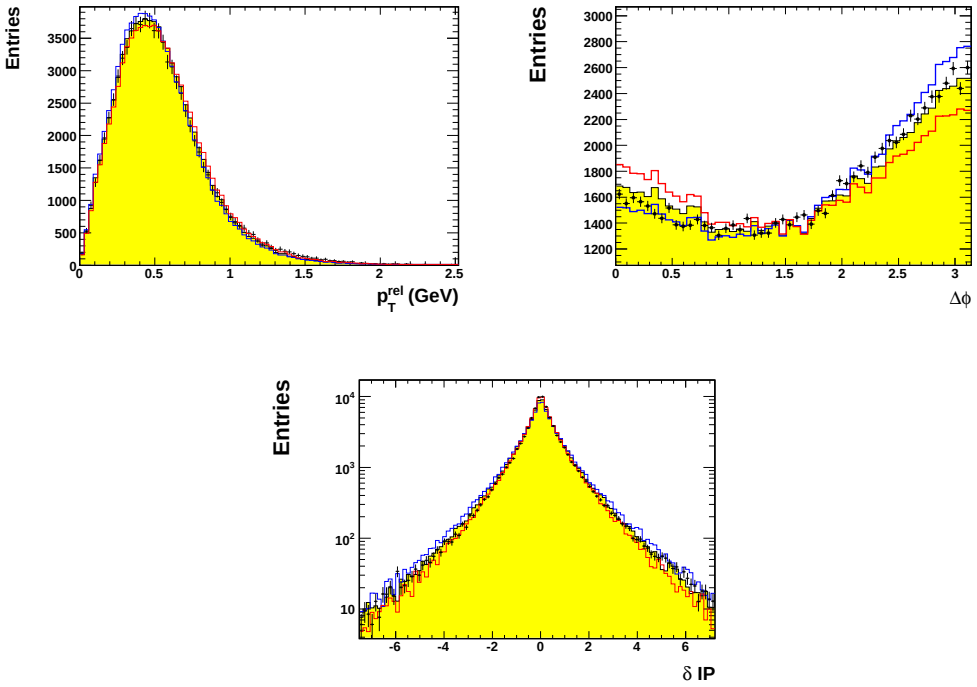


Figure 9.5: *Distribution of the p_T^{rel}, $\Delta\phi$ and δ_{IP} variables. The uncertainty due to discrepancies between data (points) and Monte Carlo (yellow) are studied by varying the Monte Carlo distributions. The variations are shown by the red and the blue histograms.*

Summing up the decay uncertainties, typical values are of the order of $\sim 10\,\%$ but can go up to $\sim 30\,\%$ in a few bins. The uncertainty due to δ_{IP} has a significant contribution increasing the systematic uncertainty of the HERA II analysis. But as the variable improves the separation power especially for the charm determination the variable has an overall positive effect.

9.1.5 Trigger Correction

Studies of the trigger performance for the HERA II analysis showed, that a trigger correction is necessary, to have the same trigger efficiency in data and Monte Carlo. The correction function was calculated as a function of the p_T^{jet2}, the momentum of the second highest energetic jet [143]:

$$w_{MC} = (0.01 * p_T^{jet2} + 0.87). \tag{9.2}$$

The trigger correction was varied by a factor of two to extract a limit on the uncertainty, which was found to be of the order of $\sim 1\,\%$ for beauty and $\sim 3\,\%$ for charm.

9.1.6 Luminosity Uncertainty

The uncertainty on the integrated luminosity has been determined for the different data taking periods. For the data taking from 1996-1997 the uncertainty is 1.8 %, 2.25 % for 1998-2000 and $\sim 3\,\%$ for 2004-2007. As this uncertainty does not depend on the kinematic range, it has been treated separately as an overall normalisation uncertainty.

9.2 Consistency Checks

In addition to the systematic checks several consistency checks were made to demonstrate the validity of the signal extraction procedure. These checks are not variation within an uncertainty but most of the time replacement of cuts or variables or a reweighting of events in order to improve variable descriptions.

9.2.1 Selection Cuts

- *Momentum Cut*: The cut on the transverse momentum of the electron candidate was varied by ±3 %, which corresponds to the momentum uncertainty for a track with $p_T = 0.9\,\text{GeV}$ (Equation 3.1). The resulting variation in the cross section was in all bins below $\sim 1\,\%$
- *Cut on* E^{EMC}/E^{CAL}: For the HERAI analysis the variable has been dropped from the likelihood and replaced by the cut which is used in the HERAII analysis. As the deviation was small and the significance of the fit was not influenced this cut was used in the HERA II analysis to minimise the

number of variables in the likelihood that could influence the simulation of the likelihood response. The variation of the cut between $[0.93, 0.98]$ had a negligible effect.

- *Jet Association*: In the HERA I analysis the jet association was done using a cut on the distance in the $\eta\phi$-plane at $R < 1.5$. This cut has been tightened as a consistency check. The effect of this variation on the measurement was found to be negligible. The few events that were found in addition, when using a cut of 1.5 instead of 1.0, are critical jets in the association to the truth heavy quark and hence reduce the correlation between the jet and quark energy. This also had the effect, that the definition of the association is closer to the alternative approach, where the tracks were associated to the jet directly by the k_T algorithm.

- *Geometry Cuts*: Several studies were made to investigate possible effects of the new geometry cuts. The effect of the variation is only visible in the background enriched region, where the description of the η distribution is improved by applying the cuts. In the signal enriched region and therefore on the cross section the effect is small.

9.2.2 Likelihood Description

- dE/dx: In addition to the variation of the dE/dx description varying the input PDFs separately or simultaneously the dE/dx information was completely dropped from the final likelihood. To be able to extract a reasonable electron purity the likelihood was split into two independent likelihoods: The electron likelihood, which was used to define a hard cut on the electron candidates at $T^e_{dE/dx} < 1$ and the e_b likelihood with the other five variables as input. The extracted cross sections had a much larger uncertainty due to the worse separation power, but were consistent with the default values. Similarly the cross sections using the dE/dx simulation for the HERA I analysis yield to a consistent result with larger uncertainties.

- E^{CAL}/p^{trk}: As the relative influence of E^{CAL}/p^{trk} on the likelihood is relatively small, the variable could be dropped from the likelihood giving a deviation of $\sim 3\,\%$ with larger uncertainties.

- d^{cell}: As described in Chapter 7 the separation power of d^{cell} is less than for the other variables in the likelihood. Therefore remaining uncertainties due to this variable do not have a large influence on the likelihood. This was confirmed, as the resulting cross sections were consistent when dropping

this variable from the likelihood. The fact that this variable has still a positive effect on the electron identification is reflected in increased fit errors after removing the variable.

- p_T^{rel}: In addition to the maximum variation of the p_T^{rel} PDF the PDF was varied using a bin-by-bin correction. The variation was found to be similar to the overall variation. It was not possible to drop this variable from the likelihood and replace it by a hard cut, as it is the main variable for beauty identification.

- $\Delta\phi$: Similar like p_T^{rel} it was not possible to drop this variable from the likelihood. By dropping the variable it was still possible to measure the beauty cross section but the sensitivity for the charm contribution was to low to measure the charm fractions. A consistent result was determined by replacing $\Delta\phi$ by $p_T^{miss\|e}$. The variable $p_T^{miss\|e}$ had a higher separation power, and therefore reduced the uncertainties on the fit results. Nevertheless this variable was only treated as a consistency check as the description was not perfect and the likelihood was found to be very sensitive to imperfections in the Monte Carlo simulation of $p_T^{miss\|e}$.

- δ_{IP}: As a cross check δ_{IP} was replaced by IP or W_{IP}. Both other variables gave consistent results, while the uncertainties were similar using W_{IP} and increased when using IP

9.2.3 Fit Range

Another check on the likelihood fit method was a variation of the fit range. The dependence of the cross section and the uncertainty was studied as a function of the cut on $-2 \ln T_{e,b}$. The fit range used for the final cross section determination is in a stable range, where the uncertainty is minimal. It was possible to vary the fit by three bins in both directions, before the uncertainties got large for the hard cut, and the $\chi^2/d.o.f.$ increases drastically for the looser cuts.

9.2.4 Vertex Description

A reasonable description of the z_{vtx} distribution of the vertex position is necessary for two reasons. First of all the event selection is based on a selection of the z_{vtx} value, which is tightened for the HERA II analysis. Secondly the luminosity measurement is also based on a sample with a cut on z_{vtx}. Therefore migrations would directly influence the selection efficiency and the overall luminosity. To

study the order of possible uncertainties, the z_{vtx} distribution in Monte Carlo has been reweighted to the data. It has been found that the uncertainty on the cross section is less then $\sim 0.1(0.5)$ % for beauty (charm).

9.2.5 Event Yield

One important check after the definition of the event selection is to test whether this selection is stable over time. In Figure 9.6 the yield, i.e. the number of accepted events per luminosity, is shown as a function of the run number for the 06/07 data. There is no major deficit or step visible, except for a small decrease in the second half of 2007 ($\sim$61500). This behaviour is one source of uncertainty for the cross section determination and is covered by the uncertainty of the luminosity measurement. On the level of event selection, the different efficiencies of

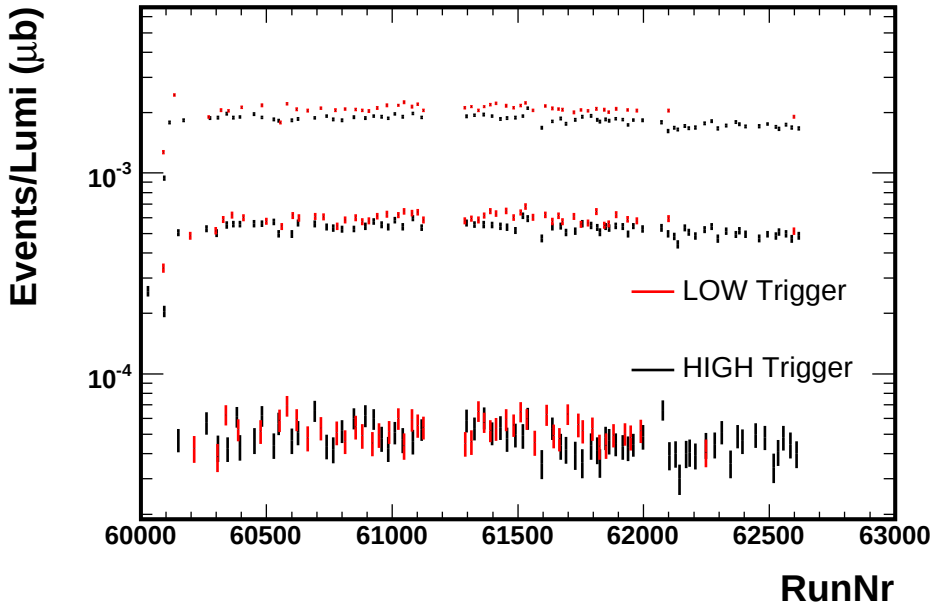


Figure 9.6: *Number of accepted events per luminosity as a function of time. The binning is chosen in a way that runs are added until the integrated luminosity exceed* $\sim 1\,\mathrm{pb}^{-1}$. *The red distributions show the yields for the low trigger configuration where the black distributions show the high trigger configuration. The three different distribution for both show the three steps of the yields at the three steps of the analysis. The high yield correspond to the number of accepted events (see Table 5.3), the middle one to the number of candidates without the cut on* $-2\ln T_e^{dE/dx}$ *(see Table 5.4), and the low one to the good candidates after final selection which were used for the signal extraction* ($-2\ln T_e^{dE/dx} < 3 \,\&\, -2\ln T_b < 10$).

the two trigger configurations are shown. For the preselection, more events pass

the trigger selection using the ”low” trigger configuration, while the yields get closer after the candidate selection and are consistent after the tight cuts for the likelihood fit.

9.3 Summary

Several consistency checks were made to study the stability of the selection and method on the cross section determination. The systematic uncertainties were added in quadrature and were found to be of the same size for beauty in the HERA I ($^{+8}_{-9}$ %) and HERA II ($^{+8}_{-9}$ %) analysis. New systematic effects due to δ_{IP} and trigger corrections had to be added while systematic effect of p_T^{rel} and dE/dx was reduced as both variables are better described in the new dataset. As the systematic uncertainty due to dE/dx had a larger effect on the charm measurement it was possible to reduce the uncertainty on the total cross section from $^{+17}_{-9}$ % for HERA I to $^{+9}_{-8}$ % for HERA II.

Due to the new procedure to handle the systematic uncertainties of the differential cross sections, and the new uncertainties, the systematic uncertainties are increased significantly in some bins. The total systematic uncertainty and the contribution from the different classes of systematics for beauty and charm are listed in Table 9.1 and compared with the statistical uncertainties in Figure 9.7. The statistic uncertainty shown by the yellow area is typical of the order of $\sim 10\,\%$ and goes up to $\sim 30\,\%$ for the bins with lower statistics. The different sources of systematic uncertainty are classified into three different classes. The uncertainties due to the likelihood variables, like dE/dx, p_T^{rel}, $\Delta\phi$ and δ_{IP}, are summarised to *likelihood uncertainties*. The cut uncertainties like the energy scale uncertainties or trigger corrections are combined as well as the background electron sources from photoconversions or Dalitz decays.

For the beauty cross sections the systematic uncertainty is clearly dominated by the likelihood uncertainty. Only in the low energy region the measurement is sensitive to the trigger and energy uncertainties. The charm measurements are more influenced by the background electrons as the likelihood distribution for background electrons has a similar shape as the charm signal. Therefore all three classes of systematic uncertainties contribute to the total uncertainty. Except for the low statistics bins the systematic uncertainty is larger than the statistical uncertainty for the charm measurement; for the beauty measurement they are of similar size.

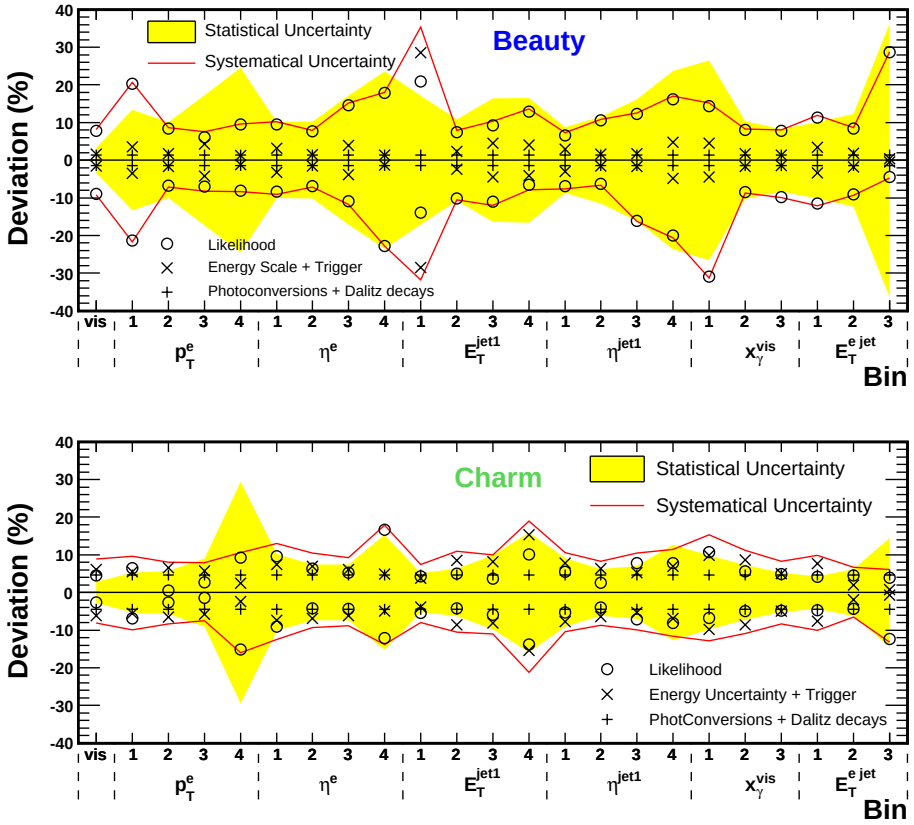


Figure 9.7: *Comparison of statistic and systematic uncertainty for the total and differential cross section measurement for beauty (upper plot) and charm (lower plot). The statistic uncertainty shown by the yellow area is typically of the order of* $\sim 10\,\%$ *and goes up to* $\sim 30\,\%$ *for the bins with lower statistics. The different sources of systematic uncertainty are classified into three different classes. The uncertainties due to the likelihood variables are summarised to likelihood uncertainties. The cut uncertainties like the energy scale uncertainties or trigger corrections are combined as well as the background electron sources from photoconversions or Dalitz decays.*

