ETHZ-IPP RP-2007-05
August 2007

Electron Identification in Heavy Quark Decays at the H1 Experiment

Roger A. Hälg

Diploma Thesis

supervised by
Prof. Dr. R. Eichler

Tutors: Dr. A. Schöning
M. Sauter

Swiss Federal Institute of Technology
Zurich

August 2007

Abstract

In this diploma thesis the development of a method to identify low energy electrons is presented. The implemented electron finder uses information from the tracking and the calorimetry system of the H1 detector. The estimators are combined in a multivariate analysis in order to train a single discriminating variable.
The potential of the chosen estimating variables in separating signal from background is studied using data samples selected from J/ψ (signal) and ρ (background) decays respectively. The description of the estimators by Monte Carlo simulation is verified.
The good separation power of the estimators in conjunction with the usage of sophisticated multivariate classifiers leads to an improved identification method compared to existing electron finders. The presented electron finder allows to opperate in the energy regime of 1-3 GeV with an efficiency and background rejection both above 90 %.

Kurzfassung

In dieser Diplomarbeit wird die Entwicklung einer Methode zur Identifikation von niederenergetischen Elektronen vorgestellt. Der implementierte Elektronenfinder benutzt Informationen des Spurkammer- und des Kalorimetriesystems des H1 Detektors. Die Estimatoren werden in einer multivariaten Analyse kombiniert um eine einzelne Diskriminierungsgrösse zu trainieren.
Das Potential der gewählten Estimatoren zur Trennung von Signal und Hintergrund wird untersucht unter der Verwendung von Datenproben selektiert in J/ψ (Signal) respektive ρ (Hintergrund) Zerfällen. Die Beschreibung der Estimatoren durch Monte Carlo Simulationen wird überprüft.
Die gute Trennwirkung der Estimatoren in Verbindung mit der Verwendung von hochentwickelten multivariaten Sortieralgorithmen führt zu einer verbesserten Identifikationsmethode verglichen mit bestehenden Elektronenfindern. Der dargelegte Elektronenfinder erlaubt es im Energiebereich von 1-3 GeV bei einer Effizienz und einer Hintergrundunterdrückung von jeweils über 90 % eingesetzt zu werden.

Die Physik ist für die Physiker eigentlich viel zu schwer.

David Hilbert

Contents

Chapter 1

Introduction

I do not know what I may appear to the world, but to myself I seem to have been only like a boy playing on the sea-shore, and diverting myself in now and then finding a smoother pebble or a prettier shell than ordinary, whilst the great ocean of truth lay all undiscovered before me.

Isaac Newton

The main purpose of this diploma thesis is to implement a method to identify low energy electrons at the H1 experiment. The presented electron finder uses information from the tracking and the calorimetry system of the H1 detector.
Low energy electrons (1-3 GeV) originate for instance from weak decays of heavy quarks (beauty- and charm-quarks at HERA). An identification method of electrons in this energy regime can make an important contribution to the understanding of heavy flavour physics.

Within this thesis a new electron finder is developed using large data samples consisting of decay electrons from elastic J/ψ for signal and pions of ρ vector mesons for background events. The data selection is based on a new electron trigger implemented for the H1 experiment in 2006. Several estimators are defined based on calorimeter and tracking information in order to distinguish between signal (electrons) and background (pions). The separation power of the specific variables is determined by means of the data samples and the description of the estimators by Monte Carlo simulation is verified. A discrimination quantity reverting to sophisticated classifier methods is derived by combining the estimators in a multivariate analysis. The discriminator is trained using the selected data samples for signal and background. The presented electron identification algorithm is supposed to be applied in a difficult kinematic regime with large hadronic background.

The following list gives the outline of the thesis:

- The next chapter gives a summary of the motivation for this study.

- In chapter 3 the HERA collider and the detector of the H1 experiment are introduced. The subsystems of the detector delivering the information used for the electron identification are presented in more detail, namely the tracking system and the liquid argon calorimeter.

- The theory behind the identification of electrons is described in chapter 4 where the relevant processes in the detector for electrons and expected background particles are presented.

- Chapter 5 gives a short overview of the basic components of the used Monte Carlo simulation.

- The chosen method and the employed quantities for the identification of electrons are discussed in chapter 6. The data selection for the studies is followed by the resulting separation power of the estimators. The estimating variables are evaluated on data and Monte Carlo simulated events. The distributions of the specific estimators are separately compared between data and Monte Carlo simulation for signal and background, in order to verify the description of the chosen variables by simulation.

- Chapter 7 starts with a general introduction to multivariate analysis. The used classifier methods and the obtained results are discussed.

- A first application of the developed finder and the corresponding results are presented in chapter 8. The implemented method to identify electrons is used to obtain an invariant mass distribution of the decay electrons of inelastic J/ψ vector mesons.

- In the last chapter the conclusions of this diploma thesis and a short outlook for further studies is given.

Chapter 2

Motivation

Measure what is measurable, and make measurable what is not so.
Galileo Galilei

In the HERA collider, located at DESY in Hamburg, high energetic electrons and protons are brought to collisions. The H1 experiment employs these *ep*-interactions to investigate for instance the production of heavy charm- and beauty-quarks. Heavy flavour physics is one of the main focuses of the H1 experiment. The investigated events are divided into two kinematic regions using the virtuality Q^2 describing the square of the momentum transfer from the electron to the proton. Events with a virtuality $Q^2 > 1\,\mathrm{GeV}^2$ are called Deep Inelastic Scattering (DIS). The regime of $Q^2 < 1\,\mathrm{GeV}^2$ is referred to as the photoproduction regime. The rate of *ep*-events decreases with an increasing virtuality since the cross section is proportional to $1/Q^4$. Therefore the DIS events happen less frequently as those in the photoproduction regime.
Heavy quarks, meaning b- and c-quarks, decay via the weak interaction. About 10 % of the decays are semileptonic, where the intermediate $W^{\pm}$-boson produces a lepton and a neutrino. The charged leptons can be detected, whereas the neutrinos are not traceable in the detector. The aim of this diploma thesis is to develop an algorithm to identify electrons produced in semileptonic decays of heavy quarks.

The most important production process of heavy quarks at HERA is boson-gluon-fusion (BGF). This process is illustrated in leading order in figure 2.1. A gluon originating from the proton forms a quark-antiquark pair. The (anti-)quark absorbs a photon, emitted by the electron, carrying the virtuality Q^2. The beauty-quark production is suppressed compared to the charm-quark production since the cross section of the BGF depends on the mass and the charge of the produced particles.

The measurement of the heavy quark production is used to determine the gluon structure of the proton. The analysis of these events allows to test the predictions of the theory of perturbative Quantum Chromodynamics (pQCD). A perturbative approach is

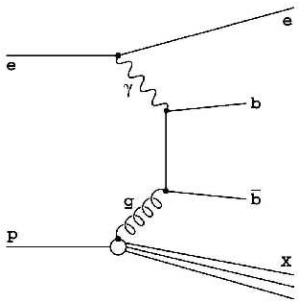

Figure 2.1: Leading order Feynman diagramm of a boson gluon fusion process.

feasible for the regime of photoproduction in particular for b-quarks since the large mass defines a hard scale.
Advantages of the semileptonic decay channel of heavy quarks for the measurement of the production cross section are the large brunching ratio of about 10 % and the possibility to trigger such events. The muonic decay channel has already been studied in different measurements since the muons give a clear signal in the detector. The information provided by electron and muon measurements is complementary and therefore the studies on the electron channel should deliver further insights.

Both experiments at HERA investigating *ep*-interactions (H1 and ZEUS) have performed analyses of the b-quark production using data from the HERA-I run period. The results for the b-quark cross section in DIS and photoproduction are shown in figure 2.2. The measured b-quark cross section is plotted versus the transversal momentum of the b-quark $p_t(b)$ relative to the predictions of pQCD. The measurements show a slightly higher cross section than the theoretical prediction. Therefore there is a great interest in verifying these results involving more statistics and to perform measurements for even lower momentum particles.

In the course of the luminosity upgrade for the HERA-II running period an additional trigger system for the H1 detector is installed. This Fast Track Trigger (FTT) opens the possibility to select events more specific. The FTT reconstructs tracks of charged particles with an accuracy comparable to the offline reconstruction and is able to identify selected topologies and event kinematics. Based on the FTT a new trigger is implemented allowing a fast and accurate identification of low momentum electron events at trigger level [1].

In the standard H1 event reconstruction software, a software package is implemented to search for electron and muon signatures. This package is called KALEP [2] and was written in 1994. It uses calorimeter and track information to identify leptons. KALEP

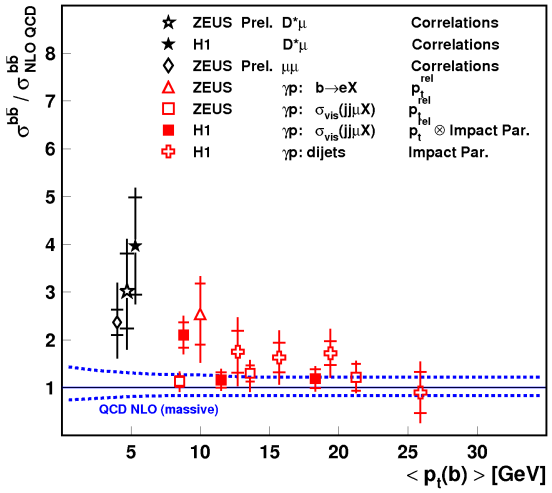

Figure 2.2: The cross section of b-quark production measured by H1 and ZEUS relative to the pQCD predictions as a function of the transversal momentum of the b-quark $p_t(b)$.

has become the standard tool to identify muons and electrons at low energy.

The performance of the KALEP electron identification is illustrated in figure 2.3. The output of the algorithm has a quality flag, which is denoted by the letter Q in the figure. The electron identification by KALEP uses four estimators, compound of the track momentum, energy measurements in the electromagnetic and hadronic part of the liquid argon calorimeter and a quantity for the shower length (see section 3.2 for the description of detector components).
No sharp cuts on the used estimators are imposed but deviations from the defined limits for each variable are summed up. The outcome of this evaluation is divided into four groups. A higher electron quality means a better background rejection but leads to a reduced efficiency.
The single data points for every quality step at the values efficiency $\epsilon = 1$ and rejection $\pi = 1$ are underflow (respectively overflow) points which are taken into account for the computation.

Although KALEP is the standard finder, the algorithm is not applicable for every analysis. One of its problems is that the electron-misidentification probability is too high for many applications like the study of $b \to eX$ decays.

Since the publication of KALEP several analyses have been done, where an individual lepton finder was developed. Unfortunately, none of these algorithms have been included in the H1 software environment and therefore the adaptions and the improvements are not directly available for new analyses.

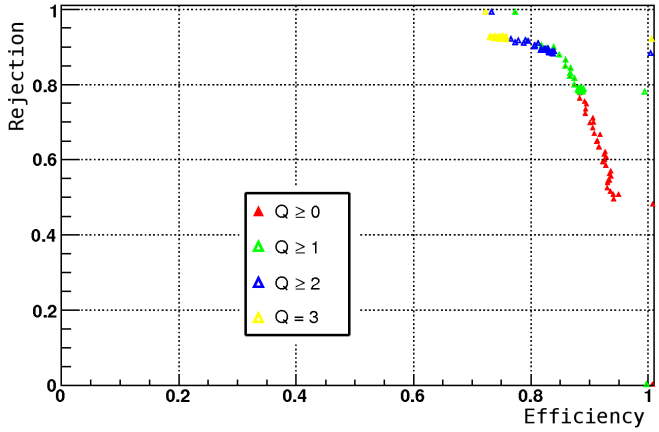

Figure 2.3: Background rejection versus signal efficiency for the electron identification by KALEP. Data selection: one KALEP-identified electron from J/ψ-candidates. [3]

One of the improved electron finders is described in a study of charm and beauty production at the H1 experiment using dilepton events [4]. This identification algorithm considers the different shower shapes for electrons and hadrons (see chapter 4) in more details. After some preselection cuts the discrimination is performed using a linear method (Fisher Discriminant) to derive a single test statistic out of a set of four estimators. The resulting performance for different preselection cuts on one of the used estimators (E/p) can be seen in figure 2.4.

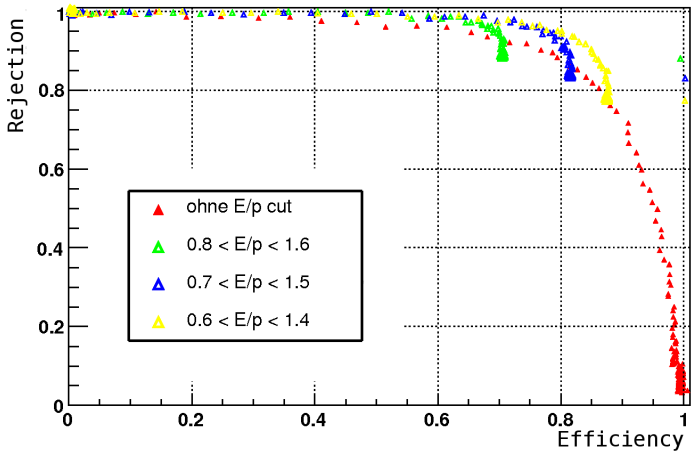

Figure 2.4: Background rejection versus signal efficiency for the electron identification in the study of charm and beauty production at H1 using dilepton events [4]. Data selection: one KALEP-identified electron from J/ψ-candidates. [3]

Another method for the identification of electrons has been developed within a study about the measurement of the beauty cross section using semileptonic decay into electrons at HERA [5]. This electron finder uses neural networks to combine calorimeter shower shape parameters with the specific energy loss $\mathrm{d}E/\mathrm{d}x$.

The figures clearly show, that for an electron finder further improvements are achievable. A new electron finder should be applicable in different environments, especially in the regime of lowest momenta (approximately 1 to 3 GeV). In this phase space an application is for instance open heavy quark production.

Chapter 3

The H1 Detector at HERA

It doesn't matter how beautiful your theory is, it doesn't matter how smart you are. If it doesn't agree with experiment, it's wrong.
RICHARD FEYNMAN

This chapter is an introduction to the H1 experiment at DESY (*Deutsches Elektronen Synchrotron*). Therefore a brief description of the HERA storage ring is given and the main components of the H1 detector are presented. A more detailed view of the calorimetry system and the tracking system is given as it is the most important part of the detector for this thesis.

3.1 HERA

The HERA (*Hadron-Elektron Ring Anlage*) collider is located at DESY in Hamburg (Germany) and consists of two seperated rings for electrons (and positrons respectively) and for protons in a common tunnel. The storage ring has a length of 6336 m and is schematically illustrated in figure 3.1. One ring accelerates protons to an energy of up to 920 GeV, whereas the other ring accelerates electrons (or positrons) to an energy of 27.6 GeV. The *ep*-collisions yield a centre of mass energy $\sqrt{s} = 318$ GeV. The protons are accelerated counterclockwise and collide at two interaction points in the middle of the straight parts of the ring, where the experiments H1 and ZEUS are situated, with the clockwise rotating electrons or positrons. The particles in HERA are accelerated in bunches and a bunch crossing takes place every 96 ns. At the other two straight parts of the storage ring, two further experiments using only one of the beams are installed. The HERMES-experiment uses the electron/positron beam to perform collision experiments on polarised gas targets. At the fourth interaction point the experiment HERA-B was studying nucleon-proton interactions until 2001.

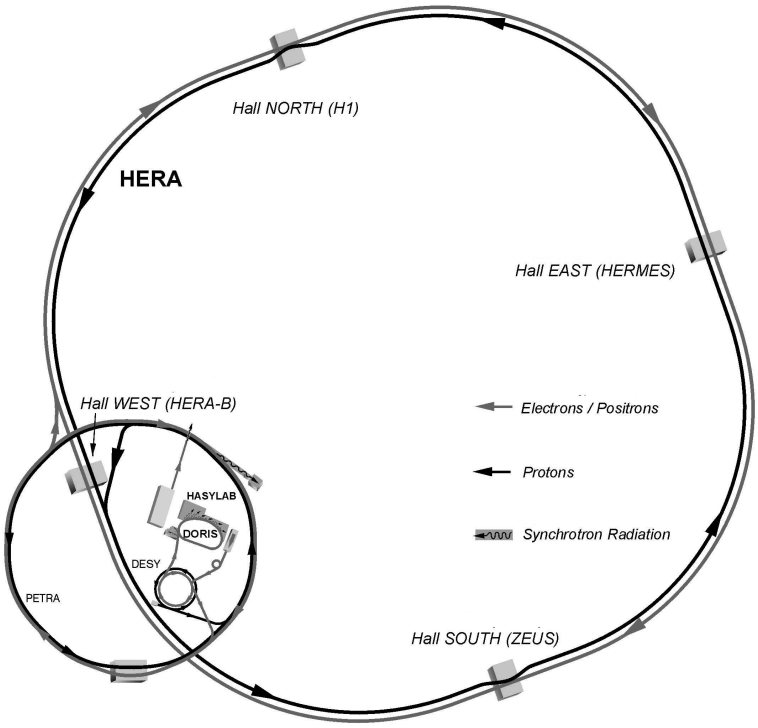

Figure 3.1: Schematic illustration of the HERA collider, the storage rings and its pre-accelerators at DESY. The experimental halls are denoted by the corresponding experiments.

The protons are accelerated in three steps to 40 GeV before they are injected from the smaller ring PETRA (*Positron Elektron Tandem Ring Anlage*) into HERA where the final acceleration step takes place. The electrons/positrons are injected from PETRA into HERA at an energy of 12 GeV after passing three pre-accelerators.

Two characteristic quantities of a high energy physics accelerator are the centre of mass energy and the luminosity. Whereas the energy defines the accessible phase space and the possible reactions, the luminosity $\mathcal{L}$ is related to the expected rate of interactions $(\frac{\mathrm{d}N}{\mathrm{d}t})$ for a given cross section σ.

$$\mathcal{L} = fn\frac{N_{\mathrm{electron}}N_{\mathrm{proton}}}{4\pi\sigma_x\sigma_y} = \frac{1}{\sigma}\frac{\mathrm{d}N}{\mathrm{d}t} \ ,$$

where f is the frequency of revolution for particles in the ring and n the number of colliding bunches in each beam. N_{electron} and N_{proton} are the number of particles in the bunches of the corresponding beams and σ_x and σ_y is the horizontal and vertical beam spread respectively.
The HERA storage ring comprehends up to 180 bunches for each beam where every

bunch consists of the order of 10^{10} particles.
The first running period of HERA (HERA-I) started in 1992 and ended in summer 2000. In the following shutdown period, lasting several months, an extensive upgrade was performed. Within the experiments H1 and ZEUS additional superconducting focussing magnets were installed, achieving a reduction in the beam size. Whereas the new bending of the beam forced other detector components to be adjusted, the luminosity seen by the experiments has been increased.
After this extended shutdown and upgrade period, HERA-II was operated until the end of june 2007.

3.2 The H1 Detector

The H1 detector is a general purpose detector located in the north of the HERA ring. It was built and maintained by a collaboration of physicists from many institutes located in different countries. The detector is capable of identifying neutral and charged particles originating from the electron proton interaction.
The protons pass the H1 detector along the z-axis, electrons/positrons along the opposite direction. The particles are brought to collision at the nominal interaction point located in the middle of the detector. The design of the detector allows to cover almost the complete solid angle of 4π around the interaction point. In order to respect the different energies of the colliding beams, the detector is built asymmetrically.
In figure 3.2 a schematic sideview of the H1 detector is shown.

The coordinate system used by the H1 experiment is orientated such that the positive z-axis points in direction of the proton beam, called the forward direction. The xy-plane is perpendicular to this direction where x points to the center of the HERA ring and y upwards. The origin of the H1 coordinate system is the nominal interaction point. The polar angle θ is defined as the angle between the trajectory of the particle and the z-axis. The azimuthal angle ϕ is defined in the xy-plane where $\phi = 0$ corresponds to the positive x-axis.

The H1 experiment consists of three main sections for different detection purposes. These sections are the tracking chambers, the calorimeter and the muon system, where every section is made up of several sub-detectors. The different components are built cylindrical around the beam line. The tracking detectors are located closest to the beam line and measure tracks of charged particles. They are surrounded by the calorimeter which measures particle energies. The outmost part is the muon system which detects muons and high energy hadrons.
A comprehensive description of the H1 detector can be found in [6].

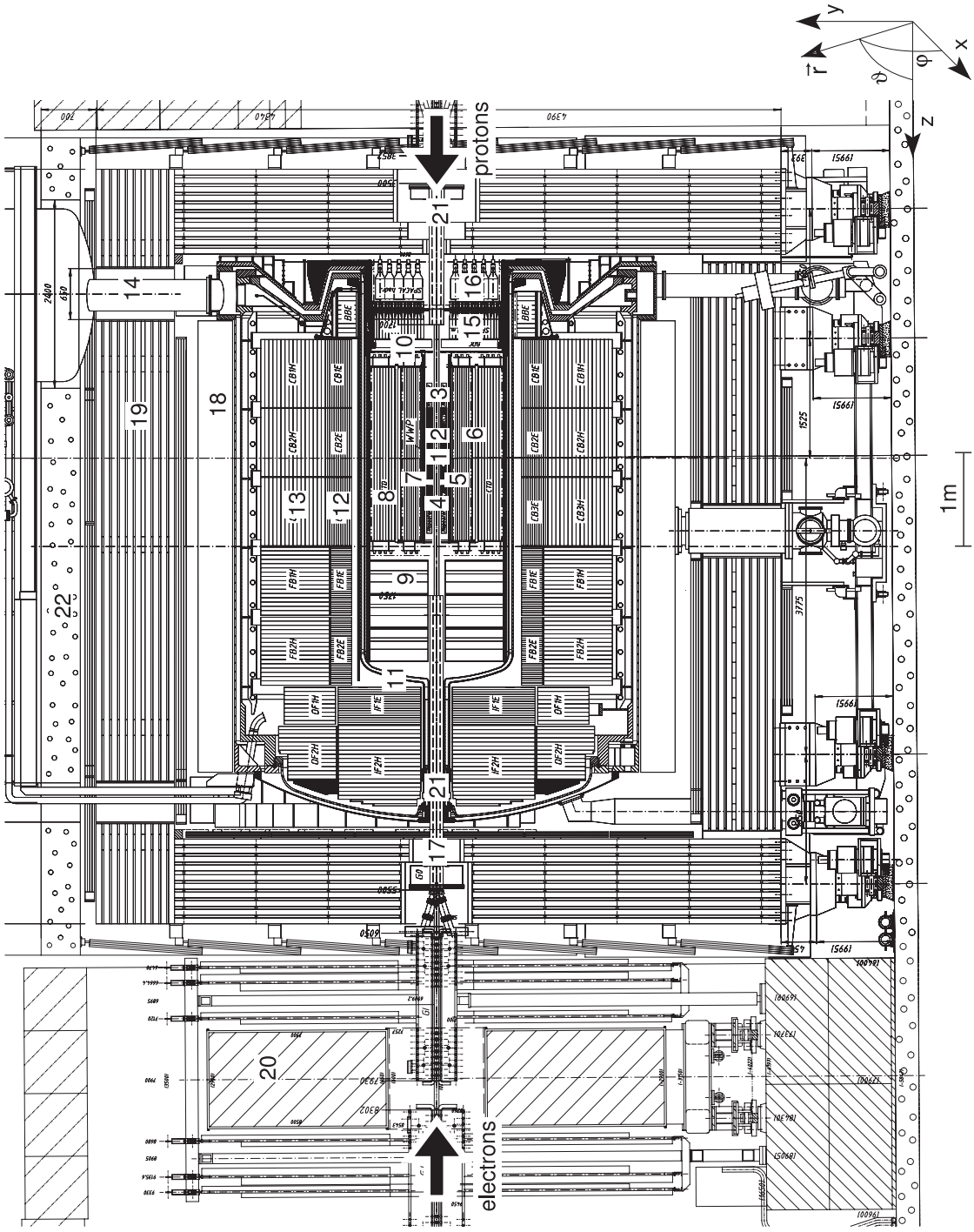

Figure 3.2: A longitudinal cut through the H1 detector showing the main components of the H1 detector. The legend is shown in table 3.1 for the different constituents. The coordinate system used in H1 is shown on the top right and its origin is located at the interaction point (WWP) denoted by (1).

	Detector component	Abbreviation
1	Nominal interaction point	IP (WWP)
	Tracking detectors	
2	Central silicon tracker	CST
3	Backward silicon tracker	BST
4	Forward silicon tracker	FST
5	Central inner proportional chamber	CIP
6	Central outer z-chamber	COZ
7	Inner central jet chamber	CJC1
8	Outer central jet chamber	CJC2
9	Forward tracking detectors	FTD
10	Backward proportional chamber	BPC
	Calorimeters	
11	Liquid argon cryostat	
12	Liquid argon electromagnetic calorimeter	LAr ECAL
13	Liquid argon hadronic calorimeter	LAr HCAL
14	Liquid argon cryogenics system	
15	Electromagnetic spaghetti calorimeter	SpaCal elm.
16	Hadronic spaghetti calorimeter	SpaCal hadr.
17	Plug calorimeter	Plug
18	Superconducting solenoid	
	Muon detectors	
19	Instumented iron (central muon / tail catcher)	CMD/TC
20	Forward muon detector (incl. toroid magnet)	FMD
21	New superconducting focusing magnets	GO/GG
22	Concrete shielding	

Table 3.1: Legend to figure 3.2: The main components of the H1 detector.

3.2.1 Tracking Detectors

The tracking system of the H1 detector consists of precision silicon solid-state detectors, multiple wire proportional chambers and drift chambers (see figure 3.3) and is divided into a Central and a Forward Tracking Detector (CTD and FTD). The silicon detectors are used to identify decay vertices from long lived particles, whereas the drift chambers are responsible for the reconstruction of the tracks of the particles. The fast signals of the proportional chambers are mainly used for triggering purposes. The precise position measurements of the drift chambers allow to determine the momentum of charged particles, as their trajectory is bent because of the applied magnetic field in the detector. The superconducting solenoid located outside of the calorimeter produces this nearly homogeneous magnetic field of 1.15 T.

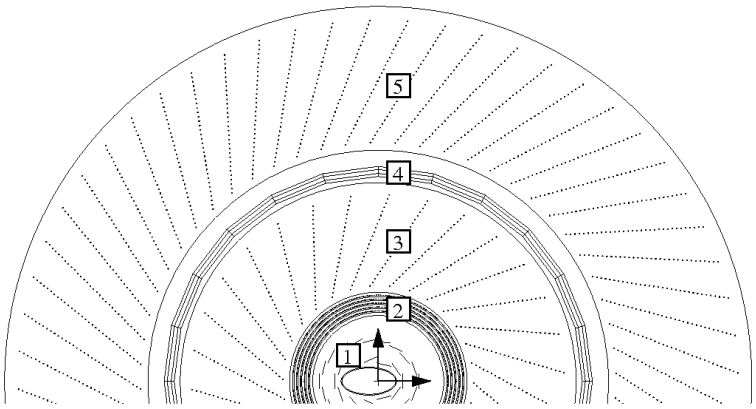

Figure 3.3: Radial view of the central tracking system of the H1 experiment. The following components are shown in radial order starting from the beam pipe: (1) Central Silicon Tracker CST, (2) Central Inner Proportional Chamber CIP, (3) Central Jet Chamber 1 CJC1, (4) Central Outer z-Chamber COZ, (5) Central Jet Chamber 2 CJC2. The signal wires in CJC1 and CJC2 parallel to the beam line are illustrated by dots.

The innermost tracking detectors are the silicon detectors where the Central Silicon Tracker (CST) consists of two layers of double-sided silicon detectors with an angular coverage of $30° < \theta < 150°$. The two layers enclosing the beam pipe have radii between 4 and 13 cm in an elliptical arrangement adapted from the beam pipe profile. In forward and backward direction there are two more silicon tracking detectors installed (FST and BST respectively) to get a better coverage for tracking in the polar angle θ.
Going from the beam pipe outward, the CST is followed by the Central Inner Proportional Chamber (CIP), which is a multi-wire proportional chamber with a pad readout. The signal pads are installed perpendicular to the beam line and the CIP surrounds the CST circularly to cover a polar angle range of $11° < \theta < 169°$. The fast tracking information delivered by this detector allows to reconstruct the position of an interaction

along the beam line and the high time resolution is used for triggering. Due to the fast signal processing the signal from the proportional chamber determines the time of the interaction.
The inner tracking devices are surrounded by the main component of the tracking system at H1 the Central Jet Chamber (CJC). This coaxial drift chamber is split into an inner ring (CJC1) and an outer ring (CJC2). The drift chambers consist of several layers of sense wires located parallel to the beam line and form 30 azimuthal drift cells for CJC1 and 60 cells for CJC2 respectively. The exact position of the particle can be determined by the known position of the wire, the drift time of the induced charges to the signal wire and the drift velocity. The position along the wire is determined from the ratio of the charges read out from both wire ends. The precise timing information determined by the CJC helps in recognising events from cosmic ray muons, where the signal from the upper segment of the track is earlier registered than the lower one. A polar angle range of $20^\circ < \theta < 160^\circ$ is covered and the information from the CJC is used by the Fast Track Trigger (FTT) as input signals. The tracks detected by the CJC have a high resolution in $r\phi$ but a worse in z-direction. Therefore the central outer z-chamber (COZ) is located between CJC1 and CJC2 to determine the exact z-position of the tracks in this direction using the drift time to the signal wires which are oriented perpendicular to the z-axis.
In the forward region the detector is instrumented with three planar drift chamber modules (FTD) measuring tracks in an acceptance region of $7^\circ < \theta < 25^\circ$ whereas in the backward region an additional proportional chamber (BPC) is mounted to measure the angle of the scattered electron from the ep-interaction.

The measurement of the specific energy loss of a particle in material $\mathrm{d}E/\mathrm{d}x$ is performed using the information of the Central Jet Chamber. The measurement is obtained by integration of the charges read out from the CJC wires. To get a useful result numerous corrections have to be applied, which will not be discussed in further details. The particle mass dependence of $\mathrm{d}E/\mathrm{d}x$ in the CJC allows to use this quantity in order to obtain a particle identification.

3.2.2 Calorimeters

The energy of both charged and neutral particles can be measured with calorimeters. The actual amount of deposited energy in a calorimeter depends on the particle type, its momentum and the calorimeter configuration. The calorimetry system of the H1 experiment consists of two main calorimeters. The forward and central region is covered by a Liquid Argon (LAr) Calorimeter ($4^\circ < \theta < 153^\circ$) which surrounds the H1 tracker. The Spaghetti Calorimeter (SpaCal) measures the energy of passing particles in the backward region of the detector.

