A Multi Track Method for b-Tagging

Dissertation

zur

Erlangung der naturwissenschaftlichen Doktorwürde
(Dr. sc. nat.)

vorgelegt der
Mathematisch-naturwissenschaftlichen Fakultät
der
Universität Zürich

von

Ilaria Foresti

aus

Italien

Begutachtet von
Prof. Dr. Peter Truöl
Dr. Felix Sefkow

Zürich 2004

Die vorliegende Arbeit wurde unter der Leitung von Dr. Stefania Xella und Prof. Dr. Peter Truöl durchgefürt und von der Mathematisch-naturwissenschaftichen Fakultät der Universität Zürich auf Antrag von Prof. Dr. Peter Truöl, Prof. Dr. Ulrich Straumann and Dr. Stefania Xella-Hansen als Dissertation angenommen.

"A life without searching is not worth living"
Socrates

Abstract

This work focuses on the development of a b-tagging tool based on a statistical method of identification. After the hadronization of each b quark into a B hadron, B hadrons are tagged on the basis of their long lifetime. The probability for the B decay products to have a common origin in the primary vertex is calculated; this variable is called *combined probability* and is used as discriminating quantity. The analysis is performed on a sample of 11.6 pb^{-1} photoproduction data collected in 1997 by the H1 experiment. Two jets events with an identified muon are selected, in the range 0.1< y <0.8, $10^o < \theta_{jet} < 170^o$, $35^o < \theta_\mu < 130^o$. The tag is tested also on an inclusive Monte Carlo sample. The combined probability yields a good tagging variable and a good variable to enrich the sample in $b\bar{b}$ events. The visible cross section is measured to be $\sigma^{vis} = (129 \pm 22(stat.)^{+22}_{-23}(syst.))$ pb, in agreement with previous measurements and above the QCD prediction.

Zusammenfassung

Die vorliegende Arbeit beschreibt die Entwicklung eines Identifikationswerkzeugs, um b Quarks zu finden. Die Methode beruht auf einer statistischer Basis. Nach Hadronizierung wird jedes b Quark in ein B Hadron umgewandelt; B Hadronen werden dann durch ihre lange Lebensdauer identifiziert. Die Wahrscheinlichkeit, dass B-Zerfallsprodukte einen gemeinsamen Ursprung im Hauptscheitelpunkt haben, wird berechent; diese kombinierte Wahrscheinlichkeit wird als unterscheidende Grösse benutzen. Die Analyse wird auf einen Datensatz von 11.6 pb^{-1} Photoproduktion Daten, die mit dem H1 Detektor in 1997 aufgezeichnet wurden, durchgeführt. Ereignissen mit zwei Jets und einem identifizierten Myon wurden ausgewählt in dem Bereich 0.1< y <0.8, $10^o < \theta_{jet} < 170^o$, $35^o < \theta_\mu < 130^o$. Weiterhin wird die Methode auf einem inklusiven Monte Carlo Datensatz getestet. Die kombinierte Wahrscheinlichkeit stellt sich als eine gute Identifikationsvariable und als gute Variable für die Datensatzanreicherung in $b\bar{b}$ Ereignissen heraus. Der gemessene Wirkungsquerschnitt ist $\sigma^{vis} = (129 \pm 22(stat.)^{+22}_{-23}(syst.))$ pb , was mit vorhergehenden Resultaten bereinstimmt und oberhalb der Vorhersagen von QCD Rechnungen liegt.

Contents

Preface

HERA (*Hadron Electron Ring Accelerator*) is a proton-lepton collider built around the DESY[1] site in Hamburg. It consists of a storage ring with 6.3 Km circumference. The proton beam and the lepton beam (which can consist of positrons or electrons; to simplify the naming, it will be always called *electron* beam) run in opposite directions inside vacuum tubes. Four experiments are assembled along the ring; two of them (H1 and ZEUS) demand the collisions of the beams, the other two (HERMES and HERA-b) have a fixed target and use only one of the beams. The beams are made of bunches of particles, coming at intervals of 96 ns (corresponding to a collision rate of 10.4 MHz). A few bunches don't have a mate to collide with: these *pilot* bunches are 'left free' to interact with the residual gas in the tube, to perform background studies.
The first proton-electron collisions in HERA occurred in October 1991. During the first phase of its life, until June 2000, HERA has run with the proton beam energy set at 820 GeV (920 GeV) and the electron beam energy set at 27.5 GeV, and has been delivering a luminosity of the order of 10^{31} cm^{-2} s^{-1}.
One of the main goals of the HERA experiments is to investigate the structure of the proton, by providing precise measurements of the structure functions, which are related to the parton densities. Heavy quark production, which is the subject of this thesis, is particularly sensitive to the gluon density. It came as surprise when several independent measurements of the heavy quark production cross sections gave results above the QCD predictions.
After a big upgrade, HERA started its second phase in April 2002 (higher luminosity run) with the proton and electron beam energy set, respectively, at 920 GeV and 27.5 GeV.
The work of this thesis consists in the analysis of data collected in 1997 with H1[2]. A statistical method for b-tagging[3] is being developed. This *multi track method*, as it is called, provides a new variable - the COMBINED PROBABILITY - which can be used in further measurements of the beauty production cross section. The multi track method doesn't rely on any specific decay channel of the B hadrons, thus it provides higher statistics event samples compared to the methods used so far.
This analysis is developed on a known data sample, to check the consistency of the results with those obtained with a different tagging method.

---

[1] Deutsches Elektronen Synchrotron

[2] In 1997, the proton beam energy was 820 GeV, while the electron beam energy was 27.5 GeV. This means a center of mass energy $\sqrt{S} \simeq 300$ GeV.

[3] The expression *b-tagging* means the identification of events where a b quark has been produced

Chapter 1

Beauty production at HERA

In the Standard Model of particle physics (see, for example, [1]) a quark is a pointlike particle, sensitive to electroweak and strong interactions. This work focuses on the production of b quarks via the interaction of a virtual photon, radiated from an electron, with a proton (γP interaction).