Bin	Beauty Likel.	Beauty Cuts	Beauty Elec.	Beauty ∑syst	Charm Likel.	Charm Cuts	Charm Elec.	Charm ∑syst
vis	$^{+7.8}_{-9.0}$%	±1.6%	±1.4%	$^{+8.1}_{-9.2}$%	$^{+5.4}_{-2.7}$%	±6.1%	$^{+4.6}_{-4.5}$%	$^{+8.9}_{-8.1}$%
1 p_T^e	$^{+20.3}_{-21.4}$%	±3.5%	”	$^{+20.7}_{-21.5}$%	$^{+6.5}_{-6.7}$%	±5.5%	”	$^{+9.6}_{-9.4}$%
2 p_T^e	$^{+8.4}_{-6.8}$%	±1.7%	”	$^{+8.7}_{-7.1}$%	$^{+0.6}_{-2.6}$%	±6.5%	”	$^{+8.0}_{-7.0}$%
3 p_T^e	$^{+6.1}_{-7.0}$%	±4.2%	”	$^{+7.6}_{-7.2}$%	$^{+2.7}_{-1.5}$%	±5.8%	”	$^{+7.8}_{-6.6}$%
4 p_T^e	$^{+9.5}_{-8.2}$%	±1.2%	”	$^{+9.7}_{-8.4}$%	$^{+9.2}_{-15.1}$%	±2.4%	”	$^{+10.6}_{-16.4}$%
1 η^e	$^{+9.5}_{-8.3}$%	±3.2%	”	$^{+10.2}_{-8.6}$%	$^{+9.7}_{-9.1}$%	±7.4%	”	$^{+13.0}_{-11.1}$%
2 η^e	$^{+7.6}_{-6.9}$%	±1.5%	”	$^{+7.9}_{-7.2}$%	$^{+6.3}_{-4.2}$%	±6.9%	”	$^{+10.4}_{-7.7}$%
3 η^e	$^{+14.6}_{-11.0}$%	±3.9%	”	$^{+15.2}_{-11.1}$%	$^{+5.2}_{-4.4}$%	±6.1%	”	$^{+9.2}_{-7.8}$%
4 η^e	$^{+17.9}_{-22.8}$%	±1.3%	”	$^{+18.0}_{-22.9}$%	$^{+16.6}_{-12.1}$%	±4.9%	”	$^{+17.9}_{-13.8}$%
1 E_T^{jet1}	$^{+21.0}_{-13.9}$%	±28.5%	”	$^{+35.4}_{-14.1}$%	$^{+4.3}_{-5.4}$%	±3.8%	”	$^{+7.3}_{-8.4}$%
2 E_T^{jet1}	$^{+7.5}_{-10.1}$%	±2.4%	”	$^{+8.0}_{-10.3}$%	$^{+5.0}_{-4.2}$%	±8.5%	”	$^{+10.9}_{-7.7}$%
3 E_T^{jet1}	$^{+9.2}_{-11.0}$%	±4.5%	”	$^{+10.3}_{-11.2}$%	$^{+3.7}_{-5.9}$%	±8.1%	”	$^{+10.0}_{-8.8}$%
4 E_T^{jet1}	$^{+12.9}_{-6.6}$%	±4.1%	”	$^{+13.6}_{-6.9}$%	$^{+10.1}_{-13.9}$%	±15.3%	”	$^{+18.9}_{-15.3}$%
1 η^{jet1}	$^{+6.6}_{-7.0}$%	±3.0%	”	$^{+7.4}_{-7.2}$%	$^{+5.4}_{-5.3}$%	±7.8%	”	$^{+10.5}_{-8.4}$%
2 η^{jet1}	$^{+10.6}_{-6.4}$%	±1.6%	”	$^{+10.8}_{-6.7}$%	$^{+2.6}_{-4.0}$%	±6.4%	”	$^{+8.3}_{-7.6}$%
3 η^{jet1}	$^{+12.3}_{-16.1}$%	±1.7%	”	$^{+12.5}_{-16.2}$%	$^{+7.8}_{-7.2}$%	±5.3%	”	$^{+10.5}_{-9.6}$%
4 η^{jet1}	$^{+16.2}_{-20.0}$%	±4.8%	”	$^{+17.0}_{-20.1}$%	$^{+7.8}_{-8.2}$%	±6.9%	”	$^{+11.4}_{-10.4}$%
1 x_γ^{obs}	$^{+14.4}_{-30.9}$%	±4.5%	”	$^{+15.1}_{-31.0}$%	$^{+10.6}_{-6.8}$%	±9.8%	”	$^{+15.3}_{-9.4}$%
2 x_γ^{obs}	$^{+8.0}_{-8.4}$%	±1.7%	”	$^{+8.3}_{-8.7}$%	$^{+5.6}_{-5.0}$%	±8.6%	”	$^{+11.2}_{-8.2}$%
3 x_γ^{obs}	$^{+7.7}_{-9.8}$%	±1.4%	”	$^{+8.0}_{-10.0}$%	$^{+4.8}_{-4.9}$%	±5.0%	”	$^{+8.3}_{-8.1}$%
1 $E_T^{e\,jet}$	$^{+11.4}_{-11.4}$%	±3.4%	”	$^{+11.9}_{-11.7}$%	$^{+4.1}_{-4.7}$%	±7.7%	”	$^{+9.8}_{-8.0}$%
2 $E_T^{e\,jet}$	$^{+8.4}_{-9.1}$%	±1.8%	”	$^{+8.7}_{-4.9}$%	$^{+4.5}_{-4.4}$%	±2.0%	”	$^{+6.7}_{-7.8}$%
3 $E_T^{e\,jet}$	$^{+28.7}_{-4.4}$%	±0.4%	”	$^{+28.8}_{-4.9}$%	$^{+3.9}_{-12.4}$%	±0.7%	”	$^{+6.0}_{-14.0}$%

Table 9.1: *Systematic uncertainties for the beauty and charm measurements shown for the HERA II data-set. For each bin the total systematic uncertainty and the contributions from the three subclasses of electron uncertainties, energy scale and trigger uncertainties and electron background uncertainties are listed. For the systematic uncertainty due to the electron background uncertainty no dependence on the variable under study has been found. Therefore this value has been treated as a global uncertainty.*

Chapter 10

Cross Sections and Comparison to Theoretical Predictions

10.1 Visible Cross Sections

Using Equation 8.5 the fractions from the fits (see Chapter 8) can directly be converted to visible cross sections. All following cross sections have been measured in the kinematic range $Q^2 < 1\,\mathrm{GeV}^2$, $0.2 < y < 0.8$, with at least two jets with $E_T^{jet} > 7(6)\,\mathrm{GeV}$, $|\eta^{jet} < 2.5|$ and an electron from a semileptonic decay with $p_T^e > 0.9\,\mathrm{GeV}$ in the range $|\eta^e| < 1.5$.

The cross sections are given for the data taking periods HERA I (96-00) and HERA II (06-07), and compared with NLO QCD calculations calculated with the FMNR program. The following beauty cross sections, given for a centre-of-mass energy of $\sqrt{s} = 318\,\mathrm{GeV}$, were obtained:

$$\begin{aligned} \sigma_b^{vis}(96-00) &= (125 \pm 11(stat.)^{+10}_{-11}(syst.))\,\mathrm{pb} \\ \sigma_b^{vis}(06-07) &= (148 \pm 5(stat.)^{+12}_{-14}(syst.))\,\mathrm{pb}. \end{aligned}$$

The HERA II cross section is a little bit higher than the HERA I, but they are consistent within the uncertainties. For HERA II the statistical uncertainty is significantly reduced, whereas the systematic uncertainties are similar. For the HERA II analysis the systematic uncertainties are still under investigation and will be reduced in the future. At the time when this thesis was written the trigger correction was still under development. Therefore the uncertainty was calculated on a rather conservative level to cover possible central values after final corrections. Additionally the variation for of the impact parameter resolution was varied over a wide range as the tuning of the MVD resolution is still ongoing and could affect the distributions. In the final analysis where the full HERA II data-set will

be analysed especially the systematic uncertainty on the beauty cross section will be reduced. The improvement in the statistical uncertainty, which is dominated by the fit error, is caused both by the increased statistics and even more by the improved separation power of the likelihood function. In comparison to the NLO calculations (see Equation 10.1) the HERA I measurement was fairly consistent within the error limits and for HERA II, the cross section is about 2.3σ higher than the prediction.

From the simultaneous fit of the beauty and the charm fractions the visible cross sections for semileptonic electrons from charm have been determined:

$$\begin{aligned}\sigma_c^{vis}(96-00) &= (278 \pm 33(stat.)^{+48}_{-24}(syst.))\,\text{pb}\\ \sigma_c^{vis}(06-07) &= (315 \pm 9(stat.)^{+28}_{-25}(syst.))\,\text{pb},\end{aligned}$$

where again the HERA II result is higher than the HERA I result, with improved precision. The improvements in the separation power of the likelihood and the reduced electron background due to the improved conversion finder caused a higher reduction of the relative uncertainty on the HERA I charm measurement compared to the beauty cross sections. The systematic uncertainty, which is under better control for the new data, is of similar size, but more symmetric compared with the HERA I analysis. The charm cross sections are consistent with the NLO QCD prediction given in Equation 10.2.

For the HERA I analysis, the total cross section was also determined separately for $\sqrt{s} = 300\,\text{GeV}$ and $\sqrt{s} = 318\,\text{GeV}$. The measurements are given in Table C.1 together with the summed cross sections and shown in Figure 10.1. The cross sections at the two centre-of-mass energies are consistent with each other and agree with the HERA II results. In the figure the cross sections are compared with the NLO QCD prediction as well as the LO+PS prediction from the PYTHIA Monte Carlo. Good agreement with the NLO calculation is observed.

To compare the shape of the PYTHIA prediction with the later measurements, the contributions from beauty and charm have been scaled with a constant scaling factor. For beauty (charm) an average factor of $\sim 2.0(1.5)$ has been taken, which describes the normalisation for both data periods reasonable well. These factors are also used to scale the PYTHIA predictions in the following figures, and were also used for the control plots at different stages of signal enrichment.

10.2 Next-to-leading Order Predictions

For the comparison with next-to-leading order QCD predictions a massive calculation, based on the program written by Frixione Mangano Nason Ridolfi (FMNR)

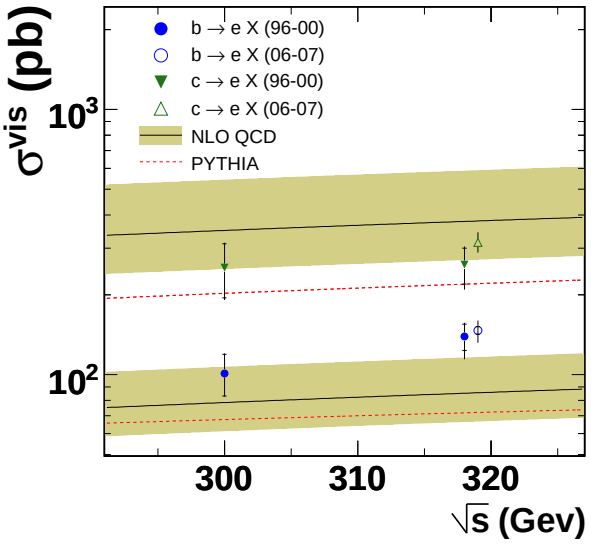
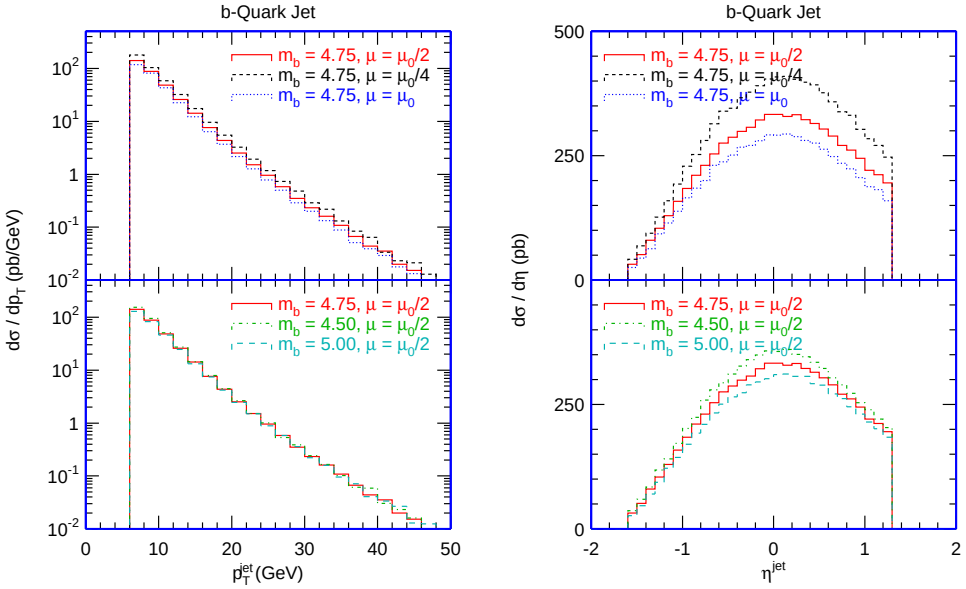


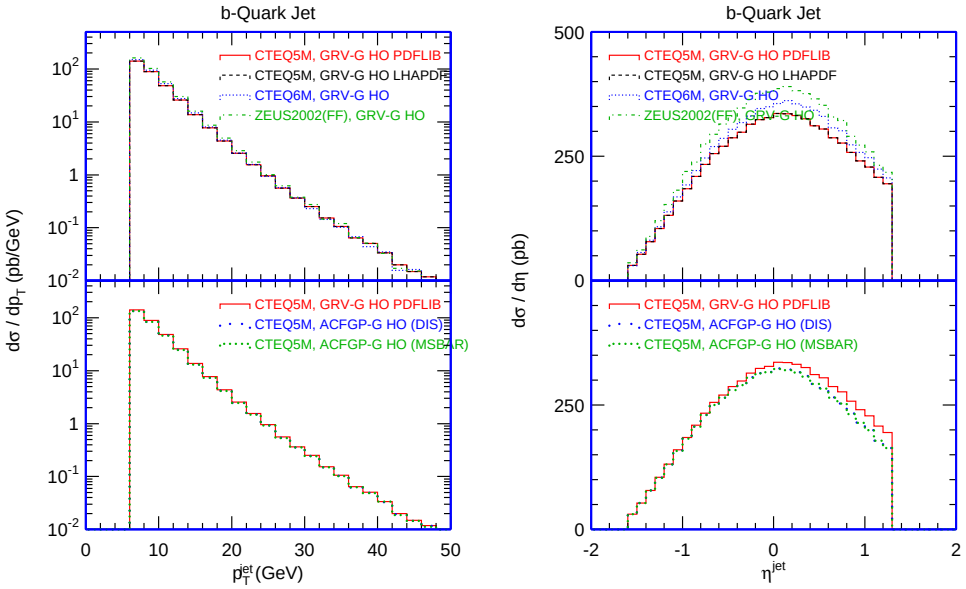
Figure 10.1: *Total cross sections for electrons from b and c quarks in photoproduction events. The measurements are shown as points, where the measurement for the HERA II analysis have been shifted away from their nominal centre-of-mass energy in order to be visible. The inner error bar shows the statistical uncertainty and the outer error bar shows the statistical and systematic uncertainties added in quadrature. The solid lines show the NLO QCD prediction after hadronisation corrections, with the theoretical uncertainties indicated by the band; the dashed lines show the prediction from PYTHIA.*

[144], was used. The calculations were done using the $\bar{MS}$ subtraction scheme, where soft and collinear divergences are cancelled by generating correlated events which have to be summed [145]. As this calculation is highly sensitive to the energy or the ratio of p_T^2/m^2 the resulting cross sections are sensitive to the choice of renormalisation and factorisation scales.

Figure 10.2 show differential cross sections as a function of p_T^e and η^e for different sets of scale parameter (Figure 10.2(a)) and PDFs (Figure 10.2(b)). The variation of the mass and the scale in the FMNR calculation is the dominant fraction the overall NLO uncertainty. The variation of PDFs from the CTEQ parametrisation as denoted in the figure is comparable small, while the ZEUS-S PDF using the fixed flavour numbering scheme results in higher cross sections. The NLO QCD calculations used for the comparison in this theses were extracted using only CTEQ PDFs.



(a) Scale and mass variation



(b) PDF variation

Figure 10.2: *The variation of the mass and the scale (a) in the FMNR calculation is the dominant fraction of the overall NLO uncertainty. The variation of the PDF is shown in (b). Courtesy of [146].*

The total visible cross sections were calculated to $\sigma_b^{NLO} = (109^{+27}_{-16})$ pb and $\sigma_c^{NLO} = (376^{+168}_{-109})$ pb, while the uncertainties were calculated by varying the parameters of this calculation. The following parameter set was used for the central value of the cross section prediction:

- Proton PDF: CTEQ5M [147]
- Photon PDF: GRV-G HO [148]
- the heavy quark masses (pole masses) were set to:
  $m_b = 4.75\,\text{GeV}$ and $m_c = 1.6\,\text{GeV}$
- the renormalisation/factorisation scales were set to:
  $\mu_R = \mu_F = \mu_0 = \sqrt{\hat{p}_T^2 + m_{b(c)}^2}$,
  where $\hat{p}_T^2$ is the average transverse momentum of the two heavy quarks
- Peterson fragmentation [149] with $\epsilon_b = 0.0035$ and $\epsilon_c = 0.035$
- $\Lambda_{QCD}^{(5)}$ was set to $0.226\,\text{GeV}$
- b (c) branching fractions to decay into electrons was set to $\mathcal{B} = 0.221(0.096)$. For beauty both the contributions from prompt and from cascade decays, excluding $b \to \tau \to e$ and $b \to J/\Psi \to e^+e^-$, are taken into account in the effective branching fractions. The spectrum of the electrons from beauty and charm has been taken from PYTHIA.

The uncertainty on the calculated cross section is dominated by the scale uncertainty, which is directly connected to the heavy quark mass. To maximise the change of cross section the masses and the scales were varied simultaneously. The mass ranges were $m_b \in [4.5, 5.0\,\text{GeV}]$ and $m_c \in [1.35, 1.85\,\text{GeV}]$ and the scales were varied within $\mu_R = \mu_F \in [\frac{1}{2}\mu_0, 2\mu_0]$.

In contrast to the effect of the variation of the Peterson parameter $\epsilon_b \in [0.02, 0.05]$ the variation of $\epsilon_c \in [0.02, 0.07]$ lead to a significant cross section variation and was added in quadrature to the systematic uncertainty. The variation of $\Lambda_{QCD}^{(5)}$ between $[0.20, 0.25]$, the variation of the PDF parametrisation, and the variation of the decay spectra did not have a significant effect on the cross section uncertainty. The decay spectra was once calculated from a simple free quark decay model, and compared in addition to spectra taken from experimental measurements [120]. The total uncertainty on the NLO QCD cross section predictions are of the order of +25 % and −15 % for beauty and +45 % and +28 % for charm.

10.2.1 Hadronisation Corrections

The NLO QCD calculations, which provide predictions on parton level, have to be corrected for hadronisation effects (see Section 2.3.2). A bin-by-bin procedure was used according to $d\sigma = d\sigma_{NLO} \cdot C^{had}$, where $d\sigma_{NLO}$ is the cross section for partons in the final state of the NLO calculation. The hadronisation correction factor, C^{had}, was defined as the ratio of dijet cross sections, extracted from the PYTHIA Monte Carlo, after and before the hadronisation process, $C^{had} = d\sigma_{MC}^{Hadrons}/d\sigma_{MC}^{Partons}$. In the PYTHIA Monte Carlo, the parton level is defined analogously to the hadron level, with the difference, that the jets were reconstructed by applying the k_T algorithm to the outgoing partons. The numbers for the hadronisation corrections are listed in Tables C.1-C.5. The corrected total visible cross section at hadron level is:

$$\sigma_b^{NLO} = (88^{+22}_{-13})\,\text{pb} \quad (10.1)$$

$$\sigma_c^{NLO} = (380^{+170}_{-110})\,\text{pb}. \quad (10.2)$$

10.3 Differential Cross Sections

The differential cross sections as a function of p_T^e, η^e, E_T^{jet1}, η^{jet1}, x_γ^{obs} and E_T^{ejet} are shown in Figures 10.3, 10.4, 10.5 and 10.6, respectively. The fractions to calculate the differential cross sections were determined bin-by-bin from the likelihood distributions given in Figures C.1-C.3. The values of the differential cross sections including the statistical and systematic uncertainties are listed together with the NLO predictions and the hadronisation corrections in Tables C.2 - C.5. Both the predictions from the NLO QCD calculations as well as the scaled PYTHIA cross sections describe the data well. In general the new results have a trend to larger values, but are consistent with the previous result. The disagreement for b quarks in the first η bin of both the electron and the jet is still under study. No systematic effect for the deviation has been found so far. Before publishing the new results, this discrepancy needs more investigation. It might be related to the disagreement at high y which is observed in the background and beauty enriched samples as shown in Figures A.14 and A.15. A possible reason could be remaining background electrons from DIS in the sample. For the HERA I analysis this disagreement was not observed.