The SpaCal is a lead-scintillating fibre calorimeter compound of an electromagnetic and a hadronic section. The polar angle range $153^\circ < \theta < 173^\circ$ is covered and due to the fine granularity the energy and the impact position of particles can be measured

with high accuracy. The reached energy resolution in the electromagnetic section is $\sigma(E)/E \simeq 0.08/\sqrt{E/\mathrm{GeV}} \oplus 0.01$ compared to $\sigma(E)/E \simeq 0.30/\sqrt{E/\mathrm{GeV}} \oplus 0.07$ in the hadronic part. Furthermore the SpaCal has an excellent time resolution of about 1 ns. According to its properties and the location in the detector, the main task of the SpaCal is to performe accurate measurements on the scattered beam electron.

The liquid argon calorimeter is a non-compensating sampling calorimeter. It is divided in an electromagnetic (ECAL) and a hadronic (HCAL) part. The electromagnetic part measures the energy of electrons and photons, whereas the energy of hadrons is measured using the electromagnetic and the hadronic part of the calorimeter. The energy of these particles is measured in the LAr calorimeter by absorption while muons deposit only a small amount of energy in both parts of the LAr calorimeter by ionisation. In case of low energy pions it is possible that they get absorbed already in the electromagnetic part of the calorimeter without reaching the hadronic part. The calorimeter signal of this pions could be mistaken as those of electrons.
The active material of this calorimeter is liquid argon which is cooled down to $-183\,^{\circ}\mathrm{C}$ by a cryostatic system around the calorimeter. The absorbing layers are made of lead in the electromagnetic part whereas stainless steel is used in the hadronic part. The width of the electromagnetic part corresponds to 20 - 30 radiation lengths, that of the hadronic part corresponds to 4.7 - 7 interaction lengths. The amount of dead material between the tracking detector and the calorimeter has been reduced to a minimum.
The calorimeter is divided into eight wheels along the z-direction. They are called according to their position in the calorimeter (see figure 3.4). The wheels in the central region are the central barrels denoted by CB1 - 3, those in the forward region are the forward barrels (FB1 - 2). The inner and outer forward (IF and OF) wheels are located at the end of the calorimeter in z-direction, whereas the backward barrel (BB) is located at the opposite end. Modules with a designation ending in 'E' form the electromagnetic section, whereas modules ending in 'H' belong to the hadronic section. The smallest unit consisting of absorber and active material which is read out is called calorimeter cell. The segmentation into cells and the orientation of the absorber plates is different for the specific wheels according to their position along the z-axis. The orientation of the absorber plates in the wheels is choosen such that the particles originating from the interaction point always pass the plates with an angle bigger than 45°. The segmentation of the liquid argon calorimeter into wheels, the orientation of the absorber plates and the electromagnetic and hadronic sections of the calorimeter are illustrated in figure 3.4. The space between the wheels is not instrumented and is therefore dead material. Every wheel of the liquid argon calorimeter is divided into eight octants in the r-ϕ-plane. This is shown for the wheel CB2 in figure 3.5. The space between the octants is not instrumented either.

The LAr calorimeter is highly segmented in cells, which collect the charges from the ionisation processes in the active material induced by the passing particles. The calorimeter reconstruction program converts the charges to energies in the calorimeter cells individ-

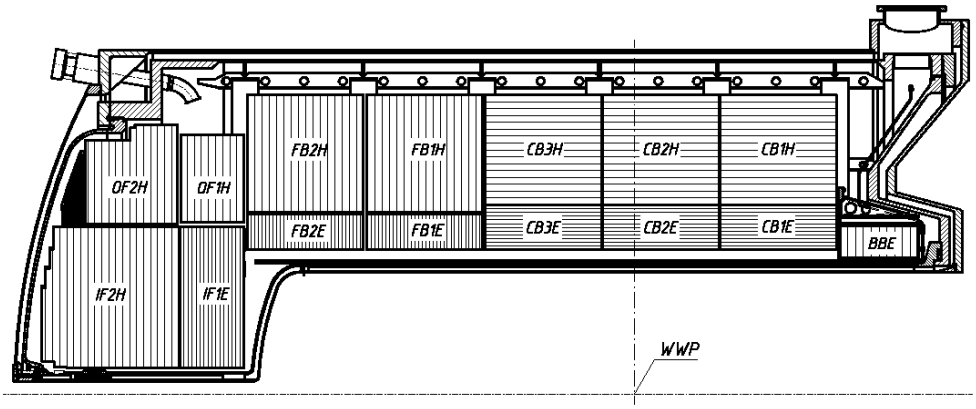

Figure 3.4: The upper half of the liquid argon calorimeter (r-z-view). The calorimeter is divided into 8 wheels. The central barrels are denoted by CB 1 - 3 and the forward barrels by FB 1 - 2. The parts in the most forward region of the calorimeter are called inner forward (IF) and outer forward (OF) that in the backward region is called backward barrel (BB). Modules with a designation ending in 'E' form the electromagnetic section, whereas modules ending in 'H' belong to the hadronic section. In every part of the calorimeter the orientation of the absorber plates is illustrated by lines.

ually for the electromagnetic and the hadronic part of the calorimeter. The electromagnetic section consists of about 31000 read out channels leading to an energy resolution of $\sigma(E)/E \simeq 0.11/\sqrt{E/\mathrm{GeV}} \oplus 0.01$, which has been determined in test beam measurements. In the hadronic part the granularity is much coarser with a total number of about 13500 readout channels. An energy resolution of $\sigma(E)/E \simeq 0.50/\sqrt{E/\mathrm{GeV}} \oplus 0.02$ is achieved in this part of the calorimeter [7].

The liquid argon calorimeter is equipped with a trigger system which delivers a signal for the first trigger level by summing the energy deposits in certain regions of the calorimeter. It is for instance possible to trigger on a high local energy deposition or a specific event configuration in the calorimeter.

3.2.3 Muon Detectors

As mentioned in the description of the calorimeters muons mainly lose their energy in ionisation processes. This leads to a higher penetration depth compared to electrons, photons and hadrons. Therefore the muon detector is located outside of the calorimtery system allowing the electrons, photons and hadrons to be absorbed before reaching the muon detector. In the H1 detector the massive magnetic field returning iron yoke of the solenoid is laminated and instrumented with limited streamer tubes to measure muon tracks. This sensitive modules are installed between the plates of the iron yoke. The muons need to have a minimum energy of 2 GeV in order to reach the muon detector.

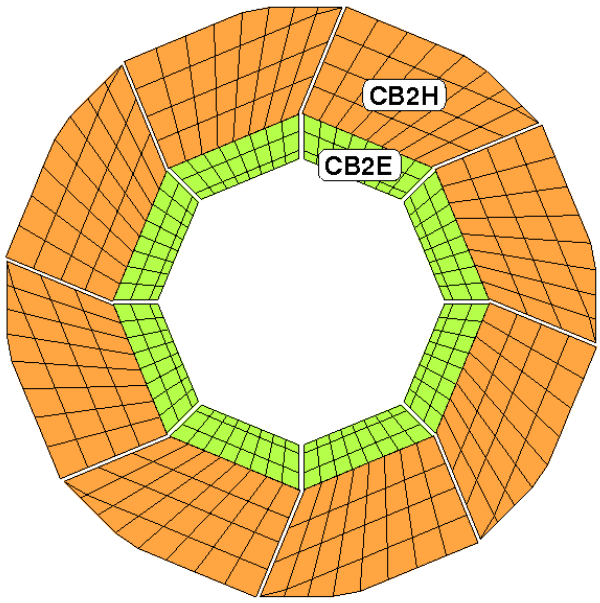

Figure 3.5: The LAr wheel CB2 in a r-ϕ-view which is composed of an inner electromagnetic section, CB2E, and an outer hadronic section, CB2H, of the calorimeter. The space between the octants of the calorimeter wheel is not instrumented.

Some of the streamer chamber layers in the iron module (since december 2006 also the ones in the forward region) are used by the instrumented iron muon trigger to select muon events.
The Central Muon Detector (CMD) consisting of the backward endcap, the backward barrel, the forward barrel and the forward endcap has an angular acceptance of $5° < \theta < 175°$. The angular acceptance of the Forward Muon Detector (FMD) starts at $\theta = 3°$, reaches full azimuthal coverage at $\theta = 5°$ and extends to $\theta = 17°$.

3.2.4 Luminosity Measurement

The luminosity $\mathcal{L}$ delivered by HERA is determined by the measurement of the rate at which the Bethe-Heitler-process occurs. The Bethe-Heitler-process $ep \to ep\gamma$ is very precisely calculable in the theory of quantum electro dynamics. Beside that its cross section is large and therefore this process is suitable to determine the luminosity of the collider. The produced photons are detected in a photon detector located at $z = -103$ m [8].

3.2.5 Trigger System

The frequency of the bunch crossing at the HERA collider is about 10.4 MHz and the expected rate of electron-proton scattering processes is about 1 kHz. Signals in the H1 detector produced by background processes are up to 1000 times more frequently than the signals from interesting physical processes. The dead-time of the H1 detector during the readout of an event limits the readout rate of the detector to 50 Hz. Therefore the trigger system must be able to prevent the readout of background processes in a reasonable manner as well as to select only physical processes which are important for further analyses.
The trigger system at the H1 experiment acts as a four level filter. Every step reduces the event rate gradually which increases the available time for the analysis of the event in the next level. In total the event rate is reduced to about $10 - 25$ Hz, permitting to store all the data permanently.
On the first level (L1) the information delivered by the sub-detectors is used on a hardware level to generate trigger signals. In case of a positive trigger decision (L1-keep signal), the information available from L1 is combined on the second level (L2) in neural networks or using topological criteria. The third trigger level was implemented within the Fast Track Trigger (FTT) project for the HERA-II running period [9]. The FTT itself works on the levels L1 to L3. The tracks of the FTT-L2 system and the data from other trigger systems are used as input signals into L3. With this information a partial event reconstruction is performed on commercial processors. The fourth level finally accomplishes a full event reconstruction, after the readout process is completed, running on standard PCs. That is to say that the L4 level does not contribute to the dead-time of the detector. At the end the detector information of the events passing all trigger levels are permanently stored.

Chapter 4

Electron Identification

Since the mathematicians have invaded the theory of relativity, I do not understand it myself anymore.

Albert Einstein

This section is about the theory behind the identification of electrons. In order to identify particles interacting in detectors, one has to understand how different particles act when passing the detector.
The main goal of this finder is to distinguish between electrons ($e^{\pm}$) and hadrons. Especially electrons from pions, since pions are the most important contribution to the background. Therefore the interactions of electrons and hadrons are discussed.

4.1 Electromagnetic Shower

A particle shower is a cascade of secondary particles produced when an incoming high energetic particle interacts with dense matter. In experiments this is normally the calorimeter in the detector. In this interaction new particles are produced with less energy than the primary particle. Each produced particle has the ability to interact with matter to repeat the process. This continues until many low-energy particles are produced whose energies are low enough to be absorbed in the detector material. The characteristic shape of this procedure leads to the concept "*Shower* ". A picture of a measured electromagnetic shower in the ICARUS liquid argon drift chamber is shown in figure 4.1.

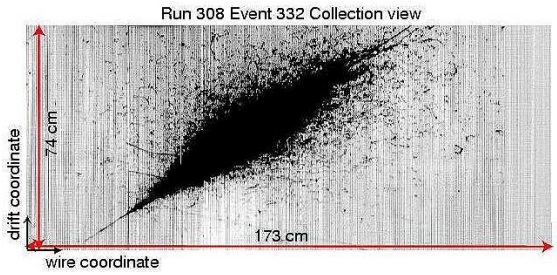

Figure 4.1: Electromagnetic shower observed in the ICARUS LAr drift chamber during the technical run with cosmic rays at Pavia, summer 2001 [10].

In case of electrons the interaction of the incoming particle with the dense matter is called electromagnetic. The most important electromagnetic processes in the calorimeter are bremsstrahlung, where the electron emits a photon ($e \rightarrow e\gamma$), and pair production, where the photon converts into an electron-positron pair ($\gamma \rightarrow e^+e^-$). The development of an electromagnetic shower is shown schematically in figure 4.2.
Underneath a certain critical energy, the electrons start to loose energy mainly through ionisation until they get absorbed by the dense matter.

Since electrons are light particles, they loose their energy in a smaller number of collisions than heavy charged particles. In fact their mass is equivalent to the mass of the orbital electrons in the matter with which they are interacting. This means that a large fraction of the energy of the incoming electron can be lost in a single interaction. If the electron interacts with a nucleus it is even possible for the electron to change direction of propagation or to get backscattered. This leads to large deviations from the incoming direction of the electron. This also means, that the paths of monoenergetic electrons do not necessarily look similar.

In order to identify a cascade in the calorimeter as an electromagnetic shower, or more precisely as an electron, the longitudinal and lateral distribution of the energy depositions is important.
The longitudinal development is controlled by the high-energy part of the shower. This means that the length of the cascade scales with the radiation length in the given material and logarithmic in the incoming energy [11]. The number of electrons in the shower drops more quickly with the depth of penetration than the energy deposition. This is because of the increasing number of photons in the shower due to bremsstrahlung in the absorber material.
The maximum number of particles in the cascade is reached, when the average energy per particle drops below the material dependent critical energy. From this point no more new particles are produced and the shower decays slowly through ionisation losses for electrons and by Compton scattering and absorption for photons (see figure 4.3).

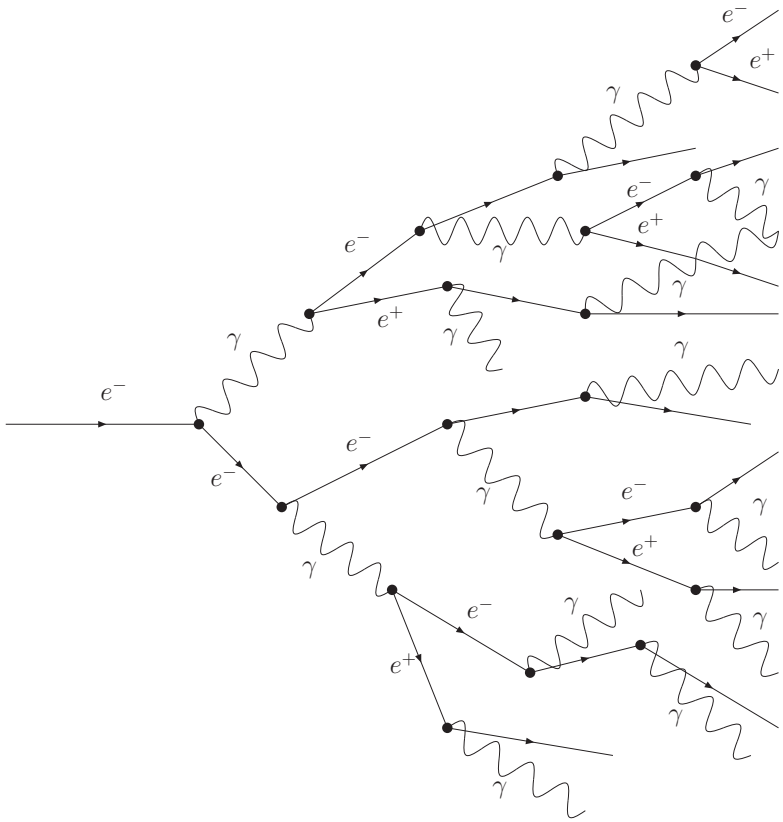

Figure 4.2: Schematical illustration of an electromagnetic shower in the calorimeter. Shown are the processes of bremsstrahlung ($e \rightarrow e\gamma$) and pair production ($\gamma \rightarrow e^+e^-$).

The lateral shower distribution is well described by a heuristic constant, the "*Moliere Radius*". Typically 95% of an electromagnetic shower is contained in a cylinder with radius $2R_{Moliere}$. The Moliere Radius is defined by $R = X_0 \cdot \frac{E_S}{E_{critical}}$, where the scale energy is $E_S = \sqrt{\frac{4\pi}{\alpha}} m_e c^2$, X_0 the radiation length and $E_{critical}$ the critical energy of the medium [11].

Recapitulatory one can say that showers of electrons are narrow and homogeneous. The sidewise enlargement is determined by the scattering processes of the electrons and positrons in the beginning of the shower origin and by Compton scattering after the phase of pair production.

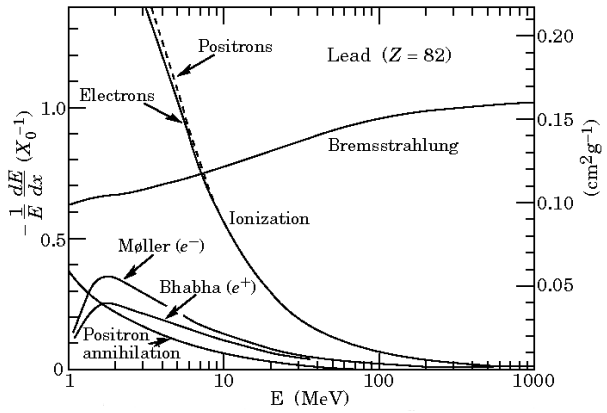

Figure 4.3: Fractional energy loss of electrons in lead as a function of electron energy. The critical energy is defined as the point where the ionisation loss is equal to the bremsstrahlung loss [11].

4.2 Hadronic Shower

As hadrons consist of quarks, the main interaction in dense matter is induced by the strong force, leading to a variety of possible processes and a different shower shape in the detector compared to electrons. This is capitalised to distinguish hadrons from other particles for the identification.
In the detector the incoming hadron interacts with the nuclei in the material and produces several lower-energy hadrons. At high energy, these interactions are mainly multiparticle production and particle emission originating from nuclear decay of excited nuclei, usually pions and nucleons. These processes continue, as in the electromagnetic case, until all particles are stopped and absorbed in the material. Due to the characteristics of the strong force this hadronisation of the incoming quark builds up a cone in the calorimeter.

The binding energy, which is needed for the release of nucleons, is too small to be detected by the H1 calorimeter. Supplementary, energy may be hidden due to muons or neutrinos or by the delay until excited nuclei emit their absorbed energy. Hadrons can also interact via the electromagnetic interaction, therefore the shower can partially be electromagnetic. For example the decay of a neutral pion ($\pi^0 \rightarrow \gamma\gamma$) induces an electromagnetic cascade. The total energy fraction of the electromagnetic sub-shower may get bigger than the hadronic fraction in special processes. All of this leads to worse energy resolution for the hadronic part of the calorimeter than for the electromagnetic, since the detectable energy of an electron shower is always larger than that of a hadronic shower. This property of the calorimeter is called non-compensating.

For high-energy hadrons the shower length depends logarithmically on the energy of the incoming particle and linearly on the interaction length λ in the corresponding mate-

rial. For the lateral distribution one finds that 95% of the hadronic shower is contained in a cylinder with radius λ.

Hadronic showers do not only occur in particle detectors but also in nature. In earth's atmosphere they origin from cosmic rays, which usually are hadrons. An illustration can be seen in figure 4.4.

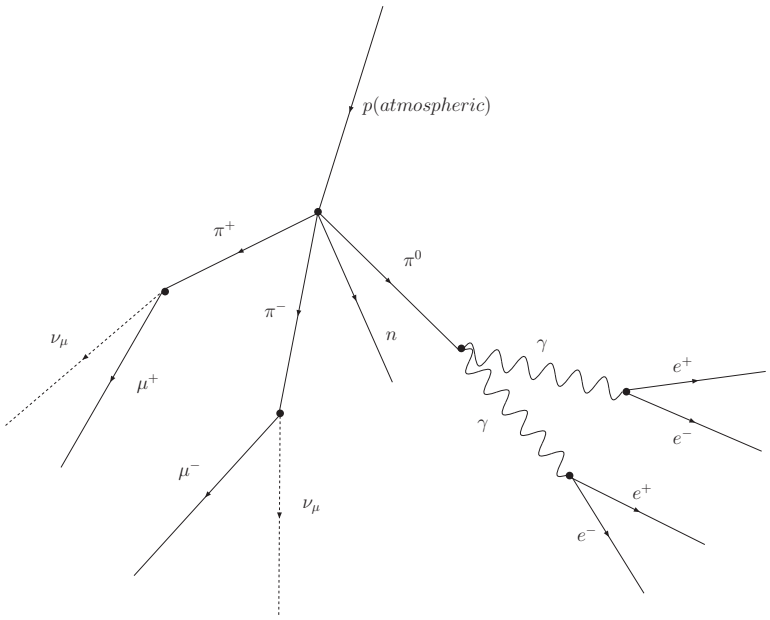

Figure 4.4: An illustration of a hadronic shower in the atmosphere. The incoming hadron interacts via the strong force. The secondary particles can interact also via the electromagnetic force (π^0) leading to a hadronic and a electromagnetic part of the shower.

To summarise again, the lateral distribution of hadronic showers is consisting of a narrow cone, emanating from the electromagnetic sub-shower, and a radially surrounding irregular halo.

4.3 dE/dx

The energy loss of a particle in matter is already used to measure the energy of an incoming particle, but the specific energy loss per path length $\frac{\mathrm{d}E}{\mathrm{d}x}$ is also very useful to identify particles. For the energy measurement it is important to collect the whole energy of the particle in the calorimeter. To identify the particle in addition, it is interesting how the energy is lost along the path through the detector, in particular which processes do occur.

The specific energy loss per path length of a particle traversing matter is described by the Bethe-Bloch formula [11]:

$$-\frac{\mathrm{d}E}{\mathrm{d}x} = \kappa z^2 \frac{Z}{A}\frac{1}{\beta^2}\left(\frac{1}{2}\ln\frac{2m_e c^2\gamma^2\beta^2 T_{max}}{I^2} - \beta^2 - \frac{\delta(\beta\gamma)}{2}\right) ,$$

where

- $\kappa = 4\pi N_A r_e^2 m_e c^2$ with
  - N_A: Avogadro's number $6.0221415 \cdot 10^{23}$ mol^{-1},
  - r_e: Classical electron radius $\frac{e^2}{4\pi\epsilon_0 m_e c^2}$,
  - m_e: Electron mass,
  - c: Speed of light,
  - e: Electron charge,
- z: Charge of incident particle in units of e,
- Z: Atomic number of absorber,
- $\beta = \frac{v}{c}$: Speed of incident particle v in units of c,
- $\gamma = \frac{1}{\sqrt{1-\beta^2}}$: Boost of incident particle,
- I: Mean excitation energy of electrons in the absorber,
- T_{max}: Maximum kinetic energy which can be imparted to a free electron in a single collision by an incident particle of mass M

$$T_{max} = \frac{2m_e c^2\beta^2\gamma^2}{1 + 2\gamma m_e/M + (m_e/M)^2} ,$$

- $\delta(\beta\gamma)$: Density effect correction to ionisation energy loss.

At the H1 experiment the specific energy loss is determined by means of the CJC. It is possible to determine the mass of the passing particle from the measurement of $\mathrm{d}E/\mathrm{d}x$ and of the momentum. Using a Likelihood-method one can then indicate a probability for a particle identification. The numeric value of the specific energy loss is proportional to the charge that is deposited by the passing particle (ionisation of the molecules in the gas chamber) and therefore proportional to the charge collected by the wire in the chamber.

A comprehensive discussion of $\frac{\mathrm{d}E}{\mathrm{d}x}$ measurement at H1 can be found in [12].

In figure 4.5 one can see the specific energy loss per unit length as a function of the momentum of different types of particles, measured at the H1 experiment.

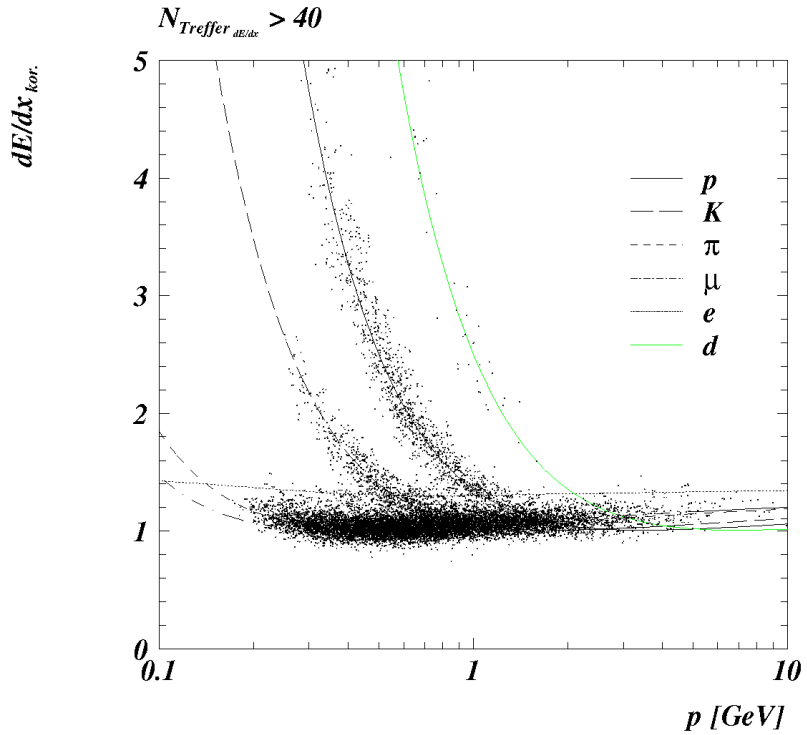

Figure 4.5: Scatterplot of the specific energy loss versus the track momentum [12].

Chapter 5

Monte Carlo Modelling

A mathematical truth is neither simple nor complicated in itself, it is.
EMILE LEMOINE

Various physical and mathematical systems, in particular in high energy physics, can be simulated by the use of computational algorithms. A widely used class for this purpose are the Monte Carlo (MC) methods. One of their mannerisms is, that they are stochastic and usually use pseudo random numbers in contrast to other simulation methods. Because of the repetition of algorithms and the large number of calculations involved, Monte Carlo is a method suited for numerical calculation using a computer.

Monte Carlo event generators are used in particle physics to model events as detailed as could be observed by a perfect detector. Three steps are performed in order to use Monte Carlo simulations for electron-proton collisions at the H1 experiment:

- *Generator*: Models and physics theories are considered and simulated by a dedicated program for the generation of events. In this thesis the event generator **DIFFVM** is used to generate the four-vectors of the attending particles by randomly sampling the physical distributions and respecting the kinematic variables. The events studied in this thesis are the decays of the vector mesons $J/\psi \to e^+e^-$ and $\rho \to \pi^+\pi^-$.

- *Simulation*: The behaviour of the passing particle and the detector response is simulated for the selected events by the use of a software describing the detector (**H1SIM**) which relies on a program package describing the different interactions with material (**GEANT**) and the electronic readout.

- *Reconstruction*: The simulated detector output is delivered to a reconstruction algorithm (**H1REC**). This step should be ideally the same for Monte Carlo and

data events, having the same output format, which then can directly be used for an analysis.

5.1 The DiffVM Generator

In order to simulate the diffractive production of vector mesons at HERA, the DiffVM Generator [13] was written. The Vector Dominance Model and the Regge theory form the core of this generator in electron-proton scattering.

First a virtual photon is generated ($e \rightarrow e\gamma$) according to the Weizsäcker-Williams approximation [14], [15]. The photon fluctuates then into a virtual vector meson prior to the diffractive interaction, by pomeron exchange with the proton.
The dependence of the γp cross section for vector meson production from the centre of mass energy and the momentum transfer at the proton vertex is handled in the framework of Regge theory.
The cross section for elastic scattering of a vector meson and a proton can be written as

$$\frac{\mathrm{d}\sigma}{\mathrm{d}t} = \left.\frac{\mathrm{d}\sigma}{\mathrm{d}t}\right|_{t=0,W=W_0} \cdot \mathrm{e}^{-b|t|} \cdot \left(\frac{W}{W_0}\right)^{4\epsilon}, \tag{5.1.1}$$

where t is the momentum transfer, W the centre of masss energy of the γp system, b the slope parameter and ϵ a free parameter.

The vector meson and the proton might be intact after the interaction. In this specific case, the vector meson is produced elastically. In the framework of this thesis, only such type of events are studied.

5.2 Detector Simulation

The particles and their four-vectors are the output of DiffVM. In order to make a meaningfull comparison between data and the simulated events, the outcome of DiffVM is given to the H1sim program, which makes the simulation of the H1 detector response. This package relies on Geant 3 [16], used to simulate the detector response and the electronic readout. After this step, the digitised information is given as an input of the event reconstruction H1Rec. This program is run on both data and Monte Carlo, issuing the same format for real and simulated events.

5.3 The Single Particle Generator

This generator allows to examine the detector response for a specific kind of particle, like electron or pion, in a given phase space. The momentum p, the polar angle θ and the azimuthal angle ϕ are uniformly distributed in a fixed interval and given as an input to the full detector simulation.

A detailed discussion of Monte Carlo modelling can be found for instance in [17] and [18].

Chapter 6

Electron Finder

We have to remember that what we observe is not nature herself, but nature exposed to our method of questioning.

Werner Heisenberg

In chapter 2 a short overview of electron finders used or developed at the H1 experiment was given. In this chapter the constitutional concept and the method of the electron finder evolved in this thesis is followed by a description of the used quantities. Moreover the selection of the data used in this study is presented. At the end a comparison of the used quantities for signal and background as well as between data and simulation is given.

6.1 Method

In this section the basic ideas of a general electron finder based on track and calorimeter information are presented.
As already mentioned in the motivation for this thesis (chapter 2), the ideas behind electron identification are not new. In this work the ideas developed in two studies at H1 ([4] and [5]) have been used as a starting point.

The main contribution to the background for an electron identification at the H1 experiment is from pions. This means that one has to be able to distinguish between electrons and pions. For this discrimination the choice of parameters with good separation power between electrons (signal) and pions (background) is required. This parameters are based on differences in the way the particles deposit their energy in the calorimeter and pass through the central tracker (tracking information).

The procedure for implementing the electron finder is to choose a set of estimators with adequate separation power according to the expected behaviour of electrons and pions in the calorimeter. This estimators are then tested on electron and pion data samples

(described in section 6.3) and finally used as input variables to a multivariate analysis (see chapter 7) in order to get a discrimination between signal (electron) and background (pion). In addition the description of the used variables by Monte Carlo simulation is verified (section 6.4).