1.1 The γP process

The general diagram describing the γP interaction is shown in fig. 1.1. A similar diagram, with the photon replaced by a Z boson, completes the description of the neutral current process; replacing the photon with a $W^{\pm}$ boson, describes the charged current process. The production of b quarks occurs in the neutral current process, and is largely dominated by the exchange of a virtual photon, because of the much bigger mass of the Z.
Referring to fig. 1.1, let k and k' be the four momentum of, respectively, the incoming and the scattered electron, and p the four momentum of the incoming proton. The quantity

$$Q^2 = -(k - k')^2 = -q^2 \tag{1.1}$$

gives the four momentum transferred to the photon (being $Q^2 \neq 0$, the photon is virtual). The quantity

$$y = \frac{p \cdot q}{p \cdot k} \tag{1.2}$$

is called inelasticity of the event; in the proton rest frame it gives the fraction of the electron energy carried by the photon. The quantity

$$x = \frac{Q^2}{2p \cdot q} \tag{1.3}$$

is called Bjorken x; in the proton rest frame, it gives the fraction of the proton momentum which takes part to the γP interaction.

Let s be the square of the center of mass energy of the γP system; the quantities Q^2, y and x are related to s via the relation

$$Q^2 = xys \tag{1.4}$$

which holds when the electron and proton mass are taken to be $\simeq 0$.

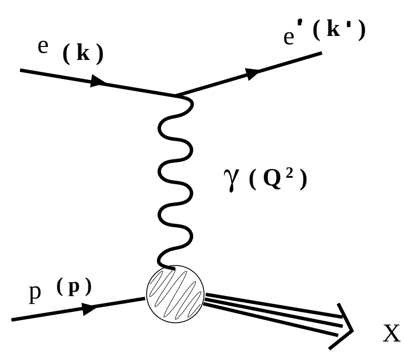

Figure 1.1: *The Feynman graph for the γP interaction.*

This work will use the unit system where c=$\hbar$=1 and therefore the quantities mass, energy, momentum are all given in the same unit (GeV).
In γP interactions the electron is preferentially scattered at small angles with respect to its incoming direction, since small values of Q^2 are preferred. A DIS[1] event is an event where the scattered electron is identified inside the detector, while in photoproduction the electron escapes the detection; typical values for Q^2 in photoproduction are below 1 GeV^2. A $b\bar{b}$ photoproduction event identified and reconstructed in H1 is shown in fig 1.2, with the longitudinal and transverse projections and the energy deposits. A $b\bar{b}$ DIS event is shown in fig. 1.3.

1.1.1 Inside the proton

The proton is a complex system of elementary particles (quarks and gluons) bound and constrained through the strong interaction. A useful approach to describe the γP interaction is to view it as the scattering of a pointlike probe against a structured system. Bound states involving the structured system are ignored, the constituents are considered as asymptotically free particles generating the scattering potential. This description corresponds to a perturbative analysis of the scattering process. Once the cross section for the scattering of a probe on a pointlike object is known, the cross section for the scattering on a non pointlike object can be written as

$$\frac{d\sigma}{d\Omega} = \frac{d\sigma}{d\Omega}|_{point} \times |F(q)|^2 \tag{1.5}$$