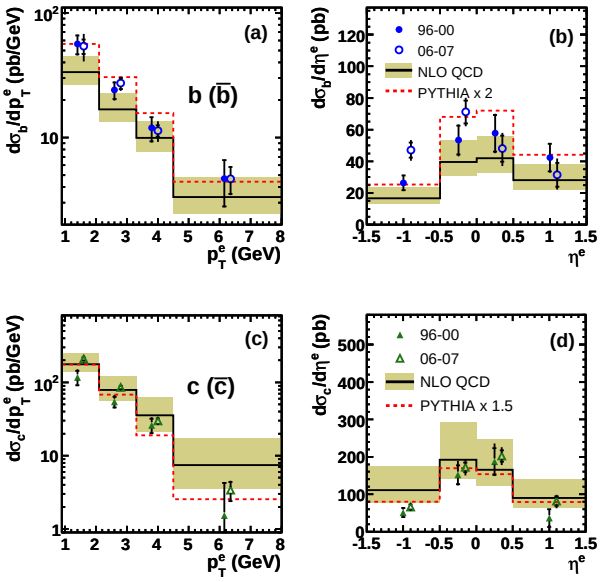


Figure 10.3: *Differential cross sections for 99-00 and 06-07 as a function of a), c) the transverse momentum and b), d) the pseudorapidity of the electrons. Plots a) and b) are for b-quark production while c) and d) are for c-quark production. The measurements are shown as points. The inner error bar shows the statistical uncertainty and the outer error bar shows the statistical and systematic uncertainties added in quadrature. The solid line shows the NLO QCD prediction after hadronisation correction, with the theoretical uncertainties indicated by the band; the dashed line shows the scaled predictions from PYTHIA.*

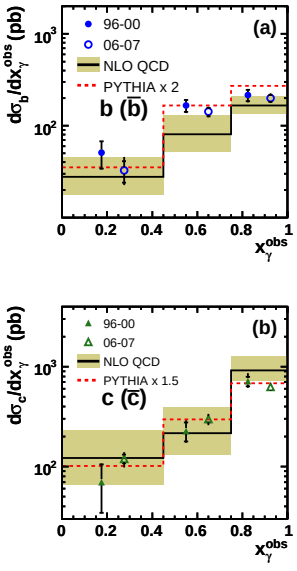


Figure 10.4: *Differential cross sections as a function of x_γ^{obs}. a) shows the distribution for electrons from b-quark production while b) shows c-quark production. Other details as in the caption of Figure 10.3.*

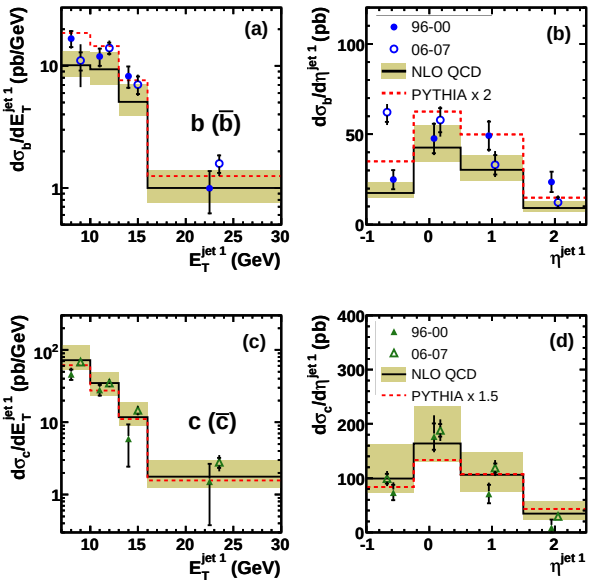


Figure 10.5: *Differential cross sections for 99-00 and 06-07 as a function of a), c) the transverse energy and b), d) the pseudorapidity of the leading jet. Plots a) and b) are for b-quark production while c) and d) are for c-quark production. Other details as in the caption of Figure 10.3.*

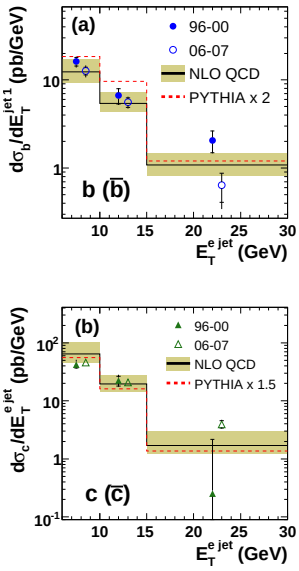


Figure 10.6: *Differential cross sections for a) b-quark and b) c-quark production as a function of the transverse energy of the jet associated to the electron, $E_T^{e\,jet}$. Other details as in the caption of Figure 10.3.*

10.4 Comparison to Other Measurements

The different kinematic regions of the heavy flavour measurements from ZEUS [46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61] make it impossible to compare the results directly with each other. For a comparison it was necessary to extrapolate the measured quantities to the kinematic ranges of inclusive data as it was shown in [109]. The disadvantage of this method is the large extrapolation uncertainty, spoiling the precision of the measurement. In another approach, the ratios of the measurements to the theoretical predictions are compared. This method is used for the beauty measurements in photoproduction as shown in Figure 2.14.

The observed good agreement to the NLO calculation (see Figure 10.6) made it possible to transform the cross sections measured as a function of $E_T^{e\,jet}$ to more general cross sections as a function of the b quark transverse momentum, p_T^b, which is independent from the event selection. The aim of this transformation is the ability to compare this measurement with other measurements in the photoproduction regime from both H1 and ZEUS. The ratios of the measurements as a function of $E_T^{e\,jet}$ to the NLO QCD calculation have been transferred bin-by-bin to ratios in the measurement as a function of p_T^b. The position of the bin measurement in the phase space of quark momentum has been determined by using the mean of the momentum distribution in the three jet energy bins. Figure 10.7 shows that the jet transverse energy is highly correlated to the transverse momentum of the initial quark, validating this procedure.

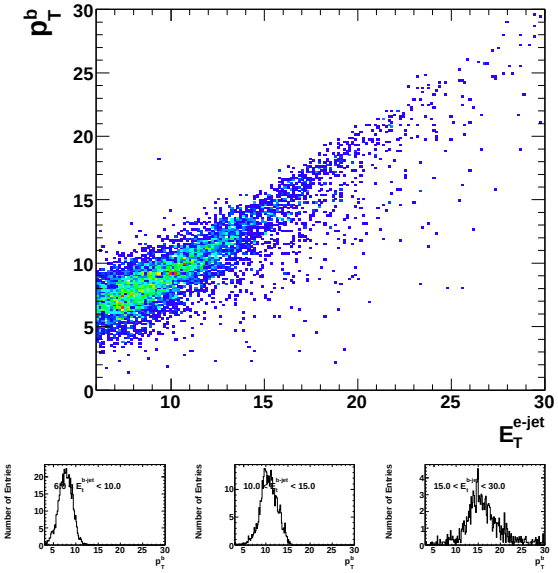


Figure 10.7: *Correlation between the jet energy, $E_T^{e\,jet}$, and the transverse momentum of the originating b quark, p_T^b. The lower plots show the projected distribution of p_T^b for the three bins of the measurement as a function of $E_T^{e\,jet}$. The mean values were used to determine the position in the phase space of the b quark where this analysis is located in comparison with previous measurements.*

From the three projections the mean value of p_T^b has been taken as the centre-of-gravity of the three measured values. In Figure 10.8 the differential cross section for b-quark production is shown as a function of its transverse momentum. The

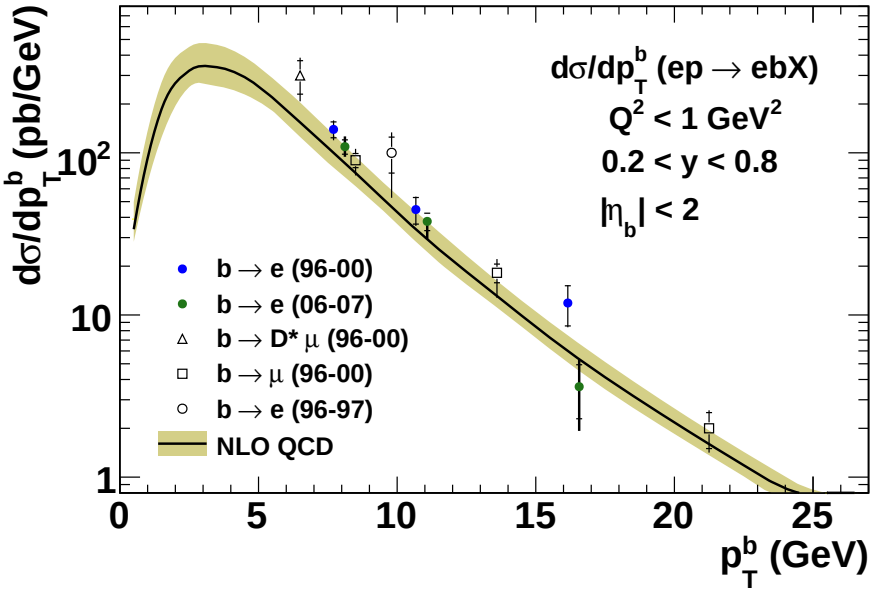


Figure 10.8: *Differential cross section for b-quark production as a function of its transverse momentum, p_T^b, compared to the results of previous ZEUS measurements as indicated in the figure. The measurements from the HERA I analysis are shown as blue points and the one from the HERA II analysis, which are shifted to the right to be separated, are shown as green points. The inner error bar shows the statistical uncertainties added in quadrature. The solid line shows the NLO QCD prediction from the FMNR program with the theoretical uncertainty shown as the shaded band.*

new points from the HERA II data set were added to Figure 10.8 taken from the publication [16]. Another comparison of the HERA I analysis with more recent results as well as with results obtained with H1 data was already shown in Figure 2.14. The results shown in both plots overlap in p_T^b with other measurements and have comparable or lower uncertainties, giving a consistent picture of b-quark production in ep collisions at low Q^2.

The ratios of the total cross sections for beauty and charm [46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73] to the NLO QCD calculation are shown in Figures 10.9 and 10.10, respectively. The

figures show the ratios for both the photoproduction and deep inelastic scattering regime. The results presented in this thesis are shown by the red circles and triangles. They are compared with various measurements from ZEUS (black) and H1 (grey). The measured ratios are of the order of ~ 1.5 for beauty, where the more precise measurements tend to have lower ratios. In the case of charm measurements the ratios fluctuate around a value of ~ 1. In addition to the NLO QCD prediction obtained using the CTEQ5M PDF the central value for the prediction using the ZEUS-S PDF [150] is shown. The calculation using ZEUS-S is in general above the central value of the CTEQ5M calculation. The presented analysis is consistent with the previous beauty measurements and at the lower side of the charm measurements.

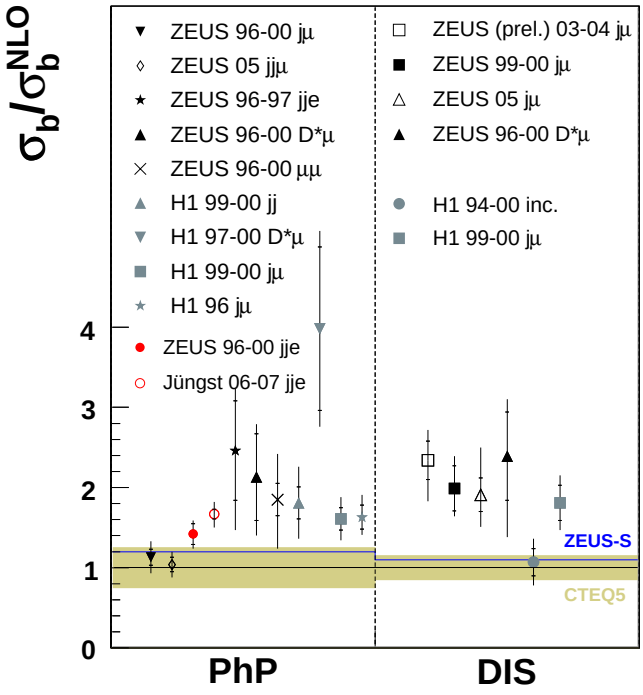


Figure 10.9: *Open beauty production in photoproduction and deep inelastic scattering. Plotted is the ratio of the measured total cross section to NLO QCD predictions. The typical uncertainties for the NLO QCD predictions using the CTEQ5 PDFs are indicated by the shaded band. In addition the obtained prediction using ZEUS-S PDF is shown by the blue line. The results presented in this thesis are shown by the red circles, where the full circle shows the HERA I and the open circle the HERA II result. They are compared with the measurements from ZEUS (black) and H1 (grey). The measured ratios are of the order of* ~ 1.5 *for beauty. The more precise measurements tend to have lower ratios.*

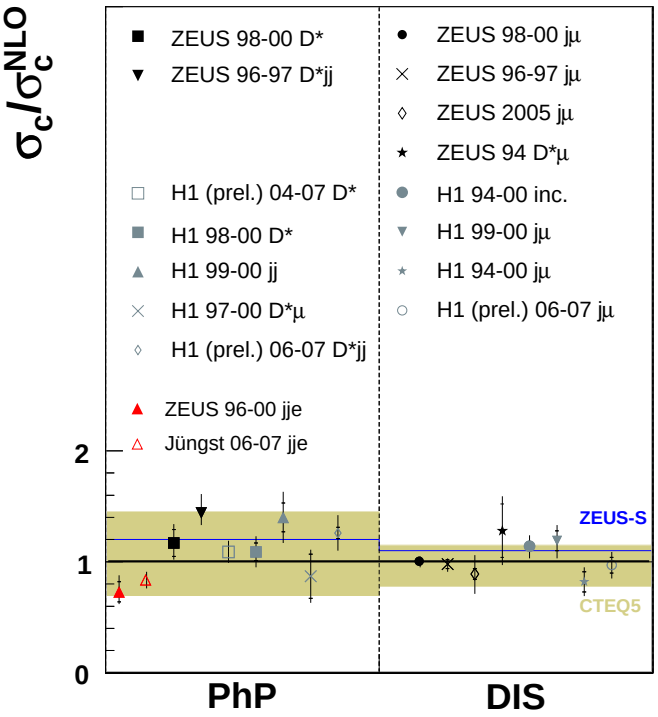


Figure 10.10: *Open charm production in photoproduction and deep inelastic scattering. Plotted is the ratio of the measured total cross section to NLO QCD predictions. Details as in Figure 10.9. The results presented in this thesis are shown by the red triangles, where the full triangle shows the HERA I and the open triangle the HERA II result. They are compared with the measurements from ZEUS (black) and H1 (grey). The ratios fluctuate around values of* ~ 1*. The presented analysis is consistent with the previous charm measurements and at the lower side of the charm measurements.*

Chapter 11

Conclusions

In this thesis measurements of heavy flavour production at HERA were presented. For the identification of the heavy quark events, the semileptonic decays of the beauty or charmed hadrons into electrons or positrons were used, giving a complementary method to previous analyses using decays into muons. For this analysis the photoproduction regime has been investigated, where the dominant process of heavy flavour production is boson gluon fusion. Due to the event signature, this processes can be enriched by the requirement of two jets with high transverse momentum, which provide, in addition to the high masses of the heavy quarks a hard scale for the comparison with NLO QCD calculations at low Q^2.

In a first step, a combined method for the identification of electrons and heavy quark decays has been developed, using the data collected by the ZEUS detector between 1996 and 2000. The variable, dE/dx, which has been corrected for several effects, has been found to be a very powerful tool for the particle identification not only for the electron selection. The most important variable for the beauty identification, p_T^{rel}, was already used in previous analyses, whereas the neutrino signature, expressed by the azimuthal angle to the electrons, $\Delta\phi$, was developed as a new variable sensitive to both heavy flavour contributions. The variable improved the sensitivity for the beauty identification and made the simultaneous extraction of the beauty and charm fractions possible. The signal fractions were determined by a fit of a likelihood test function, in which the sensitive variables have been combined into one single discriminating variable.

After the publication of the HERA I measurement, the analysis was extended to the HERA II data. Using the data collected in 2006 and 2007 the analysis method was further improved, benefitting from the larger statistics and the improved tracking using the MVD, the new detector component. Especially the improved tracking had a deep impact on the analysis. It was possible to improve the conversion

finder (see Section 5.4) for the background rejection and use the new sensitivity of the lifetime information. Several variables were tested and the significance of the track displacement was included as a new variable into the likelihood.

The HERA II visible and differential cross sections were calculated and compared with the HERA I result. The HERA II cross sections were found to be tentatively a bit higher, but comparable except for the lowest η bin, where a disagreement was observed. The statistical uncertainty was significantly reduced, while the systematical uncertainty is still comparable. For the final publication of the HERA II data, it is necessary to understand the worse description of the jet energy and the kinematic variable y and solve the systematic effect, that caused the disagreement in η. The quality of the result can be improved further by adding the data of the remaining years 2003–2005, which would double the statistics, and by a better estimation of the systematic effects due to δ_{IP} and $w_{trigger}$, which have been determined in a rather conservative way. This would especially increase the precision of the beauty cross section. Using the full HERA II data-set it might also be possible to benefit from the higher separation power of the variables δ_{DL} and m_{vtx}.

The differential cross sections were found to be well described in shape by the leading order plus parton showering Monte Carlo PYTHIA and by the NLO QCD calculations made with the FMNR program. The beauty cross sections are at the upper edge of the NLO calculations, while the charm cross sections agree very well with the predictions. By combining all measurements a consistent picture of beauty and charm production at low Q^2 was obtained.

The measurements were compared with previous HERA measurements. They are competitive with the measurements using the muonic channel. The muonic channel, however, implies a minimal momentum of $\sim 2\,\text{GeV}$ compared to the $0.9\,\text{GeV}$ of electrons. In addition to the central part the muon analyses cover also the forward direction of the direction.

The method presented in this analysis was found to be a very flexible tool for particle and decay identification, which has also been adopted for other analyses. At the moment an analysis in the DIS regime is ongoing, where at most the same variables are used. Another idea is use the electron identification, in a double-tag analysis, where both hadrons decay semileptonically. This type of analysis was performed on di-muon events so far, but would benefit from the additional available di-electron and electron-muon channels.

Appendix A

Control Plots

In this appendix additional control plots are given for the background and signal enriched regions as in Section 7.4. Because some variables, like E_T^{jet} or η^e, showed discrepancies between data and Monte Carlo after the preselection, it was checked whether their discrepancies were caused by imperfection in the background description or if they are caused by problems in the detector description, which also influence the signal distribution. In the context of the estimation of the systematic uncertainties it was studied whether the discrepancies had larger impact on the cross section determination.