Estimators

As discussed in section 4.2, hadrons only deposit a fraction of their energy in the electromagnetic calorimeter and in addition some energy deposition is lost to the calorimetry. Hence the measured energy in the electromagnetic part of the calorimeter is expected to be smaller than the energy deduced from the momentum assigned by the tracking chamber. For electrons in contrast, the induced electromagnetic shower is usually fully contained in the electromagnetic part of the liquid argon calorimeter and therefore this energy should be comparable to the measured momentum.
The ratio of the energy deposited in the electromagnetic calorimeter in a cylinder around the elongated track (radius $R = 30\,\text{cm}$) and the momentum of the track $\frac{E}{p}$ leads to a first estimator to separate electrons from hadrons.

The hadronic shower is supposed to have a wider lateral extension in the calorimeter than the electromagnetic shower. This is the idea behind a further estimator. For every track there are two cylinders defined around the continuation of the trajectory in the calorimeter with different radii. The position and orientation of the cylinders is denoted in figure 6.1. The used variable is then introduced as the fraction of the energy in the smaller ($R_i = 15\,\text{cm}$) cylinder divided by the energy collected in the wider ($R_o = 30\,\text{cm}$) cylinder $I = \frac{E_{\text{inner}}}{E_{\text{outer}}}$. This estimator describes the lateral fraction of the measured shower in the calorimeter. For electrons one expects more or less all the energy deposition in the smaller cylinder, leading to $I \approx 1$.
For the usage of this estimator in the context of electron identification, this variable serves also as an isolation criterion against hadronic showers. The hadronic shower consisting of the electromagnetic subshower and the surrounding halo leads to a value for the isolation criterion $I \leq 1$, as the hadronic energy deposition is spread over a wider lateral region.

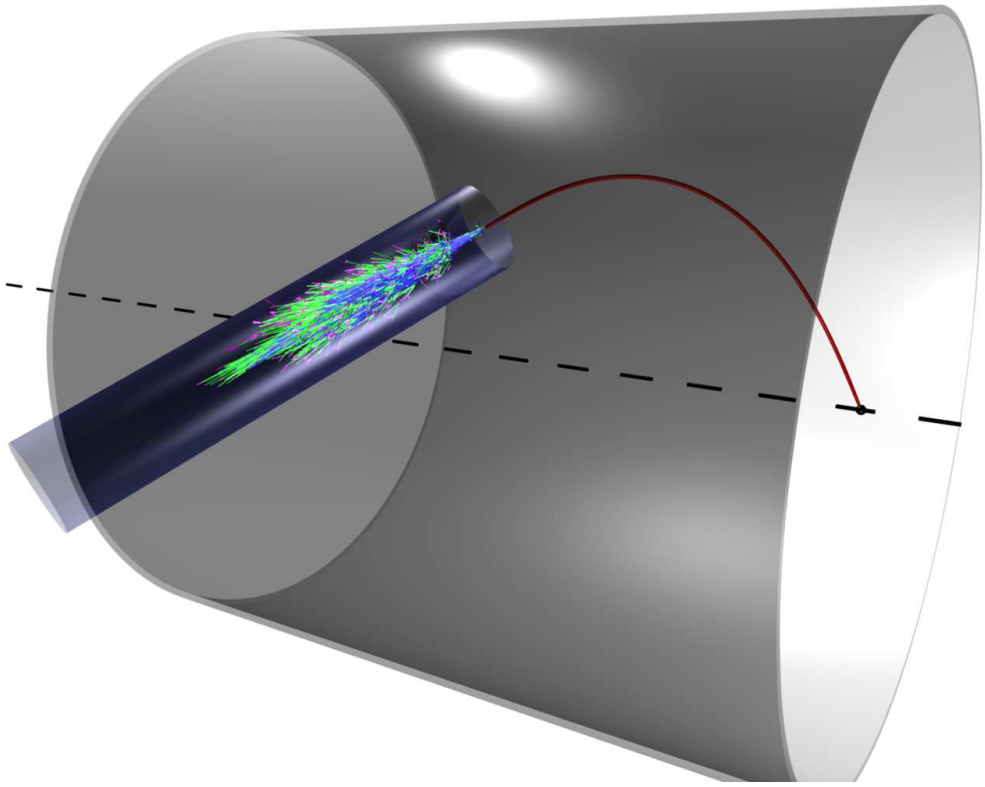

Figure 6.1: Illustration of the defined cylinder around the elongated track in the calorimeter. The dashed line represents the beam line, the solid line the track of the produced particle. The induced shower in the calorimeter is indicated in the small cylinder.

The direct lateral distribution of the shower is exploited for the estimator S_{rad}, a measure for the lateral shower distribution. For this a shower direction, defined by the track momentum vector at the impact point into the electromagnetic part of the calorimeter ($\vec{p}_{\text{atcalo}}$), is introduced. The perpendicular distance from this axis to a calorimeter cell ($dist$) is calculated (see figure 6.2). The distance is then weighted by the squareroot of the deposited energy in the corresponding cell ($\sqrt{E_{\text{cell}}}$). The distances to all the cells in a cylinder of radius $R = 30$ cm around the trajectory are summed up with the mentioned weights:

$$S_{\text{rad}} = \sum_{\text{cells}} \frac{\sqrt{E_{\text{cell}} \cdot dist^2}}{\sum_{\text{cells}} \sqrt{E_{\text{cell}}}} \ .$$

The distance is defined by the following expression:

$$dist = \frac{|\vec{r}_{\text{cell}} \times \vec{p}_{\text{at calo}}|}{|\vec{p}_{\text{at calo}}|}.$$

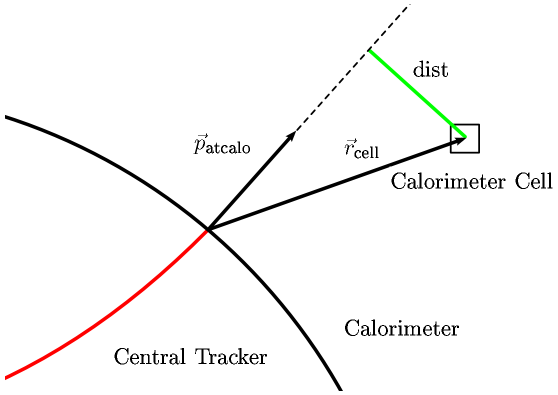

Figure 6.2: Illustration of the used distance for the calculation of the estimator S_{rad}. The perpendicular distance of the calorimeter cell to the momentum vector of the particle at the beginning of the calorimeter is denoted by *dist*.

The differences between electrons and hadrons in the longitudinal energy distribution in the calorimeter are used for another estimator. Electrons deposit a large amount of their energy close to the beginning of the calorimeter, as the bremsstrahlung starts immediately after entering the calorimeter. On the other hand, hadrons deposit the energy over a longitudinally more extended region and the electromagnetic sub-shower starts deeper inside the calorimeter.
Therefore the used measure for the shower length considers the position of the energy deposition along the path through the calorimeter and the amount of deposited energy. S_{len} consists of the sum over the calorimeter cells in a cylinder of radius $R = 30\,\text{cm}$ around the trajectory, where the layer number is summed up and weighted by the measured energy of the cell. The layers and the partitioning of the calorimeter is illustrated in figure 6.3.

$$S_{\text{len}} = \sum_{\text{cells}} \frac{Layer \cdot E_{\text{cell}}}{\sum_{\text{cells}} E_{\text{cell}}} \ .$$

This quantity defines the longitudinal shower centre which is also a measure for the shower length.

The estimator which uses the information from the Central Jet Chamber (CJC), describes the specific energy loss. The H1 software environment provides the Likelihood-value for the specific energy loss of a particle in material $\frac{\mathrm{d}E}{\mathrm{d}x}$, measured in the gas chamber (see section 4.3). This Likelihood-value is derived by calculating the difference between

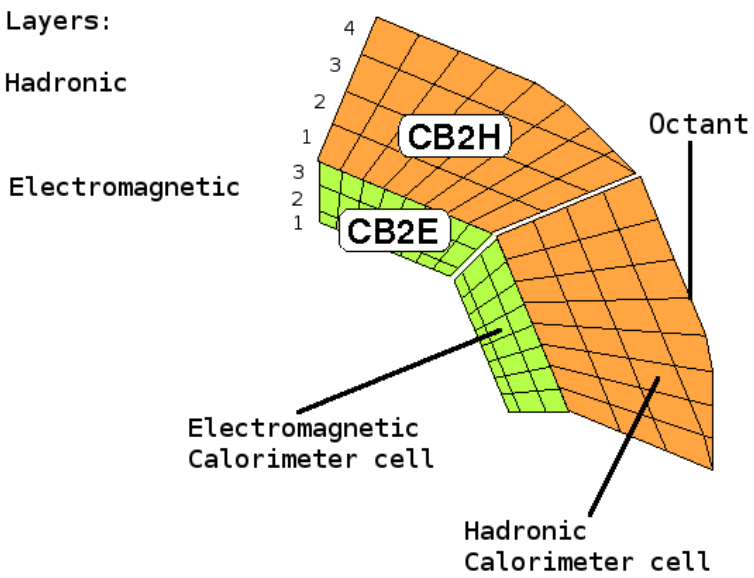

Figure 6.3: The layer and cell structure of the liquid argon calorimeter (r-ϕ view of the CB2 wheel). The layers are divided into electromagnetic and hadronic layers and numbered.

the expectation value for a specific particle and the actual measurement. This value follows a χ^2-distribution [17]. The particle probability Likelihood $L^{\mathrm{d}E/\mathrm{d}x}_{\mathrm{particle}}$ for electrons and pions is used to define the estimator. It is derived using the assumption that the particle is either an electron or a pion. This gives another separation quantity. For the discrimination the normalised Likelihood is defined by:

$$L^{\mathrm{d}E/\mathrm{d}x}_{\mathrm{norm}} = \frac{L^{\mathrm{d}E/\mathrm{d}x}_{\mathrm{electron}}}{L^{\mathrm{d}E/\mathrm{d}x}_{\mathrm{electron}} + L^{\mathrm{d}E/\mathrm{d}x}_{\mathrm{pion}}} \,.$$

The last two estimators are the total measured energy in a cylinder around the track of radius $R = 30$ cm in the electromagnetic and hadronic part of the liquid argon calorimeter, E_{elmag} respectively E_{had}.

In addition to the mentioned estimators, two more variables are given as input variables to the training of the multivariate analysis. Namely the transverse momentum p_{T} and the polar angle θ. This happens in order to take into account the energy dependences of the given estimators and to respect the differences of the calorimeter for different θ (different regions of the calorimeter, i.e. the barrels). This is further discussed in section 6.4 for the simulation and chapter 7 for the training process of the multivariate analysis.

To summarise, the presented electron identification is based on the following quantities:

- $\frac{E_{\mathrm{elmag}}(\mathrm{calo})}{p(\mathrm{track})}$
- $\frac{E_{\mathrm{inner}}(\mathrm{calo})}{E_{\mathrm{outer}}(\mathrm{calo})}$
- $S_{\mathrm{rad}} = \sum_{\mathrm{cells}} \frac{\sqrt{E_{\mathrm{cell}} \cdot dist^2}}{\sum_{\mathrm{cells}} \sqrt{E_{\mathrm{cell}}}}\,(\mathrm{calo})$
- $S_{\mathrm{len}} = \sum_{\mathrm{cells}} \frac{Layer \cdot E_{\mathrm{cell}}}{\sum_{\mathrm{cells}} E_{\mathrm{cell}}}\,(\mathrm{calo})$
- $\frac{\mathrm{d}E}{\mathrm{d}x}\,(\mathrm{track})$
- $E_{\mathrm{elmag}}\,(\mathrm{calo})$
- $E_{\mathrm{had}}\,(\mathrm{calo})$.

Where calo (track) refers to a measurement of the quantity in the calorimeter (tracker).

The separation power of these variables is then tested on electron and pion data samples. Moreover the estimators are compared to Monte Carlo simulation and then used as input variables for a multivariate analysis to determine a discriminator.

6.2 Test Samples

In this section the data selection used for testing the estimators and the training of the discriminator is presented. There are two possibilities to build test samples. The following paragraphs describe the selection of the measured data in order to get data samples. The discussion of the results achieved by these samples is presented in section 6.3. The second possibility is to use Monte Carlo samples. This allows to study the desired quantities in a simulated environment. These results are shown in section 6.4.

In order to qualify an estimator as a good separator, it has to be tested separately on a preferably clean sample for signal (electrons) and background (pions). These samples can be used to train the discriminator on signal and background too.
In the following two paragraphs the selection criteria for signal and background data is

described. Two well understood decays are chosen to select events containing electrons ($J/\psi \to e^+e^-$) and pions ($\rho \to \pi^+\pi^-$) respectively.

6.2.1 Electron Sample

As a source of isolated electrons the decay of elastic J/ψ is used ($J/\psi \to e^+e^-$, branching ratio (5.94 ± 0.06) % [11]).
The data used for this sample were taken in the years 2006 and 2007, when protons with energies of 920 GeV were brought to collision with electrons or positrons of 27.6 GeV.

Standard Selection

The event selection is done by requesting a diffractive J/ψ-candidate with two central tracks only. All triggers were used.
The electron candidates from the J/ψ decay are selected by the KALEP [2] finder under the restriction that the reconstructed invariant mass of the J/ψ lies between 1.3 GeV and 3.5 GeV.

This selection leads to the invariant J/ψ-mass peak shown in figure 6.4.

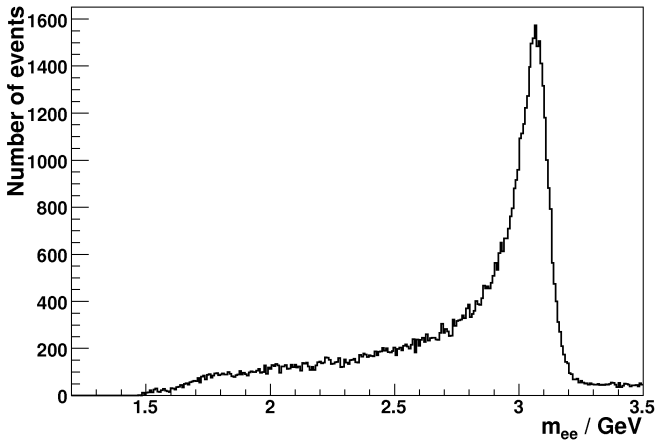

Figure 6.4: Distribution of the reconstructed invariant mass m_{ee} of the decay $J/\psi \to e^+e^-$ for the standard selection.

Since not only the efficiency of the KALEP electron finder is not so good but also its background rejection is always below 90 %, one expects that there are still other particles left in the sample, especially pions. For higher efficiencies (lower KALEP-electron quality) the background rejection drops considerably (see figure 2.3). Therefore the sample is

analysed with the help of $\frac{dE}{dx}$ to separate the electrons in the sample from other particles. The $\frac{dE}{dx}$-distribution in figure 6.5 shows the amount of pions in the sample. The fraction is about 23 %.

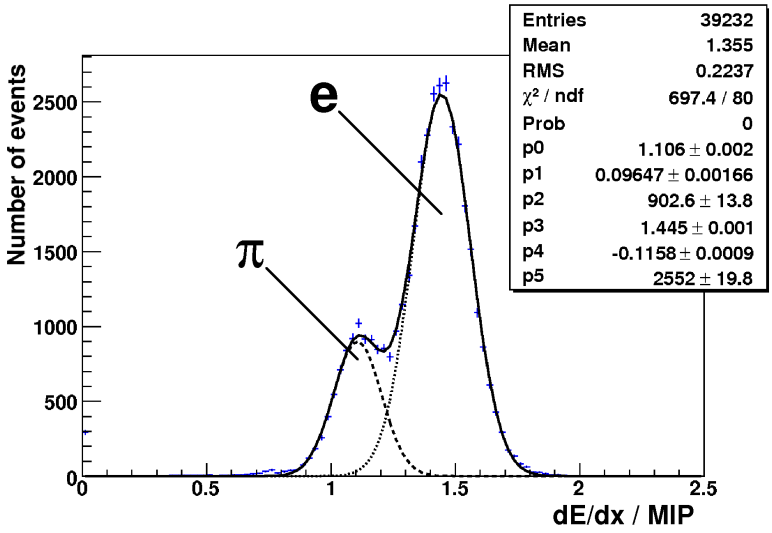

Figure 6.5: Output of the $\frac{dE}{dx}$-analysis of the J/ψ electron sample using the standard selection. The $\frac{dE}{dx}$ distribution is given in Minimum Ionising Particle (MIP). The data points are fitted to two gaussian functions for the electron and the pion fraction. The solid line represents the sum of the functions.

Improved Selection

In order to get an electron enriched sample and to reduce the bias introduced by the use of the KALEP finder, a new data sample is selected. The following criteria are applied:

- Number of diffractive J/ψ-candidates > 0
- Two central tracks
- Two tracks in total
- Reconstructed invariant mass: 1.3 GeV $< m_{ee} <$ 3.5 GeV
  The lower boundary of the invariant mass is chosen such that the radiative tail (see figure 6.4) towards lower energy is included in the selection.

- One track is identified by KALEP as an electron, the other is not probed by KALEP and to reduce the bias used for the sample. The identified track is not selected for the sample.

- Likelihood-value of $\frac{\mathrm{d}E}{\mathrm{d}x}$ under an electron assumption of both tracks is $L_{\mathrm{norm}}^{\mathrm{d}E/\mathrm{d}x} \geq 0.5$.

The cut on $\frac{\mathrm{d}E}{\mathrm{d}x}$ is useful to reduce the amount of pions. This can be illustrated by examining the effect of the $\frac{\mathrm{d}E}{\mathrm{d}x}$-cut on a data sample without a $\frac{\mathrm{d}E}{\mathrm{d}x}$-selection. The data selection applied is the same as specified for the improved selection except of the $\frac{\mathrm{d}E}{\mathrm{d}x}$-cut. In figure 6.6 on the left side the normalised $\frac{\mathrm{d}E}{\mathrm{d}x}$-electron Likelihood distribution, $L_{\mathrm{norm}}^{\mathrm{d}E/\mathrm{d}x}$, of tracks that are not probed with the KALEP finder is shown, whereas the other track is identified by KALEP as an electron. On the right hand side the same quantity is plotted but with the additional requirement that $L_{\mathrm{norm}}^{\mathrm{d}E/\mathrm{d}x}$ of the KALEP-identified track, thus the other track from the selected event, lies above 0.5.

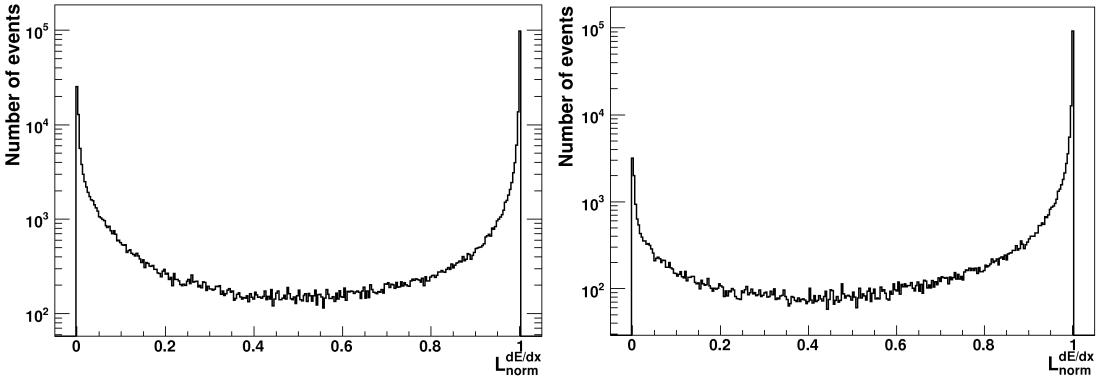

Figure 6.6: Normalised $\frac{\mathrm{d}E}{\mathrm{d}x}$-electron Likelihood distribution $L_{\mathrm{norm}}^{\mathrm{d}E/\mathrm{d}x}$ of the not KALEP-probed track without (left) and with (right) cut on $L_{\mathrm{norm}}^{\mathrm{d}E/\mathrm{d}x}$ of the KALEP-identified track.

Figure 6.6 (left plot) again shows the contamination of the electron sample. The cut on the normalised $\frac{\mathrm{d}E}{\mathrm{d}x}$-electron Likelihood ($L_{\mathrm{norm}}^{\mathrm{d}E/\mathrm{d}x} > 0.5$) on the KALEP-identified track (right plot) reduces the background approximately by a factor of three, where every event with $L_{\mathrm{norm}}^{\mathrm{d}E/\mathrm{d}x} < 0.5$ is defined as background. The reduction in the number of background-like events is clearly visible and therefore shows the usefullness of this cut. Under the assumption that the background can be reduced by a factor of three, the cut on both selected tracks leads to a remaining background in the electron sample of about 5 %.

This revised event selection leads to a rather pure electron sample, which is used for further tests. The invariant mass peak for this sample is shown in figure 6.7.

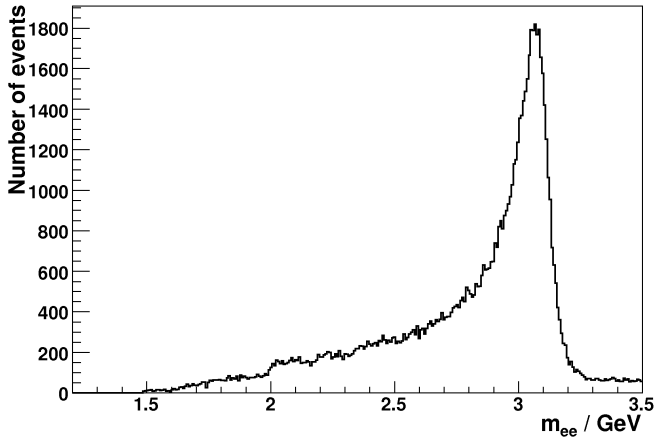

Figure 6.7: Distribution of the reconstructed invariant mass m_{ee} of the decay $J/\psi \rightarrow e^+e^-$ for the improved selection.

6.2.2 Pion Sample

A data sample of pions is used to represent the background in an event. It provides a basis to explore the separation power of the chosen estimators and it is needed for the multivariate analysis to train the discriminator.
The selected process for pion production is the decay of a ρ meson ($\rho \rightarrow \pi^+\pi^-$, branching ratio ~ 100 % [11]). The used data is from the same period as for the electron sample, the years 2006 and 2007.

The event selection is done by requesting a diffractive light vectormeson candidate, two central tracks and no other tracks. The distance from the z-vertex position to the nominal interaction point has to be smaller than 25 cm. Both tracks have to come from the primary vertex, the lower p_t pion should have $p_\mathrm{t}(\mathrm{track}) > 0.12$ GeV and the higher p_t pion a transverse momentum larger than 0.7 GeV. In addition, two more cuts are applied, an angle restriction and a boundary for the mass. The polar angle should be in the interval $20° < \theta < 160°$ whereas the reconstructed invariant mass is limited by $0.6\ \mathrm{GeV} < m_{\pi\pi} < 1.1\ \mathrm{GeV}$.

This selection gives the invariant mass distribution depicted in figure 6.8.

After the promising outcome of the investigation of the electron sample with the help of $\mathrm{d}E/\mathrm{d}x$, the described pion sample is also analysed. In figure 6.9 the output of the analysis is shown. The left side of the figure shows the distribution for the presented data selection, the right side after the optimisation with the help of $\mathrm{d}E/\mathrm{d}x$.

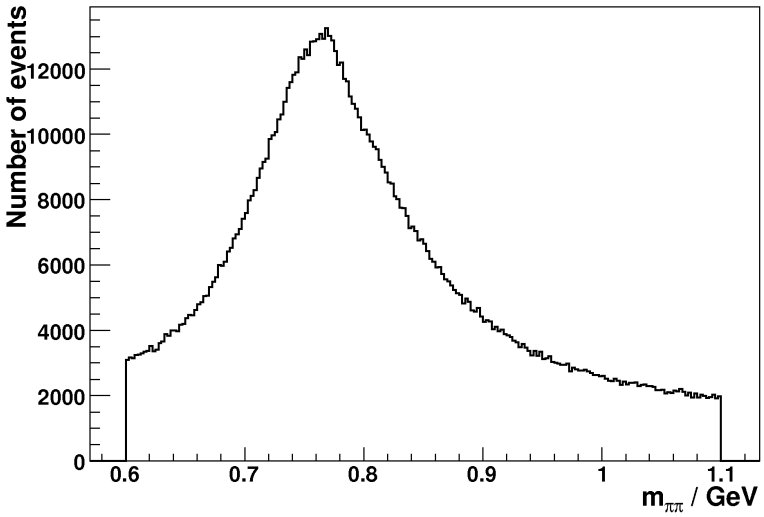

Figure 6.8: Reconstructed invariant mass peak, $m_{\pi\pi}$, for the presented event selection.

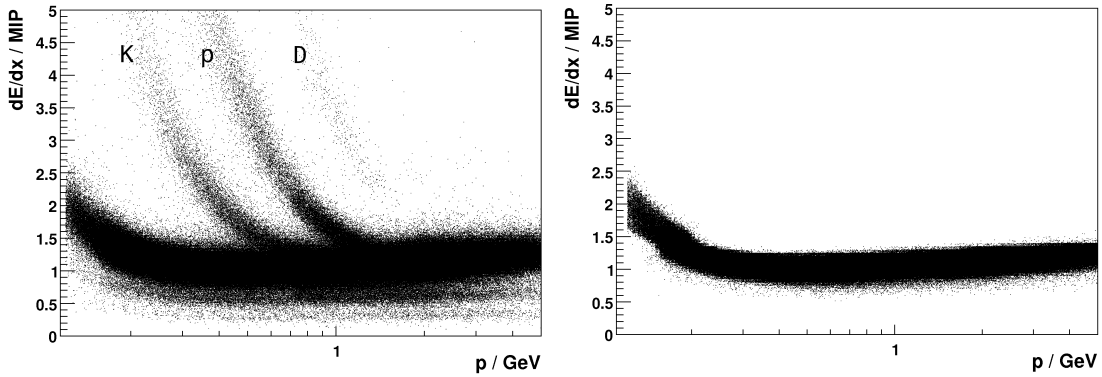

Figure 6.9: Scatter plot of the $\mathrm{d}E/\mathrm{d}x$ distribution for the pion hypothesis versus the track momentum. Left: presented event selection, right: after the applied cuts.

The scatter plot of $\mathrm{d}E/\mathrm{d}x$ under the assumption of a pion versus the momentum of the track, clearly shows the contamination of the sample, mainly by other hadrons.
In order to enrich the pion sample, three cuts are applied simultaneously: on $L_{pion}^{\mathrm{d}E/\mathrm{d}x}$, $L_{norm}^{\mathrm{d}E/\mathrm{d}x}$ with an electron-pion assumption, as in the case of the electron sample, and on the timing information of the event from the CJC. The distribution of the $\mathrm{d}E/\mathrm{d}x$-Likelihood value of pions before and after the applied cuts is shown in figure 6.10. The cut on this quantity is definded by $L_{pion}^{\mathrm{d}E/\mathrm{d}x} > 0.05$.

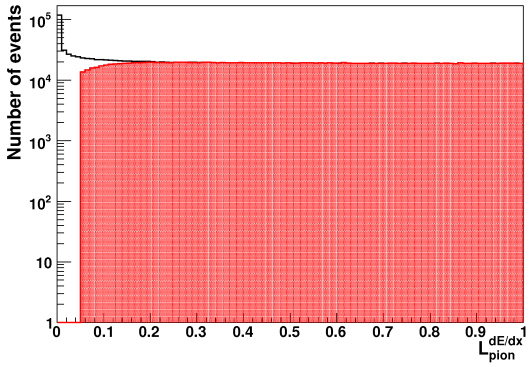

Figure 6.10: $L_{pion}^{\mathrm{d}E/\mathrm{d}x}$ distribution of the pion sample. The shaded histogram shows the distribution after all applied cuts.

Although this is a loose cut, it reduces the contamination considerably.

The normalised $\mathrm{d}E/\mathrm{d}x$-Likelihood using the electron and the pion hypothesis is used to reduce the amount of electrons in the pion sample. This quantity is defined in the same way as for the electron sample in section 6.2.1. The distribution is presented in figure 6.11.

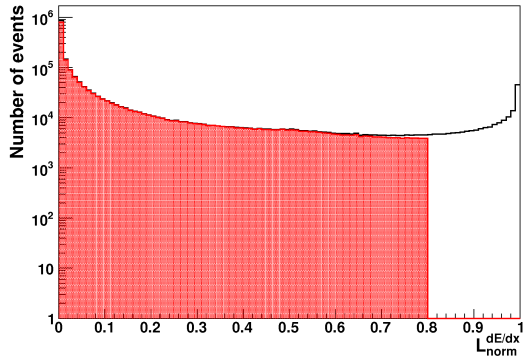

Figure 6.11: $L_{norm}^{\mathrm{d}E/\mathrm{d}x}$ distribution of the pion sample. The shaded histogram shows the distribution after all applied cuts.

The distribution of the normalised Likelihood clearly shows that the pion data sample contains a noticeable amount of electrons. For the new data selection a cut at $L_{norm}^{\mathrm{d}E/\mathrm{d}x} = 0.8$ is applied to reduce the fraction of the electrons.

In order to improve the background rejection the timing information of the gas chamber is used. The measured distribution and the applied cuts are shown in figure 6.12.

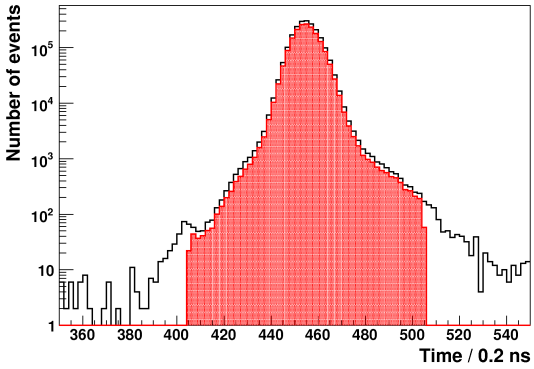

Figure 6.12: Timing information of the events used for the pion sample. The shaded histogram shows the distribution after all applied cuts.

The timing information helps to select the information which actually belongs to the observed event. The used cut on this quantity is defined by $405 \cdot 0.2\ \text{ns} < t < 505 \cdot 0.2\ \text{ns}$, where the time zero point is not defined by the *ep*-interaction.

For further tests this new event selection is used as the background (pion) sample. The invariant mass peak for this sample is shown in figure 6.13.