[1] *Deep Inelastic Scattering*

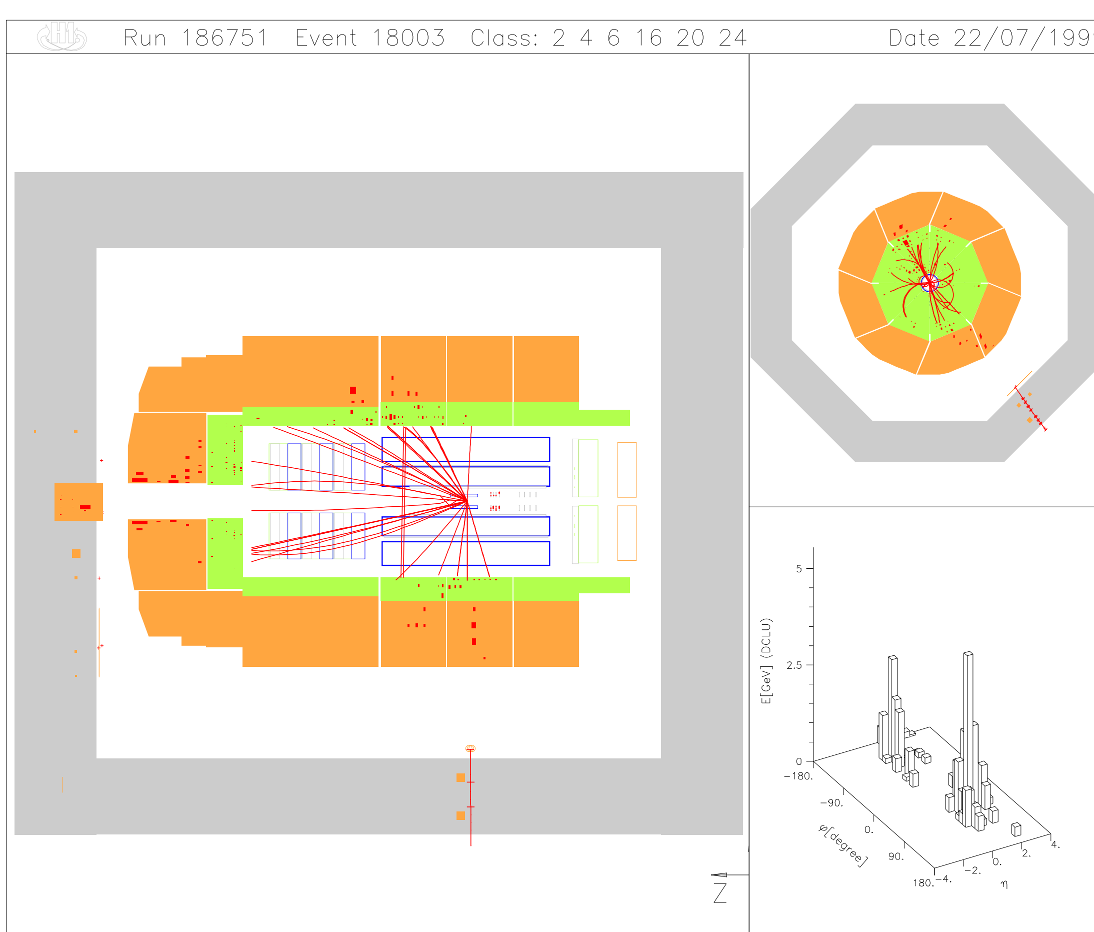

Figure 1.2: *A $b\overline{b}$ photoproduction event recorded by the H1 detector, in the longitudinal (left side) and transverse (upper right) projections; two jets and a central muon are visible. The energy deposits are shown in the bottom right picture.*

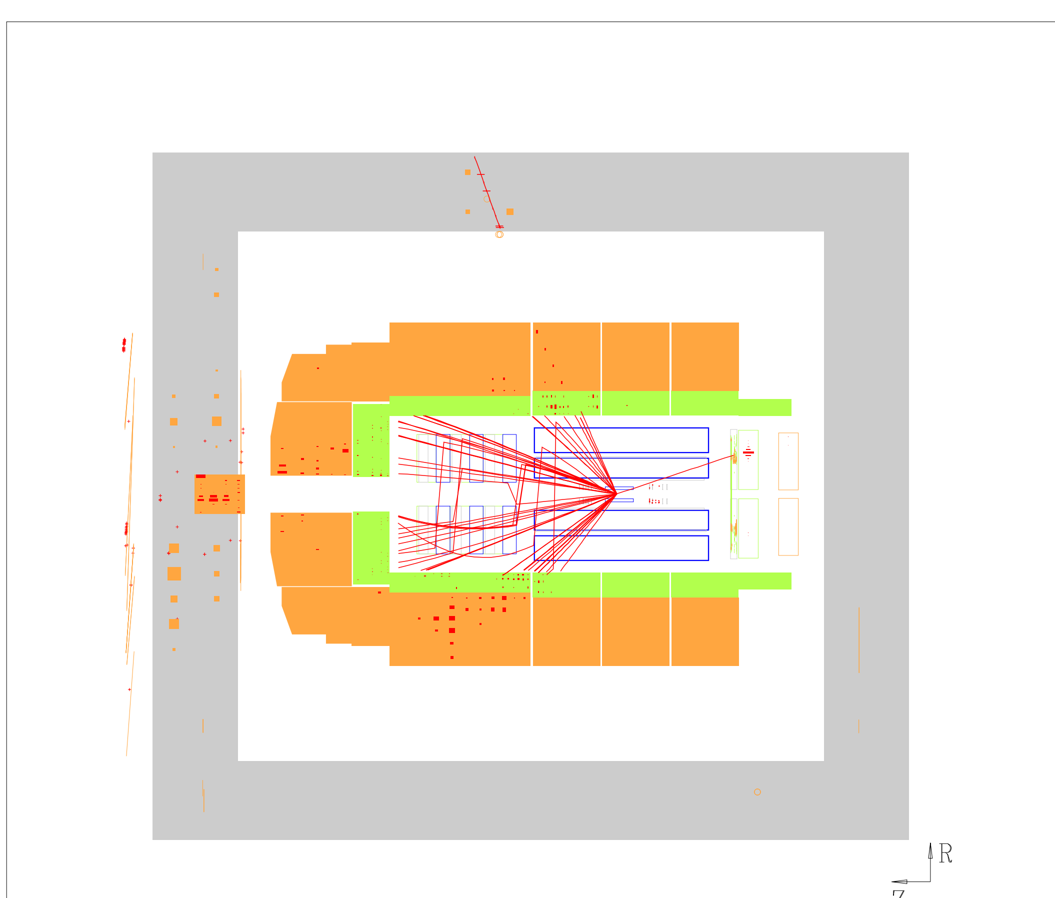

Figure 1.3: *A $b\bar{b}$ DIS event recorded by the H1 detector, in the longitudinal projection. The scattered electron is visible on the right side.*

where q is the transferred momentum; $F(q)$ is called *structure function* and carries the information about the structure of the system.
One of the first descriptions of the proton was given in the Parton Model [2], in which the proton is thought to consist of pointlike particles, called *partons*, each carrying a part of the momentum of the proton. The cross section can be calculated as the incoherent sum of the cross sections for the scattering of the probe on each parton, weighted with a function giving the probability for the parton to have a given momentum and energy; the constituents must then recombine into hadrons with a probability equal to 1. In other words, when the probe particle has enough momentum, it will see the system named proton as a bunch of independent particles. This means that the partons have a transverse momentum which is small, compared to the proton momentum or, equivalently, the interaction between the partons occurs at a very long time scale, compared to the time of the scattering interaction. To make this description correct, it has to be made in a frame where the proton is moving with the speed of light, so that the masses can be neglected. Time dilatation ensures a long interaction time between partons.
A more general and detailed theory describing the interaction between quarks and gluons is QCD[2] (see, for example, [3]). The QCD lagrangian is invariant under local gauge transformations. The colour interaction is described on the basis of the symmetry group SU(3); the generators of the group are eight colour fields, representing the gluons. To define the gluon propagator it is necessary to fix a gauge, thus QCD is non abelian. In other words, the field contains a homogeneous part, similarly to the wave function describing an asymptotically free particle, and this implies that the field can interact with itself. The crucial difference with respect to the electromagnetic field, described in QED[3] (see, for example, [4]), is that the gluons, messengers of the colour field, carry the colour quantum number, while the photon is electrically neutral.