In the following several control plots are shown for background, signal, beauty and charm enriched samples. The sample selection was done by cutting on the beauty and charm hypothesis test (see figure 6.15). The background enriched region corresponds to the sample after preselection without the cut on the dE/dx likelihood. The signal and beauty enriched region are selected by cutting on the beauty hypothesis at $T_b < 5$ and $T_b < 1.5$, respectively. The charm hypothesis was used for the charm enriched sample using a cut at $T_c < 5$. Figures A.1-A.3 are control plots for the variables used in the likelihood, except for the δ_{IP} which was already shown in Figure 7.7. The description of the signal enriched samples is significantly improved with respect to the background enriched sample, validating the procedure to correct the distributions of $\Delta\phi$ and p_T^{rel} for the background Monte Carlo. The different populated regions in the variables for the selection stages, show the influence of the variable in the likelihood. The kinematic variables p_T^e, η^e and ϕ^e are shown in Figures A.5-A.7. The discrepancy in η is mostly visible in the background region, while the other two variables are described quite well for all selections. Similar conclusions can be drawn from the jet variables for the highest, second highest and electron associated jet (see Figure A.9-A.11). The distribution of the pseudorapidity is under control for the signal region, where a remaining shift is still visible for the jet energy. In the last four figures (A.13-

A.15) the kinematic variables y, x_γ^{obs} and z_{vtx} are shown. The discrepancies in the background enriched region for x_γ^{obs} disappeared after the tight selection, whereas y is improved, but shows remaining problems in the description.

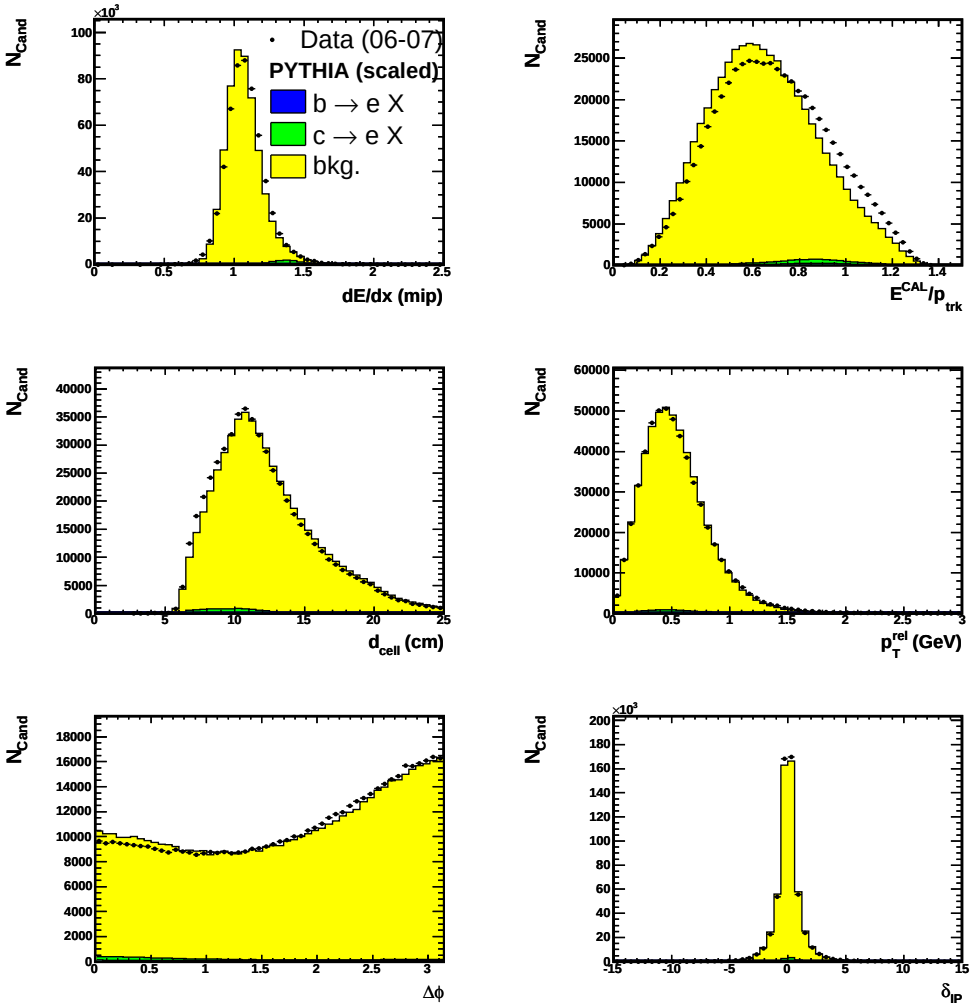


Figure A.1: *Distribution of the likelihood input variables for the background enriched sample (see text). The Monte Carlo contributions have been scaled according to their scaling factors (see Chapter 10). In the background enriched samples remaining deficits in the simulation are still visible. Especially E^{CAL}/p_{trk} and $\Delta\phi$ are not well described.*

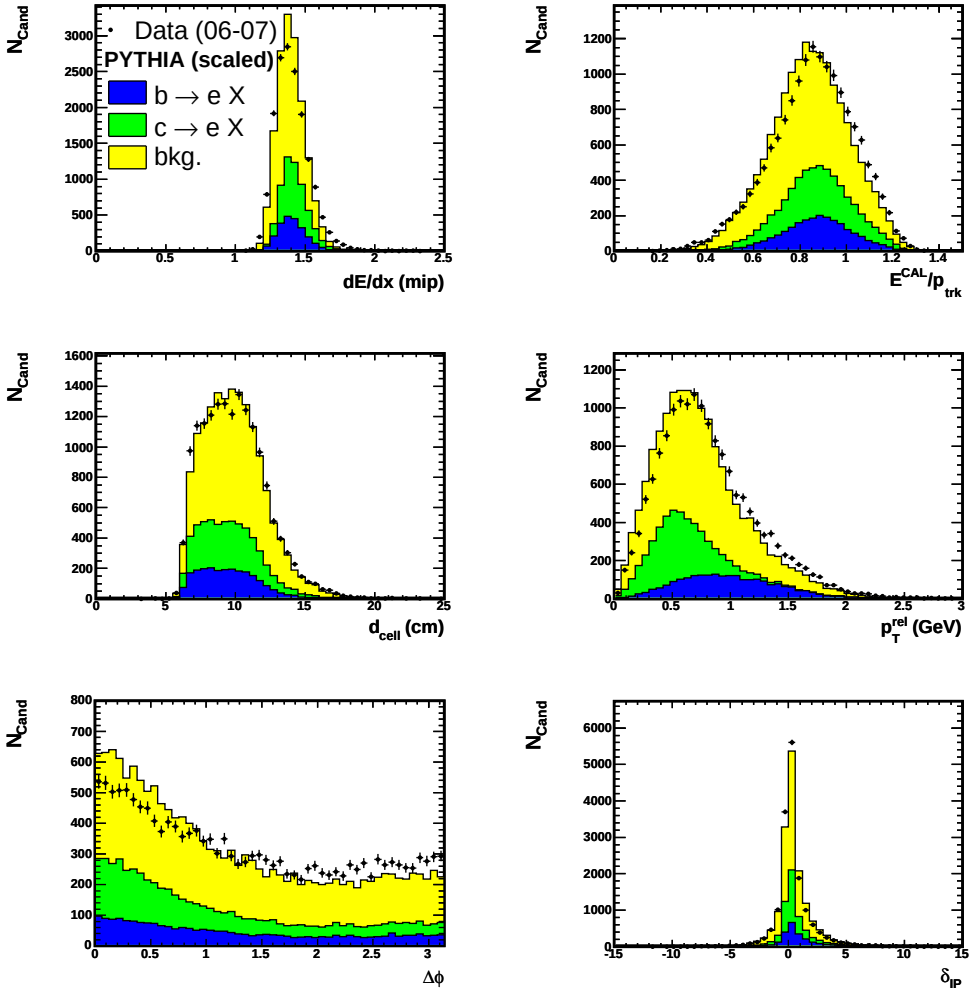


Figure A.2: *Distribution of the likelihood input variables for the medium enriched sample (see text). The Monte Carlo contributions have been scaled according to their scaling factors (see Chapter 10). In the medium enriched samples the description is improved, but disagreements in $\Delta\phi$ made it necessary to correct for.*

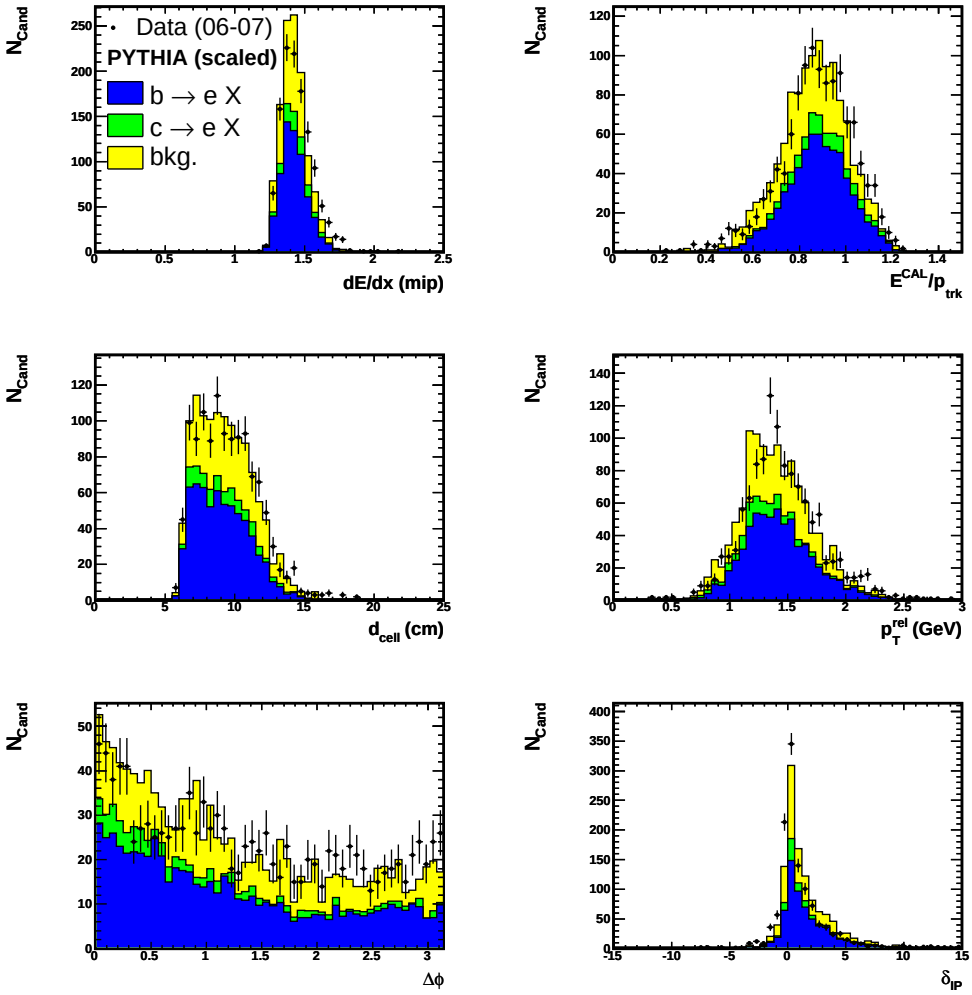


Figure A.3: *Distribution of the likelihood input variables for the beauty enriched sample (see text). The Monte Carlo contributions have been scaled according to their scaling factors (see Chapter 10). In this selection the variables are reasonable well described justifying the correction procedure for the background contribution.*

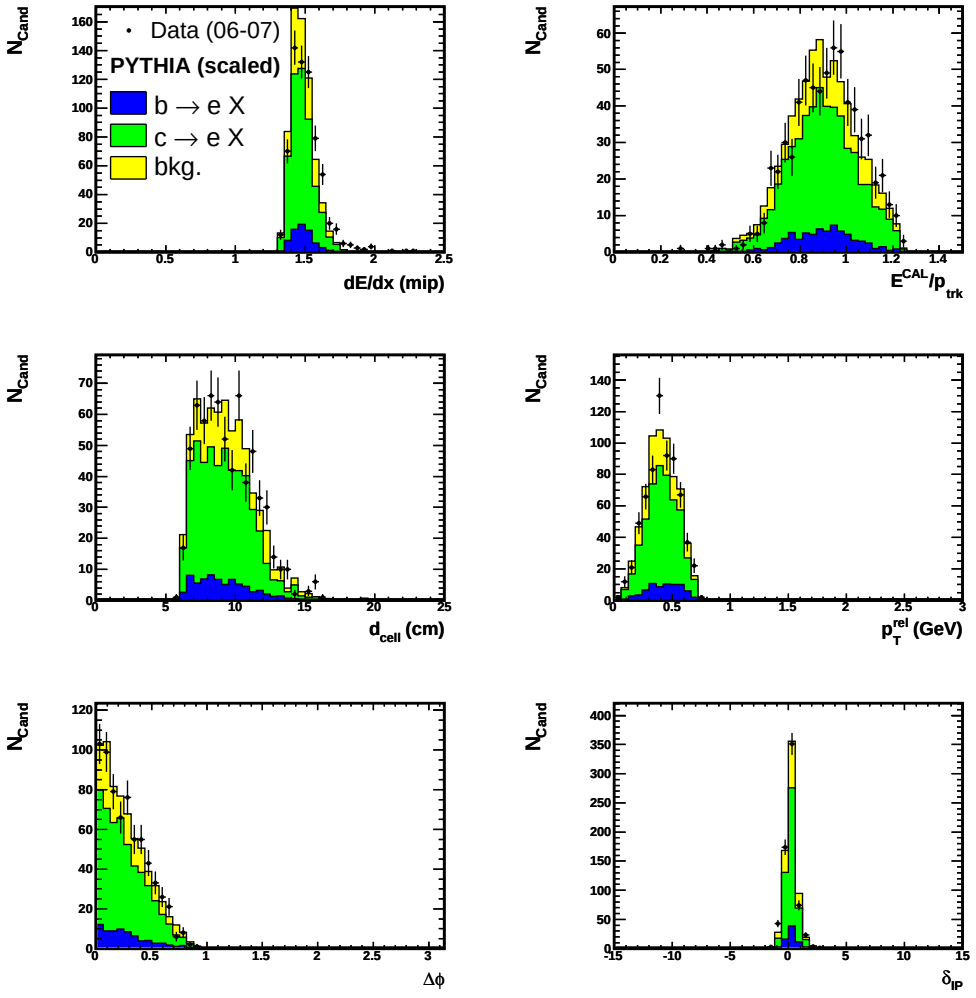


Figure A.4: *Distribution of the likelihood input variables for the charm enriched sample (see text). The Monte Carlo contributions have been scaled according to their scaling factors (see Chapter 10). In this selection the variables are reasonable well described justifying the correction procedure for the background contribution.*

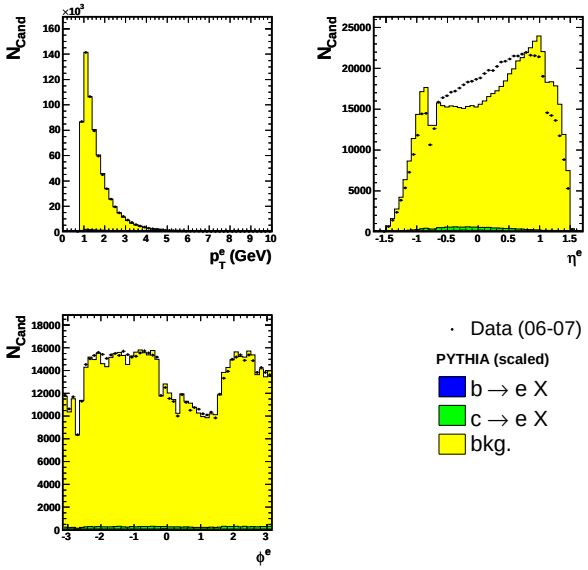


Figure A.5: *Distribution of p_T^e, η^e and ϕ^e for the background enriched sample (see text). The Monte Carlo contributions have been scaled according to their scaling factors (see Chapter 10). In the background enriched samples p_T^e and η^e are described very well, whereas the η distribution is completely off.*

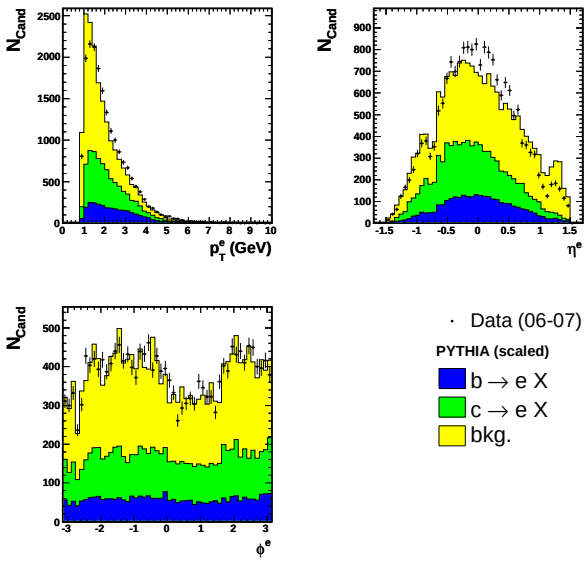


Figure A.6: *Distribution of p_T^e, η^e and ϕ^e for the signal enriched sample (see text). The Monte Carlo contributions have been scaled according to their scaling factors (see Chapter 10). In the signal enriched samples the description of η^e is significantly reduced.*

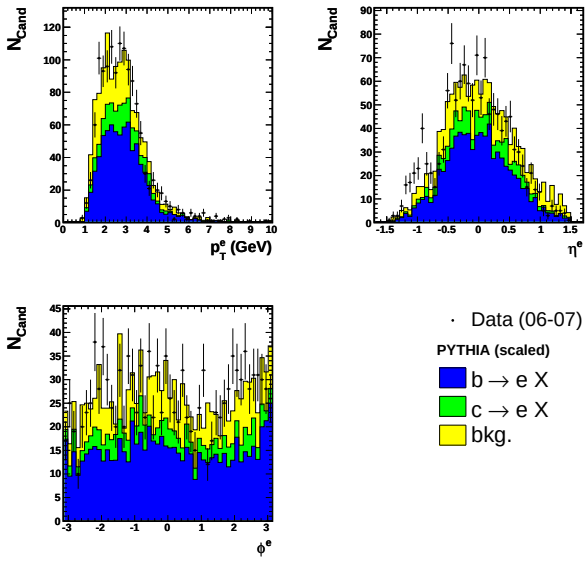


Figure A.7: *Distribution of p_T^e, η^e and ϕ^e for the beauty enriched sample (see text). The Monte Carlo contributions have been scaled according to their scaling factors (see Chapter 10). In the beauty enriched samples all three variables are well described within their statistical uncertainty.*