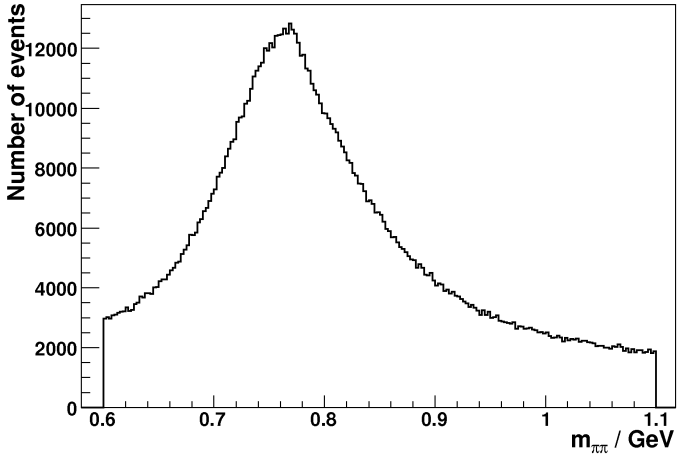

Figure 6.13: Reconstructed invariant mass distribution, $m_{\pi\pi}$, for the new event selection after the usage of the information of $\mathrm{d}E/\mathrm{d}x$.

With these two samples it is now possible to examine the separation power of the chosen estimators and further on to train the discriminator using a method of a multivariate analysis.

The signal sample consists of the decay electrons of about 80'000 J/ψ vector mesons whereas the background sample contains pions from approximately 1'100'000 ρ mesons.

6.3 Comparison of Estimators for Signal and Background

In this section a comparison of the estimators for the electron and the pion sample is presented. The estimators are analysed concerning the separation power between the two samples. This results give a first indication of the potential of the electron finder using the quantities introduced in section 6.1.

The tests of the estimators are divided into intervals regarding the momentum and the polar angle of the track. This allows to study the separation double differentially in the momentum and the polar angle. The polar angle θ is divided into three intervals according to the regions of the calorimeter: central barrel (CB), forward barrel (FB) and inner forward (IF). This is done by identifying the wheel of the calorimeter cell which measured an energy deposition. The segmentation of the liquid argon calorimeter into wheels is illustrated in figure 3.4 in section 3.2.2.

As the electron and the pion samples do not contain the same number of events, every histogram is normalised to the number of electron events in the corresponding interval.

In order to study the separation power of the different estimators in the observed distributions for electrons and pions the polar angle θ and the transversal momentum p_t are reweighted. This reweighting is necessary to exclude effects from different distributions in the kinematical variables.
In figures 6.14 and 6.15 the results of the reweighting for θ and p_t are illustrated.

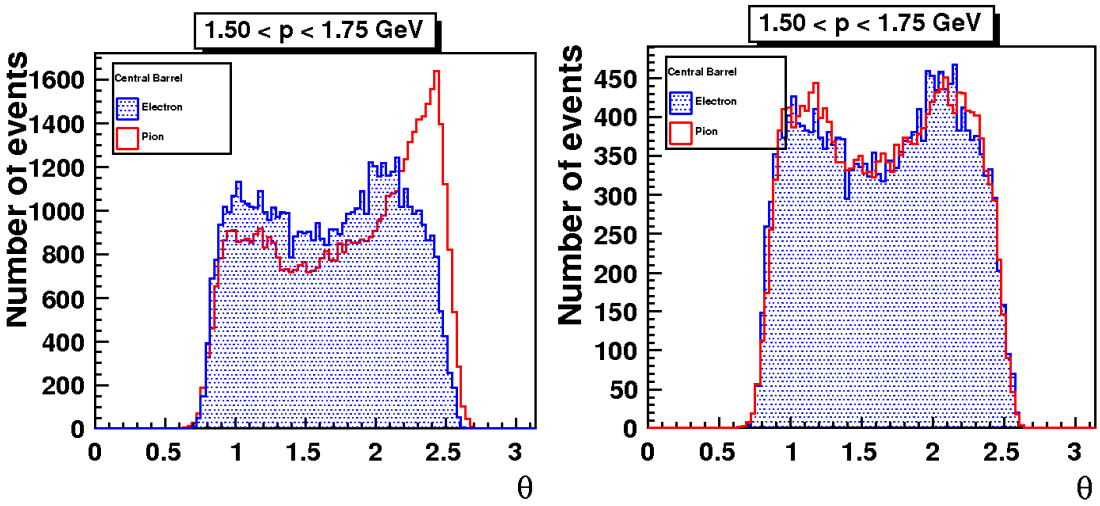

Figure 6.14: Distribution of the polar angle θ in the central barrel, 1.5 GeV $< p <$ 1.75 GeV. Left: generated, right: reweighted.

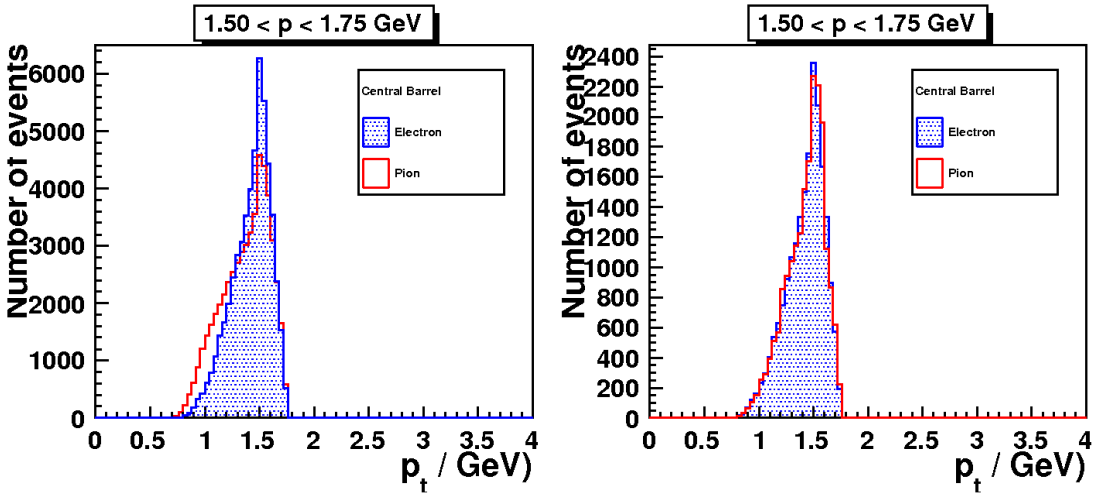

Figure 6.15: Distribution of the transversal momentum in the central barrel, 1.5 GeV $< p <$ 1.75 GeV. Left: generated, right: reweighted.

In the following the results for the first estimator, the ratio of energy to momentum, is presented.
Electrons usually deposit all their energy in the electromagnetic calorimeter leading to a value of $\frac{E_{\text{elmag}}(\text{calo})}{p(\text{track})} \approx 1$. The value for pions is lower as they do not deposit all the energy in the electromagnetic part of the calorimeter.

Figure 6.16 shows the distribution of $\frac{E_{\text{elmag}}}{p}$ for two different radii. The result for the electrons is shown, as in all the following figures, in the shaded histogram. The left side shows that the smaller cylinder ($R_i = 15$ cm) is too small to fully contain the hadronic shower in lateral direction. In case of the wider cylinder ($R_o = 30$ cm), a higher fraction of the hadronic shower is contained as the number of zero-entries is smaller. The chosen momentum interval is defined by 1.5 GeV $< p <$ 1.75 GeV and the region of the calorimeter is the forward barrel.
Figure 6.17 shows the same quantity for electrons and pions in a different momentum interval (3.25 GeV $< p <$ 3.5 GeV) and another detector region (central barrel). This momentum interval is the upper boundary of the examined momentum spectrum in this thesis, as the statistics of pions originating from ρ decays in higher momentum regimes is too low for a comparison.

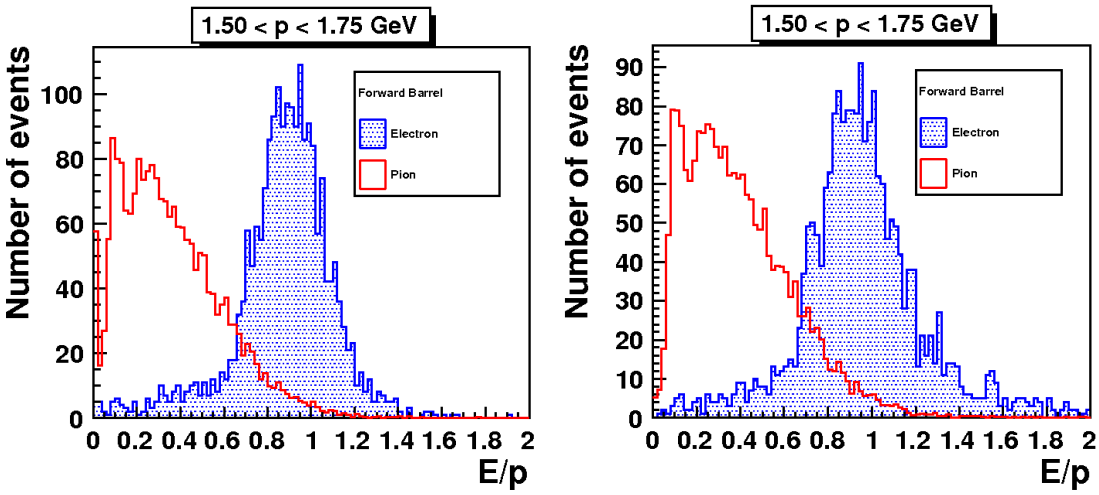

Figure 6.16: E/p distribution of electrons and pions in the forward barrel for two different cylinder radii, left: $R_i = 15$ cm, right: $R_o = 30$ cm. Momentum interval: 1.5 GeV $< p <$ 1.75 GeV.

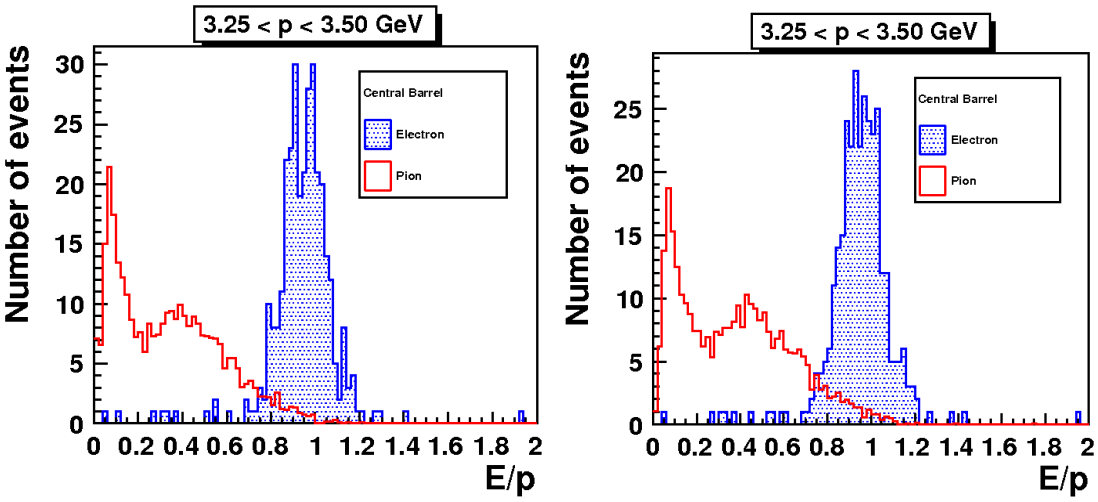

Figure 6.17: E/p distribution of electrons and pions in the central barrel for two different cylinder radii, left: $R_i = 15$ cm, right: $R_o = 30$ cm. Momentum interval: 3.25 GeV $< p <$ 3.5 GeV.

The figures show the already mentioned behaviour of electrons, that they usually deposit all their energy in the electromagnetic part of the liquid argon calorimeter. The peak around $E/p \approx 1$ is clearly visible. For higher momentum electrons, the fraction of electrons with a value of E/p obviously below 1, is even smaller. This could be explained by the smaller fraction of energy lost in dead material for higher energetic electrons.
The fraction of the deposited energy is in case of pions definitely smaller, as the examined momentum interval for pions and electrons is the same. The peak value of E/p is clearly shifted towards smaller values compared to the electrons.
Especially for higher momenta (figure 6.17) the overlap of the curves is not big. A direct cut on this quantity allows a good separation between electrons and pions. It is possible to reject a reasonable amount of pions while most electrons are kept. Therefore the separation power of the estimator E/p is good.

The isolation criterion $I = \frac{E_{\text{inner}}}{E_{\text{outer}}}$ is shown in figure 6.18 for the central barrel in the momentum interval 2.25 GeV $< p <$ 2.5 GeV.

Pions tend to have slightly lower values for I than electrons due to the wider lateral distribution of the shower in the calorimeter. But both distributions clearly peak at $I = 1$. Compared to the result of E/p, the separation is not as pronounced.

In case of S_{rad}, the measure for the shower radius (see figure 6.19), the separation power depends on the energy of the particles. For energies of about 1 GeV S_{rad} looks similar for electrons and pions. For higher energies, about 3 GeV, the measure for the shower radius of pions is shifted towards higher values. For electrons, the curve does not

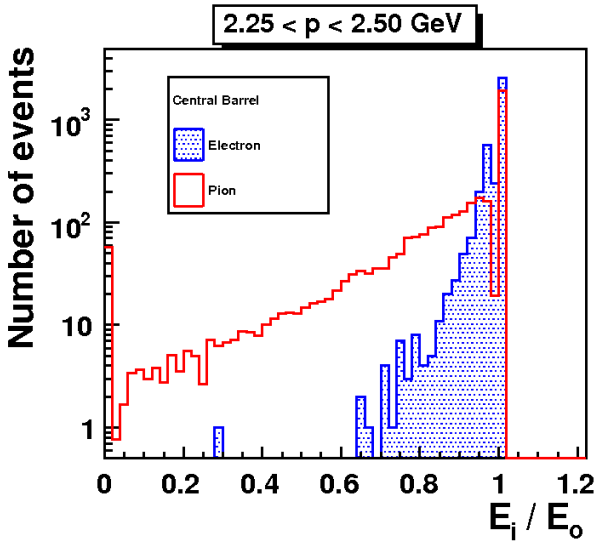

Figure 6.18: $I = \frac{E_{\text{inner}}}{E_{\text{outer}}}$ for electrons and pions in the central barrel, 2.25 GeV $< p <$ 2.5 GeV.

change significantly. In this energy sector the separation is visible.

The estimator describing the longitudinal shower distribution shows a completely different behaviour between electrons and pions. As pions usually deposit only a small fraction of their energy in the electromagnetic part of the calorimeter, this deposition is often measured in a single layer. In contrast to electrons, pions do not necessarily deposit their energy at the beginning of the calorimeter. Therefore the signal can be measured on any layer. This behaviour leads, according to the definition of the measure for the shower length, to spikes in the distribution of S_{len} for pions as can be seen in figure 6.20. The weighting of the layer by the energy is canceled, which gives an integer number for S_{len}. Both depicted distributions are shown for the central barrel but for different momentum intervals.

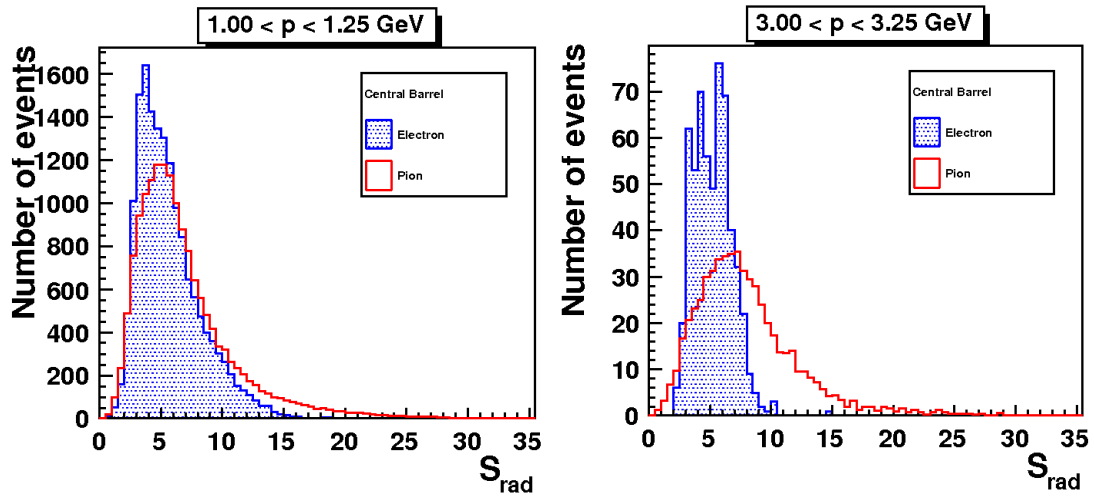

Figure 6.19: S_{rad} for electrons and pions in the central barrel, left: 1.0 GeV < p < 1.25 GeV, right: 3.0 GeV < p < 3.25 GeV.

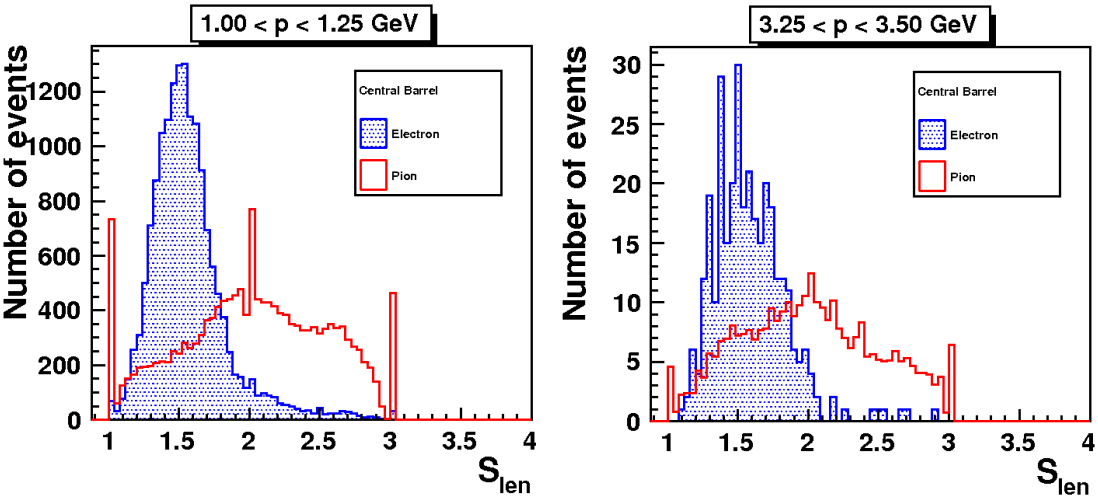

Figure 6.20: S_{len} for electrons and pions in the central barrel, left: 1.0 GeV < p < 1.25 GeV, right: 3.25 GeV < p < 3.5 GeV.

As the energy of the pions increases, the trend to build spikes decreases (right hand side). For electrons in contrast, the deposition starts usually on the first layer and is mainly contained in the first two layers. This leads to a narrower peak for S_{len} of electrons. In general the distribution gets wider for higher energetic electrons. The shown distributions do not contain values above $S_{\text{len}} = 3$, as the electromagnetic part of the liquid argon calorimeter consists of only three layers in the central and forward barrel.
For this estimator a clear separation between electrons and pions is visible towards higher values of S_{len}, especially for larger momenta.

The last two used quantities are the measured energy in the electromagnetic and the hadronic part of the liquid argon calorimeter in a cylinder ($R_o = 30$ cm) around the selected tracks for electrons and pions. The energy-dependence of the separation is shown in figure 6.21.
In case of the electromagnetic energy, the separation between electrons and pions depends on the examined energy interval. The separation gets clearly better with increasing energy. The distribution of the deposited energy for pions gets wider for higher energies. But the fraction of pions where more than 75 % (approximately rise of the peak for electrons with $p \geq 2$ GeV) of the energy of the track is deposited in the electromagnetic part of the calorimeter, is very small.
As the amount of deposited energy in case of electrons is nearly 100 %, the peakvalue of the measured energy is directly proportional to the track momentum. This different behaviour of electrons and pions leads to the visible separation for higher momenta.

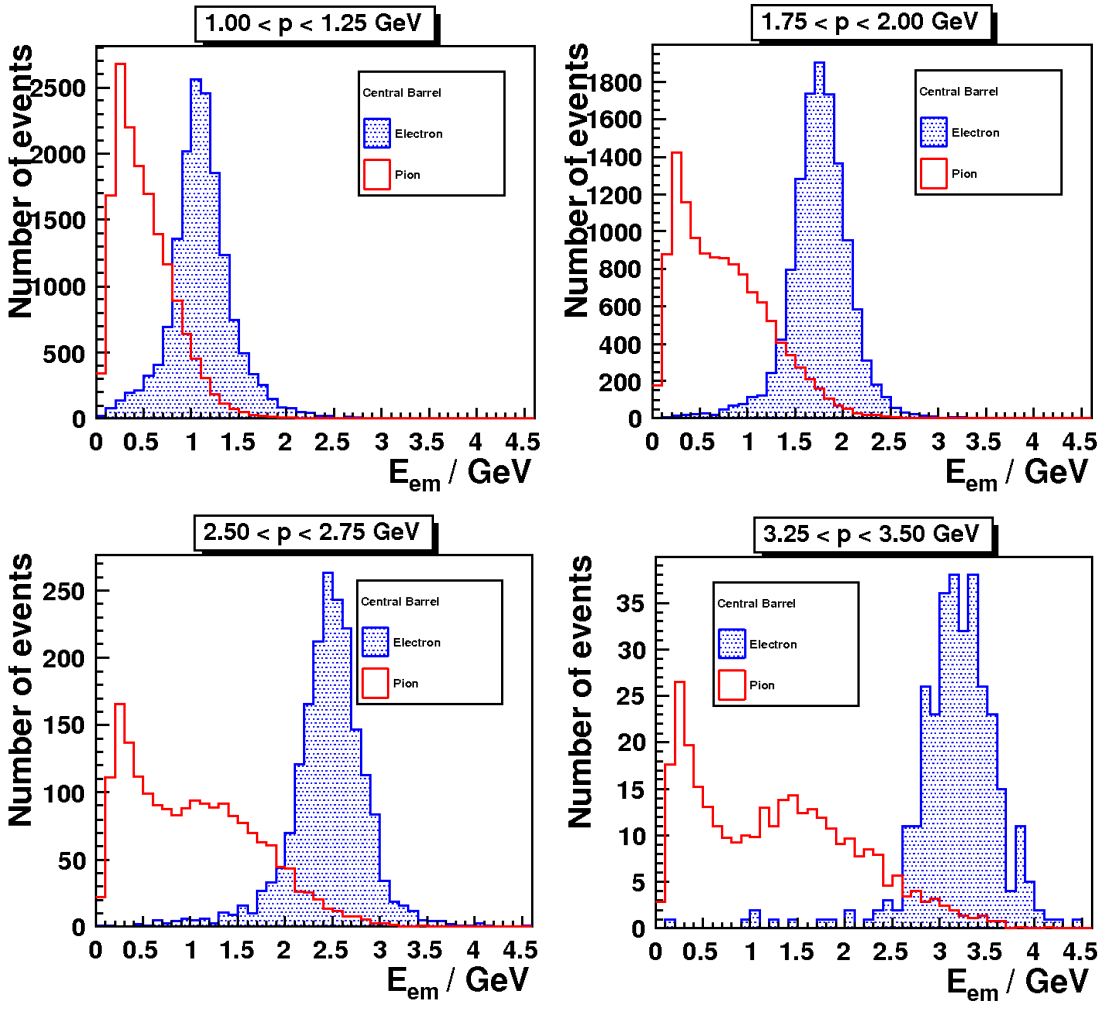

Figure 6.21: Measured electromagnetic energy of electrons and pions in a cylinder around the track in the central barrel, top left: 1.0 GeV $< p <$ 1.25 GeV, top right: 1.75 GeV $< p <$ 2.0 GeV, lower left: 2.5 GeV $< p <$ 2.75 GeV, lower right: 3.25 GeV $< p <$ 3.5 GeV.

The same behaviour can be observed in the forward barrel of the calorimeter and is depicted in figure 6.22. The separation is not as good as for the central barrel, but still visible.

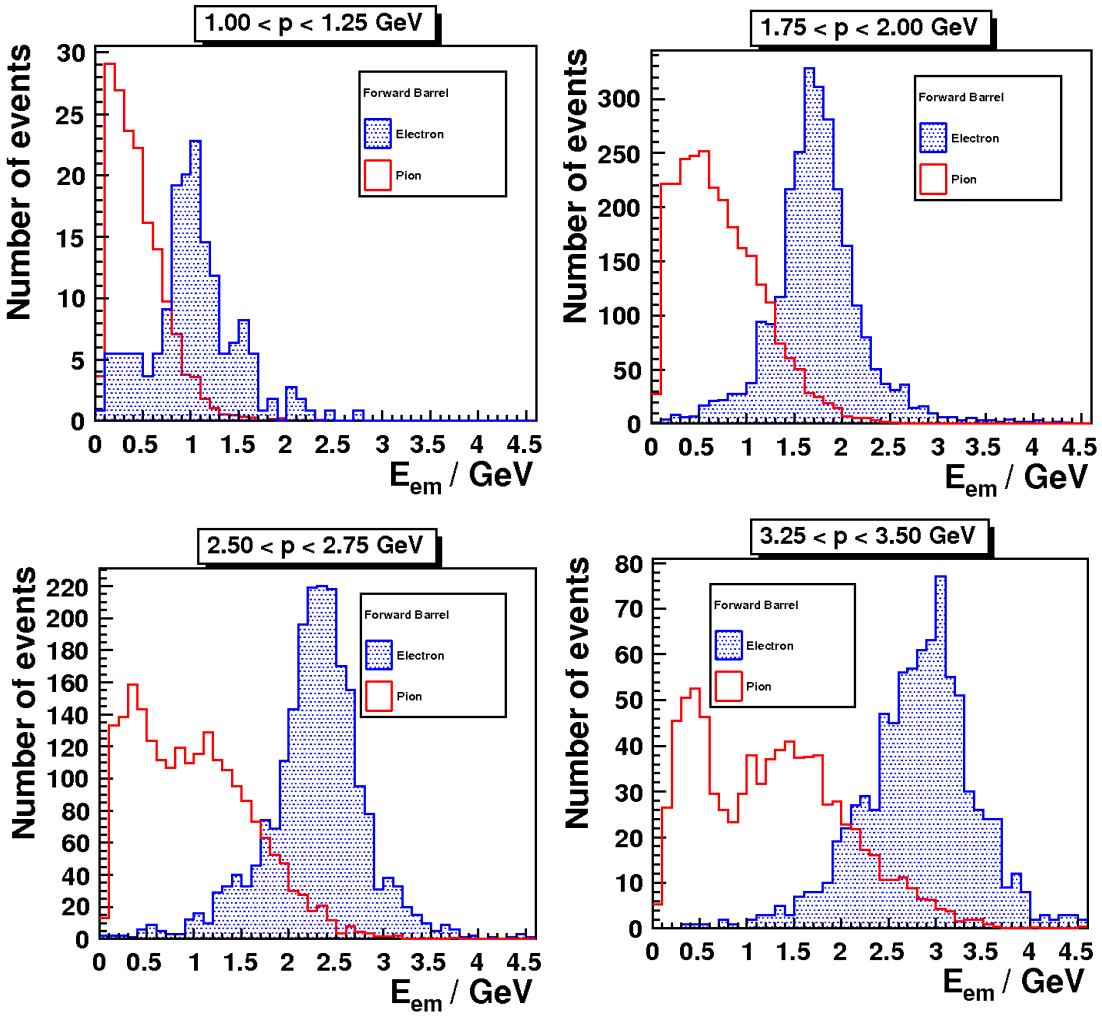

Figure 6.22: Measured electromagnetic energy of electrons and pions in a cylinder around the track in the forward barrel, top left: 1.0 GeV $< p <$ 1.25 GeV, top right: 1.75 GeV $< p <$ 2.0 GeV, lower left: 2.5 GeV $< p <$ 2.75 GeV, lower right: 3.25 GeV $< p <$ 3.5 GeV.

The measurement of the deposited energy in the hadronic part of the calorimeter shows, that electrons very rarely reach the hadronic calorimeter in the studied energy regime. Therefore no energy is measured in general.
The pions do indeed reach this part of the detector and deposit a measurable amount of

energy in the hadronic cells. This leads to a clear difference in the distribution of the measured energy and is depicted in figures 6.23 and 6.24 for the central and the forward barrel respectively. Two different momentum intervals are shown in each figure.

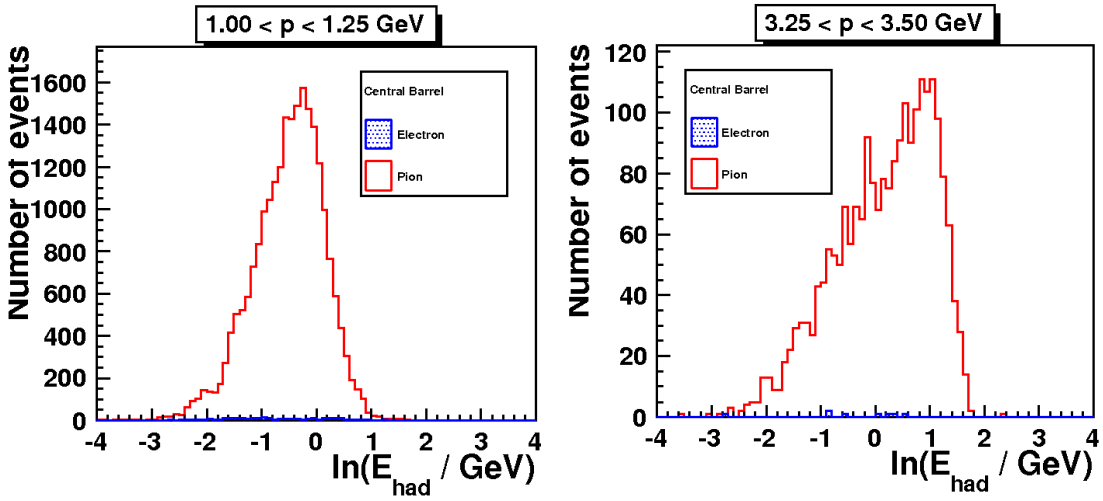

Figure 6.23: Measured hadronic energy of electrons and pions in a cylinder around the track in the central barrel, left: 1.0 GeV $< p <$ 1.25 GeV, right: 3.25 GeV $< p <$ 3.5 GeV.