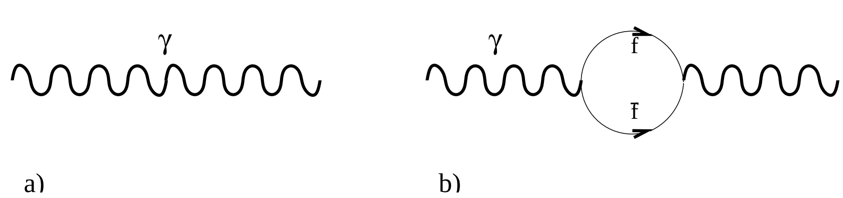

Figure 1.4: *a) The Feynman graph for the photon, and b) the diagram for the one loop correction (vacuum polarization).*

In QED the photon propagator is represented by the graph shown in fig. 1.4 a). In the perturbative approach, the expression of the propagator can be expanded in a power series; the graphs corresponding to the subsequent powers contain fermion loops (not observable in the final state); fig. 1.4 b) shows the diagram at the first order (*vacuum*

[2] Quantum Chromodynamics
[3] Quantum Electrodynamics

polarization). Each vertex is characterized by the coupling α, which is a function of Q^2. In order to avoid divergences, a cutoff in the energy must be introduced, that is, one has to choose a value μ_R for Q^2 and subtract the divergences. α becomes a function of Q^2 and μ_R and is called *renormalized running coupling.* Any observable quantity which depends on Q^2 is re-expressed as a function of $\alpha(Q^2)$.
In QCD, an analogous procedure is applied. Fig. 1.4 has to be replaced by fig. 1.5, which includes the gluon loop. The dependence on Q^2 is also different: while α increases with increasing Q^2, α_S decreases: in the limit $Q^2 \to \infty$ quarks and gluons are free (*asymptotic freedom*). Higher orders in the perturbative expansion give additional contributions to the running coupling.

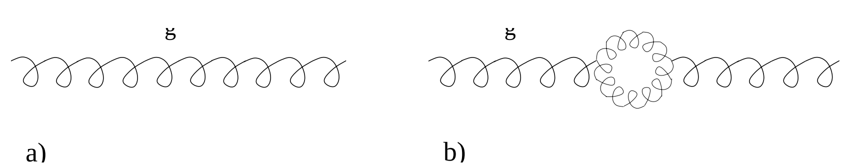

Figure 1.5: *a) The Feynman graph for the gluon, and b) the diagram for the one loop correction.*

For small values of Q^2 (less than 1 GeV^2) the perturbative expansion cannot be applied anymore. The coupling gets very large (*infrared slavery*), and quarks and gluons are found only in hadrons, which have an overall colour charge equal to zero. Quarks and gluons produced with low Q^2 recombine into hadrons with probability equal to 1 (*hadronization process*).
Even for large values of Q^2, the emission of a gluon of low energy or small angle off a parton is a process that cannot be described with perturbative QCD; this process introduces logarithmic divergences in the serial expansion. A way to handle the problem is the *factorization* of the process in a short distance part (high Q^2) and a long distance part (low Q^2), absorbing the divergences in a factor which has to be found experimentally; the energy scale μ_F separates the two regions.
To describe a process of inelastic scattering on a hadron of given momentum, one needs to follow the involved parton from the initial state up to the scattering interaction. The parton, which carries the fraction x of the hadron momentum, is expected to branch emitting gluons. The parton evolution can be described with the DGLAP[4] [5] equation, through the branching of each parton to a new state, which will subsequently branch.

[4] *Dokshitzer-Gribov-Lipatov-Altarelli-Parisi*

1.1.2 Inside the photon

The photon is described in the Standard Model as a pointlike particle, with electric charge and mass equal to zero. But, in hard scattering processes with hadrons the photon looks like a structured system; only a fraction x_γ of the photon momentum is involved in the scattering.
Similarly to what is done for the proton, the photon structure at high energies can be described with parton density functions, in addition to the lepton density functions calculated in QED.
Differently from what happens on the proton side, the photon can still interact as a pointlike particle, behaving like a parton with x_γ=1. In this case the photon is called *direct*; in the other case it is called *resolved*.
Differently from the proton, the photon can fluctuate into a $q\overline{q}$ pair; this is accounted for by an additional term in the parton evolution equation, giving the splitting of the photon directly into $q\overline{q}$. This fluctuation of the photon had already been described, for lower energy scattering, by the VMD[5] model [6]. Lower energy means in this case Q^2 in the order of 1 GeV^2. The VMD model lets the photon fluctuate between two possible states: a bare, massless photon and a hadronic photon. The latter is a massive particle with the same quantum numbers as the photon, e.g. an ω or a ρ meson.

1.2 The γP beauty production

In photoproduction (γP) events at HERA Q^2 is quite small; this means that γP involves a long distance interaction, which occurs in a short time interval compared to the time scale of the interaction of the partons inside the photon. In LO[6] photoproduction is dominated by direct photon interaction. The coupling to a gluon from the proton is the dominant process for $q\overline{q}$ production. The graph for this *boson-gluon fusion* process is drawn in fig. 1.6.
The calculation of the γP cross section follows techniques used in hadron-hadron scattering [7]. The cross section for eP scattering can be deduced from the γP cross section, since the incoming photon is treated as electromagnetic radiation off the incoming electron.

$$\sigma_{eP} = \int f_\gamma^e(y)\ \sigma_{\gamma P}\ dy \tag{1.6}$$

where $f_\gamma^e(y)$ is the flux of the quasi-real photons off the electron; in the Weizsäcker-Williams approximation [8], valid for $\mid Q^2 \mid < m_P^2$, $f_\gamma^e(y)$ reads:

$$f_\gamma^e(y) = \frac{\alpha}{2\pi}(\frac{1+(1-y)^2}{y} ln\frac{Q^2_{max}(y)}{Q^2_{min}(y)} - \frac{2(1-y)}{y}(1-\frac{Q^2_{max}(y)}{Q^2_{min}(y)})) \tag{1.7}$$