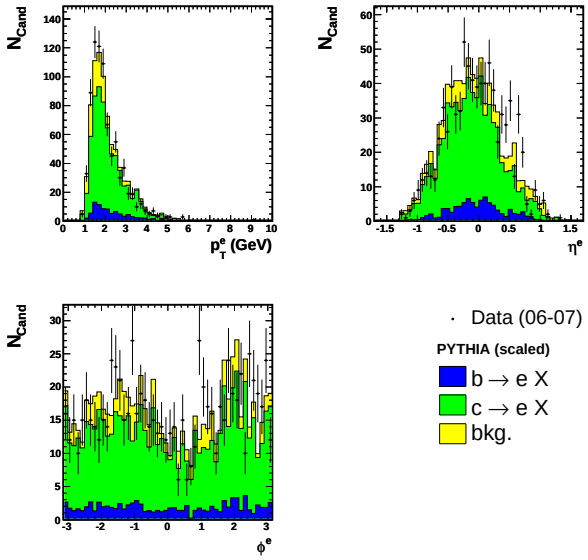


Figure A.8: *Distribution of p_T^e, η^e and ϕ^e for the charm enriched sample (see text). The Monte Carlo contributions have been scaled according to their scaling factors (see Chapter 10). In the charm enriched samples all three variables are well described within their statistical uncertainty.*

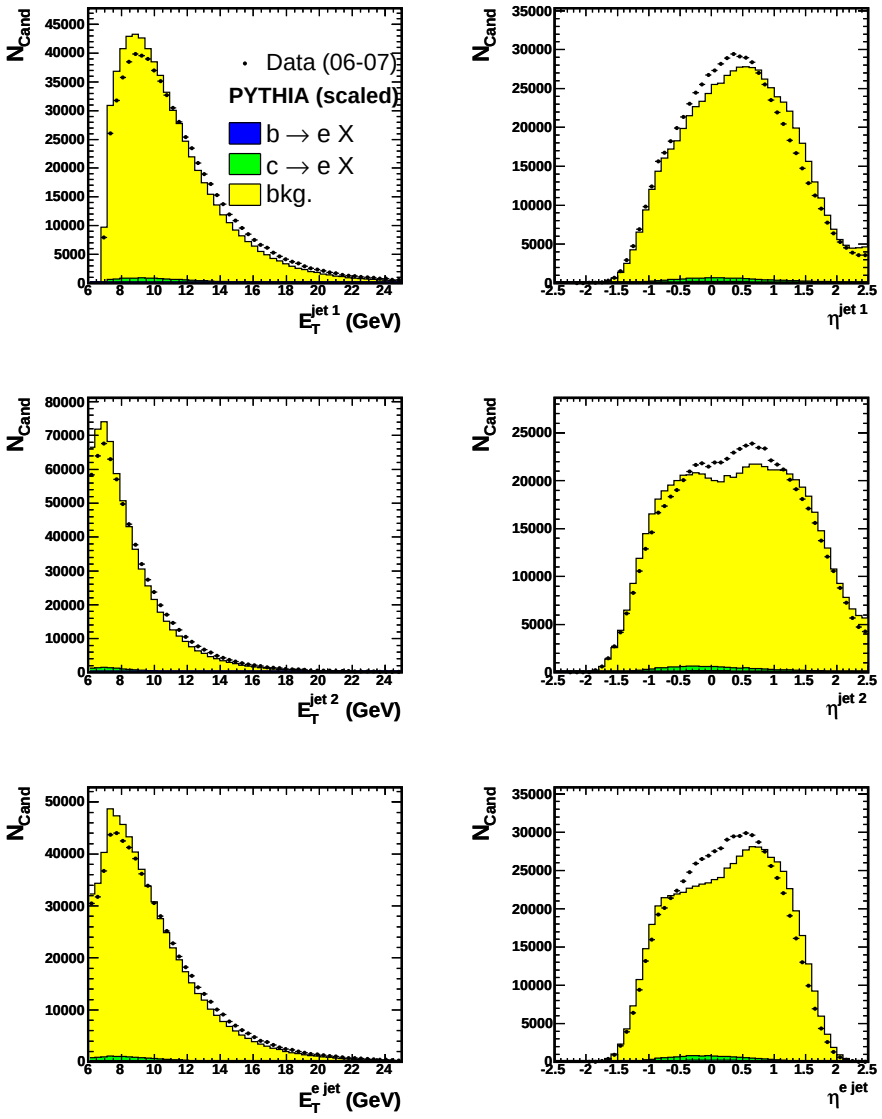


Figure A.9: *Distribution of E_T^{jet} and η^{jet} for the highest, second highest and electron associated jet for the background enriched sample (see text). The Monte Carlo contributions have been scaled according to their scaling factors (see Chapter 10). In the background enriched samples all the variables are reasonable described, where the energy distributions are shifted in the Monte Carlo and the η distributions show some deficits in the central region.*

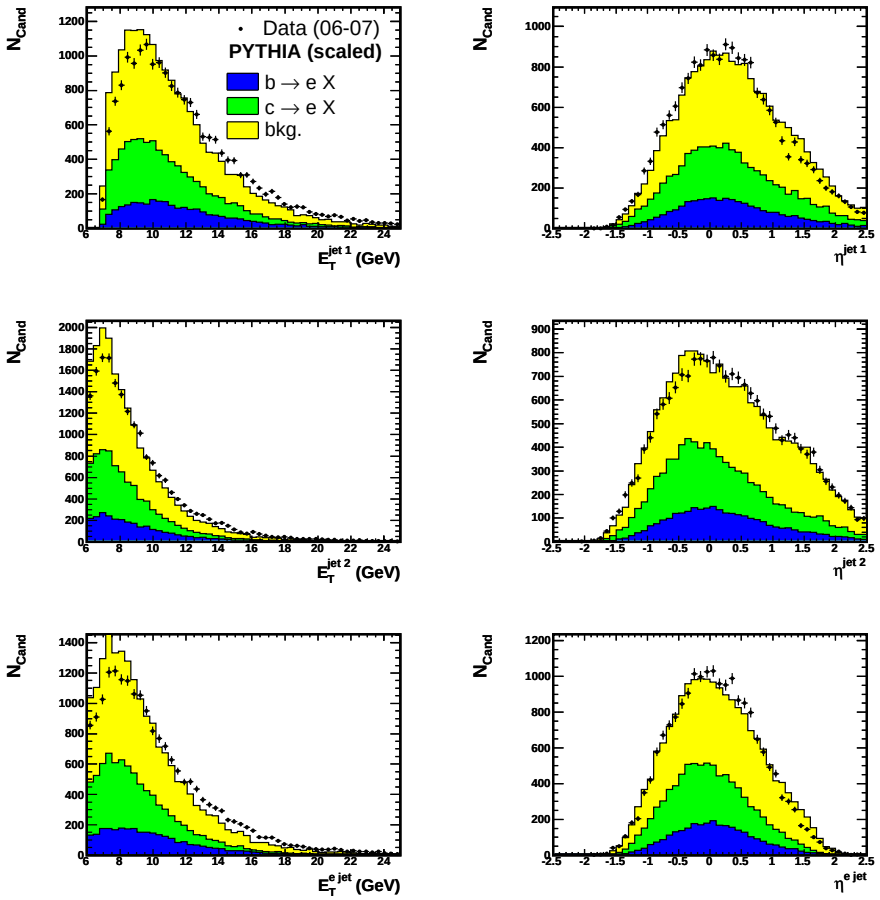


Figure A.10: *Distribution of E_T^{jet} and η^{jet} for the highest, second highest and electron associated jet for the signal enriched sample (see text). The Monte Carlo contributions have been scaled according to their scaling factors (see Chapter 10). In the signal enriched samples the discrepancies in η^{jet} observed for the background enriched region have mostly disappeared, where the shift in the jet energies is still visible.*

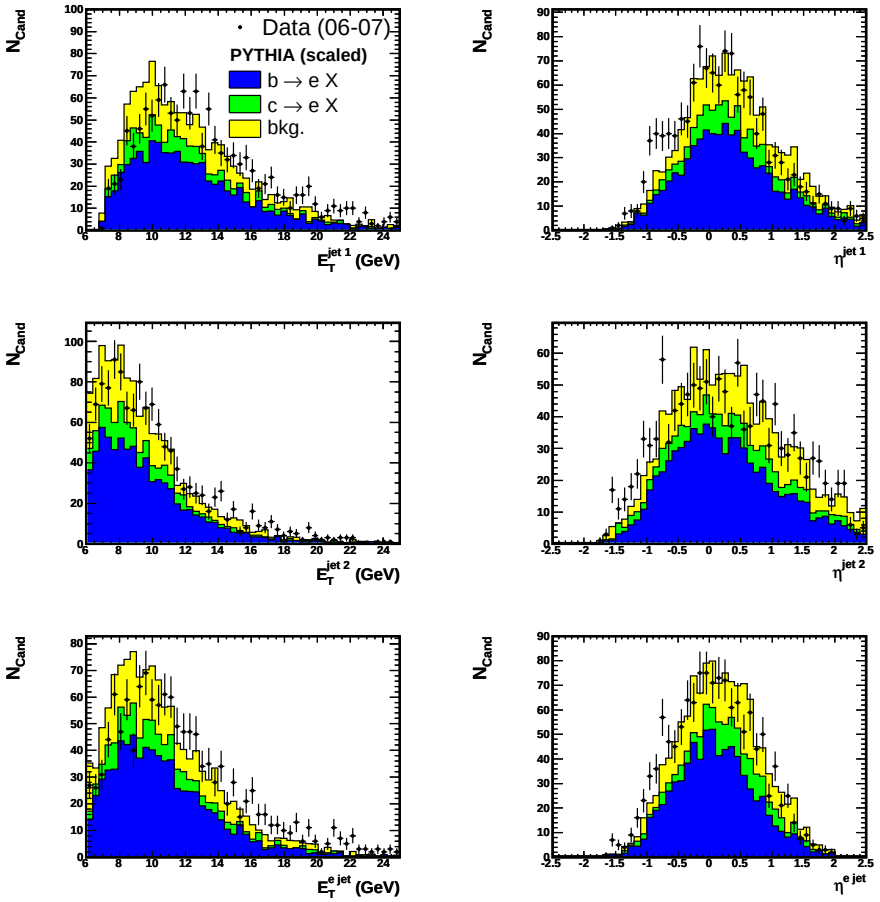


Figure A.11: *Distribution of E_T^{jet} and η^{jet} for the highest, second highest and electron associated jet for the beauty enriched sample (see text). The Monte Carlo contributions have been scaled according to their scaling factors (see Chapter 10). In the beauty enriched samples the discrepancies in η^{jet} observed for the background enriched region have mostly disappeared, where the shift in the jet energies is still visible.*

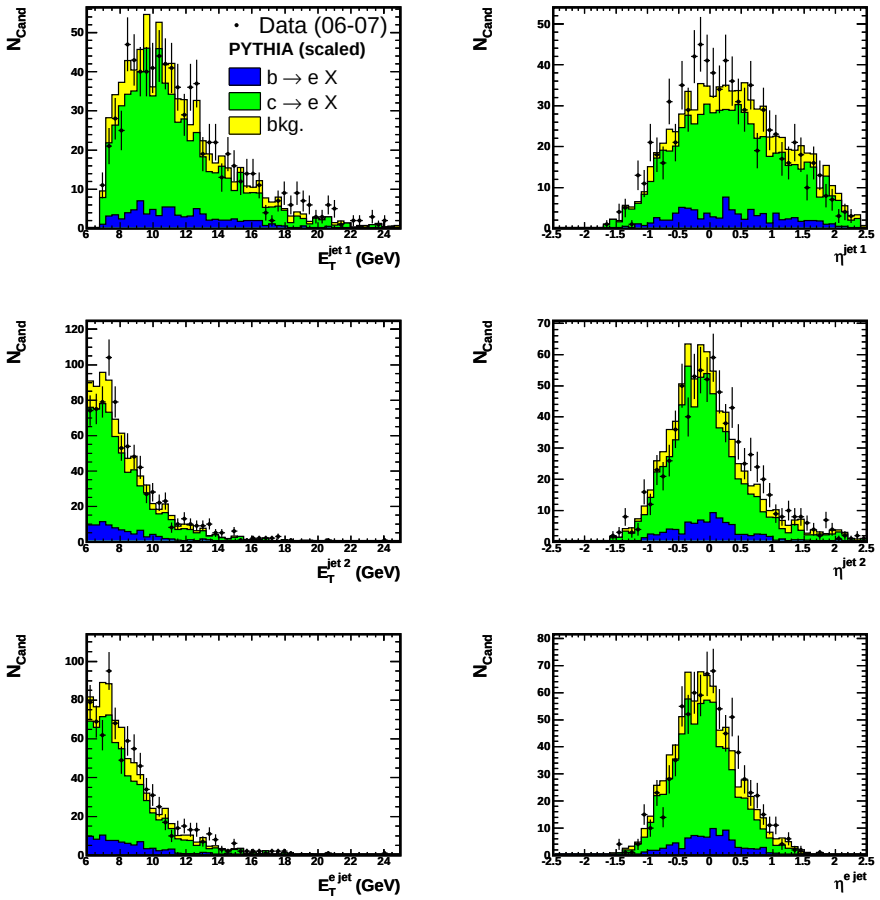


Figure A.12: *Distribution of E_T^{jet} and η^{jet} for the highest, second highest and electron associated jet for the charm enriched sample (see text). The Monte Carlo contributions have been scaled according to their scaling factors (see Chapter 10). In the charm enriched samples the discrepancies in both η^{jet} and E_T^{jet} observed for the background enriched region have disappeared.*

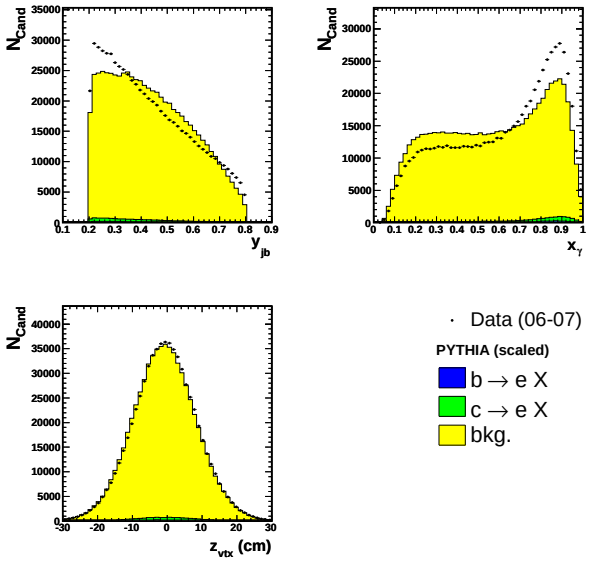


Figure A.13: *Distribution of y_{jb}, x_γ^{obs} and z_{vtx} for the background enriched sample (see text). The Monte Carlo contributions have been scaled according to their scaling factors (see Chapter 10). In the background enriched samples the kinematic variables y_{jb} and x_γ^{obs} in data are not described by the Monte Carlo.*

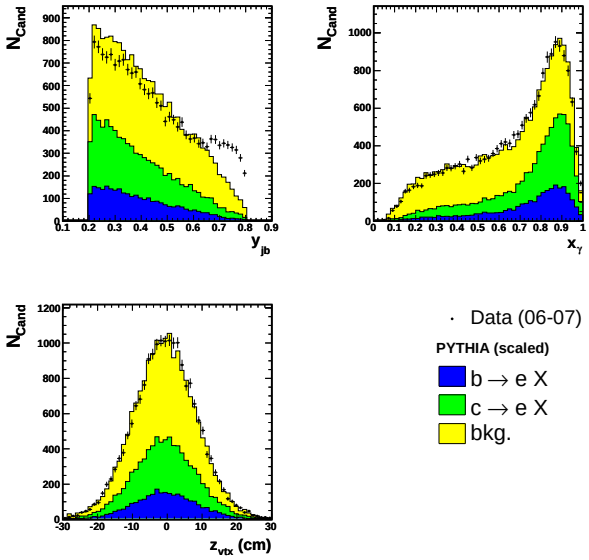


Figure A.14: *Distribution of y_{jb}, x_γ^{obs} and z_{vtx} for the signal enriched sample (see text). The Monte Carlo contributions have been scaled according to their scaling factors (see Chapter 10). In the signal enriched samples the description of y_{jb} and x_γ^{obs} has been substantially improved by rejecting the background candidates.*

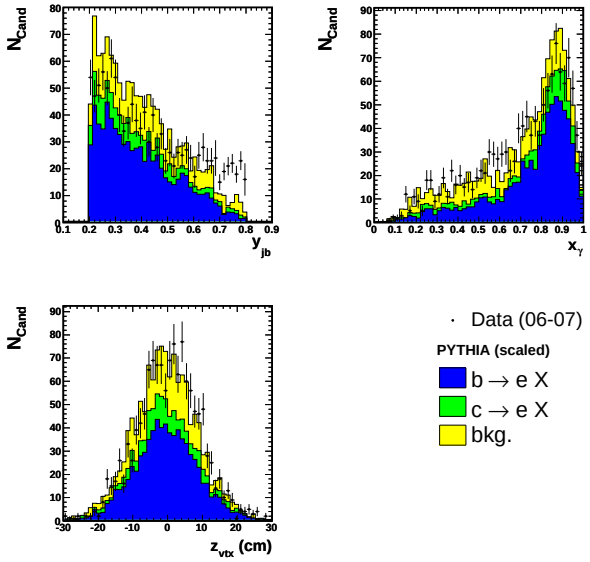


Figure A.15: *Distribution of y_{jb}, x_γ^{obs} and z_{vtx} for the beauty enriched sample (see text). The Monte Carlo contributions have been scaled according to their scaling factors (see Chapter 10). In the beauty enriched samples a similar quality of data description is seen as in the signal enriched region is observed.*

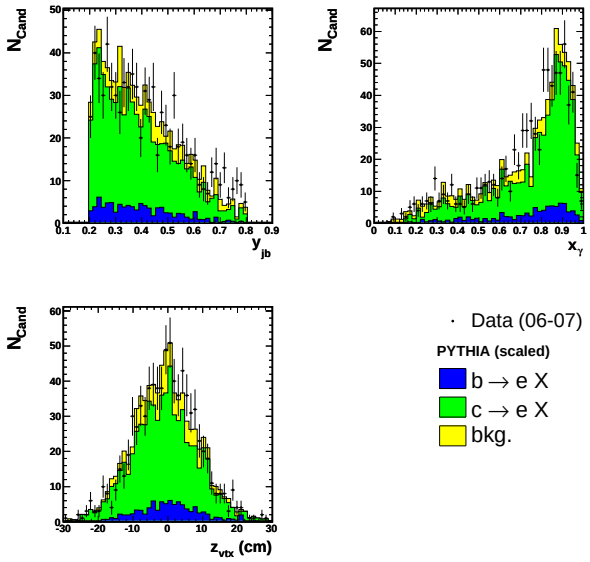


Figure A.16: *Distribution of y_{jb}, x_γ^{obs} and z_{vtx} for the charm enriched sample (see text). The Monte Carlo contributions have been scaled according to their scaling factors (see Chapter 10). In the charm enriched samples all three variables are well described.*

Appendix B

Acceptances

In this appendix the acceptances and purities are listed for all bins of the differential cross sections. The beauty (blue) and the charm (green) signals have similar acceptances and purities, where the purity for e_b tends to be a little bit higher. The binning has been chosen for the HERA I analysis in a way, that the typical purities are of the order of $\sim 70\,\%$ and do not drop much below $\sim 50\,\%$. The acceptances are usually between $10 - 15\,\%$ without having too large steps.