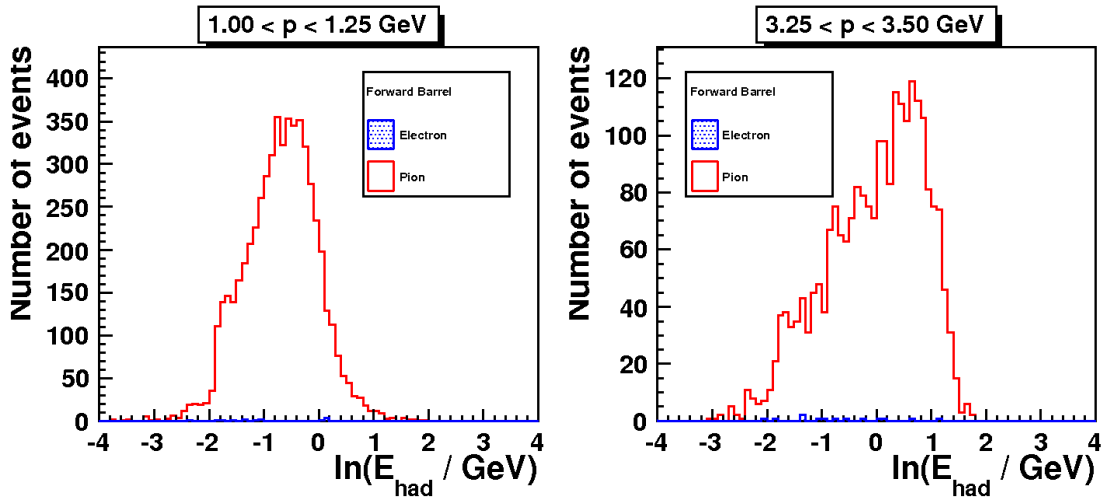

Figure 6.24: Measured hadronic energy of electrons and pions in a cylinder around the track in the forward barrel, left: 1.0 GeV $< p <$ 1.25 GeV, right: 3.25 GeV $< p <$ 3.5 GeV.

6.4 Comparison of Estimators in Data and Monte Carlo Simulation

In this section the description of the estimators using simulated data is presented. The same quantities, discussed in the previous section 6.3, are studied for Monte Carlo simulation and compared to the data samples for electrons and pions. The data selection criteria described in section 6.2 are also applied to the MC samples.

For this work the most relevant part is the energy distribution of electromagnetic and hadronic calorimeter showers, since $\mathrm{d}E/\mathrm{d}x$ is not calibrated for 2006/07 and therefore not usable in Monte Carlo simulation.

In order to compare the resulting distributions for the estimators, a J/ψ Monte Carlo simulation is chosen for the generation of the Monte Carlo signal (electron) sample. A ρ Monte Carlo simulation is suitable for the background (pion) sample. The available ρ Monte Carlo sample did not contain enough high momentum pions to allow for a meaningful comparison to data. Consequently, a sample of inline generated single pions was simulated. The momentum interval was adapted to include enough high momentum pions for a comparison.

The generators of Monte Carlo simulation often produce distributions of the kinematic variables which differ from the distributions observed in the data, especially for single particle Monte Carlo. To get the best achievable description of the actually observed distributions, the variables of the Monte Carlo events are reweighted to the data. In this study the reweighting is done in the polar angle θ and the transversal momentum p_t. As described in the previous section 6.3 the distributions of θ and p_t for the pion data sample are reweighted to them of the electrons. The distributions of the kinematic quantities for the Monte Carlo samples are as well adapted to those of the electrons in the data sample. This assures that all the samples have equal distributions of the kinematic quantities.
In figures 6.25 and 6.26 the result of the reweighting for θ and p_t is illustrated. As an example, the central barrel and the momentum interval $1.5\,\mathrm{GeV} < p < 1.75\,\mathrm{GeV}$ are chosen. The distribution is shown for pions (data and inline generated pions). Since for electrons the difference between data events and events simulated by DiffVM is not very pronounced, no reweighting is needed.

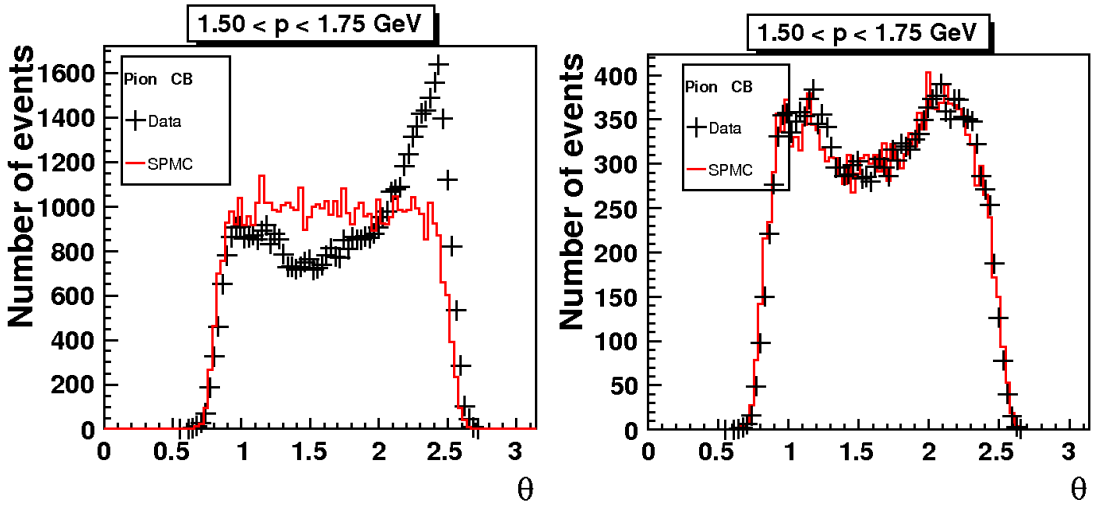

Figure 6.25: Distribution of the polar angle θ in the central barrel, 1.5 GeV $< p <$ 1.75 GeV. Left: generated, right: reweighted.

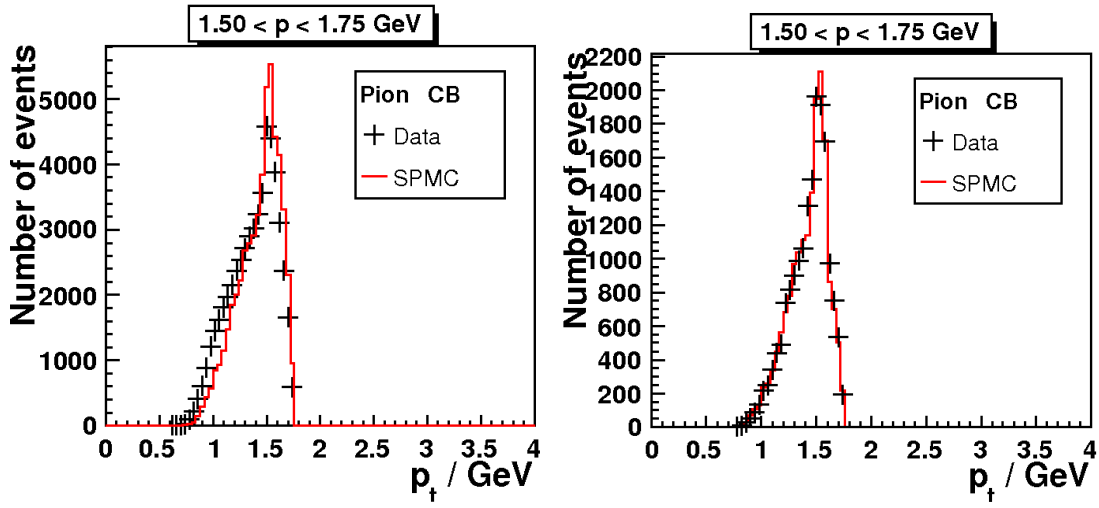

Figure 6.26: Distribution of the transversal momentum in the central barrel, 1.5 GeV $< p <$ 1.75 GeV. Left: generated, right: reweighted.

The figures of the comparison between data and Monte Carlo are presented and discussed in the following paragraphs.

The description by simulation of the first estimator, introduced as the measured electromagnetic energy divided by the track momentum $\frac{E_{\text{elmag}}}{p}$, is shown in figures 6.27 and 6.28 for two different momentum intervals in the central barrel of the detector. The figure shows on the left hand side the result for electrons and on the right hand side that for pions. The chosen momentum intervals are defined by 1.5 GeV $< p <$ 1.75 GeV and 2.5 GeV $< p <$ 2.75 GeV respectively.
The description of signal and background is good, as the shapes of the curves are nearly identical.
In the forward barrel the distribution of $\frac{E_{\text{elmag}}}{p}$ is slightly different for data and simulation, as can be seen in figure 6.29 for 2.5 GeV $< p <$ 2.75 GeV. The Monte Carlo prediction is, especially in case of electrons, shifted towards higher values of E/p. A possible explanation for this could be that in the forward barrel of the detector, the amount of passed dead material is bigger than in the central barrel. Probably the amount of dead material included in the simulation is too low. Therefore the reconstruction process allocates a higher energy to the simulated particles compared to real particles. The length scale for the energy loss of electrons is the radiation length, whereas that of pions is the interaction length. As the radiation length of electrons is much smaller than the interaction length of pions, the effect is more pronounced for electrons than for pions.

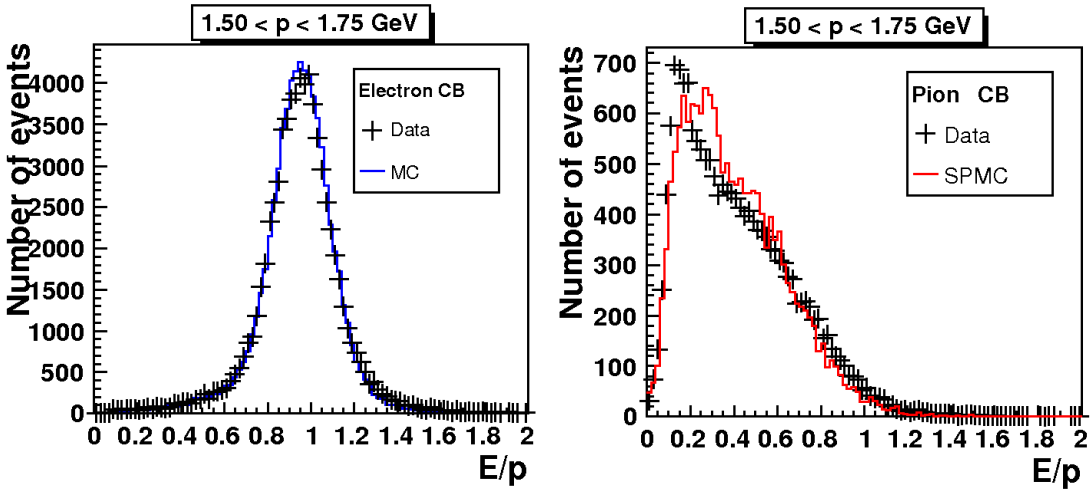

Figure 6.27: Comparison of $\frac{E_{\text{elmag}}}{p}$ in a cylinder around the track with radius $R =$ 30 cm in the central barrel. The plot on the left side shows the distribution of electrons, that on the right side that of pions. The chosen momentum interval is 1.5 GeV $< p <$ 1.75 GeV.

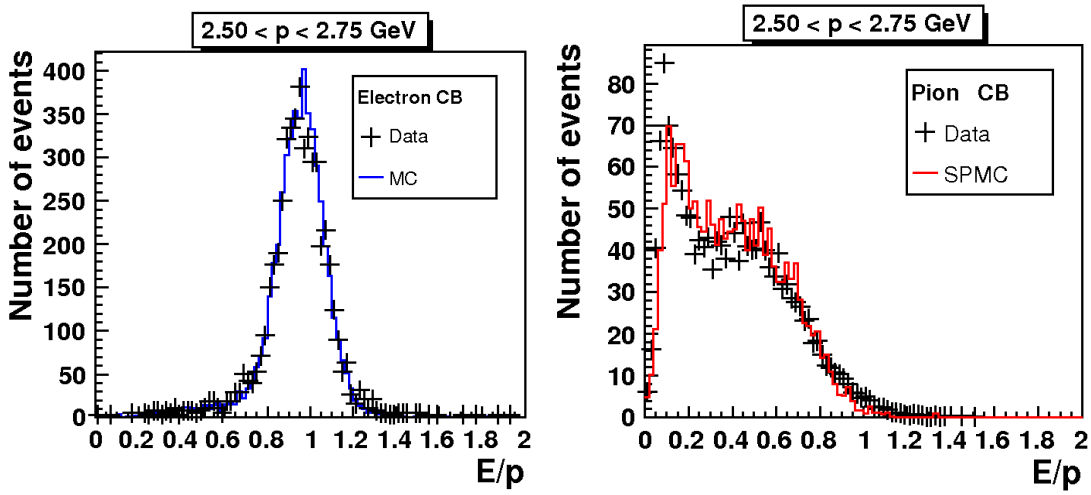

Figure 6.28: Comparison of $\frac{E_{\mathrm{elmag}}}{p}$ in a cylinder around the track with radius $R = 30$ cm in the central barrel. The plot on the left side shows the distribution of electrons, that on the right side that of pions. The chosen momentum interval is 2.5 GeV $< p <$ 2.75 GeV.

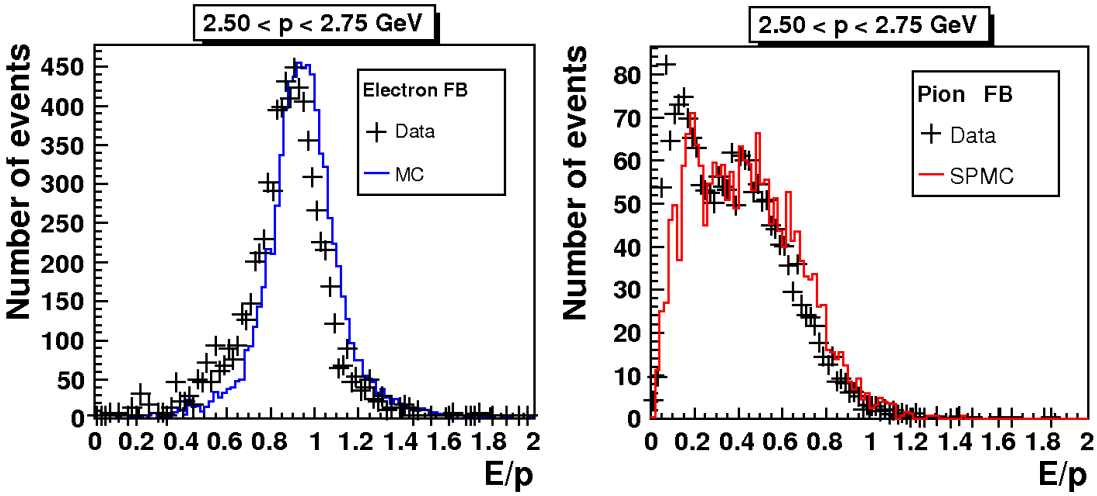

Figure 6.29: Comparison of $\frac{E_{\mathrm{elmag}}}{p}$ between data and Monte Carlo in the forward barrel for electrons and pions, 2.5 GeV $< p <$ 2.75 GeV.

The comparison between simulation and data for $I = \frac{E_{\text{inner}}}{E_{\text{outer}}}$ shows a good description as can be seen in figures 6.30 and 6.31 for two different momentum intervals. The shown detector region is the central barrel. The prediction for the proportion of the energy in the smaller cylinder to the energy in the wider cylinder for electrons and pions is good. The distributions of $I = \frac{E_{\text{inner}}}{E_{\text{outer}}}$ for electrons and pions in the forward barrel is presented in figure 6.32. The particles have a momentum of about $p \approx 2$ GeV.

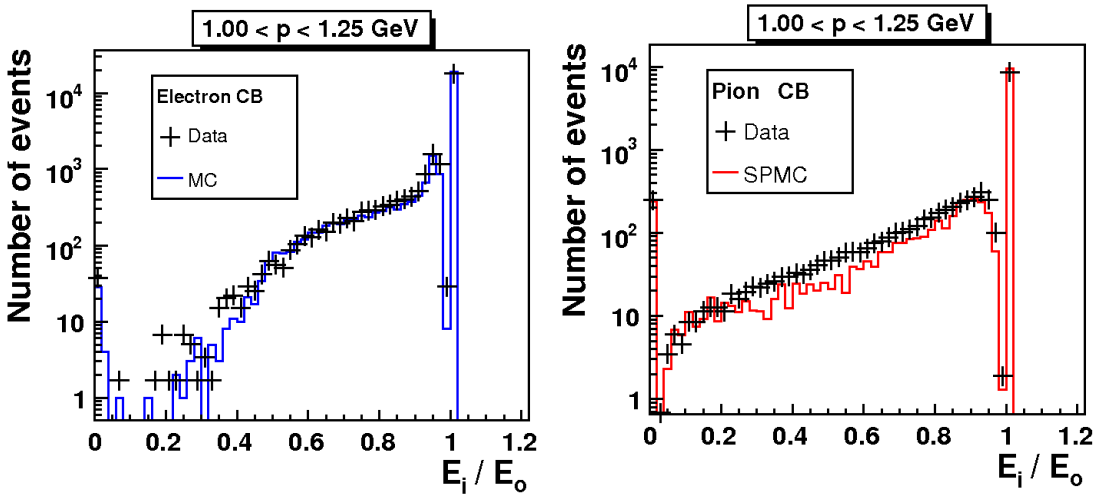

Figure 6.30: Data-MC comparison of $I = \frac{E_{\text{inner}}}{E_{\text{outer}}}$ in the central barrel for electrons and pions, 1.0 GeV $< p <$ 1.25 GeV.

The peak at $I = 1$ in the data distribution appears also in the simulation and is pronounced similarly. The agreement below $I = 1$ is also very good.
Therefore the description of this estimator by simulation for the electron and pion samples is correct.

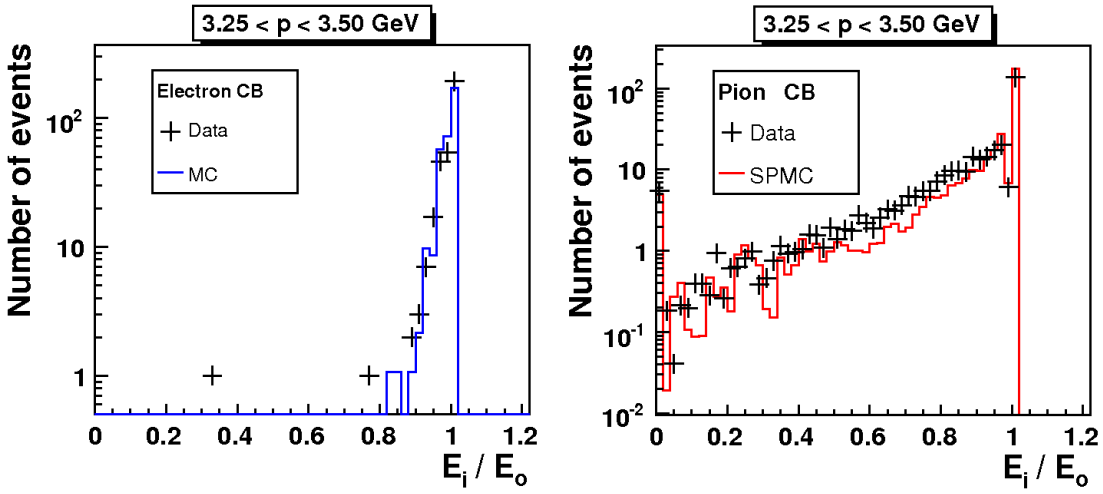

Figure 6.31: Data-MC comparison of $I = \frac{E_{\text{inner}}}{E_{\text{outer}}}$ in the central barrel for electrons and pions, 3.25 GeV $< p <$ 3.5 GeV.

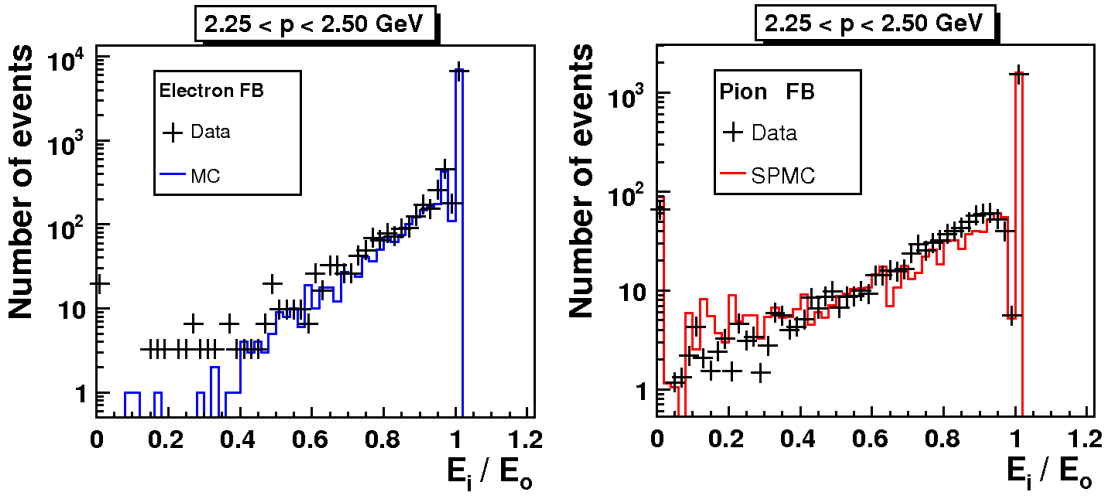

Figure 6.32: Data-MC comparison of $I = \frac{E_{\text{inner}}}{E_{\text{outer}}}$ in the forward barrel for electrons and pions, 2.25 GeV $< p <$ 2.5 GeV.

The comparison between data and simulation of the measure for the shower radius is shown in figure 6.33 for lowest momenta (1.0 GeV $< p <$ 1.25 GeV) electrons and pions.

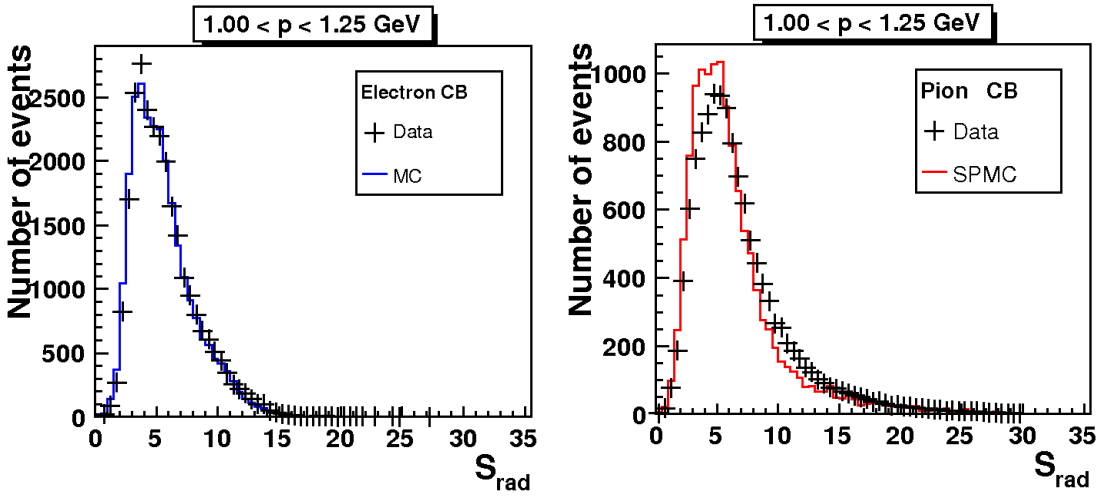

Figure 6.33: Comparison of S_{rad} between data and Monte Carlo in the central barrel for electrons and pions, 1.0 GeV $< p <$ 1.25 GeV.

The distribution for the electron data sample is well described by the prediction of the simulation. No difference in the curve shape is visible. In case of pions, the agreement is slightly worse.
Figure 6.34 shows the comparison between the data distribution and the simulation in the forward barrel. The depicted momentum interval is 2.25 GeV $< p <$ 2.5 GeV.

In this region of the calorimeter, S_{rad} is in good agreement with Monte Carlo expectation. Therefore this estimator again is well described by Monte Carlo simulation.

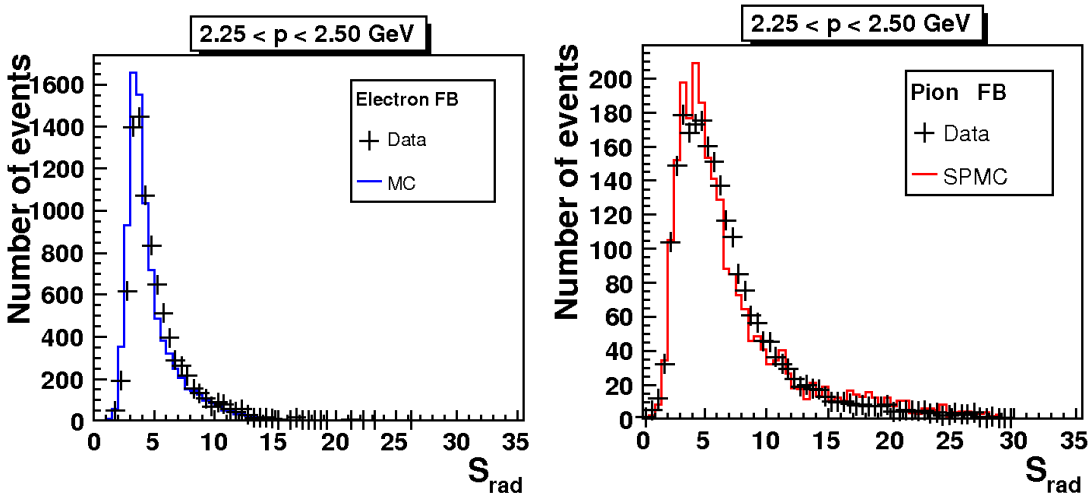

Figure 6.34: Comparison of S_{rad} between data and Monte Carlo in the forward barrel for electrons and pions, 2.25 GeV $< p <$ 2.5 GeV.

The measure for the shower length S_{len} is discussed next. The comparison between data and simulation for electrons and pions is presented in figure 6.35.

The Monte Carlo prediction for electrons is again in good agreement with the data. The shape of the curves show no difference. The more complex structure of the S_{len}-distribution for pions is not properly described. The occurrence of spikes at integer values of the estimator, as described in section 6.3, is predicted by the Monte Carlo simulation. The data distribution tends to have more entries for higher values of S_{len} than the simulation. Although the agreement is not as good as for the already presented estimators, the shape of the distribution is similar.
The same behaviour is observed in the forward barrel and in other momentum intervals. An example can be seen in figure 6.36.

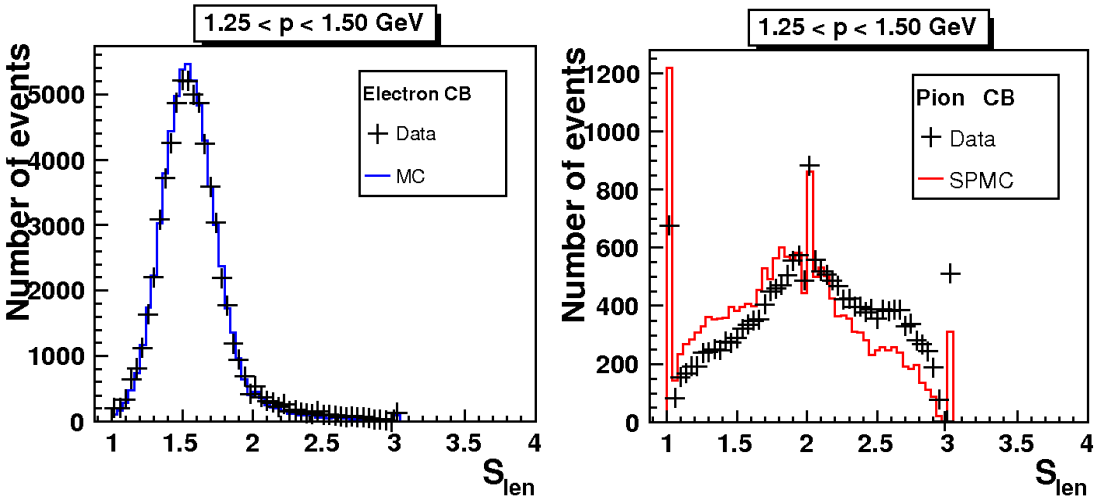

Figure 6.35: Comparison of S_{len} between data and Monte Carlo in the central barrel for electrons and pions, 1.25 GeV $< p <$ 1.5 GeV.

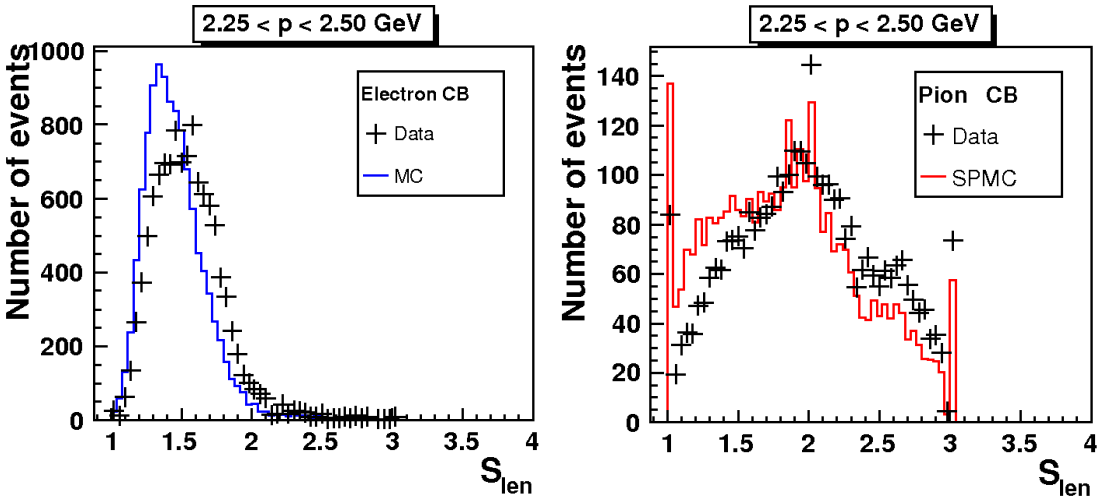

Figure 6.36: Comparison of S_{len} between data and Monte Carlo in the central barrel for electrons and pions, 2.25 GeV $< p <$ 2.5 GeV.