[5]Vector Meson Dominance
[6]Leading Order

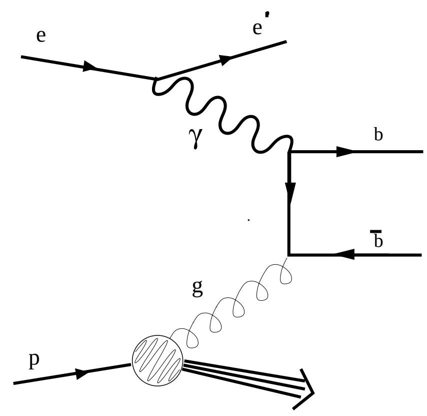

Figure 1.6: *The Feynman graph for the boson-gluon fusion process.*

According to the factorization hypothesis, the hadronic cross section for the production of heavy quarks can be calculated from the subtracted short distance cross section, while the long distance effects are accounted for in *parton density functions*. The cross section can be written as:

$$d\sigma^{Q\overline{Q}X}_{H_1H_2}(\sqrt{s}) =$$
$$\Sigma_{i,j}\int dx_i dx_j d\hat{\sigma}_{ij}(x_i p_1, x_j p_2, M, \alpha(\mu_R), \mu_R, \mu_F) F_i^{H_1}(x_i, \mu_F) F_j^{H_2}(x_j, \mu_F) \tag{1.8}$$

where $d\hat{\sigma}_{ij}$ in the subtracted short distance cross section, F are the parton density functions for the hadrons H_1 and H_2, p_1, p_2 are the hadron momenta, x_i is the fraction of the hadron momentum carried by the parton i, M is the heavy quark mass, μ_F and μ_R are the factorization and renormalization scales.
In the γP scattering, the direct photon contribution has to be added, and the cross section (LO) can be written as:

$$d\sigma^{Q\overline{Q}X}_{\gamma P}(\sqrt{s}) = d\sigma^{point}_{\gamma P} + d\sigma^{hadronic}_{\gamma P} =$$
$$\Sigma_i \int dx_i F_i^P(x_i, \mu_F) d\hat{\sigma}_{\gamma i}(p_\gamma, x_i P, M, \alpha(\mu_R), \mu_R, \mu_F, \mu_\gamma) + \tag{1.9}$$
$$\Sigma_{ij}\int dx_i dx_j F_j^\gamma(x_j, \mu_\gamma) F_i^P(x_i, \mu_F) d\hat{\sigma}_{ij}(x_j p_\gamma, x_i P, M, \alpha(\mu_R), \mu_R, \mu_F, \mu_\gamma)$$

where p_γ and P are respectively the photon and proton momentum and μ_γ is the scale of the electromagnetic vertex.
For the production of b quarks, the energy threshold in the $b\bar{b}$ rest frame is set by $E \geq 2m_b$ - with $m_b \simeq 4.5$ GeV the mass of the b quark. Since m_b is quite large compared to the c quark mass, and even larger compared to the lighter quarks (u, d and s), the cross section

for the production of b quarks is two orders of magnitudes smaller than that for c quarks, and three orders of magnitudes smaller than that for light quarks.
Given the relatively high threshold, the contribution of the resolved photon, where only a fraction x_γ of the small photon energy is used for $b\bar{b}$ production, is expected to be very small; the graphs in the LO are shown in fig. 1.7.

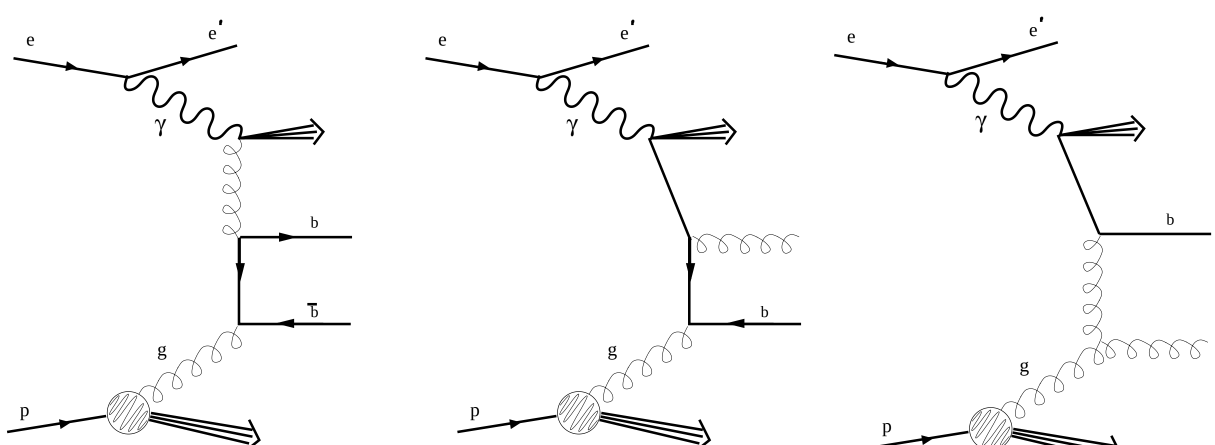

Figure 1.7: *The Feynman graphs for the γP with a resolved photon.*

In the photoproduction regime, the production of a $b\bar{b}$ pair occurs mainly at the threshold. The b quarks then convert into B hadrons with an energy of a few GeV.