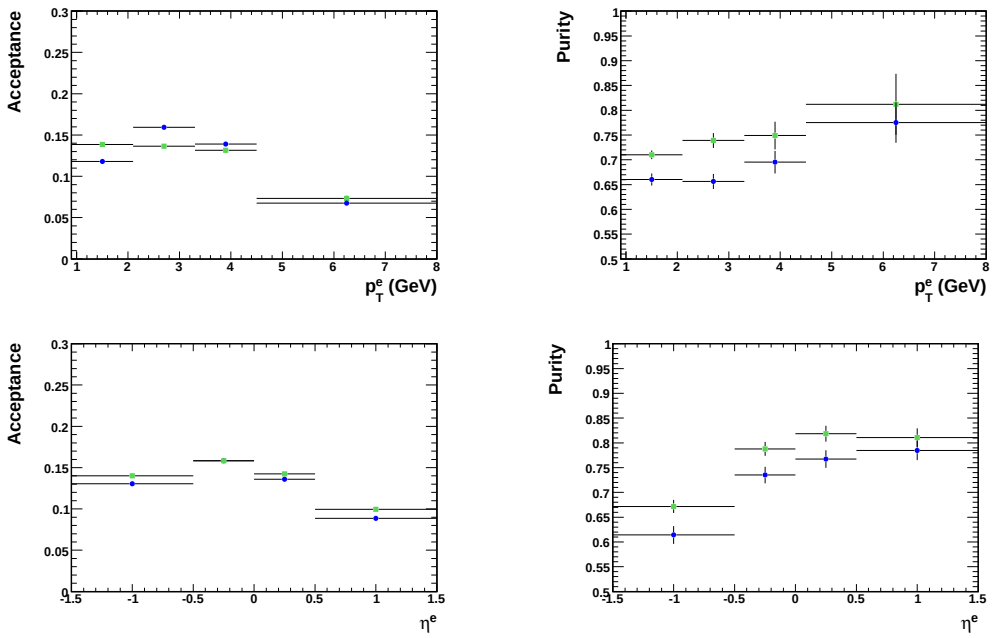


Figure B.1: *Acceptances (left) and purities (right) for beauty (blue) and charm (green) for the differential variables p_T^e and η^e.*

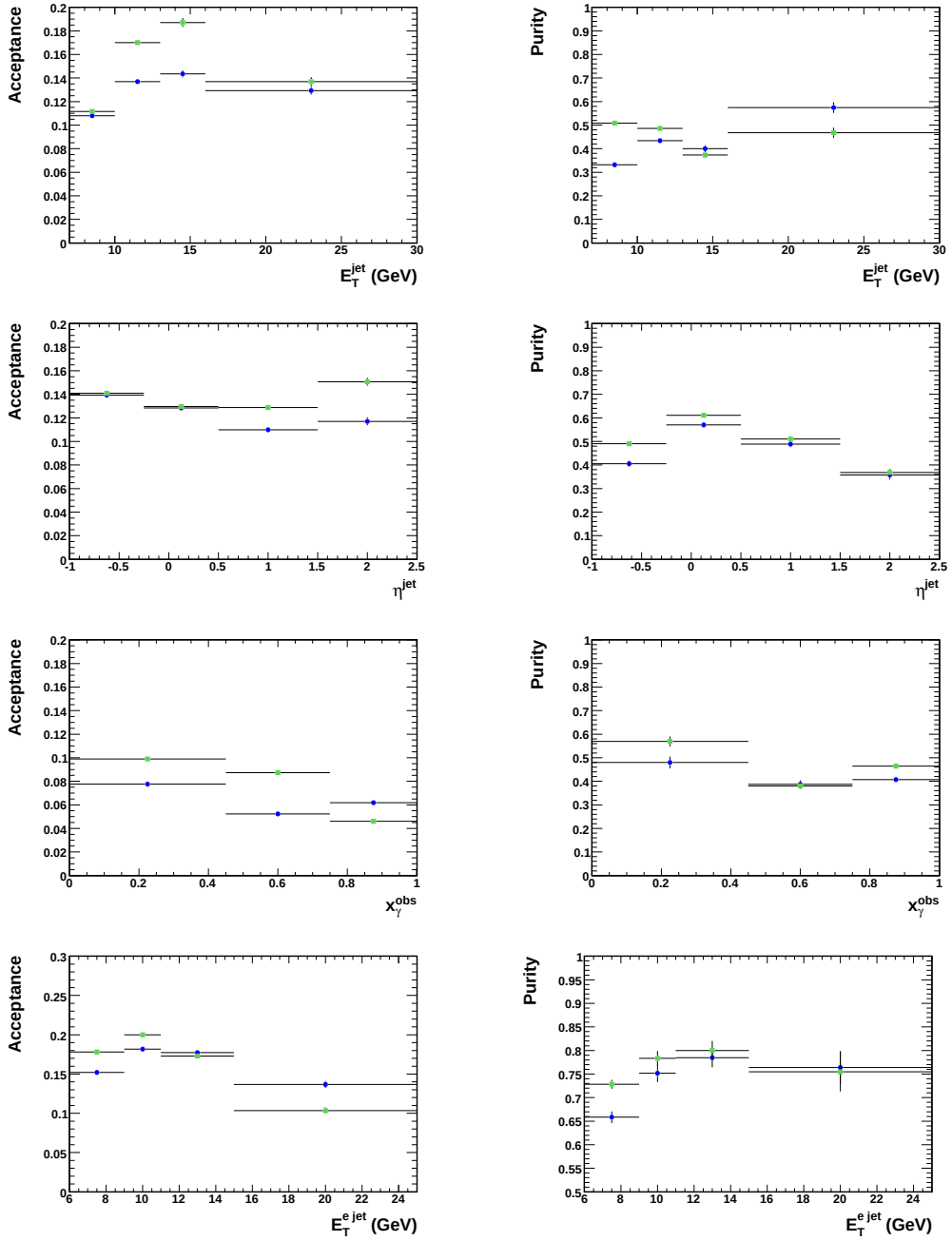


Figure B.2: *Acceptances (left) and purities (right) for beauty (blue) and charm (green) for the differential variables E_T^{jet1}, η^{jet1}, x_γ and $E_T^{e\,jet}$.*

Appendix C

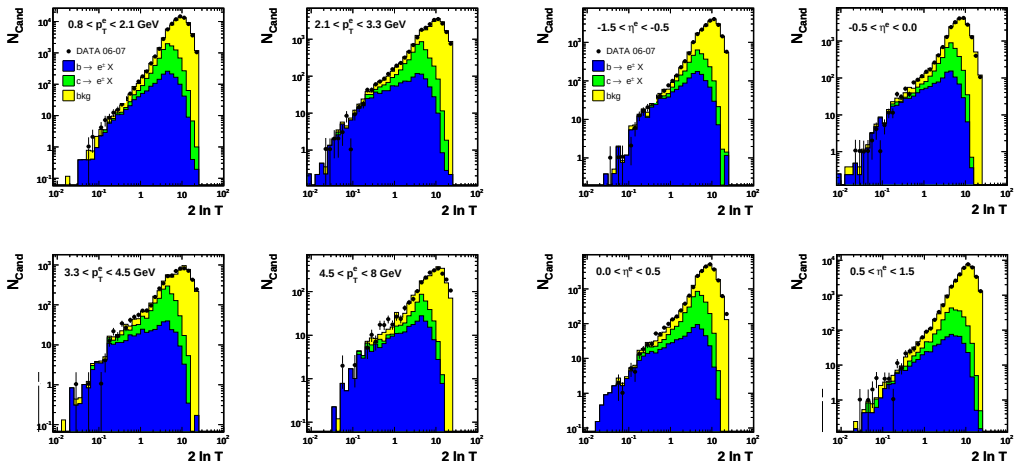
Fit Details

In this appendix the fit details for the (differential) cross sections are listed. The total cross sections were determined from the fraction fit to the likelihood distribution as shown in Figures 8.1 and 8.2. From the fractions the cross sections were determined as described in Chapter 8. The visible beauty and charm cross sections including the statistical and systematic uncertainty are listed in Table C.1 separately for the two data set. They are compared with the NLO QCD calculation which has been corrected for the hadronisation effects by $C_{b/c}^{had}$. The HERA I data set was also split into the two ranges of different centre-of mass energies.

	σ_b^{vis} (pb)	σ_b^{NLO} (pb)	C_b^{had}	σ_c^{vis} (pb)	σ_c^{NLO} (pb)	C_c^{had}
96—97	$101 \pm 18^{+8}_{-9}$	81^{+20}_{-12}	0.81	$253 \pm 58^{+44}_{-22}$	360^{+160}_{-100}	1.00
98—00	$139 \pm 16^{+11}_{-12}$	88^{+22}_{-13}	0.81	$260 \pm 40^{+45}_{-23}$	380^{+170}_{-110}	1.01
96—00	$125 \pm 11^{+10}_{-11}$	88^{+22}_{-13}	0.81	$278 \pm 33^{+48}_{-24}$	380^{+170}_{-110}	1.01
06—07	$147 \pm 5^{+12}_{-14}$	88^{+22}_{-13}	0.81	$315 \pm 9^{+28}_{-25}$	380^{+170}_{-110}	1.01

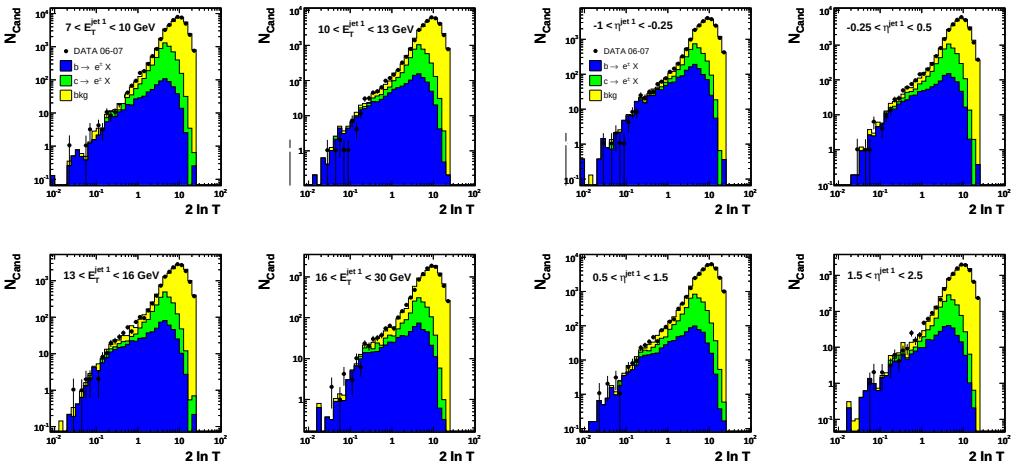
Table C.1: *Total cross sections for electrons from b or c quarks in photoproduction events, $Q^2 < 1 GeV^2$ and $0.2 < y < 0.8$, with at least two jets with $E_T^{jet} > 7(6) GeV$, $|\eta^{jet}| < 2.5$ and the subsequent semileptonic decay to an electron with $p_T^e > 0.9 GeV$ and $|\eta^e| < 1.5$. The values are given for (96-00) and (06-07), where the (96-00) data set was extrapolated to $\sqrt{s} = 318 GeV$ as it has a period of $\sqrt{s} = 300 GeV$ (96—97) and $\sqrt{s} = 318 GeV$. The first error is statistical and the second is systematic. In addition, the NLO QCD prediction and its uncertainty is given, after applying the appropriate hadronisation correction (C_b^{had}, C_c^{had}).*

For the differential cross sections the likelihood distribution was split into the bins of the variable under study. The data distributions are shown for all bins together with the scaled Monte Carlo contributions in Figures C.1-C.3. The numbers of the central values and the range of uncertainty given by the statistical and systematic uncertainties are listed in Tables C.2-C.5, together with the NLO QCD calculations and the hadronisation corrections.



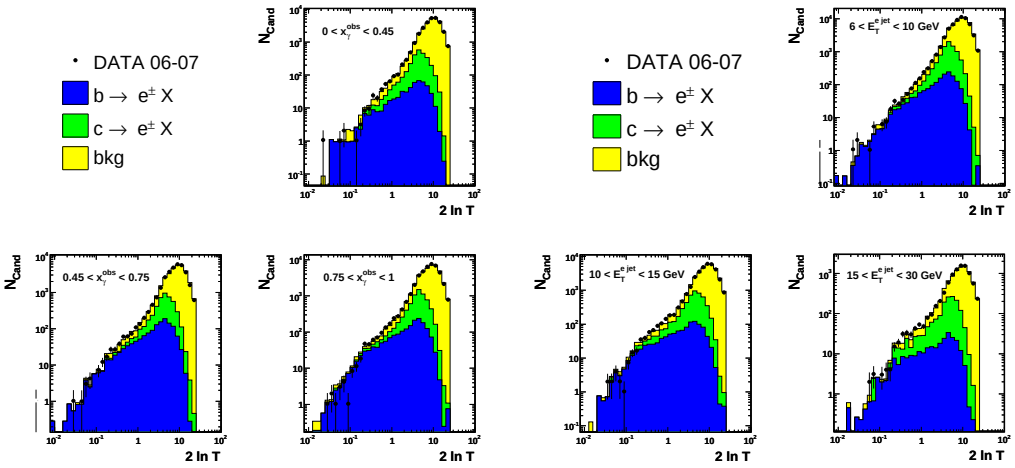
(a) Fit in bins of p_T^e (b) Fit in bins of η^e

Figure C.1: *Distribution of the likelihood test function to extract the beauty and charm fractions. The four plots in (a) show the distribution for the four bins in p_T^e, and the ones in (b) show the distribution for the four bins in η^e. The histograms show the e_b (blue), e_c (green) and background (yellow) contributions, where the Monte Carlo histograms have been scaled according to the fit result in each bin. The sum of the Monte Carlo is overlayed with the data distribution (points)*



(a) Fit in bins of E_T^{jet1} (b) Fit in bins of η_{jet1}

Figure C.2: *Distribution of the likelihood test function to extract the beauty and charm fractions. The four plots in (a) show the distribution for the four bins in E_T^{jet1}, and the ones in (b) show the distribution for the four bins in η^{jet1}. For further details see the caption of Figure C.1.*



(a) Fit in bins of x_γ^{obs} (b) Fit in bins of $E_T^{e\,jet}$

Figure C.3: *Distribution of the likelihood test function to extract the beauty and charm fractions. The three plots in (a) show the distribution for the four bins in x_γ^{obs}, and the ones in (b) show the distribution for the three bins in $E_T^{e\,jet}$. For further details see the caption of Figure C.1.*

p_T^e (GeV)	$d\sigma_b/dp_T^e$ (pb/GeV)	$d\sigma_b^{\mathrm{NLO}}/dp_T^e$ (pb/GeV)	C_b^{had}	$d\sigma_c/dp_T^e$ (pb/GeV)	$d\sigma_c^{\mathrm{NLO}}/dp_T^e$ (pb/GeV)	C_c^{had}
0.9 : 2.1	$56.3\pm9.6^{+4.3}_{-5.0}$	34^{+11}_{-7}	0.78	$117\pm26^{+20}_{-10}$	177^{+71}_{-38}	1.02
	$54.0\pm7.2^{+11.2}_{-11.7}$			$209\pm11^{+20}_{-21}$		
2.1 : 3.3	$24.0\pm3.7^{+1.8}_{-2.1}$	$16.8^{+5.9}_{-3.5}$	0.79	$54.4\pm9.0^{+9.5}_{-4.8}$	80^{+42}_{-23}	0.98
	$27.3\pm2.7^{+2.4}_{-2.0}$			$85.1\pm4.8^{+6.8}_{-7.1}$		
3.3 : 4.5	$11.9\pm2.6^{+0.9}_{-1.1}$	$9.9^{+3.6}_{-2.3}$	0.84	$26.0\pm5.8^{+4.5}_{-2.3}$	36^{+27}_{-14}	0.99
	$11.3\pm2.0^{+0.9}_{-2.0}$			$29.7\pm2.7^{+2.3}_{-2.2}$		
4.5 : 8.0	$4.7\pm1.9^{+0.4}_{-0.4}$	$3.3^{+1.4}_{-0.9}$	0.94	$1.5\pm2.7^{+0.3}_{-0.1}$	$7.5^{+9.5}_{-4.0}$	0.99
	$4.7\pm1.1^{+0.5}_{-0.4}$			$3.4\pm1.0^{+0.4}_{-0.5}$		
η^e	$d\sigma_b/d\eta^e$ (pb)	$d\sigma_b^{\mathrm{NLO}}/d\eta^e$ (pb)	C_b^{had}	$d\sigma_c/d\eta^e$ (pb)	$d\sigma_c^{\mathrm{NLO}}/d\eta^e$ (pb)	C_c^{had}
-1.5 : -0.5	$26.4\pm4.6^{+2.0}_{-2.4}$	$16.7^{+6.6}_{-3.6}$	0.75	$51\pm12^{+9}_{-4}$	111^{+66}_{-33}	0.98
	$47.0\pm4.7^{+4.8}_{-4.2}$			$65.4\pm6.7^{+8.5}_{-8.2}$		
-0.5 : 0.0	$53.4\pm9.1^{+4.1}_{-4.8}$	$39.5^{+13.8}_{-8.3}$	0.81	$152\pm25^{+26}_{-13}$	192^{+100}_{-53}	1.01
	$71.2\pm7.2^{+5.6}_{-5.1}$			$172\pm13^{+18}_{-18}$		
0.0 : 0.5	$57.7\pm11.6^{+4.4}_{-5.1}$	$41.9^{+13.9}_{-9.0}$	0.82	$187\pm36^{+33}_{-16}$	165^{+82}_{-43}	1.02
	$48.0\pm8.2^{+7.3}_{-5.6}$			$202\pm15^{+19}_{-18}$		
0.5 : 1.5	$42.4\pm8.7^{+3.2}_{-3.8}$	$28.1^{+10.1}_{-6.3}$	0.84	$36\pm24^{+6}_{-3}$	90^{+51}_{-26}	1.02
	$31.5\pm7.4^{+5.6}_{-7.2}$			$82\pm12^{+15}_{-11}$		

Table C.2: *Differential electron cross sections as a function of p_T^e and η^e for both data sets. The two rows of cross sections listed for each bin show the results for the HERA I and the HERA II data set. For further details see the caption of Table C.1.*