The last two discussed quantities are the measured energies in the electromagnetic and hadronic part of the liquid argon calorimeter. As described before, the deposited energy in a cylinder of radius $R = 30$ cm around the prolongation of the track in the calorimeter is summed up and allocated to the electromagnetic or hadronic part of the calorimeter. The result of the comparison is shown in figure 6.37.

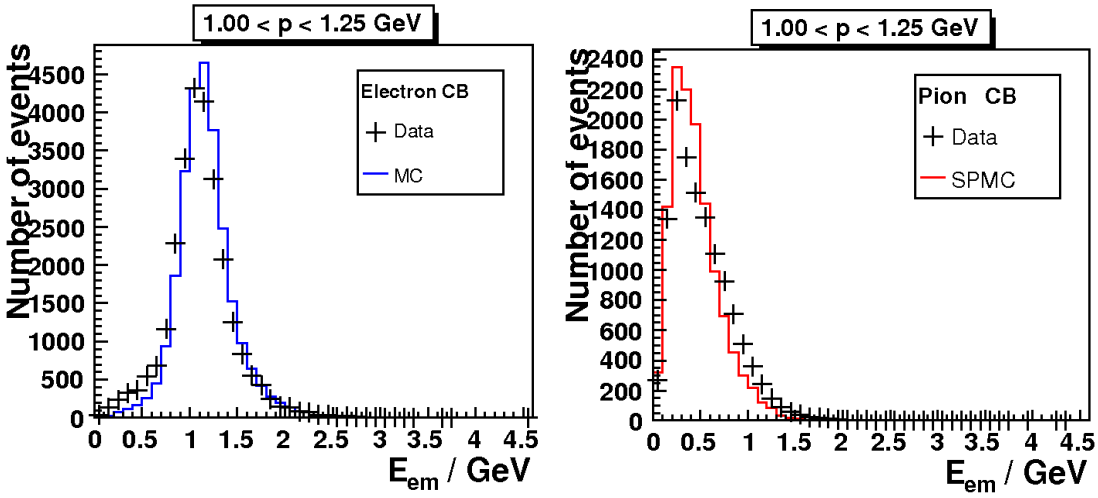

Figure 6.37: Data-MC comparison of the electromagnetic energy in the central barrel for electrons and pions, 1.0 GeV $< p <$ 1.25 GeV.

The agreement between data and Monte Carlo is slightly better for the electron sample. But in both cases the shape of the histograms are similar and the agreement is good in general.

Figure 6.38 shows a shift between the distributions of data and simulation. This is probably the same shifting-effect in the forward barrel as mentioned for the first estimator $\frac{E_{\text{elmag}}}{p}$ (see figure 6.29).

The data distribution is slightly shifted towards lower values of the estimator. The reason could be the worse correction of dead material in the detector by Monte Carlo compared to the data correction. This leads to a prediction of higher energy values than measured. As the shifting effect occurs for $\frac{E_{\text{elmag}}}{p}$ and E_{elmag}, this explanation is plausible.

Finally the outcome of the comparison between data and Monte Carlo for the hadronic energy is presented in figure 6.39.

The description of the hadronic energy for pions is good in general. The shape of the distribution is predicted very well by the simulation.

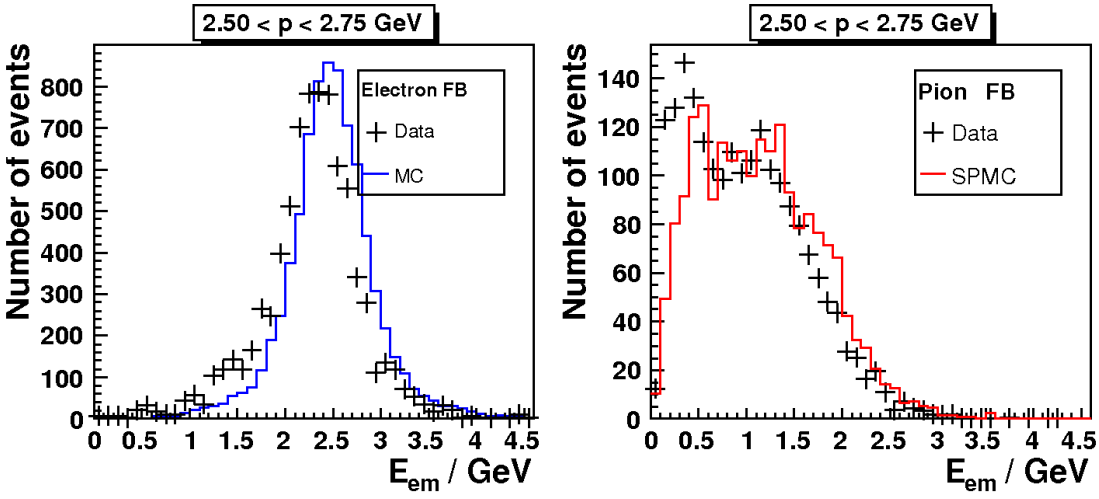

Figure 6.38: Data-MC comparison of the electromagnetic energy in the forward barrel for electrons and pions, 2.5 GeV $< p <$ 2.75 GeV.

The agreement between data and Monte Carlo for electrons is worse than for pions. The values for the hadronic energy in case of electrons is lower than for pions and the number of events for this estimator is low. The energy deposition of electrons should be fully contained in the electromagnetic part of the calorimeter. The cracks between the wheels and the ϕ-octants in the liquid argon calorimeter are not instrumented (see section 3.2 for the

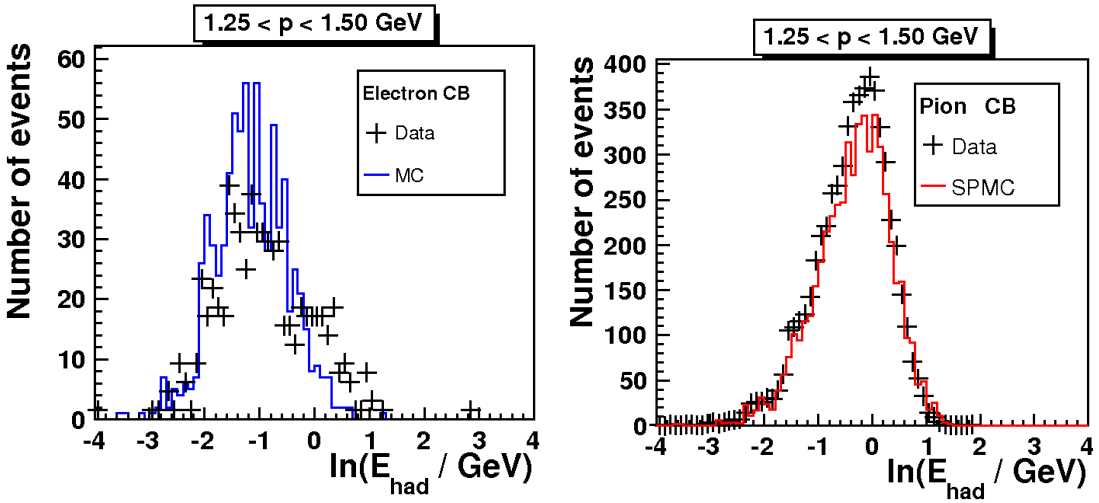

Figure 6.39: Data-MC comparison of the hadronic energy in the central barrel for electrons and pions, 1.25 GeV $< p <$ 1.5 GeV.

layout of the calorimeter). This allows the electrons to probably reach the hadronic region of the calorimeter and therefore induce a weak signal which is seen in this estimator.

6.5 Summary

The comparison between the electron and pion data samples has shown that the chosen estimators exhibit a large separation power and are suitable to distinguish between electrons and pions using calorimeter and tracking information. In particular the separation power of the estimator describing the ratio of energy to momentum $\frac{E_{\text{elmag}}}{p}$ is evident.

The agreement between Monte Carlo simulation and the data allows the use of all the presented variables in the next chapter where the quality criteria in context of a multivariate analysis will be discussed (chapter 7).

Chapter 7

Multivariate Analysis

There are 10^{11} stars in the galaxy. That used to be a huge number. But it's only a hundred billion. It's less than the national deficit! We used to call them astronomical numbers. Now we should call them economical numbers.

Richard Feynman

A multivariate analysis (MVA) is based on the statistical principle of multivariate statistics, which involves simultaneous observation and analysis of more than one statistical variable. The coaction and the dependence structure of the involved quantities are analysed using specific procedures. A multivariate analysis is suitable, as soon as a decision respecting multiple possibly correlated quantities has to be made. The aim of such an analysis is to obtain a clear picture of the system and to allow a reasonable decision, taking correlations into account.

The concrete use for multivariate analyses in high-energy physics is to extract the maximum available information from the available data. This has become more and more demanding in order to search for smaller signals in larger data samples. The computing power and its availableness have grown larger in recent years and therefore the decision guidance by multivariate methods based on machine learning techniques has become an important constituent of many studies in particle physics, like Higgs searches at LEP [19] or top mass measurement at the Tevatron [20].
Due to this increased request the multivariate classifiers themselves have been significantly improved and new ways to tune and to combine classifiers have been developed.

7.1 Software Toolkit and Methods

The software environment for the H1 experiment at DESY is based on the analysis framework ROOT [21]. Since ROOT version 5.11/06 a toolkit for multivariate analysis

is included. This framework is new and still in development. It holds a large variety of multivariate classification algorithms and is called Toolkit for Multivariate Data Analysis with ROOT (TMVA) [22]. TMVA is a collection of different models to build a discriminator based on the information of several input variables. The algorithms range from well known Likelihood estimators over linear discriminants to more recent classifiers.
This software package covers all the needed steps for a complete multivariate analysis. This fact and the possible embedding into the given software environment at H1 are the reasons why TMVA is chosen as the toolkit for the multivariate analysis in this thesis. The most important processes for this work are the training of the chosen discriminator, the testing of discriminator and the evaluation of the procedure.

The first step of the analysis, after having chosen a multivariate classifier (for instance neural networks or decision trees), is the training of the discriminator. Therefore the discussed estimators in sections 6.1 and 6.3 are given to the software as input variables. According to the algorithm of the classifier, the estimators are evaluated using a signal and a background sample. For this purpose the selected data samples, discussed in section 6.2, are used. The discriminator then is trained to distinguish between signal and background, in this case between electrons and pions, using the information given by the estimators.
Consecutively to this machine learning technique the discriminator is tested. This process again needs pure samples for signal (electrons) and background (pions). Therefore the used samples are divided into two parts by TMVA, already for the training process. In this work the partitioning of the events in the samples in training and testing happens randomly. This phase assures the operativeness of the discriminator and allows to determine its performance.
In a last step the result of the trained and tested discriminator is evaluated.

TMVA is specifically designed to the needs of high-energy physics applications and therefore manages the simultaneous training, testing and performance evaluation of all the included classifiers with a user-friendly interface. Moreover the results are visualised and the toolkit provides an interface for the application of the trained classifiers to data. This allows to use the discriminator, adjusted to a given problem, directly on data samples in the H1 software environment.

During the study of the multivariate analysis for this thesis and the work with TMVA, it was possible to contribute to the improvements of the TMVA toolkit. Beside understanding better the used software and contributing to the progress by reporting some bugs, the interesting conversations with the developers of the toolkit brought some more insights for this thesis in general.

As already mentioned TMVA delivers a multitude of classifier methods. For this thesis only two of them are studied in more details: Multilayer Perceptron (MLP) and Boosted Decision Tree (BDT). These algorithms lead to the best discrimination (see section 7.2)

and their basic principles are therefore presented in the following sections. For other classifiers just the results are shown.

7.1.1 Artificial Neural Networks

An artificial neural network [23] is a mathematical model or computational model which interconnects artificial neurons. The fundamental idea is based on biological neural networks. Each neuron in the network produces a certain response at a given set of input signals. This external signal applied to the input neurons, puts the network in a characteristic state that can be measured from the response of the output neurons. The network structure is build up because the response of a specific neuron depends on the output of other (connected) neurons. An artificial neural network can be implemented as an adaptive system that changes its structure based on the processed information during the learning phase. For the application of discrimination between signal and background one can see the artificial neural network as a mapping of the input variables (estimators) to a single output neuron, the output variable of the network. The usage of an artificial neural network is a non-linear statistical data modeling tool if at least one neuron has a non-linear response to the input signal.
Although the specific neurons are simple processing elements, the network itself can show a complex global behaviour determined by the connections between the neurons. During the learning process the strength (weights) of the connections in the network are altered to produce a signal flow according to the input of the network.

In this work a special kind of artificial neural network is used. The characteristic traits of this networks are the organisation of the neurons in layers where only connections from one layer to the immediate next one is allowed. This kind of artificial neural network is called *multilayer perceptron*. This limitation reduces the complexitivity of the network and fixes its structure. The first layer is the input layer and the last one the output layer. All other layers are hidden layers. The alignment of a multilayer perceptron is shown in figure 7.1. Each directional connection between the output of one neuron and the input of another is given an individual weight. The input value of a neuron is calculated by multiplying the weight of the connection with the output value of the previous neuron.

For the method of electron identification using an MLP a single neuron in the output layer is required. This resulting value in the output neuron is normalised to a given interval, usually $[0, 1]$. A value of 0 represents a background like track, whereas 1 means a signal like track. This continuous distribution of the classifier allows to cut on a specific value in order to achieve a desired identification efficiency or background rejection.

The weights for the input of the neurons are determined during the training phase of the multivariate analysis. The most common algorithm for adjusting the weights that optimise the classification performance of a neural network is the so-called *back propagation*. The test statistics given at the output of a network with one hidden layer and with

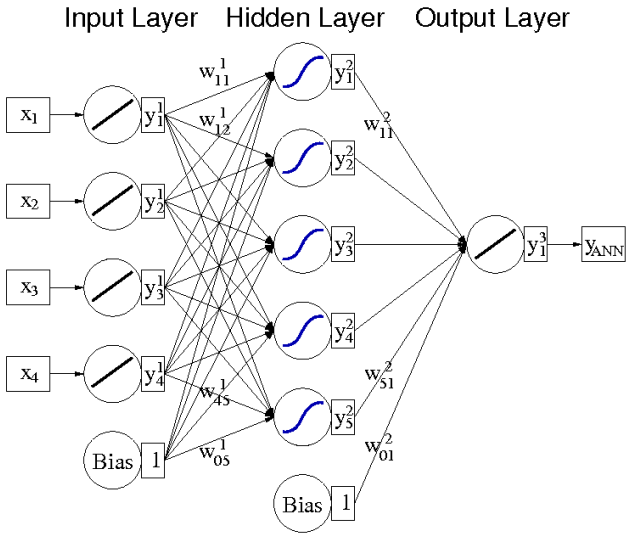

Figure 7.1: Illustration of a multilayer perceptron with one hidden layer and a single output neuron [23]. x_1 - x_4 are the input variables to the neurons y_i^j, where j labels the layer of the network structure. The weights between the neurons are denoted by w_{kl}^j.

a single output node is determined by

$$y_{\mathrm{ANN}} = s_2 \left(a_2 + \sum_{j=1}^{n_\mathrm{h}} w_{j1}^{(2)} \; y_j^{(2)} \left(\vec{x} \right) \right) ,$$

where

$$y_j^{(2)} \left(\vec{x} \right) = s_1 \left(a_1 + \sum_{i=1}^{n_\mathrm{var}} w_{ij}^{(1)} \; x_i \right) .$$

The input variables $\vec{x} = (x_1, \ldots, x_{n_\mathrm{var}})$ are represented as a set of nodes, which together constitute the input layer $(y_1^{(1)}, \ldots, y_{n_\mathrm{var}}^{(1)})$. The nodes in the hidden layer are denoted by $y_1^{(2)}, \ldots, y_{n_\mathrm{h}}^{(2)}$. The functions $s_k \left(\cdot \right)$ are in general called the activation functions of a layer. The weight between input-layer neuron i and hidden-layer neuron j is denoted by $w_{ij}^{(1)}$, whereas $w_{j1}^{(2)}$ is the weight between the hidden-layer neuron j and the output neuron. The offset (or threshold) of the activation functions are denoted by a_1 and a_2.

During the learning process the neural network output y_{ANN} is computed for each training event and compared to the desired output $\hat{y} \in \{0, 1\}$ according to the sample (background or signal) the event belongs to. An error function is defined which measures the deviation of the network response y_{ANN} to the theoretical value $\hat{y} = 1$ for signal

and $\hat{y} = 0$ for background events. The set of the finally used weights is derived by minimising the error function, using iterative methods starting from a random set of weights.

7.1.2 Boosted Decision Trees

A *decision tree* is a special display format of a multitude of binary decision rules. It illustrates successive hierarchical decisions on the input variables in a tree structure. An example is shown in figure 7.2. An input quantity is passed through the tree structure where at every branch a yes or no decision is performed. This left/right (yes/no) turnoffs are repeated until some stop criterion is reached. These final leaf nodes are classified during the training as signal ($y_{\mathrm{BDT}} = 1$) or background ($y_{\mathrm{BDT}} = -1$) depending on the majority of test events from the signal or background sample which end up in this leaf. In the ideal case this allows to separate the phase space of the input variables into regions of signal and background events. In contrast to cut-based methods the decision tree is able to split the phase space into a large number of hypercubes, each of them is assigned to consist of either signal or background events.
The assignment of the estimating variables to the branches is implemented in a manner such that at a specific node the estimator with the best separation power between signal and background at this point is inserted. This means that every splitting criterion is based on a single variable. This assignment is done during the training starting with the root node. Therefore the decision tree "grows" during the training phase.
The last step is the pruning of the decision tree. The tree is cut back from the bottom up after it has reached its maximum size. In this process statistically insignificant parts of the tree are removed which reduces computation time and the overtraining.

In this thesis an extension to the presented decision tree is used: a *Boosted Decision Tree* [23]. Several decision trees (a forest) are derived from the same sample. The same classifier is trained several times using a successively reweighted training event sample. This trees then are combined to form a single classifier, a boosted decision tree. The classifier is given by a (weighted) majority vote of the individual decision trees. Boosting increases the statistical stability of the classifier with respect to fluctuations in the training sample and typically also improves the separation performance compared to a single decision tree.
The boosting algorithm used for this analysis allocates every event that is misclassified during the training a higher weight for the training of the following tree, starting with the original event weights. This boost weight α is derived from the misclassification rate err of the previous tree,

$$\alpha = \frac{1 - err}{err}.$$

The resulting event classification for the boosted decision trees is given by

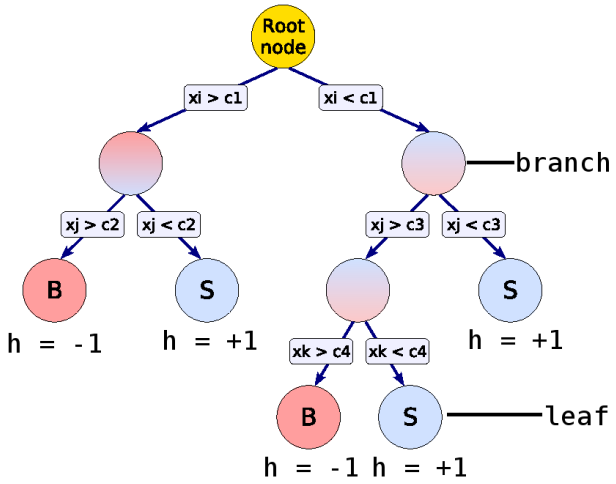

Figure 7.2: Schematic view of a decision tree. Starting from the root node, a sequence of binary splits using the discriminating variables x_i, x_j and x_k is performed. Each split uses the variable that at this node gives the best separation between signal and background when being cut on. The same variable may thus be used at several nodes, while others might not be used at all. The leaf nodes at the bottom end of the tree are labeled "S" for signal and "B" for background depending on the majority of events that end up in the respective nodes. Adapted from [23].

$$y_{\mathrm{BDT}}(\vec{x}) = \sum_{i \in \mathrm{forest}} \ln(\alpha_i) \cdot h_i(\vec{x}),$$

where the sum is over all individual trees $h_i(\vec{x})$ in the forest and $\vec{x}$ being the tuple of input variables. Small (large) values for y_{BDT} indicate a background-like (signal-like) event, since an individual tree is encoded for signal and background as $h(\vec{x}) = +1$ and -1 respectively.

The output format of the analysis by TMVA allows to include the trained discriminator in the software environment H1OO of the H1 experiment. Due to the large training samples the default configuration for boosted decision trees produces too extensive code to be directly implemented into H1OO (about 7 MB of C code). Therefore the training of the boosted decision tree has to be tuned. The discussions with the developers of TMVA pointed out the necessary options to be optimised. After comprehensive tests the output of the training is suitable to be included in programs in the H1OO environment. A first application of this electron finder is presented in chapter 8.

7.2 Results

In this section the results of the multivariate analysis using the TMVA toolkit are presented and discussed. The input variables for the training of the classifiers are the estimators introduced in section 6.1 and 6.3. The training of the discriminators is done twice: One training is done with the information of $\mathrm{d}E/\mathrm{d}x$ and one without, as the information of $\mathrm{d}E/\mathrm{d}x$ is not present in Monte Carlo simulations. At the end the results of the training using the Monte Carlo samples are shown.

As mentioned in section 6.3 and 6.4 for a proper comparison the kinematic variables of the signal and background samples have to be adapted. Therefore a reweighting of the estimators for electrons and pions in the polar angle θ and in the transversal momentum p_t is done. For the training of the discriminators the same samples are used as for the analysis of the separation power of the different estimators, discussed in section 6.3 (see figures 6.14 and 6.15).

The following list contains all the variables that are given to the training of the discriminators as input quantities:

- $\frac{E_{\mathrm{elmag}}(\mathrm{calo})}{p(\mathrm{track})}$; Electromagnetic LAr energy divided by the momentum of the track
- $\frac{E_{\mathrm{inner}}(\mathrm{calo})}{E_{\mathrm{outer}}(\mathrm{calo})}$; Ratio of the electromagnetic LAr energy in a smaller cylinder and a wider cylinder
- $S_{\mathrm{rad}} = \sum_{\mathrm{cells}} \frac{\sqrt{E_{\mathrm{cell}} \cdot dist^2}}{\sum_{\mathrm{cells}} \sqrt{E_{\mathrm{cell}}}}$ (calo); Measure for the lateral shower distribution
- $S_{\mathrm{len}} = \sum_{\mathrm{cells}} \frac{Layer \cdot E_{\mathrm{cell}}}{\sum_{\mathrm{cells}} E_{\mathrm{cell}}}$ (calo); Measure for the shower length
- $\frac{\mathrm{d}E}{\mathrm{d}x}$ (track) (only for the first training); Specific energy loss in the tracking chamber
- E_{elmag} (calo); Electromagnetic energy measured in the LAr calorimeter
- E_{had} (calo); Hadronic energy measured in the LAr calorimeter
- θ (track); Polar angle of the track
- p_t (track); Transversal momentum of the track.

The distributions of the transversal momentum p_t for electrons and pions allow to apply a preselection cut ($p_t > 0.5$ GeV) since there are no electrons in the sample with such a low transversal momentum. This preselection cut is applied before the classifiers are trained.

Correlations of the Estimators

The discrimination power between signal and background depends on the quality of the used input variables. The separation power of each estimator is important but also the correlations between the specific estimators have impact on the discrimination. If two used variables are highly correlated no additional information can be extracted. Therefore it is not expedient to use both variables. On the other hand if all the variables are not correlated to each other, a simple cut based discrimination can be applied. These dependences can be visualised using the correlation matrix of the input variables. A high absolute value of a matrix entry means a strong correlation between the two specific variables. The correlation coefficients are given in percent. The correlation matrices for signal (electron) and background (pion) are shown in figures 7.3 and 7.4.

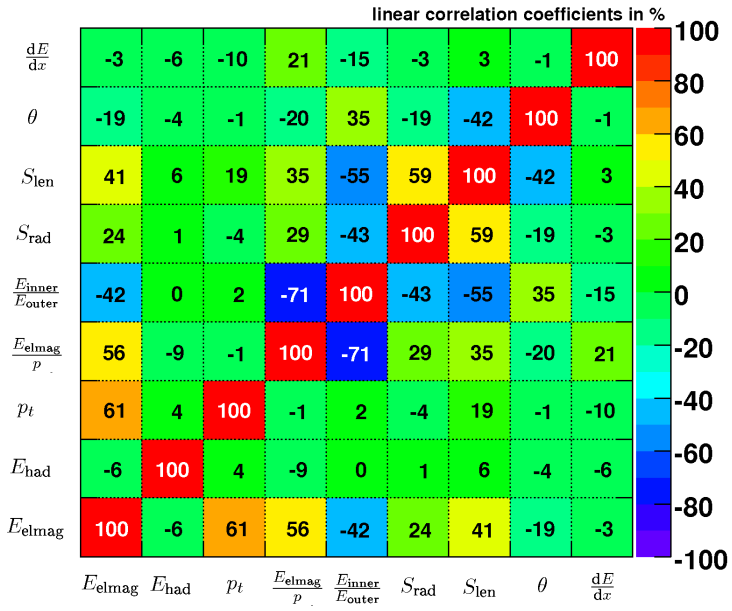

Figure 7.3: Correlation matrix for the estimating variables of the signal (electron) sample.

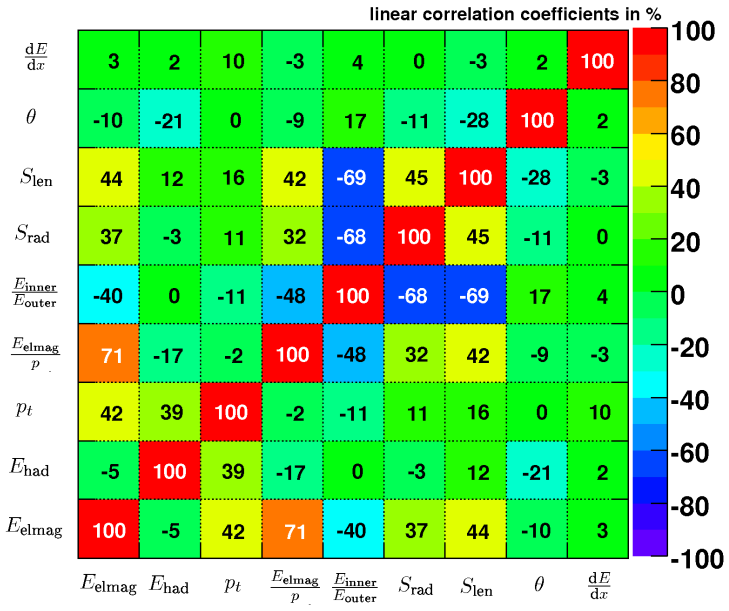

Figure 7.4: Correlation matrix for the estimating variables of the background (pion) sample.

A perfect correlation matrix of estimators would be diagonal. This means that no variable depends on another and all the information given by the estimators can be used to full capacity. A nearly ideal behaviour with respect to the other variables can be seen for the specific energy loss $\mathrm{d}E/\mathrm{d}x$. The correlation coefficients for this estimator are always low and therefore its information is complementary to that of the other estimators.
There are some offdiagonal elements with significantly high correlation values. It is therefore probable that not much information would be lost with less variables. In the following some correlations between specific variables are discussed. The correlation between S_{rad} and S_{len} is expected as the length of a particle shower is always correlated to its lateral distribution via the energy. The dependence between E_{elmag} and $\frac{E_{\mathrm{elmag}}}{p}$ follows directly from the definition. The same applies for E_{elmag} and $\frac{E_{\mathrm{inner}}}{E_{\mathrm{outer}}}$. The momentum of an electron or a pion is always related to the energy of the particle, therefore the correlation coefficient for E_{elmag} and p_t is not small.
The difference between the correlation matrix for signal and background is not big. Although one can see some differences. For instance the correlation between E_{had} (E_{elmag}) and p_t is bigger (smaller) for the background sample. This is the expected behaviour for pions, as the momentum p is correlated to the total energy ($E_{\mathrm{had}} + E_{\mathrm{elmag}}$) and for electrons $E_{\mathrm{had}} \sim 0$.

Training with dE/dx

The separation power of the combined estimators can be rated according to the trained classifier. The resulting distributions of the discriminators MLP, BDT, Likelihood and Rulefit are shown in figures 7.5 and 7.6.

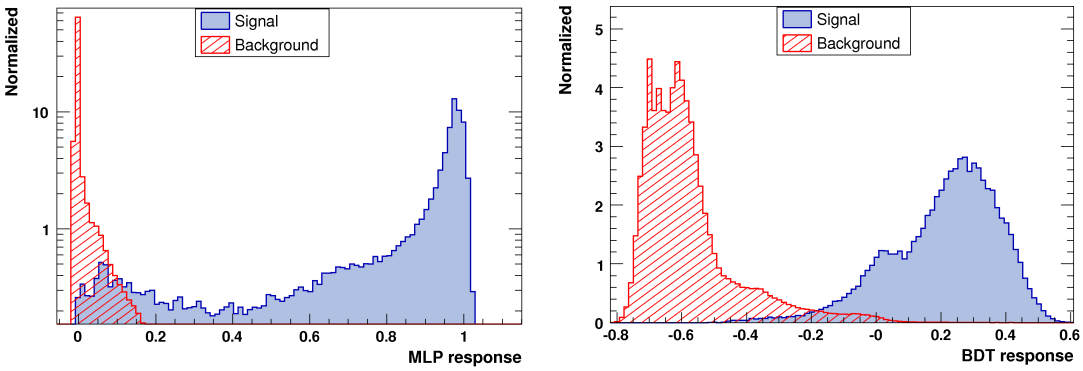

Figure 7.5: Distributions of the trained discriminators including dE/dx in the set of input variables. Left: MLP, right: BDT.

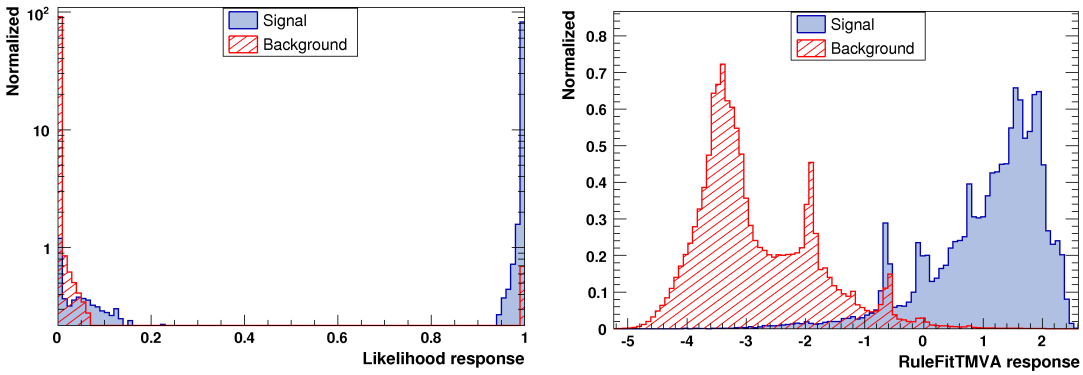

Figure 7.6: Distributions of of the trained discriminators including dE/dx in the set of input variables. Left: Likelihood, right: RuleFit.