1.3 The hadronization process

To consider a γP process (see fig. 1.8, representing the boson-gluon fusion process) one needs the evolution of partons inside the proton (1) and eventually in the photon, until the interaction of two partons occurs at the scale Q^2. Then (2), two quarks (or a quark and a gluon) are emitted, and each one will create a jet of particles. The quark undergoes successive branching (3) producing a parton shower. When branching stops, partons have recombined into hadrons. A phenomenological model for this process, called hadronization, has to be chosen and applied. The Monte Carlo algorithm used in this work, JETSET [9] applies the Lund *string* model: the energy of the $q\bar{q}$ system creates a colour string (4), which is subsequently stretched until it breaks; quark pairs are created from the energy, and at the breaking (5) a number of quarks recombine into observable hadrons. Instable final state particles, like the B hadrons, can then decay. Stable particles and decay products contribute to the final state jet. In this work the jet reconstruction

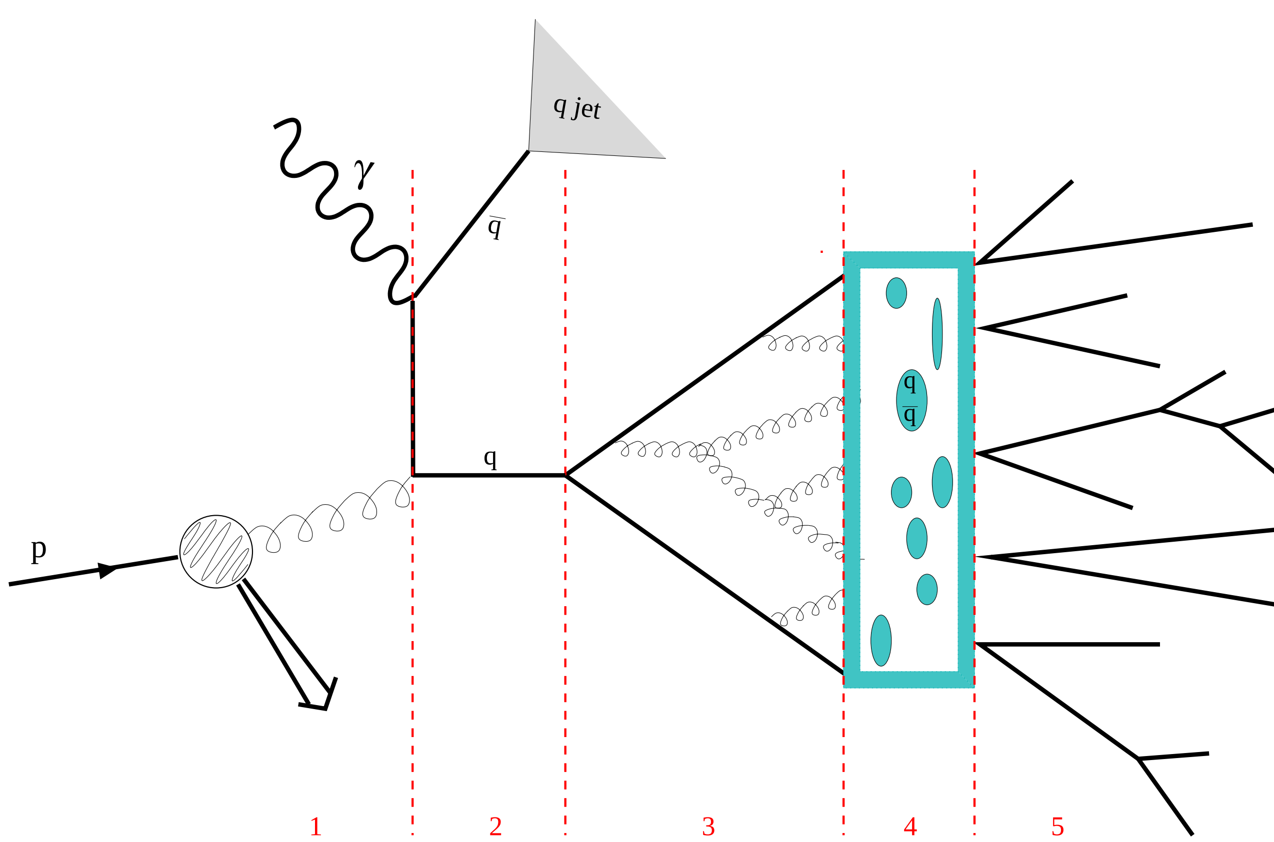

Figure 1.8: *1) The γP interaction (boson-gluon fusion diagram) and subsequent processes: 2) production of the quarks, 3) quark branching and gluon emission, 4) string formation and 5) production of hadrons in the final state. The phases 3, 4, 5 are shown only for the quark; the same picture has to be repeated for the antiquark.*

is made through the $K_\perp$ [10] algorithm.

1.4 The decay of B hadrons

The large mass of B hadrons opens a wide range of decay channels. Table 1.1 lists characteristics and the main decay mode of the two lightest B hadrons (mesons).

	Mass (MeV)	mean life (ps)	semil. decays	branching ratio
$B^{\pm}$	5279.0±0.5	1.653±0.028	$l^{\pm}\nu_l$ + anything	(10.2 ± 0.9) %
B^0	5279.4±0.5	1.548±0.032	$l^{\pm}\nu_l$ + anything	(10.5 ± 0.8) %

Table 1.1: *Characteristics and semileptonic decay channels of the B^0 and $B^{\pm}$ mesons.*

Given the large amount of energy available in the decay, B hadrons tend to produce high track multiplicity events. Among the number of possibilities, the muon channels can be studied with a detector in an easier and cleaner way.
The decay of a B hadron can be described assuming the b quark decays as a free particle. This description is called *spectator model*; the graph is shown in fig. 1.9, where the second quark of the B meson is the spectator quark. The computation is developed referring to the semileptonic decay of a free muon.