$E_T^{\text{jet 1}}$ (GeV)	$d\sigma_b/E_T^{\text{jet 1}}$ (pb/GeV)	$d\sigma_b^{\text{NLO}}/E_T^{\text{jet 1}}$ (pb/GeV)	C_b^{had}	$d\sigma_c/E_T^{\text{jet 1}}$ (pb/GeV)	$d\sigma_c^{\text{NLO}}/E_T^{\text{jet 1}}$ (pb/GeV)	C_c^{had}
7 : 10	$16.8\pm2.5^{+1.3}_{-1.5}$	$10.1^{+3.2}_{-1.9}$	0.59	$45.9\pm7.3^{+8.0}_{-4.0}$	72^{+43}_{-19}	0.99
	$11.1\pm1.9^{+3.9}_{-3.5}$			$67.7\pm3.5^{+5.0}_{-5.4}$		
10 : 13	$12.0\pm1.9^{+0.9}_{-1.1}$	$9.4^{+3.7}_{-2.3}$	0.97	$28.0\pm4.7^{+4.9}_{-2.4}$	35^{+14}_{-12}	1.07
	$14.0\pm1.5^{+1.1}_{-1.5}$			$34.9\pm2.2^{+3.8}_{-3.7}$		
13 : 16	$8.3\pm1.6^{+0.6}_{-0.7}$	$5.1^{+2.0}_{-1.1}$	1.18	$5.9\pm3.4^{+1.0}_{-0.5}$	$11.7^{+7.0}_{-2.9}$	1.03
	$7.0\pm1.2^{+0.7}_{-0.8}$			$14.6\pm1.4^{+1.5}_{-1.6}$		
16 : 30	$1.00\pm0.39^{+0.08}_{-0.09}$	$1.00^{+0.39}_{-0.08}$	1.22	$1.5\pm1.1^{+0.3}_{-0.1}$	$1.8^{+1.2}_{-0.5}$	0.89
	$1.59\pm0.26^{+0.22}_{-0.13}$			$2.78\pm0.44^{+0.53}_{-0.59}$		
$\eta^{\text{jet 1}}$	$d\sigma_b/d\eta^{\text{jet 1}}$ (pb)	$d\sigma_b^{\text{NLO}}/d\eta^{\text{jet 1}}$ (pb)	C_b^{had}	$d\sigma_c/d\eta^{\text{jet 1}}$ (pb)	$d\sigma_c^{\text{NLO}}/d\eta^{\text{jet 1}}$ (pb)	C_c^{had}
-1.0 : -0.25	$24.9\pm5.2^{+1.9}_{-2.2}$	$17.5^{+6.1}_{-2.7}$	0.82	$73\pm14^{+13}_{-6}$	99^{+64}_{-26}	0.95
	$62.2\pm5.4^{+4.6}_{-4.8}$			$99.1\pm9.0^{+10.4}_{-10.4}$		
-0.25 : 0.5	$47.6\pm8.2^{+3.7}_{-4.2}$	$42.6^{+12.7}_{-7.7}$	1.01	$177\pm24^{+31}_{-15}$	164^{+75}_{-35}	1.05
	$57.9\pm6.7^{+6.3}_{-3.9}$			$188\pm12^{+16}_{-17}$		
0.6 : 1.5	$49.3\pm7.8^{+3.8}_{-4.4}$	$30.4^{+7.9}_{-6.1}$	0.91	$71\pm17^{+12}_{-6}$	106^{+41}_{-32}	1.04
	$33.1\pm5.3^{+4.2}_{-5.4}$			$119\pm12^{+18}_{-15}$		
1.5 : 2.5	$23.7\pm5.5^{+1.8}_{-2.1}$	$9.2^{+3.6}_{-2.4}$	0.76	$8\pm15^{+1}_{-1}$	35^{+23}_{-12}	1.01
	$12.2\pm2.9^{+2.1}_{-2.5}$			$29.6\pm3.8^{+3.4}_{-3.4}$		

Table C.3: *Differential cross sections for the most energetic jet as a function of E_T^{jet} and η^{jet} for both data set. The two rows of cross sections listed for each bin show the results for the HERA I and the HERA II data set. For further details see the caption of Table C.1.*

x_γ^{obs}	$d\sigma_b/dx_\gamma^{\text{obs}}$ (pb)	$d\sigma_b^{\text{NLO}}/dx_\gamma^{\text{obs}}$ (pb)	C_b^{had}	$d\sigma_c/dx_\gamma^{\text{obs}}$ (pb)	$d\sigma_c^{\text{NLO}}/dx_\gamma^{\text{obs}}$ (pb)	C_c^{had}
0.00 : 0.45	$51\pm17^{+4}_{-5}$	28^{+18}_{-10}	1.07	$70\pm35^{+12}_{-6}$	122^{+108}_{-56}	1.16
	$32.5\pm8.6^{+4.9}_{-10.2}$			$120\pm12^{+18}_{-15}$		
0.45 : 0.75	$166\pm25^{+13}_{-15}$	81^{+50}_{-28}	2.27	$227\pm49^{+40}_{-20}$	216^{+178}_{-85}	1.32
	$142\pm15^{+12}_{-12}$			$299\pm22^{+34}_{-33}$		
0.75 : 1.00	$216\pm31^{+17}_{-19}$	166^{+47}_{-30}	0.55	$715\pm79^{+124}_{-63}$	920^{+370}_{-190}	0.90
	$199\pm17^{+16}_{-20}$			$528\pm28^{+44}_{-44}$		

Table C.4: *Differential cross sections as a function of x_γ^{obs} for both data set.The two rows of cross sections listed for each bin show the results for the HERA I and the HERA II data set. For further details see the caption of Table C.1.*

$E_T^{\text{e jet}}$ (GeV)	$d\sigma_b/E_T^{\text{e jet}}$ (pb/GeV)	$d\sigma_b^{\text{NLO}}/E_T^{\text{e jet}}$ (pb/GeV)	C_b^{had}	$d\sigma_c/E_T^{\text{e jet}}$ (pb/GeV)	$d\sigma_c^{\text{NLO}}/E_T^{\text{e jet}}$ (pb/GeV)	C_c^{had}
6 : 10	$16.1\pm1.8^{+1.2}_{-1.4}$	$12.3^{+5.1}_{-3.0}$	0.67	$42.2\pm5.2^{+7.3}_{-3.7}$	64^{+38}_{-18}	1.00
	$12.6\pm1.3^{+1.5}_{-1.5}$			$45.9\pm2.0^{+4.5}_{-4.6}$		
10 : 15	$6.6\pm1.3^{+0.5}_{-0.6}$	$5.4^{+1.8}_{-1.1}$	1.00	$22.3\pm4.2^{+3.9}_{-2.0}$	$19.6^{+7.5}_{-5.5}$	1.06
	$5.58\pm0.69^{+0.48}_{-0.52}$			$20.8\pm1.3^{+1.4}_{-1.4}$		
15 : 30	$2.1\pm0.6^{+0.2}_{-0.2}$	$1.08^{+0.40}_{-0.26}$	1.21	$0.3\pm1.9^{+0.1}_{-0.1}$	$1.7^{+1.2}_{-0.5}$	0.87
	$0.64\pm0.23^{+0.18}_{-0.03}$			$3.95\pm0.57^{+0.24}_{-0.52}$		

Table C.5: *Differential cross sections of $E_T^{e\ jet}$ for the jet associated to the electron from beauty or charm decays for both data set. The two rows of cross sections listed for each bin show the results for the HERA I and the HERA II data set. For further details see the caption of Table C.1.*

List of Figures

List of Tables

Bibliography

[1] Thomson, J. J., *Cathode Rays*. Phil. Mag. **44**, 293 (1897).

[2] R. Rhodes, *The Making of the Atomic Bomb*. 1995.

[3] W. Pauli to M. Delbrück. 6. Oct. 1958 [3075], 'Scientific Correspondence', 1.c., Volume II:1930-1939, p38.

[4] Rochester, G. D. and Butler, C. C., *Evidence for the Existence of New Unstable Elementary Particles*. Nature **160**, 855 (1947).

[5] Fermi, E. and Yang, C. N., *Are Mesons Elementary Particles?* Phys. Rev. **76**, 1739 (1949).

[6] Gell-Mann, Murray, *A Schematic Model of Baryons and Mesons*. Phys. Lett. **8**, 214 (1964).

[7] Zweig, G, *An SU_3 Model for Strong Interaction Symmetry and its Breaking; Part II*. (1964).

[8] Weinberg, Steven, *Conceptual Foundations of the Unified Theory of Weak and Electromagnetic Interactions*. Rev. Mod. Phys. **52**, 515 (1980).

[9] Salam, Abdus, *Gauge Unification of Fundamental Forces*. Rev. Mod. Phys. **52**, 525 (1980).

[10] Glashow, Sheldon Lee, *Towards a Unified Theory: Threads in a Tapestry*. Rev. Mod. Phys. **52**, 539 (1980).

[11] Halzen, Francis and Martin, Alan Douglas, *Quarks and Leptons: An Introductory Course in Modern Particle Physics*. New York, NY, 1984.

[12] J.C. Collins and D.E. Soper and G. Sterman, *Perturbative Quantum Chromodynamics*, A.H. Mueller (ed.). World Scientific, Singapore (1989).

[13] Particle Data Group, C. Amsler et al., *The Review of Particle Physics*. J. Phys **B 667**, 1 (2008).

[14] M. Jüngst, *Elektronidentifikation mit dem ZEUS-Detektor und Bestimmung des Beauty-Produktionsquerschnitts*. Diploma Thesis, Universität Bonn, Bonn, Germany, Report BONN-IB-05-15, 2005, available on `http://brock.physik.uni-bonn.de/zeus_pub.phd`.

[15] O.M. Kind, *Production of Heavy Flavours with Associated Jets at HERA*. Ph.D. Thesis, Universität Bonn, Bonn, Germany, Report BONN-IR-2007-04, 2007, available on `http://brock.physik.uni-bonn.de/zeus_pub.php`.

[16] S. Chekanov and others, *Beauty photoproduction using decays into electrons at HERA*. Phys. Rev. **D 78**, 072001 (2008), 0805.4390.

[17] Monica Turcato, *Measurement of Beauty Photoproduction at HERA*. Ph.D. Thesis, Università degli studi di Padova, December 2002. `http://www-library.desy.de/preparch/desy/thesis/desy-thesis-03-039.ps.gz`.

[18] ZEUS Collaboration, *The ZEUS webpage*, 2009. `http://www-zeus.desy.de/`.

[19] R.P. Feynman, *Very High-Energy Collisions of Hadrons*. Phys. Rev. Lett. **23**, 1415 (1969).

[20] C.G. Callan and D.J. Gross, *High-energy electroproduction and the constitution of the electric current*. Phys. Rev. Lett. **22**, 156 (1969).

[21] R. Brandelik et al., *Evidence for planar events in e^+e^- annihilation at high-energies*. Phys. Lett. **B 86**, 243 (1979).

[22] R.K. Ellis and H. Georgi and M. Machacek and H.D. Politzer and G.G. Ross. Nucl. Phys. **B**, 285 (1979).

[23] Elliot Leader and Enrico Predazzi, *An Introduction to Gauge Theories and the New Physics*. Cambridge University Press, 1982.

[24] C. Glasmann, *Jet production in deep inelastic ep scattering at HERA*. Hep-ex/0312011.

[25] Alan D. Martin, *Proton structure, Partons, QCD, DGLAP and beyond*. ACTA PHYS.POLON.B **39**, 2025 (2008).

[26] Li, Gang, *Extraction of the proton parton density functions using a NLO-QCD fit of the combined H1 and ZEUS inclusive DIS cross sections*. In *Hamburg 2008, Multiparticle dynamics (ISMD08)* 31-35.

[27] V.N. Gribov and L.N. Lipatov, *Deep inelastic ep scattering in perturbation theory*. Sov. J. Nucl. Phys. **15**, 438 (1972).

[28] L.N. Lipatov, *The parton model and perturbation theory*. Sov. J. Nucl. Phys. **20**, 94 (1975).

[29] Yu.L. Dokshitzer, *Calculation of the structure functions for deep inelastic scattering and e^+e^- annihilation by perturbation theory in Quantum Chromodynamics [in Russian]*. Sov. Phys. JETP **46**, 641 (1977).

[30] G. Altarelli and G. Parisi, *Asymptotic freedom in parton language*. Nucl. Phys. **B 126**, 298 (1977).

[31] Ya.Ya. Balitskiĭ and L.N. Lipatov, *The Pomeranchuk singularity in Quantum Chromodynamics*. Sov. J. Nucl. Phys. **28**, 822 (1978).

[32] E.A. Kuraev, L.N. Lipatov and V.S. Fadin, *Multi-reggeon processes in the Yang–Mills theory*. Sov. Phys. JETP **44**, 443 (1976).

[33] E.A. Kuraev, L.N. Lipatov and V.S. Fadin, *The Pomeranchuk singularity in nonabelian gauge theories*. Sov. Phys. JETP **45**, 199 (1977).

[34] A. Gabareen Mokhtar, *Study of Neutral and Charged Current Cross Sections at High Q^2 at HERA*. Ph.D. Thesis, Tel Aviv University, Report DESY-THESIS-2006-005, 2006.

[35] M. Glück, E. Reya and A. Vogt, *Parton structure of the photon beyond the leading order*. Phys. Rev. **D 45**, 3986 (1992).

[36] S. Frixione, P. Nason and G. Ridolfi, *Differential distributions for heavy flavor production at HERA*. Nucl. Phys. **B 454** (1995), hep-ph/9506226;

S. Frixione et al., *Total cross sections for heavy flavour production at HERA*. Phys. Lett. **B 348**, 633 (1995).

[37] B. Kahle, *Measurement of beauty-production in deep inelastic scattering at HERA II*. Ph.D. Thesis, Universität Hamburg, 2005.

[38] A.C. Bawa and W.J. Stirling. J. Phys. **G**, 1339 (1989).

[39] T. Sjöstrand and M. Bengston. Comput. Phys. Comm. **43**, 367 (1987).

[40] T. Sjöstrand et al., *High-energy-physics event generation with* PYTHIA *6.1*. Comp. Phys. Comm. **135**, 238 (2001), hep-ph/0010017;
E. Norrbin and T. Sjöstrand, *Production and hadronization of heavy quarks*. Eur. Phys. J. **C 17**, 137 (2000), hep-ph/0005110;

T. Sjöstrand, L. Lönnblad, and S. Mrenna, *Pythia 6.2: Physics and manual*. Preprint hep-ph/0108264, 2001.

[41] H.L. Lai et al., *Improved parton distributions from global analysis of recent deep inelastic scattering and inclusive jet data*. Phys. Rev. **D 55**, 1280 (1997).

[42] M. Glück, E. Reya and A. Vogt, *Photonic parton distributions*. Phys. Rev. **D 46**, 1973 (1992).

[43] R. Brun et al., GEANT3, Technical Report CERN-DD/EE/84-1, CERN, 1987.

[44] Kniehl, Bernd A. and Kramer, M. and Kramer, G. and Spira, M., *Cross-sections for charm production in e p collisions: Massive versus massless scheme*. Phys. Lett. **B 356**, 539 (1995), hep-ph/9505410.

[45] Kobayashi, Mokoto and Maskawa, Toshihide, *CP Violation in the Renormalizable Theory of Weak Interaction*. Prog. Theor. Phys. **49**, 35 (1960).

[46] ZEUS Coll., S. Chekanov et al., *Measurement of open beauty production at HERA in the $D^*\mu$ final state*. Eur. Phys. J. **C 50**, 1434 (2007), hep-ex/0609050v2.

[47] ZEUS Coll., S. Chekanov et al., *Measurement of beauty production in deep inelastic scattering at HERA*. Phys. Lett. **B 599**, 173 (2004).

[48] ZEUS Coll., S. Chekanov et al., *Bottom photoproduction measured using decays into muons in dijet events in ep collisions at $\sqrt{s} = 318GeV$*. Phys. Rev. **D 70**, 12008 (2004), hep-ex/0312057.

[49] ZEUS Coll., J. Breitweg et al., *Measurement of open beauty production in photoproduction at HERA*. Eur. Phys. J. **C 18**, 625 (2001), hep-ex/0011081.

[50] ZEUS Coll., S. Chekanov et al., *Measurement of $D^{*\pm}$ meson production in $e^{\pm}p$ scattering at low Q^2*. Phys. Lett. **B 649**, 111 (2007), hep-ex/0501069.

[51] ZEUS Coll., S. Chekanov et al., *Measurement of D mesons production in deep inelastic scattering at HERA*. Preprint DESY-07-52 (arXiv:0704.3562v1 [hep-ex]), 2007. Accepted by JHEP.

[52] ZEUS Coll., S. Chekanov et al., *Inclusive jet cross sections and dijet correlations in $D^{*\pm}$ photoproduction at HERA*. Nucl. Phys. **B 729**, 492 (2005), hep-ex/0507089.

[53] ZEUS Coll., S. Chekanov et al., *Measurement of charm fragmentation ratios and fractions in photoproduction at HERA*. Eur. Phys. J. **C 44**, 351 (2005), hep-ex/0509019v2.

[54] ZEUS Coll., S. Chekanov et al., *Measurement of $D^{*\pm}$ production in deep inelastic $e^{\pm}p$ scattering at DESY HERA*. Phys. Rev. **D 69**, 012004 (2004), hep-ex/0308068.

[55] ZEUS Coll., S. Chekanov et al., *Dijet angular distributions in photoproduction of charm at HERA*. Phys. Lett. **B 565**, 87 (2003), hep-ex/0302025.

[56] ZEUS Coll., J. Breitweg et al., *Measurement of $D^{*\pm}$ production and the charm contribution to F_2 in deep inelastic scattering at HERA*. Eur. Phys. J. **C 12**, 35 (2000), hep-ex/9908012.

[57] ZEUS Coll., J. Breitweg et al., *Measurement of inclusive $D_s^{\pm}$ photoproduction at HERA*. Phys. Lett. **B 481**, 213 (2000), hep-ex/0003018.

[58] ZEUS Coll., J. Breitweg et al., *Measurement of inclusive $D^{*\pm}$ and associated dijet cross-sections in photoproduction at HERA*. Eur. Phys. J. **C 6**, 67 (1999), hep-ex/9807008.

[59] ZEUS Coll., J. Breitweg et al., *Differential cross sections of $D^{*\pm}$ photoproduction in ep collisions at HERA*. Phys. Lett. **B 401**, 192 (1997), hep-ex/9704011.

[60] ZEUS Coll., J. Breitweg et al., *D^* production in deep inelastic scattering at HERA*. Phys. Lett. **B 407**, 402 (1997), hep-ex/9706009.

[61] ZEUS Coll., M. Derrick et al., *Study of $D^*(2010)^{\pm}$ production in ep collisions at HERA*. Phys. Lett. **B 349**, 225 (1995), hep-ex/9502002.