Both discriminators in figure 7.5 give good results, meaning a clear separation between the distribution for signal and background events. Noticeable is the small overlap of the signal and the background histograms in case of the BDT and the nonexisting overlap in case of the MLP classifier already for rather small values of the discriminator.
The discriminator trained by the Likelihood method shows a good separation. The disadvantage is the probability of misidentification (signal entries at the very left (background like) and background entries at the right end of the scale (signal like)). The overlap for

the RuleFit classifier is worse but it is possible to get a cleaner sample compared to the Likelihood classifier by loosing efficiency.

The performance of the different classifiers can be summarised in a plot where the background rejection is plotted versus the signal efficiency. This is shown in figure 7.7 for five methods. In this figure the discriminators can directly be compared to each other.

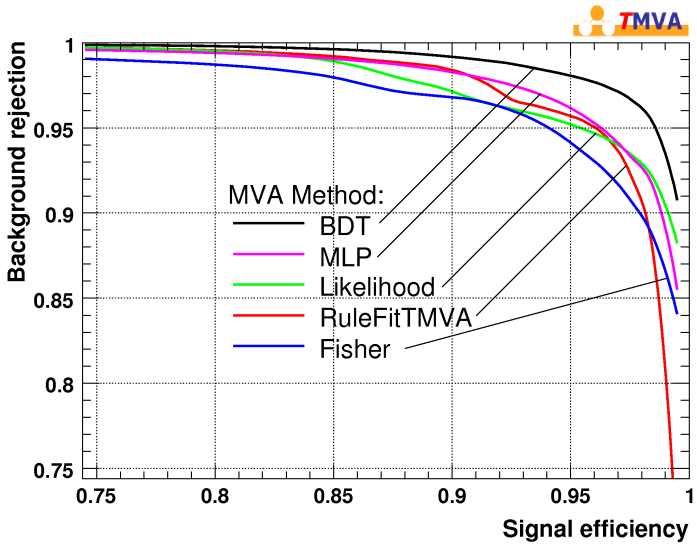

Figure 7.7: Background rejection versus signal efficiency obtained for the various classifiers after evaluating the events from the data samples selected for testing.

The boosted decision tree gives clearly the best result for this application. The background rejection is better than for any other classifier independent of the signal efficiency. It can be used for applications where the purity is important but also for studies where the efficiency is more fundamental. The disadvantage is the large output and the corresponding problems for the further usage of the code (see end of section 7.1.2).
The background rejection of the multilayer perceptron is compareable to that of the BDT, except that it drops steeper towards high efficiencies.
The comparison of the curves for the Likelihood and the RuleFit methods shows again the better background rejection of RuleFit in the region where the efficiency is below about 95 %.

In general a high background rejection can be achieved at a low cost of efficiency.

A more detailed analysis of the discriminating classifiers can be made by means of the corresponding cut efficiency graphs. There the run of the curves for signal and background efficiency as well as for the purity is shown. The number of events for signal and

background is normalised for this figures. The quantity $Q = P \cdot \epsilon$, where P is the signal purity and ϵ the signal efficiency is an indicator for the quality of the cut value on the discriminator. For the boosted decision tree and the multilayer perceptron the curves are shown in figures 7.8 and 7.9. Figure 7.10 shows the cut efficiencies for the Likelihood and RuleFit methods.

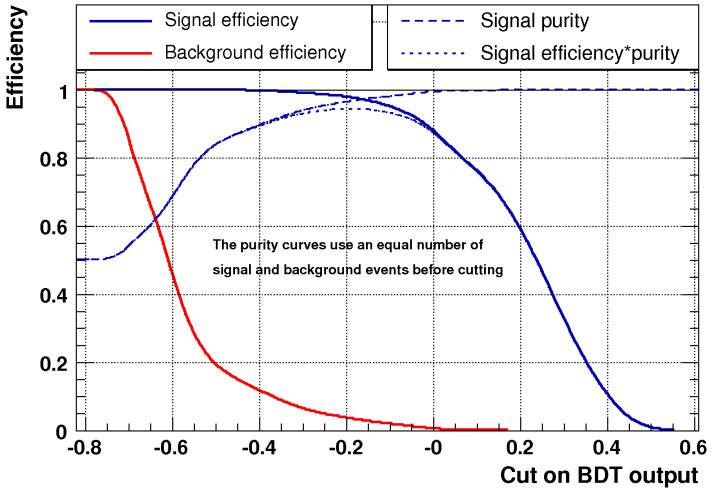

Figure 7.8: Cut efficiency plot for the BDT classifier. The signal and background efficiencies are shown versus the cut values of the discriminator as well as the signal purity.

These figures are helpful to select a working point for a given analysis. They allow to choose a suitable cutting point on the discriminator for the specific needs. The expected amount of background can be weighed up against the achievable signal efficiency.
The signal efficiency and the signal purity curves of the BDT classifier show possible cut values for nearly optimal efficiency or purity. The run of the curve for the quality $Q = P \cdot \epsilon$ shows a maximum value for a single cut value. For the MLP discriminator a steep decrease of the background efficiency and a flat distribution of Q is visible. This shows that a looser cut on the discriminator decreases the signal efficiency whithout increasing the purity much.
The quality Q of the Likelihood classifier shows a flat distribution allowing a low cut value. Compared to the BDT and MLP methods the quality and the signal purity do not reach the same high values. The RuleFit algorithm shows similar curve shapes compared to the BDT classifier. This reflects the similarity of the underlying algorithms. But the efficiency is lower in general.

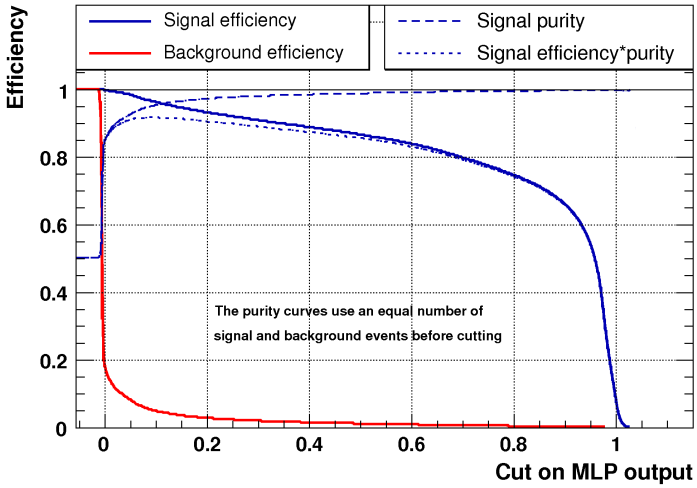

Figure 7.9: Cut efficiency plot for the MLP classifier. The signal and background efficiencies are shown versus the cut values of the discriminator as well as the signal purity.

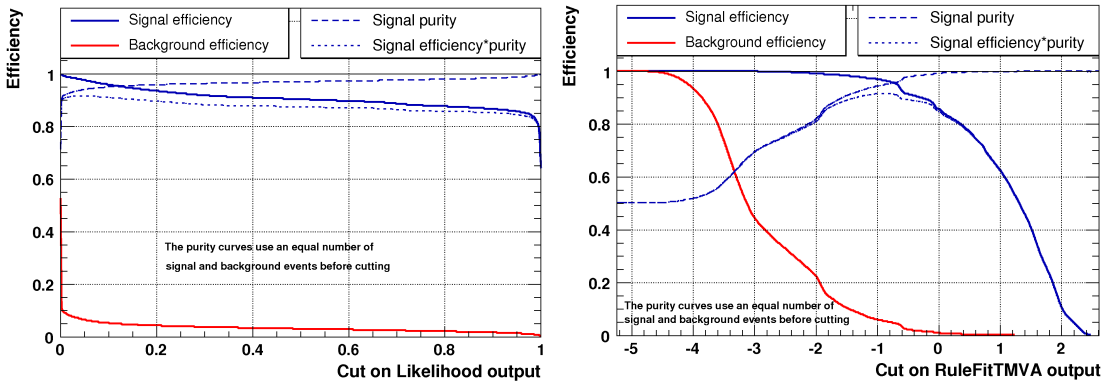

Figure 7.10: Cut efficiency plot for the Likelihood (left) and the RuleFit (right) classifiers. The signal and background efficiencies are shown versus the cut values of the discriminator as well as the signal purity.

Training without dE/dx

The whole process of training, testing and performance evaluation is done a second time for the same samples. The information about the specific energy loss dE/dx from the CJC is not given to the training as an input variable. At this time the information of dE/dx is not present in the Monte Carlo simulations. Therefore without this estimator one has the opportunity to compare the results of the multivariate analysis from the data sample with those of the training with the Monte Carlo samples.
A second reason for this step is the fact, that the information of dE/dx has already been used to select the data for the training samples. By using the specific energy loss of the particles a bias might be introduced. A way to avoid this bias would be to train a discriminator without the information of dE/dx on the present data samples. With this finder one could select new samples and then train a new discriminator with dE/dx as an estimator.

In the following the results for this training process are presented. Figures 7.11 and 7.12 illustrate the distribution of the trained classifiers.

The corresponding correlation coefficients are identical to them in figures 7.3 and 7.4 resulting from the first training using dE/dx.

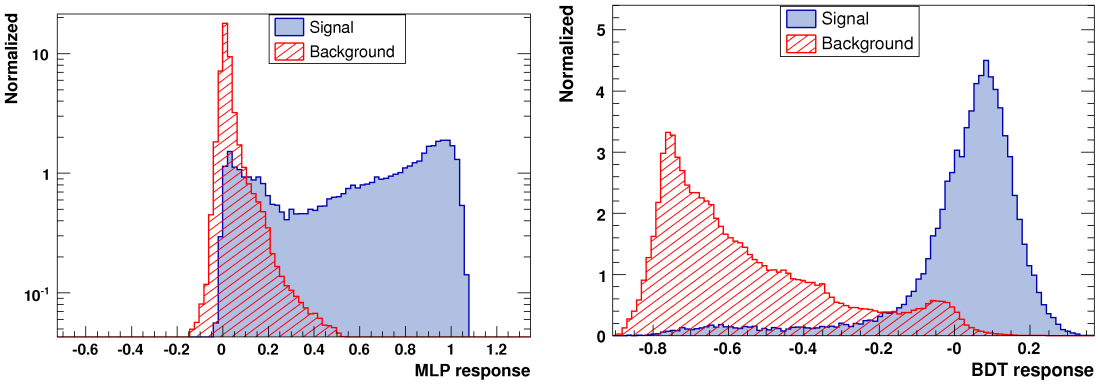

Figure 7.11: Distributions of the trained discriminators without dE/dx. Left: MLP, right: BDT.

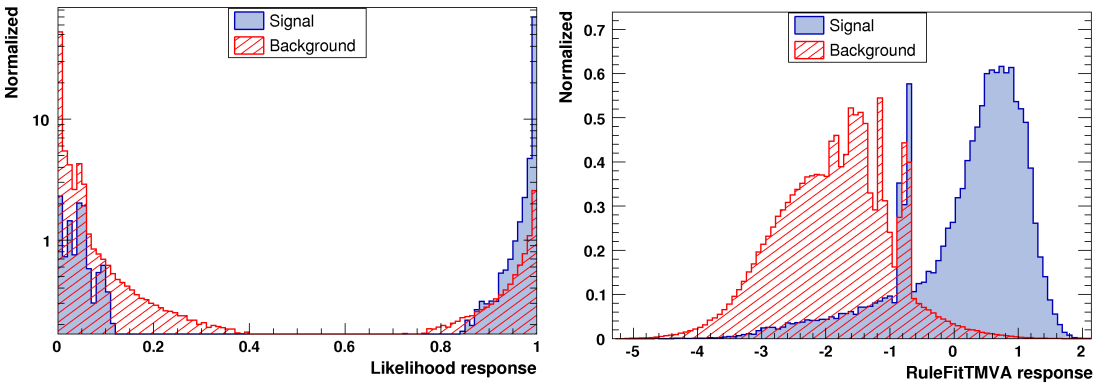

Figure 7.12: Distributions of of the trained discriminators without $\mathrm{d}E/\mathrm{d}x$. Left: Likelihood, right: RuleFit.

Compared to the distributions obtained from the first training using the information of $\mathrm{d}E/\mathrm{d}x$, the separation between the signal and background histograms is considerably worse. The overlap is in general more pronounced and the probability of misidentification has increased. Although with the MLP and the BDT discriminators it is still possible to get a pure sample, but with clearly decreased efficiency. The advantages of the MLP and BDT methods are more obvious than in case of the first training.

The comparison between the performance of the discriminators without the use of $\mathrm{d}E/\mathrm{d}x$ is shown in figure 7.13. The decrease in efficiency and background rejection compared to the curves from the first training is visible.

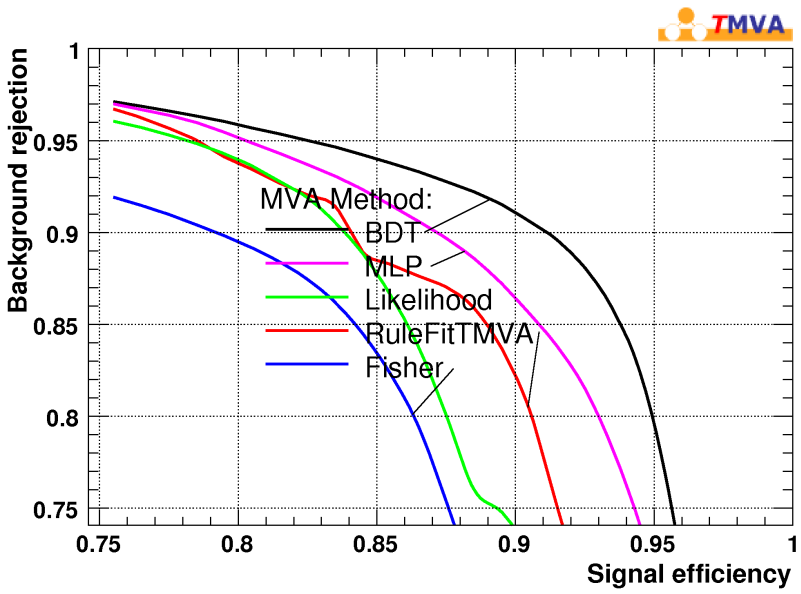

Figure 7.13: Background rejection versus signal efficiency obtained for the various classifiers without the information of $\mathrm{d}E/\mathrm{d}x$ after evaluating the test samples. To compare the results with the previous training, the same scale is shown as in figure 7.7.

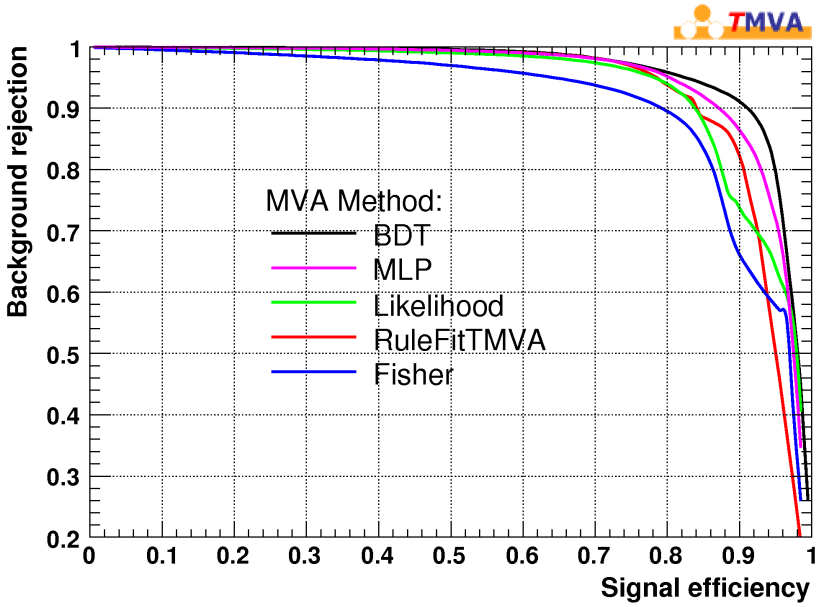

Figure 7.14: Unzoomed version of Background rejection versus signal efficiency obtained for the various classifiers without the information of $\mathrm{d}E/\mathrm{d}x$ after evaluating the test samples.

The cut efficiency graphs for the training without the specific energy loss of a particle in material are shown in figures 7.15 and 7.16.

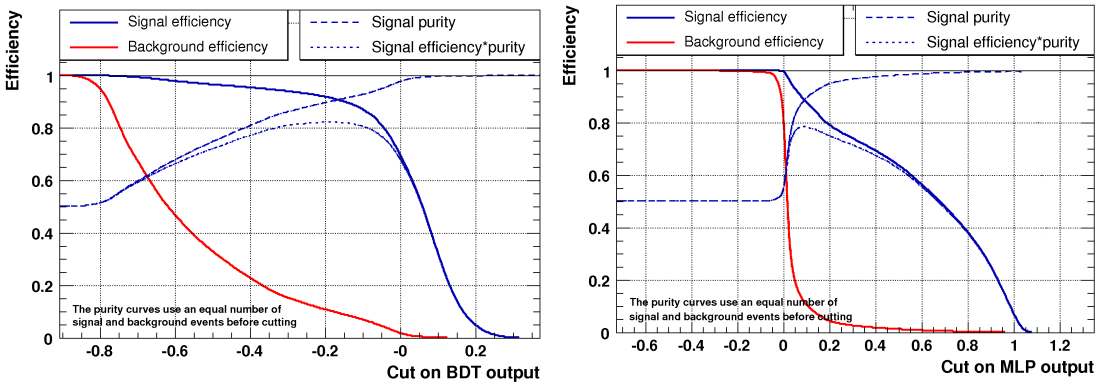

Figure 7.15: Cut efficiency plot for the BDT (left) and the MLP (right) classifier. The signal and background efficiencies are shown versus the cut values of the discriminator as well as the signal purity.

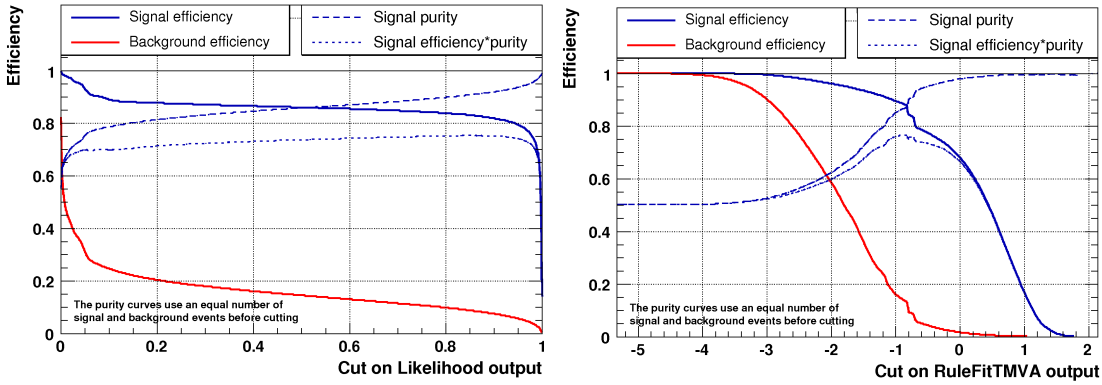

Figure 7.16: Cut efficiency plot for the Likelihood (left) and the RuleFit (right) classifiers. The signal and background efficiencies are shown versus the cut values of the discriminator as well as the signal purity.

The efficiencies and therefore the quality measure are lower for all discriminators compared to the results from the previous training.

This comparison of the two different trainings shows the importance of the estimator describing the specific energy loss of particles in order to get a clear separation between the signal from the electrons and from the pions.

Training using the Monte Carlo samples

In the following paragraphs the outcome of the multivariate analysis of the chosen estimators using the signal and background samples generated by Monte Carlo simulation is given. The specific energy loss in material can not be used for this training, as $\mathrm{d}E/\mathrm{d}x$ is not implemented in the available Monte Carlo programs. Therefore the results can be compared to the training discussed before (Training without $\mathrm{d}E/\mathrm{d}x$).

Figures 7.17 and 7.18 show the correlation coefficients of the input variables for the samples generated by Monte Carlo simulation.

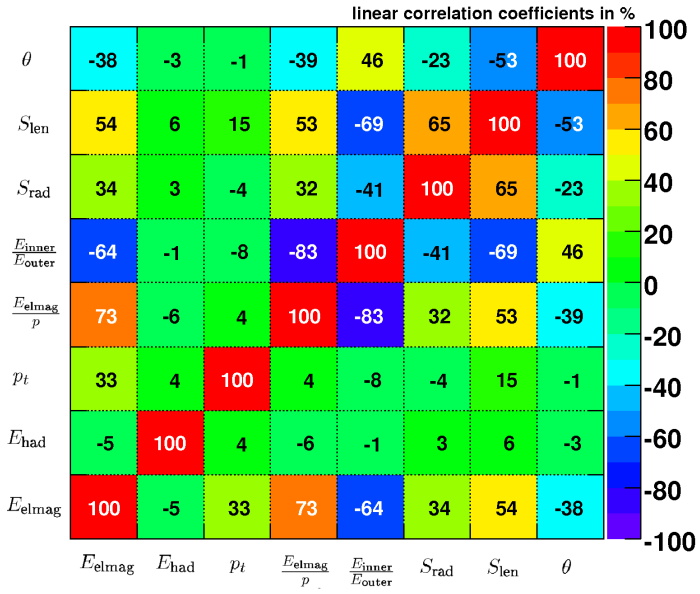

Figure 7.17: Correlation matrix for the estimating variables of the Monte Carlo signal (electron) sample.

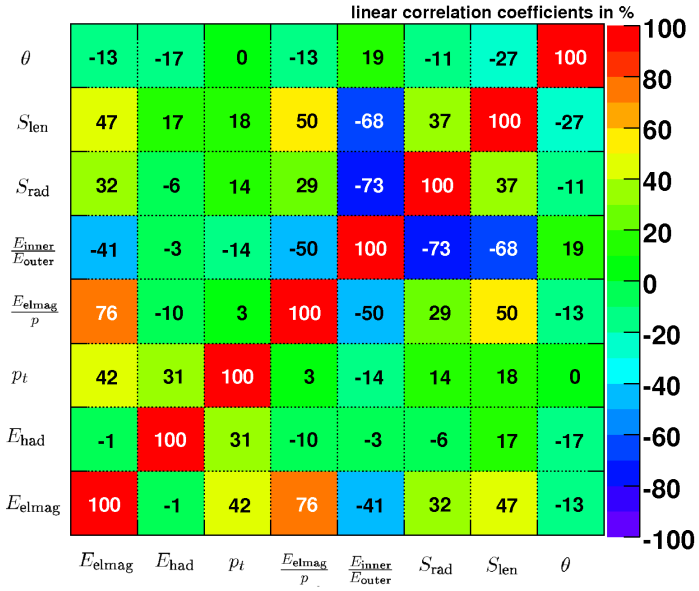

Figure 7.18: Correlation matrix for the estimating variables of the Monte Carlo background (pion) sample.

The correlation coefficients of the input variables for the training of the simulated data samples show the same tendencies as can be seen for the data samples (figures 7.3 and 7.4). In general the correlation coefficients are slightly higher or nearly equal for the Monte Carlo samples compared to the data, except for the correlation between the transversal momentum p_t and the electromagnetic energy E_{elmag}.

The resulting distributions of the discriminators MLP and BDT for the Monte Carlo simulated samples are shown in figure 7.19.

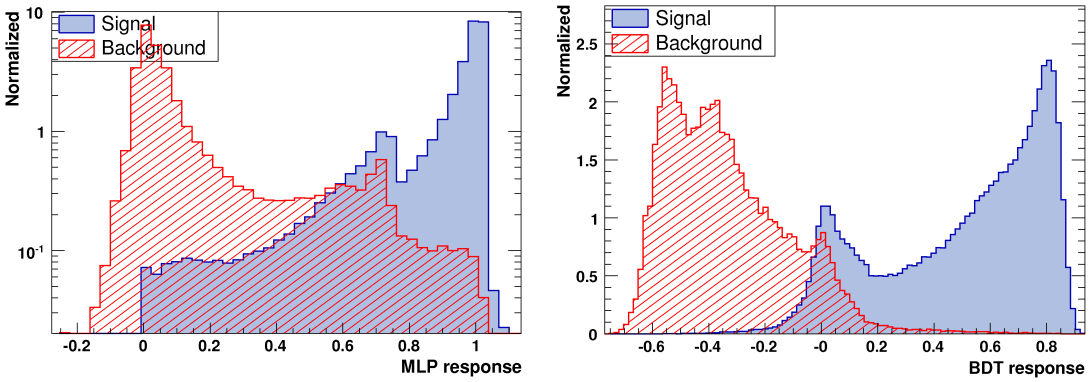

Figure 7.19: Distributions of the trained discriminators for the samples generated by Monte Carlo simulation. Left: MLP, right: BDT.

Although the agreement of the estimators between data and Monte Carlo simulation is good (section 6.4), the result of the multivariate analysis for the MLP and the BDT classifier using the Monte Carlo samples differs from that of the training using the data samples.
The overlap of the signal and background distributions for both classifiers is significantly higher for the Monte Carlo samples compared to the training using the data samples. The amount of signal like (MLP > 0.5) background events for the Monte Carlo samples is considerable. For the data sample no such events are visible. The distributions of the BDT classifier for data and Monte Carlo only differ significantly in the region around BDT ≈ 0. The peak in the Monte Carlo distribution is very pronounced compared to the data distribution.
These differences in the distributions are not understood yet.

Figure 7.20 shows the background rejection versus the identification efficiency resulting from the training of the Monte Carlo samples.

Although the separation between signal and background events in the distribution of the trained classifiers is worse for Monte Carlo simulation the resulting signal efficiency and background rejection seems to be better than for the data samples. The run of the curves are comparable but the Monte Carlo curve lies above the data curve.
This result is not comprehensible at the moment.

The cut efficiency graphs of the MLP and BDT classifiers for the training using the Monte Carlo samples are shown in figure 7.21.

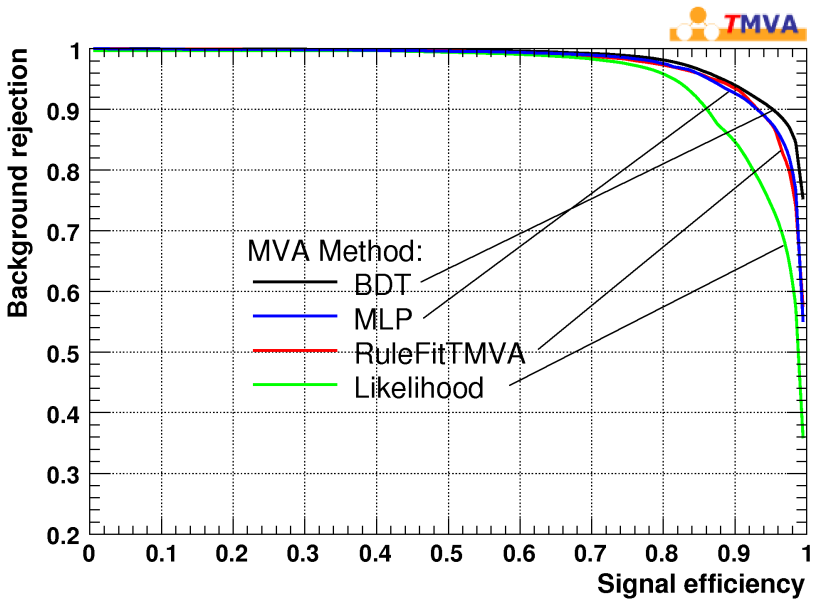

Figure 7.20: Background rejection versus signal efficiency obtained for the various classifiers trained on the Monte Carlo samples.

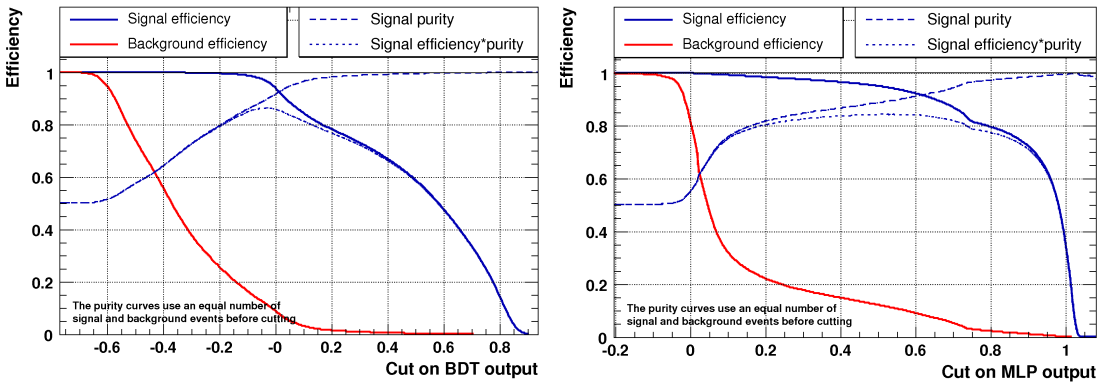

Figure 7.21: Cut efficiency plot for the BDT (left) and the MLP (right) classifier. The signal and background efficiencies are shown versus the cut values of the discriminator as well as the signal purity.

Chapter 8

First Application of the Finder

Il n'est pas certain que tout soit incertain.
Blaise Pascal

The intention of this diploma thesis is to develop an algorithm to identify electrons at low energy in data measured by the H1 detector. The identification is based on information collected by the tracking detector and the liquid argon calorimeter (see section 3.2). This information is used to define estimators which show different distributions for signal or background like tracks. These variables are evaluated and analysed using data and Monte Carlo samples (chapter 6). A discriminating variable is trained by means of a software toolkit (TMVA), which allows to set a working point for the electron identification according to the specific needs for an analysis (chapter 7).

In this chapter a first application of the developed electron finder is presented and the result is compared to that of another finder.