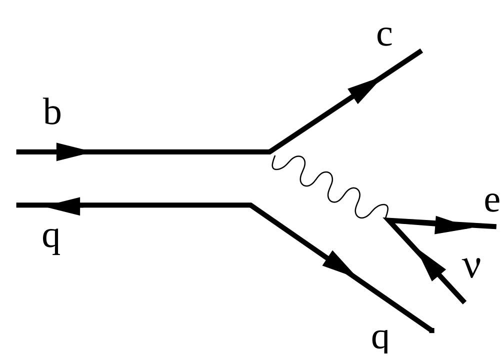

Figure 1.9: *Decay of a B hadron described with the spectator model.*

The decay width for semileptonic decays can be written as

$$\Gamma_{sl} = \Gamma_0(m_b) f(\frac{m_c}{m_b}) |V_{cb}|^2 \eta_0 \tag{1.10}$$

where $\Gamma_0 = \frac{G_F^2 m_b^5}{192\pi^3}$ is the muon decay rate, rescaled to the b quark, m_c is the mass of the c quark, V_{cb} is the element from the CKM[7] quark mixing matrix and η_0 is a correction

[7] Cabibbo Kobayashi Maskawa

term arising from strong interaction corrections. f is the *fragmentation function*. The fragmentation function gives the fraction of the quark momentum which is transferred to the hadron; several definition of f are used, following slightly different phenomenological models. Most often, as we do here, the Peterson function [11] is used.

1.5 The b cross section

Several independent measurement of σ have been performed. The surprising result is that the measured σ exceeds by more than a factor two the theoretical predictions.
The comparison of the measured cross sections with the theoretical predictions shows, that beauty production is still an open issue in high energy interactions. Not all aspects seem completely understood. This statement (see e.g. [12] for a review) also holds for photon-photon collisions measured at LEP at the energy $\sqrt{s} = 190$ GeV [13] (fig. 1.10) and proton-antiproton annihilations at the Tevatron [14].

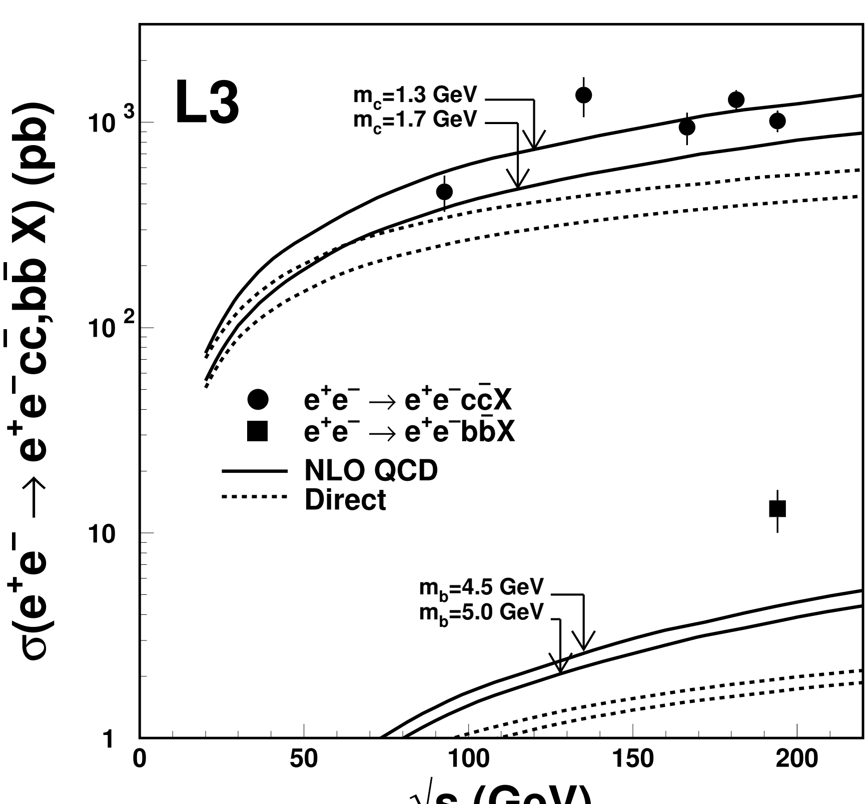

Figure 1.10: *Open charm and beauty cross section in photon-photon collisions, measured by the L3 experiment.*