[62] H1 Coll., A. Aktas et al., *Measurement of charm and beauty dijet cross sections in photoproduction at HERA using the H1 vertex detector*. Eur. Phys. J. **C 47**, 597 (2006).

[63] H1 Coll., A. Aktas et al., *Measurement of $F_2^{c\bar{c}}$ and $F_2^{b\bar{b}}$ at low Q^2 and x using the H1 vertex detector at HERA*. Eur. Phys. J. **C 45**, 23 (2006), hep-ex/0507081.

[64] H1 Coll., A. Aktas et al., *Measurement of charm and beauty photoproduction at HERA using $D^*\mu$ correlations*. Phys. Lett. **B 621**, 56 (2005), hep-ex/0503038.

[65] H1 Coll., A. Aktas et al., *Measurement of $F_2^{c\bar{c}}$ and $F_2^{b\bar{b}}$ at high Q^2 using the H1 vertex detector at HERA*. Eur. Phys. J. **C 40**, 349 (2005), hep-ex/0411046.

[66] H1 Coll., A. Aktas et al., *Measurement of beauty production at HERA using events with muons and jets*. Eur. Phys. J. **C 41**, 453 (2005), hep-ex/0502010.

[67] H1 Coll., C. Adloff et al., *Measurement of open beauty production at HERA*. Phys. Lett. **B 467**, 156 (1999), hep-ex/9909029.

[68] H1 Coll., A. Aktas et al., *Production of D^* mesons with dijets in deep-inelastic scattering at HERA*. Eur. Phys. J. **C 51**, 271 (2007), hep-ex/0701023.

[69] H1 Coll., A. Aktas et al., *Inclusive $D^{*\pm}$ meson cross sections and $D^{*\pm}$-jet correlations in photoproduction at HERA*. Eur. Phys. J. **C 50**, 251 (2006), hep-ex/0608042.

[70] H1 Coll., A. Aktas et al., *Inclusive production of D^+, D^0, D_s^+ and D^{*+} mesons in deep inelastic scattering at HERA*. Eur. Phys. J. **C 38**, 447 (2005), hep-ex/0408149.

[71] H1 Coll., C. Adloff et al., *Measurement of $D^{*\pm}$ meson production and F_2^c in deep inelastic scattering at HERA*. Phys. Lett. **B 528**, 199 (2002), hep-ex/0108039.

[72] H1 Coll., C. Adloff et al., *Measurement of D^* meson cross-sections at HERA and determination of the gluon density in the proton using NLO QCD*. Nucl. Phys. **B 545**, 21 (1999), hep-ex/9812023.

[73] H1 Coll., S. Aid et al., *Photoproduction of $D^{*\pm}$ mesons in electron–proton collisions at HERA*. Nucl. Phys. **B 472**, 32 (1996), hep-ex/9604005.

[74] Deutsches Elektronensynchrotron Hamburg, *HERA — A proposal for a large electron proton colliding beam facility at DESY*, Technical Report.

[75] ZEUS Coll., U. Holm (ed.), *The ZEUS Detector*. Status Report (unpublished), DESY (1993), available on `http://www-zeus.desy.de/bluebook/bluebook.html`.

[76] B. Wiik, *HERA status*, *Physics at HERA*, W. Buchmüller and G. Ingelman (eds.), Vol. 1, pp. 1–16. Hamburg, DESY (1991).

[77] P. Waloschek. HERA bulletin Nº4, 1995.

[78] U. Schneekloth (ed.), *The hera luminosity upgrade*. Preprint HERA-98-05, DESY, 1998, available on `http://www.desy.de/~ahluwali/HERA-98-05`.

[79] ZEUS Collaboration, *The ZEUS evtake and lumi page*, 2006. `http://www-zeus.desy.de/physics/lumi/`.

[80] E. Hilger, *ZEUS coordinate system*, Internal note 86-017, ZEUS, Hamburg, 1986.

[81] K. Gather and D.H. Saxon, *Axes, units, conventions*, Internal note 88-012, ZEUS, Hamburg, 1988.

[82] ZEUS Collaboration, *The ZEUS detector status report*. DESY, Hamburg, February 1993.

[83] J.A. Crittenden et al., *The C5 Upgrade* (unpublished). ZEUS-01-002, internal ZEUS note, 2001.

[84] A. Garfagnini, *The ZEUS microvertex detector*. Nucl. Instr. Methods **A 435**, 34 (1999).

[85] C.B. Brooks and others, *Development of the ZEUS central tracking detector*. Nucl. Instr. Methods **A283**, 477 (1998).

[86] B. Foster and J. Malos and D.H. Saxon and others, *The design and construction of the ZEUS central tracking detector*. Nucl. Instrum. Methods **A338**, 254 (1994).

[87] G.F. Hartner and Y. Iga and J.B. Lane and N.A. McCubbin and M. Wing, *VCTRAK (3.07/04): Offline Output Information*, Internal note 97-064, ZEUS, Hamburg, September 1997. http://www-zeus.desy.de/ZEUS_ONLY/zeus_notes/ZEUS_NOTES/ZEUS-97-064.ps.

[88] G.F. Hartner, *VCTRAK Briefing: Program and Math*, Internal note 98-058, ZEUS, Hamburg, August 1998. http://www-zeus.desy.de/ZEUS_ONLY/zeus_notes/ZEUS_NOTES/ZEUS-98-058.ps.

[89] R. Hall–Wilton and N. McCubbin and P. Nylander and M. Sutton and M. Wing, *The CTD Tracking Resolution*, Internal note 99-024, ZEUS, Hamburg, May 1999. http://www-zeus.desy.de/ZEUS_ONLY/zeus_notes/ZEUS_NOTES/ZEUS-99-024.ps.

[90] A. Andresen and others, *Construction and beam test of the ZEUS forward and rear calorimeter*. Nucl. Instr. Methods **A309**, 101 (1991).

[91] B. Bernstein. Nucl. Instr. Methods **A336**, 23 (1993).

[92] A. Caldwell and I. Gialas and S. Mishra and others, *Design and implementation of a high precision readout system for the ZEUS calorimeter*. Nucl. Instr. Methods **A321**, 356 (1992).

[93] M. Derrick and D. Gacek and N. Hill and others, *Design and construction of the ZEUS barrel calorimeter*. Nucl. Instr. Methods **A309**, 77 (1991).

[94] J. Krüger. Habilitationschrift, DESY, 1992. F35-92-02.

[95] H. Bethe and W. Heitler, *On stopping of fast particles and the creation of positive electrons*. Proc. Roy. Soc. **A146**, 83 (1934).

[96] K. Piotrzkowski and M. Zachara, *Determination of the ZEUS luminosity in 1993*, Internal note 94-167, ZEUS, Hamburg, 1994.

[97] J. Andruszków et al., *Luminosity measurement in the ZEUS experiment*. Acta Phys. Pol. **B 32**, 2025 (2001).

[98] M. Helbich et al., *The spectrometer system for measuring ZEUS luminosity at HERA*. Preprint physics/0512153, 2005. Submitted to Nucl. Inst. Meth.

[99] L.W. Wiggers, *Monte Carlo study of dead–time in the ZEUS DAQ system*, Internal note 87-056, ZEUS, Hamburg, 1987.

[100] S.M. Fisher and others, *ADAMO entity–relationship programming system: Reference Manual*. CERN Programming Techniques Group, ECP Division, 1993. http://adamo.web.cern.ch/Adamo/ADAMO_ENTRY.html.

[101] René Brun and others, *GEANT 3.13*. CERN, 1987. CERN DD/EE/84-1.

[102] G.F. Hartner and Y. Iga and J.B. Lane and N.A. McCubbin and M. Wing, *VCTRAK (3.07/04): Offline Output Information*, Internal note 97-064, ZEUS, Hamburg, September 1997.
http://www-zeus.desy.de/ZEUS_ONLY/zeus_notes/ZEUS_NOTES/ZEUS-97-064.ps;

G.F. Hartner, *VCTRAK Briefing: Program and Math*, Internal note 98-058, ZEUS, Hamburg, August 1998.
http://www-zeus.desy.de/ZEUS_ONLY/zeus_notes/ZEUS_NOTES/ZEUS-98-058.ps.

[103] E. Maddox, *A Kalman filter track fit for the ZEUS microvertex detector*, Internal note 03-008, ZEUS, Hamburg, July 2003.
http://www-zeus.desy.de/ZEUS_ONLY/zeus_notes/ZEUS_NOTES/ZEUS-03-008.ps;
R. Frühwirth, *Application of Kalman filtering to track and vertex fitting*. Nucl. Instr. Methods **A 262**, 444 (1987).

[104] Bethe and Bloch, *Energieverlust durch Ionisation*. Zeitschrift für Physik (1933).

[105] L.D. Landau, *On the energy loss of fast particles by ionization*. J. Phys. USSR, p. 201 (1944).

[106] W.W.M. Allison and J.H. Cobb, *Relativistic charged particle identification by energy loss*. Annual Review in Nuclear & Particle Physics, p. 253 (1980).

[107] D. Bartsch, *Energy-loss measurement with the ZEUS Central Tracking Detector*. Ph.D. Thesis, Universität Bonn, Bonn, Germany, Report BONN-IR-2007-05, 2007, available on http://brock.physik.uni-bonn.de/zeus_pub.php.

[108] Walter Blum and Luigi Rolandi, *Particle Detection with Drift Chambers*. Springer, 1993.

[109] Oliver Maria Kind, *Open Beauty Production at HERA*. Ph.D. Thesis, Physikalisches Institut der Universität Bonn, 2006.

[110] R. Zimmermann, *Kalibrierung und Charakterisierung der dE/dx-Information der Zentralen Driftkammer bei ZEUS*. Diploma Thesis, Universität Bonn, Bonn, Germany, Report BONN-IB-2007-10, 2007, available on http://brock.physik.uni-bonn.de/zeus_pub.php.

[111] P. Billoir, R. Frühwirth and M. Regler, *Track element merging strategy and vertex fitting in complex modular detectors*. Nucl. Instr. Methods **A 241**, 115 (1985).

[112] R. Frühwirth and A. Strandlie. Comp. Phys. Communication **120**, 197 (1999).

[113] S. Miglioranci, *Beauty photoproduction at HERA II with the ZEUS detector*. Ph.D. Thesis, University College London, 2006.

[114] V. Schönberg, private communication, 2009.

[115] Verena Schönberg for the ZEUS Collaboration, *Measurement of beauty photoproduction from inclusive secondary vertexing at HERA II*. (2009). Submitted to the DIS09 conference `http://www-zeus.desy.de/~vschoenb/beauty/IncAnalysis_writeupDIS09.pdf`.

[116] N. Tuning, *ZUFOs: hadronic final state reconstruction with calorimeter, tracking and backsplash correction*, Internal note 01-021, ZEUS, 2001. `http://www-zeus.desy.de/ZEUS_ONLY/zeus_notes/ZEUS_NOTES/ZEUS-01-021.ps`.

[117] S. Catani and Y.L. Dokshitzer and B.R. Webber. Phys. Lett. **B285**, 291 (1992);

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

[118] W. Buchmueller and G. Ingelmann, *Physics at HERA*. DESY, p. 35 (1991).

[119] F. Jaquet and A. Blondel, *Proceedings of the study of an ep facility for Europe*, p. 391. (1979). DESY–79/48.

[120] Adam, N. E. and others, *Absolute Branching Fraction Measurements for D^+ and D^0 Inclusive Semileptonic Decays*. Phys. Rev. Lett. **97**, 251801 (2006), hep-ex/0604044.

[121] A. López–Duran viani and S. Schlenstedt, *Electron finder efficiencies and impurities. A comparison between SINISTRA, EM and EMNET*, Internal note 99-077, ZEUS, Hamburg, 2000. `http://www-zeus.desy.de/ZEUS_ONLY/zeus_notes/ZEUS_NOTES/ZEUS-00-077.ps`.

[122] R. Sinkus and T. Voss, *Particle identification with neural networks using a rotational invariant momentum representation*. Nucl. Instr. Methods **A391**, 360 (1997).

[123] P. Günter, *Auf der Suche nach semileptonischen Charm-Zerfällen im Vorwärtsbereich von ZEUS*. Diploma Thesis, Universität Bonn, Bonn, Germany, Report BONN-IB-99-07, 1999, available on `http://brock.physik.uni-bonn.de/zeus_pub.php`.

[124] R. Shehzadi, private communication, 2009.

[125] Graziano Bruni, *P/Pion separation using the ZEUS calorimeter*, Internal note 04-018, ZEUS, Hamburg, September 2004. http://www-zeus.desy.de/ZEUS_ONLY/zeus_notes/ZEUS_NOTES/ZEUS-04-018.pdf.

[126] Olaf Deppe, *Measurement of $D^{*\pm}$ Electroproduction at HERA*. Ph.D. Thesis, Universität Hamburg, 1999. http://www-library.desy.de/preparch/desy/thesis/desy-thesis-00-006.ps.gz.

[127] M. Bell, *Measurement of Heavy Quark Production in Deep Inelastic Scattering at HERA-II*. Ph.D. Thesis, University of Oxford, 2007.

[128] M. Turcato, *Measurement of Beauty Photoproduction at HERA*. Ph.D. Thesis, Università degli Studi di Padova, Report DESY-THESIS-03-039, 2002, available on http://www-library.desy.de/preparch/desy/thesis/desy-thesis-03-039.ps.gz.

[129] R. Brun and others, *The ROOT System Homepage*. (1995-2009).

[130] K. Böhmer, *Spline-Funktionen*. B. G. Teubner Stuttgart, 1974.

[131] C. de Boor, *A Practical Guide to Splines*. Springer-Verlag, München, 1978.

[132] A. Hocker and others, *TMVA: Toolkit for multivariate data analysis*. PoS **ACAT**, 040 (2007), physics/0703039.

[133] Chekanov, S. and others, *Measurement of charm and beauty production in deep inelastic ep scattering from decays into muons at HERA*. (2009), 0904.3487.

[134] R. Barlow and C. Beeston, *Fitting using finite Monte Carlo samples*. Comput. Phys. Commun. **77**, 219 (1993).

[135] René Brun and others, *The ROOT System Homepage*. CERN, 1995—2006. http://root.cern.ch.

[136] René Brun and others, *The ROOT User's Guide*. CERN, 2005. http://root.cern.ch/root/doc/RootDoc.html.

[137] M. Corradi, private communication, 2009.

[138] F. James and M. Roos., *Minuit: A System for Function Minimization and Analysis of the Parameter Errors and Correlations*. Comput. Phys. Commun. **10**, 343 (1975).

[139] A. Caldwell, D. Kollar and K. Kroninger, *BAT - The Bayesian Analysis Toolkit*. Comput. Phys. Commun. **180**, 2197 (2009), 0808.2552.

[140] R. Barlow, *Evaluating systematic errors*. (1993). MAN-HEP-93-9.

[141] R. Barlow, *Systematic errors: Facts and fictions*. (2002), hep-ex/0207026.

[142] Matthew Wing, *Setting the jet energy scale for the ZEUS calorimeter*, *10th International Conference on Calorimetry in High Energy Physics*. Pasadena, CALTECH (2002).
http://de.arxiv.org/ps/hep-ex/0206036.

[143] A. Yagues, private communication, 2009.

[144] Stefano Frixione and Michelangelo L. Mangano and Paolo Nason and Giovanni Ridolfi, *Total cross sections for heavy flavour productions at HERA*, Technical Report CERN-TH.7527/94, CERN Theory Division, Geneva, December 1994.
http://de.arxiv.org/pdf/hep-ph/9412348.

[145] B. Mele and P. Nason and G. Ridolfi. Nucl. Phys. **B357**, 409 (1991).

[146] I. C. Brock, private communication, 2009.

[147] CTEQ Coll., H.L. Lai et al., *Global QCD analysis of parton structure of the nucleon: CTEQ5 parton distributions*. Eur. Phys. J. **C 12**, 375 (2000).

[148] M. Glück, E. Reya and A. Vogt, *Parton distributions for high energy collisions*. Z. Phys. **C 53**, 127 (1992).

[149] C. Peterson and others. Phys. Rev. **D27**, 105 (1983).

[150] ZEUS Coll., S. Chekanov et al., *Zeus next-to-leading-order qcd analysis of data on deep inelastic scattering*. Phys. Rev. **D 67**, 012007 (2003), hep-ex/0208023.

Acknowledgement

The results of this thesis would have never been achieved without the collaboration and the support of a large group of people who contributed in various ways. I am definitely not able to list all aspects and if something, or even somebody is missing, they are not forgotten beyond these few in the following unordered selection.

First I would like to thank Professor Ian Brock for all aspects of his supervision. Both the direct advice and inspirations but also the freedom in my studies gave me the possibility to gain the necessary knowledge without losing the motivation at any time during the analysis. Especially it was great to have the freedom to discuss, present and defend my results at various occasions. Secondly I would like to thank Dr. Jürgen Kroseberg, Professor Herbert Dreiner and Dr. Gregor Kirfel for the time and interest they spent on the thesis and the disputation. From these three I would especially like to thank Herbi again for the common non scientific activities. I had a lot of fun in the preparation and of course also in the trips with the community of the physics show.

Following the direct creation of the thesis I would like to thank Professor Ewald Paul, Michał Własenko, Jan Stillings and Sebastian Mergelmeyer for reading parts of the manuscript; especially Ewald for the help structuring the theory chapter. Their effort in the formulation of this thesis hopefully increased the reading-flow of the text. All remaining mistakes were probably inserted after their reading. A large contribution in my understanding of physics and the simulation is originated in endless discussions and hundreds of mails with my former supervisor Oliver Kind and Massimo Corradi who was already open for my ideas and questions at the beginning of my diploma thesis and never stopped giving advise. Mentioning with Massimo one of the collaboration members I met in Hamburg, I would like to thank the whole ZEUS collaboration and especially the members of the heavy flavour group. Beyond the heavy flavour group I have to thank to Krzysztof Wrona and Janusz Szuba for their strong support on technical details that occur during such an analysis and for their help when entering the offline group for the event display project. The ZeVis common task was very interesting and it was a pleasure benefitting from the knowledge of Oliver and Igor Rubinsky. From all colleagues in Bonn I thank especially the ZEUS Bonn group members, Verena Schönberg, Ramoona Shehzadi and the former ones Thomas Loddenkötter and Robert Zimmermann. Not to forget also the calorimeter group where especially the very profitable discussions with Detlef Bartsch and Professor Erwin Hilger helped me a lot. From the other members of the Bonn group I would like to emphasis Michał and Jan for the long physics but also private discussions. It was very important for me to have colleagues that also enjoyed some discussion right after work or sometimes even longer at some other places.

Last but by all means not least, I would like to thank my parents and my brother who have the largest also if not directly visible contribution to this thesis. They always supported and motivated me in any possible way and gave me the opportunity to focus on the work when necessary as well as also relax when possible.