8.1 Inelastic Production of J/ψ Vector Mesons

In order to test a newly developed electron finder a suitable process to analyse is chosen. An interesting process for this purpose are inelastically produced J/ψ mesons, where the regime of photo-production (low Q^2) is chosen. The reasons for this choice are the clear signal of the peak, the similarity of the process to the decay $b \to eX$ and the fact that this process is challenging and difficult.
In contrast to the data selection for the signal sample this time the inelastic process is chosen. In the elastic process two isolated tracks are involved which improves the identification probability. As the number of tracks in an inelastic process is larger this application is more demanding to a finder.

The identification of the decay electrons of the inelastic J/ψ mesons is performed for two classifiers. The methods of boosted decision trees and multilayer perceptrons are

chosen, as they give the best results and have been discussed in more details in this thesis. These results are followed by those of the KALEP [2] finder as a comparison to the existing methods of electron identification. The implementation of this study on inelastic J/ψ and the figures presented in the following sections are constituents of a PhD thesis analysing B-physics [24]. This analysis studies the decay chain $ep \to b\bar{b}X \to e^+e^-X'$ in the H1 experiment at HERA. The reconstruction of an invariant mass peak m_{ee} is a test for the operational reliability of the electron identification implemented in this thesis. Another reason for this test is the fact, that the events of J/ψ decays are background events in a beauty analysis. This is illustrated for instance in a paper of the ZEUS collaboration about the measurement of beauty production from dimuon events at HERA [25]. The invariant mass distribution and the breakdown into the expected contributions from different processes is shown in figure 8.1.

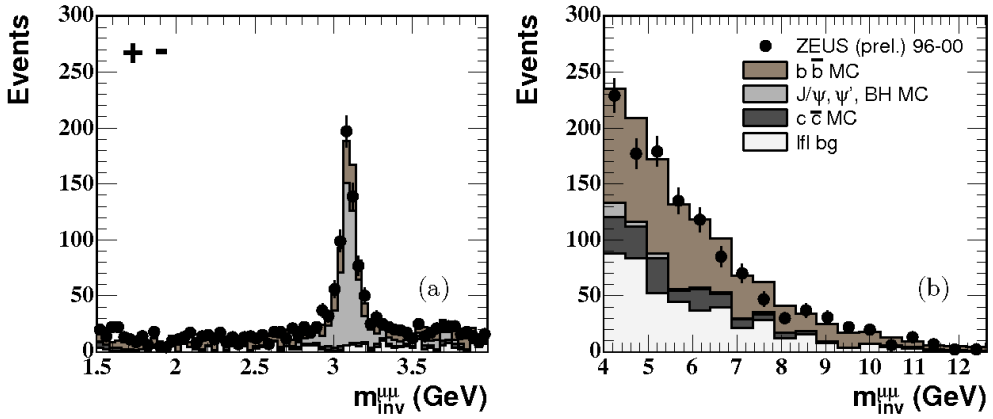

Figure 8.1: Dimuon mass distribution of unlike sign dimuon pairs in separated low and high mass regions and the breakdown into the expected contributions from different processes [25].

The following paragraphs show the distributions of the reconstructed invariant mass m_{ee} with different requirements on the discriminator and other used quantities. The used data was taken during the high energy run in 2007 while the new electron trigger [1] was active. The trigger and vertex requirements are shown in table 8.1. No further track quality requirements are applied to the starting sample.

8.1.1 Boosted Decision Tree

Figure 8.2 shows the distributions of the reconstructed invariant mass for the identified electrons for two different threshold values of the BDT discriminator. The event selection is done by requiring two leptons.

Trigger	Transversal momentum	Number of electron candidates
S29	$p_t^e > 1.5$ GeV	1 (single tag)
S32	$p_t^e > 1.2$ GeV	2 (double tag)
S34	$p_t^e > 2$ GeV	1 (single tag)

Vertex requirement
$\lvert z_{\text{vertex}} \rvert < 35$ cm

Table 8.1: Requirements on the electron trigger for the study of inelastic J/ψ events.

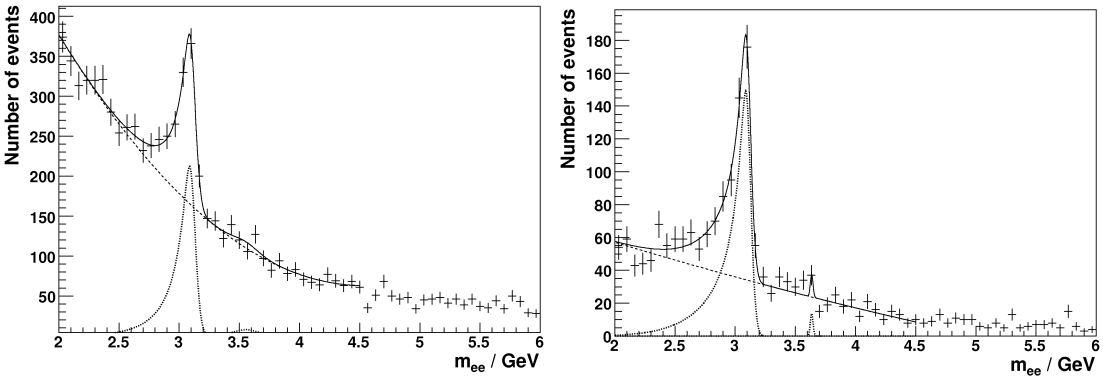

Figure 8.2: Distributions of the reconstructed invariant mass of inelastic J/ψ mesons in dilepton events for two threshold values of the BDT classifier. Left: BDT > 0.15, right: BDT > 0.3 [24].

Both distributions of the reconstructed invariant mass show a clear peak above 3 GeV without any further requirements on the tracks. A threshold value of 0.3 of the BDT classifier leads to a reduction of the background and underlines the peak structure.

The selection criteria for the chosen events are further improved by applying cuts on different quantities, motivated by a preliminary of the H1 collaboration about inelastic J/ψ production at HERA [26] (see also section 8.1.4). The first kinematic variable is the fractional energy of the J/ψ meson. It is defined by $z = (p_\psi \cdot p) / (q \cdot p)$, where p_ψ denotes the J/ψ-, p the proton- and q the virtual photon four-momentum. The fractional J/ψ energy is bounded by $0.3 < z < 0.9$. The transversal momentum squared of the J/ψ meson p_t^2 has to be above a threshold of $p_t^2 > 1$ GeV^2, respectively $p_t^{*2} > 1$ GeV^2 for the transversal momentum squared of the J/ψ in the $\gamma^* p$ center-of-mass frame. A further requirement is that the charges of the two leptons have opposite signs ($c_1 \cdot c_2 < 0$).

Two more cuts are applied in order to improve the peak structure. The used quantities describe the event properties. The centre of mass energy of the photon-proton system has to be in the interval $50 < W_{\gamma p} < 225$ GeV and the restriction on the track multiplicity

is done by requiring at least four tracks ($n_{\text{track}} \geq 4$).

The improvements of these cuts can be seen in figure 8.3.

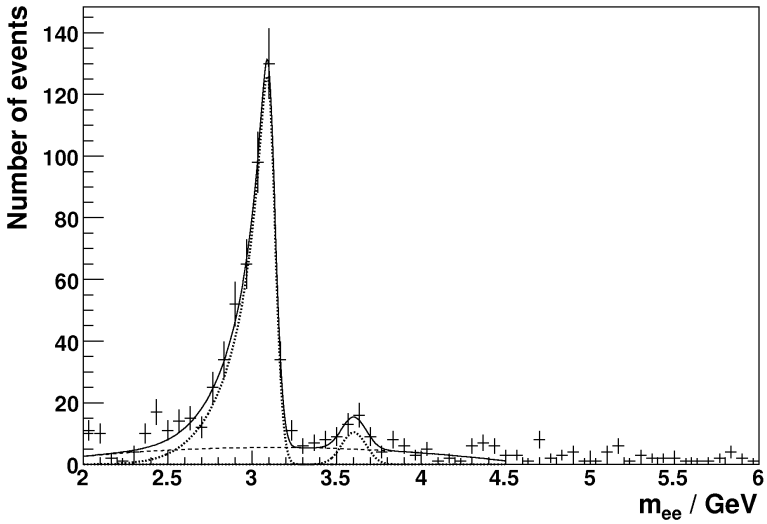

Figure 8.3: Distribution of the reconstructed invariant mass of inelastic J/ψ mesons in dilepton events with cuts on tracking and event information for a threshold value of 0.15 of the BDT classifier: $0.3 < z < 0.9$, $p_t^2 > 1\,\text{GeV}^2$, $p_t^{*2} > 1\,\text{GeV}^2$, $c_1 \cdot c_2 < 0$, $50 < W_{\gamma p} < 225\,\text{GeV}$, $n_{\text{track}} \geq 4$ [24].

The ratio of signal to background has improved and the amount of background is low. Below $m_{ee} = 3$ GeV a slight radiative tail can be seen. At about $m_{ee} = 3.6$ GeV a small peak is visible originating from the decay of ψ' mesons.
The presented preselection cuts in combination with the trained BDT discriminator show a clear peak of the reconstructed invariant mass for the inelastic J/ψ meson in photoproduction.

8.1.2 Multilayer Perceptron

The electron identification can also be done using another classifier. The outcome for the method of multilayer perceptrons is now presented. The approach is the same as for the boosted decision tree discriminator. The results are shown in the following figures. The applied preselection cuts are identical to the already presented ones.

Figure 8.4 shows the distributions of the reconstructed invariant mass for the MLP discriminator. Two different cuts are applied on the discriminator and the event selection is done by requiring two leptons.

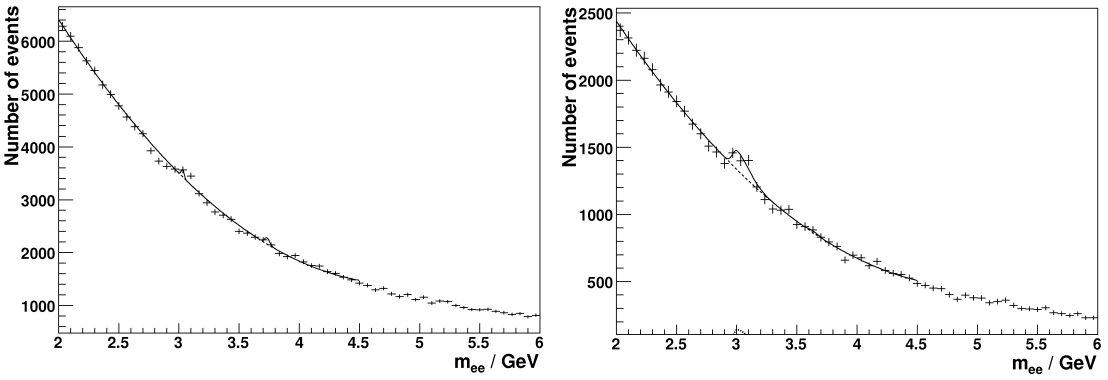

Figure 8.4: Distributions of the reconstructed invariant mass of inelastic J/ψ mesons in dilepton events for two threshold values of the MLP classifier. Left: MLP > 0.75, right: MLP > 0.9 [24].

For both cut values the peak is weakly visible although the cut values of the discriminator are high. Compared to the BDT classifier the result is clearly worse.

The event selection is improved by applying the same cuts as for the BDT classifier. The result for all applied cuts is shown in figure 8.5.

This distribution of the reconstructed invariant mass shows a clear peak above 3 GeV. Moreover the peak of the electrons deriving from the ψ' decay is slightly visible. The applied cuts reduce the background considerably and the radiative tail is visible. The signal to background ratio is clearly worse compared to the result of the BDT classifier with a smaller threshold value.

The comparison of figures 8.3 and 8.5 shows the advantage of the BDT classifier and justifies its choice.

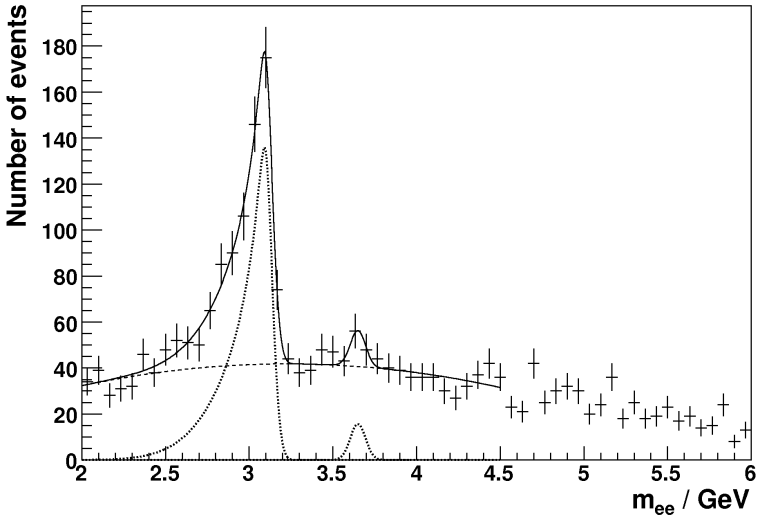

Figure 8.5: Distribution of the reconstructed invariant mass of inelastic J/ψ mesons in dilepton events with cuts on tracking and event information for a threshold value of 0.75 of the MLP classifier: $0.3 < z < 0.9$, $p_t^2 > 1\ \mathrm{GeV}^2$, $p_t^{*2} > 1\ \mathrm{GeV}^2$, $c_1 \cdot c_2 < 0$, $50 < W_{\gamma p} < 225$ GeV, $n_{\mathrm{track}} \geq 4$ [24].

8.1.3 KALEP Finder

In order to compare the results of the electron finder developed in this thesis to the existing electron identification algorithms, the same procedure is done using the KALEP finder. The applied preselection cuts on the track and event information are identical. The presented distributions illustrate the different performances of the chosen methods.

Figure 8.6 shows the reconstructed invariant mass distribution for the KALEP finder requiring two electrons of the highest quality.

The chosen preselection requiring two KALEP electrons of high quality does not reveal a peak structure. Further cuts are necessary to reduce the background.

In order to compare the results of the different electron identification methods the same cuts for the event selection are applied to the KALEP algorithm as presented for the BDT and MLP classifier in the previous paragraphs. The resulting distribution of the reconstructed invariant mass is presented in figure 8.7.

The invariant mass distribution after applying the additional cuts shows the peak of the J/ψ vector meson slightly above 3 GeV. The amount of background is clearly higher compared to the previous classifiers.

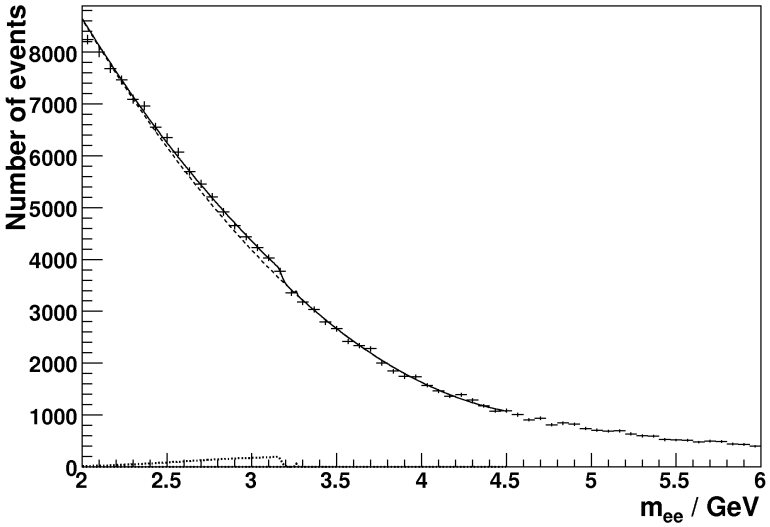

Figure 8.6: Distributions of the reconstructed invariant mass of inelastic J/ψ mesons in dilepton events for the KALEP finder. The electrons are selected with the highest KALEP quality [24].

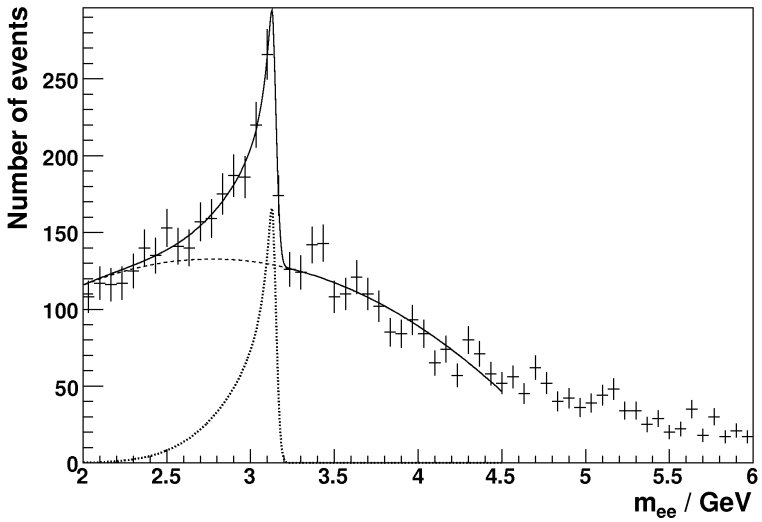

Figure 8.7: Distribution of the reconstructed invariant mass of inelastic J/ψ mesons in dilepton events with cuts on tracking and event information for high quality KALEP electrons: $0.3 < z < 0.9$, $p_t^2 > 1$ GeV^2, $p_t^{*2} > 1$ GeV^2, $c_1 \cdot c_2 < 0$, $50 < W_{\gamma p} < 225$ GeV, $n_{\mathrm{track}} \geq 4$ [24].

Recapitulatory one can say that the performance of the electron identification developed within this thesis is considerably higher compared to the existing electron finder.

8.1.4 H1 Preliminary

In springtime of 2007 a H1 preliminary result was released about inelastic electroproduction of J/ψ mesons in ep-scattering at HERA [26]. In this preliminary, beside other quantities, the reconstructed invariant mass distribution is studied by identifying electrons in the regime of deep inelastic scattering (DIS). Therefore this is another possibility to compare the performance of the electron finder presented in this thesis.
The invariant mass distribution of the preliminary is shown in figure 8.8.

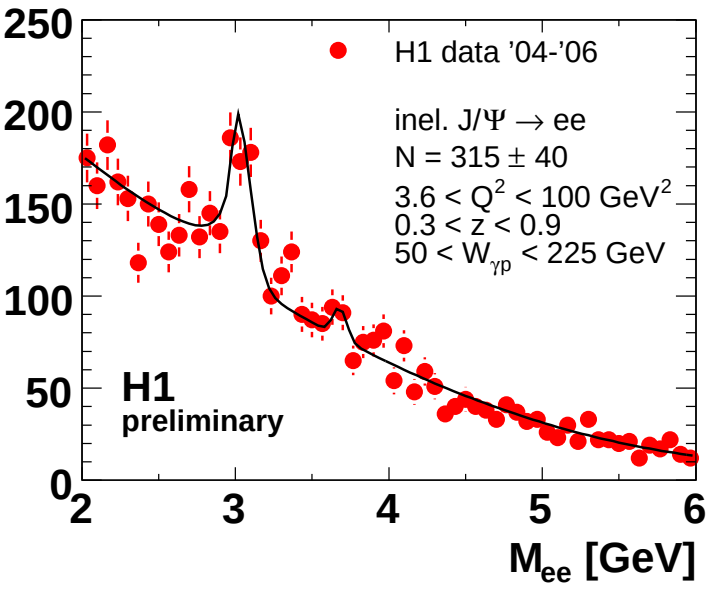

Figure 8.8: Invariant mass spectrum of two oppositely charged electrons. All selection cuts are applied: $3.6 < Q^2 < 100\ \mathrm{GeV}^2$, $0.3 < z < 0.9$, $50 < W_{\gamma p} < 225$ GeV, $p_{t,e} > 0.8$ GeV, $p^*_{t,\psi} > 1$ GeV. The line shows the result of a fit to signal and background.

Chapter 9

Summary and Conclusions

Science cannot solve the ultimate mystery of nature. And that is because, in the last analysis, we ourselves are a part of the mystery that we are trying to solve.
MAX PLANCK

An identification method for low energy electrons at the H1 experiment was developed within this thesis. The algorithm of this electron finder uses tracking and calorimeter information.
The different properties of electromagnetic and hadronic showers in detector material were employed to define estimating variables with good separation power between signal and background. The separation power of the estimators was investigated in measured data and the description of these variables by the detector simulation was verfied.
In a multivariate analysis the estimators were combined to train a discriminator on signal and background data samples using different classifier methods. The performance of the used classifiers was analysed and compared.
The resulting electron finder was implemented in the software environment of H1 and tested on inelastic J/ψ-events. The resulting electron identification was compared to other electron finders.

The chosen variables describing the different behaviour of electrons compared to pions in the detector are in good agreement with the prediction of Monte Carlo simulations. Moreover the separation power of the estimators is good and the correlation coefficients are significantly high only for a few variables.
The results from the multivariate analysis show that boosted decision trees clearly are the best classifiers for an electron finder using the estimators based on tracking and calorimeter information. A very powerful variable for discriminating between electrons and pions is the specific energy loss in the tracking chamber $\mathrm{d}E/\mathrm{d}x$.
The studies of decay electrons of inelastically produced J/ψ vector mesons and the comparison to an existing electron finder show the increase in the identification efficiency

and in the background rejection which has been achieved within this thesis.

There are however still a few topics to be studied in more detail.
Some of the deviations between the distributions of the estimators in data and simulation are not fully understood. The effects for different energies and detector regions could be further analysed.
The usage of $\mathrm{d}E/\mathrm{d}x$ for the data selection of the training samples and as an estimating variable as well could introduce a bias. A possible way to investigate this problem is to train a discriminator without using $\mathrm{d}E/\mathrm{d}x$ as an estimator. In a second step one could select data samples using this discriminator and then train a new discriminator on these samples now using the information of $\mathrm{d}E/\mathrm{d}x$. This would allow to analyse the impact of the $\mathrm{d}E/\mathrm{d}x$-estimator on the discriminator used in this thesis.

The number of used variables for the training of the discriminators could be studied in more detail. The correlation coefficients show that the information provided by the estimators is not completely complementary. A reduction in the number of used variables might reduce the complexity of the training at a low cost of identification efficiency. One of the following estimators is probably redundant: $\frac{E_{\mathrm{elmag}}}{p}$, θ, p_t and E_{elmag}. According to the presented distributions and the correlation coefficients the absence of E_{elmag} would likely not influence the resulting classifier much.
Another way to reduce the number of variables would be to use the polar angle θ and the transversal momentum p_t only for reweighting and not as an estimator. This would allow to respect the kinematic properties without affecting the training.
The number of used calorimeter variables could be reduced by defining only one energy variable $E = \frac{E_{\mathrm{had}}}{E_{\mathrm{elmag}}+E_{\mathrm{had}}}$. This quantity is especially suitable for a multivariate analysis as the codomain is defined by $0 < E < 1$, which improves the performance.
The separation power of the specific variables clearly differs. The estimator with the smallest separation power and the highest correlation to other variables is probably the measure for the lateral shower distribution S_{rad}. A further study could show the effect on the training without this estimator.

Beside the study of the used estimators the performance of the implemented electron finder could be discussed in more detail. The identification performance could be studied in dependence of the transversal momentum p_t and the polar angle θ. It is expected that this would show the impact of the z-cracks in the calorimeter (transition between the calorimeter wheels) on the efficiency. Moreover it would be interesting to know how the identification performes in the forward region of the detector, as the forward tracker does not provide a $\mathrm{d}E/\mathrm{d}x$ measurement.
The lower boundary for the electron identification in the transversal momentum p_t is another quantity which could be determined.

The algorithm to identify electrons implemented within this diploma thesis is being included in a PhD thesis analysing the decay chain $ep \to b\bar{b}X \to e^+e^-X'$ at the H1

experiment [24].
Hopefully this work and the impulses given in this outlook may contribute to a successfull development of the PhD thesis.

List of Figures

Bibliography

[1] Lea Caminada, *Implementation of a Trigger for the Decay $b \to eX$ on the Third Trigger Level at the H1 Experiment*, Diploma thesis, Swiss Federal Institute of Technology Zurich, August 2006.

[2] B. Naroska, S. Schiek and G. Schmidt, *"Lepton identification in the H1 detector at low momenta"*, H1 internal note, 1997, H1-05/97-518.

[3] Richard Kastner, Master's thesis, Universität Hamburg, in preparation.

[4] Martin Göttlich, *Study of Charm and Beauty Production at HERA/H1 Using Dilepton Events*, PhD thesis, FB Physik, Universität Hamburg, April 2007.

[5] Bengt Wessling, *Measurement of the Beauty Cross Section using the Semileptonic Decay into Electrons at HERA*, PhD thesis, FB Physik, Universität Hamburg, December 2004.

[6] I. Abt et al., *"The H1 detector at HERA"*, Nucl. Instrum. Meth., **A386** 310–347, 1997.

[7] B. Andrieu et al. (H1 Calorimeter Group Collaboration), *"The H1 liquid argon calorimeter system"*, Nucl. Instrum. Meth., **A336** 460, 1993.

[8] H1 Collaboration, *"Proposal for an Upgrade of the H1 Luminosity System and its Associated Eletronics for HERA 2000"*, 1998, Proposal submitted to the DESY PRC 98/05.

[9] A. Baird et al., *"A Fast High Resolution Track Trigger for the H1 Experiment"*, IEEE Trans. Nucl. Sci., **48** 1276–1285, 2001.

[10] ETH Zurich The Institute of Particle Physics, Poster gallery, `http://www.ipp.phys.ethz.ch/aboutus/?file=poster`, 2002.

[11] W.-M. Yao et al., *"Review of Particle Physics"*, Journal of Physics G, **33** 1+, 2006.

[12] Jörn Steinhart, *Die Messung des totalen c anti-c Photoproduktions-Wirkungsquerschnittes durch die Rekonstruktion von Lambda(c)-Baryonen unter der Verwendung der verbesserten dE/dx Teilchenidentifikation am H1 Experiment bei HERA*, PhD thesis, FB Physik, Universität Hamburg, October 1999, desy-thesis-99-029.

[13] B. List and A. Mastroberardino, *"DIFFVM: A Monte Carlo generator for diffractive processes in e p scattering"*, (DESY-PROC-1999-02) 396, April 1998, Prepared for Workshop on Monte Carlo Generators for HERA Physics (Plenary Starting Meeting).

[14] C. F. Von Weizsäcker, *"Ausstrahlung bei Stößen sehr schneller Elektronen (Radiation emitted in collisions of very fast electrons)"*, Z. Phys, **88** 612–625, 1934.

[15] E. J. Williams, *"Nature of the high-energy particles of penetrating radiation and status of ionization and radiation formulae"*, Phys. Rev., **45** 729–730, 1934.

[16] CERN Program Library Long Writeup W5013, *GEANT - Detector Description and Simulation Tool*, 1993.

[17] Niklaus E. Berger, *Measurement of Diffractive ϕ Meson Photoproduction at HERA with the H1 Fast Track Trigger*, PhD thesis, Swiss Federal Institute of Technology Zurich, May 2007.

[18] R. M. Weber, *Diffractive ρ^0 Photoproduction at HERA*, PhD thesis, Swiss Federal Institute of Technology Zurich, June 2006.

[19] Jason A. Nielsen, *Observation of an Excess in the Search for the Standard Model Higgs Boson at ALEPH*, PhD thesis, University of Wisconsin - Madison, September 2001.

[20] D0 Collaboration and V. Abazov, *"Production of Single Top Quarks via Flavor-changing Neutral Currents at the Tevatron"*, Phys. Rev. Lett, 2007, arXiv:hep-ex/0702005.

[21] R. Brun, F. Rademakers, P. Canal, N. Buncic, V. Fine and S. Panacek, `http://root.cern.ch/`, ROOT - An Object-Oriented Data Analysis Framework.

[22] A. Höcker, J. Stelzer, F. Tegenfeldt, H. Voss and K. Voss, `http://tmva.sourceforge.net/`, TMVA - Toolkit for Multivariate Data Analysis with ROOT.

[23] A. Höcker, J. Stelzer, F. Tegenfeldt, H. Voss and K. Voss, *TMVA - Toolkit for Multivariate Data analysis with ROOT Users Guide*, Cern, June 2007, arXiv physics/0703039.

[24] Michel Sauter, PhD thesis, Swiss Federal Institute of Technology Zurich, In preparation.

[25] ZEUS Collaboration, *"Measurement of beauty production from dimuon events at HERA"*, 2005, XXII International Symposium on Lepton-Photon Interactions at High Energy.

[26] H1 Collaboration, *"Inelastic Electroproduction of Charmonium at HERA"*, 2007.

Acknowledgment

As a physics student at the Swiss Federal Institute of Technology in Zurich the diploma thesis is intended to be the final step in becoming a physicist, but it is just the first step to become a scientist. I am very thankful that this thesis has not only been demanding but also a great experience. Beside getting to know the spirit of high energy physics, where nothing can be achieved without a collaboration between many like-minded people, I have been introduced to many interesting people in a marvellous city.

I would like to thank Professor Ralph Eichler for making this diploma thesis possible and for supervising me. Special thanks go to Christoph Grab who has always been an important contact during my course of studies. I enjoyed the teamwork during my practical at Paul Scherrer Institute and I am grateful for the guidance concerning my diploma thesis in general and especially for elating me to go to DESY in Hamburg.

I would like to express my gratitude to the following people for their aid while I was working on my diploma thesis:

Michel Sauter who beside introducing me to the daily business of a particle physicist never hesitated to answer my questions and to contribute to the success of my thesis.
Niklaus Berger and Guillaume Leibenguth have always had a sympathetic ear for my questions, problems and demands.
Tobias Zimmermann has been a convenient office mate and a challenging adversary in playing table tennis.
Martin Hoffmann who has been sharing his computer skills with me unhesitatingly.
Sincere thanks are due to André Schöning for asking the right questions and his innumerable ideas to get further insights into the subjects. I have always benefited a lot of his outstanding explanations.

I have met a lot of other interesting people in the past months. I especially enjoyed the time discussing with the people from the University of Heidelberg and Dortmund and the physicists working at the ZEUS experiment. I am glad that I have had the opportunity to compare notes with people from the University of Hamburg, namely Richard Kastner.

On a more personal note I want to thank my family, especially my parents and my sister, who encouraged me throughout my studies, financially and with love. Without them my stay in Hamburg would not have been possible.
Last but not least I want to thank Mirjam Wagner for taking my mind away from work when needed and reminding me that there is a world beyond physics.