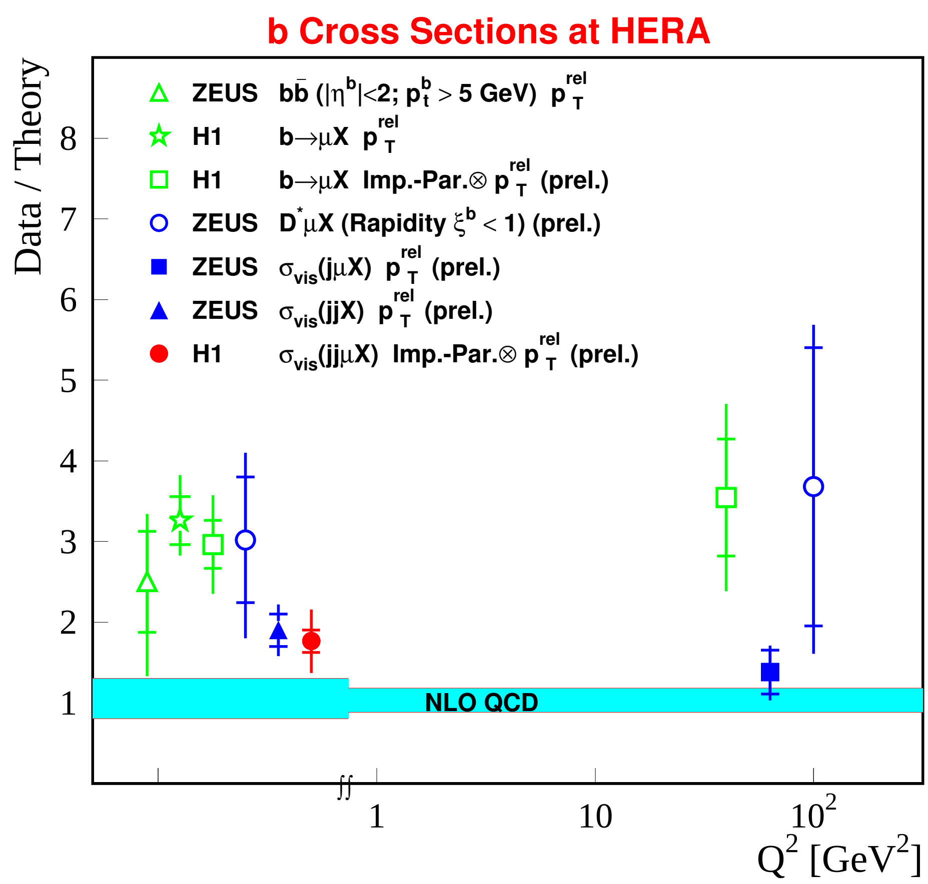

Figure 1.11: *Visible cross section at the NLO; the ratio of the measured to the predicted value is shown.*

Result	presented in	channel	method	# selected events
△	[29]	$e^+P \rightarrow e^+jje^-X$	electron $P_\perp^{rel}$	$\simeq 1500$
★	[26]	$eP \rightarrow ejj\mu X$	muon $P_\perp^{rel}$	$\simeq 1000$
○	[30]	$eP \rightarrow eD^-\mu X$	correlation D$^-\mu$	$\simeq 60$
■	[31]	$eP \rightarrow ej\mu X$	muon $P_\perp^{rel}$	$\simeq 1000$
▲	[32]	$eP \rightarrow ejj\mu X$	muon $P_\perp^{rel}$	$\simeq 2700$
●	[28]	$eP \rightarrow ejj\mu X$	muon $P_\perp^{rel} \otimes \delta$	$\simeq 1500$

Table 1.2: *Decay channel and analysis method chosen in the measurements of the cross section (cfr. fig. 1.11). The number of selected events is also reported.*

In other cases the predictions fall considerably short off the measurements. Hence a careful investigation of all assumptions entering the calculation is called for. HERA data tend itself particularly to test two key aspects, namely the role of resolved photons, i. e. photon structure, the second term in equation (1.9) and the perturbative QCD approximation of gluon radiation, i. e. proton structure. Existing NLO[8] photoproduction calculations (FMNR [17,18]) generate the b quarks exclusively in the hard process via boson-gluon fusion, assuming that the proton contains only light quarks. The renormalization and factorization scales μ_r and μ_f (see equation 1.9) are set to the transverse mass m_b=$\sqrt{m_b^2 + (p_\perp^b)^2}$ with m_b=4.75 GeV. In the FMNR program the parton distributions GRV-GHO [15] and CTEQ5D [16] for the proton are used.
A summary of the HERA results is shown in fig. 1.11. Tab. 1.2 shows the numbers of events selected for the measurements in fig. 1.11 . The statistics available for these results is low, more events would be needed for deeper investigations. The natural step would be to get away from the muon or electron requirement, whose branching ratios are in the order of 10%. This thesis aims at the development of a b-tagging method which is independent from the decay channel of the B hadrons.

1.6 Monte Carlo generators

The FMNR program provides parton level event distributions, i. e. two or three outgoing partons - a $b\bar{b}$ quark pair and possibly a gluon. In order to compare with data this should be embedded in an event generator which includes the hadronization process described in section 1.3. The inclusive events studied in this work were generated using the PYTHIA event generator [9], which is dominated by the light quark background events. PYTHIA generates direct and resolved processes, uses the same parton densities as mentioned above, but contains only LO QCD matrix element. The high statistics b quark sample was generated with the AROMA event generator [19], which includes direct production and LO matrix elements only. Needless to say, that all event generators are forced to match the experimental resolutions. These samples suffice to test the usefulness of the tagging method, but a comparison of the resulting cross section to theoretical predictions requires hadronization corrections to the parton level predictions, which may however be deduced with sufficient accuracy even from incomplete LO models.

[8] *Next to Leading Order*

Chapter 2

The H1 detector

The data collection of the H1 experiment relies on a composite detector. Figure 2.1 gives an overview of its main elements. The quality of the identification of an event and the precision of the measurement of its properties depend on the characteristics of each subsystem, which is involved for a particular type of event. In this chapter the detectors are described from the point of view of their efficiency and precision, keeping an eye on their relevance for the analysis of this work. A full description of the H1 detector can be found elsewhere [20].
The reference frame used in H1 is defined in figure 2.2; the forward direction along the z axis is the direction of the incoming proton. The x axis points towards the center of HERA, the y axis points upward. The angle θ is the polar angle with the positive direction of the z axis. The transverse plane, where the angle ϕ is defined, is called the (r-ϕ) plane. The interaction point is known with a finite precision. The beam spot in the (r,ϕ) plane is an ellipse with half widths of σ_X= 150 μm, σ_Y= 30 μm.
Coming from the collision point, the outgoing particles meet at first the silicon vertex detectors, then the tracking system, then the calorimeters and at last the muon detectors. The chain of detectors is repeated three times: in the central region (cylindrical symmetry around the interaction point, aligned along the beam axis), the forward region (with the exception of the vertex detector) and the backward region (with the exception of the muon chamber). The eP interaction is strongly asymmetric, the events are boosted in the direction of flight of the proton (i.e. forward direction). The forward and backward detectors are designed under this consideration.
Far from the interaction point of the beams, in the forward direction, there are the luminosity system and the very forward detectors (the proton remnant tagger, the forward photon spectrometer and the forward neutron counter).
Scintillator veto walls eliminate events which are not created by a beam-beam collision and hence do not have the correct flight time from the interaction point to the scintillators. This veto helps in particular against beam gas interactions. Timing information by the ToF detectors also helps to discriminate against cosmic rays, which show no correlation

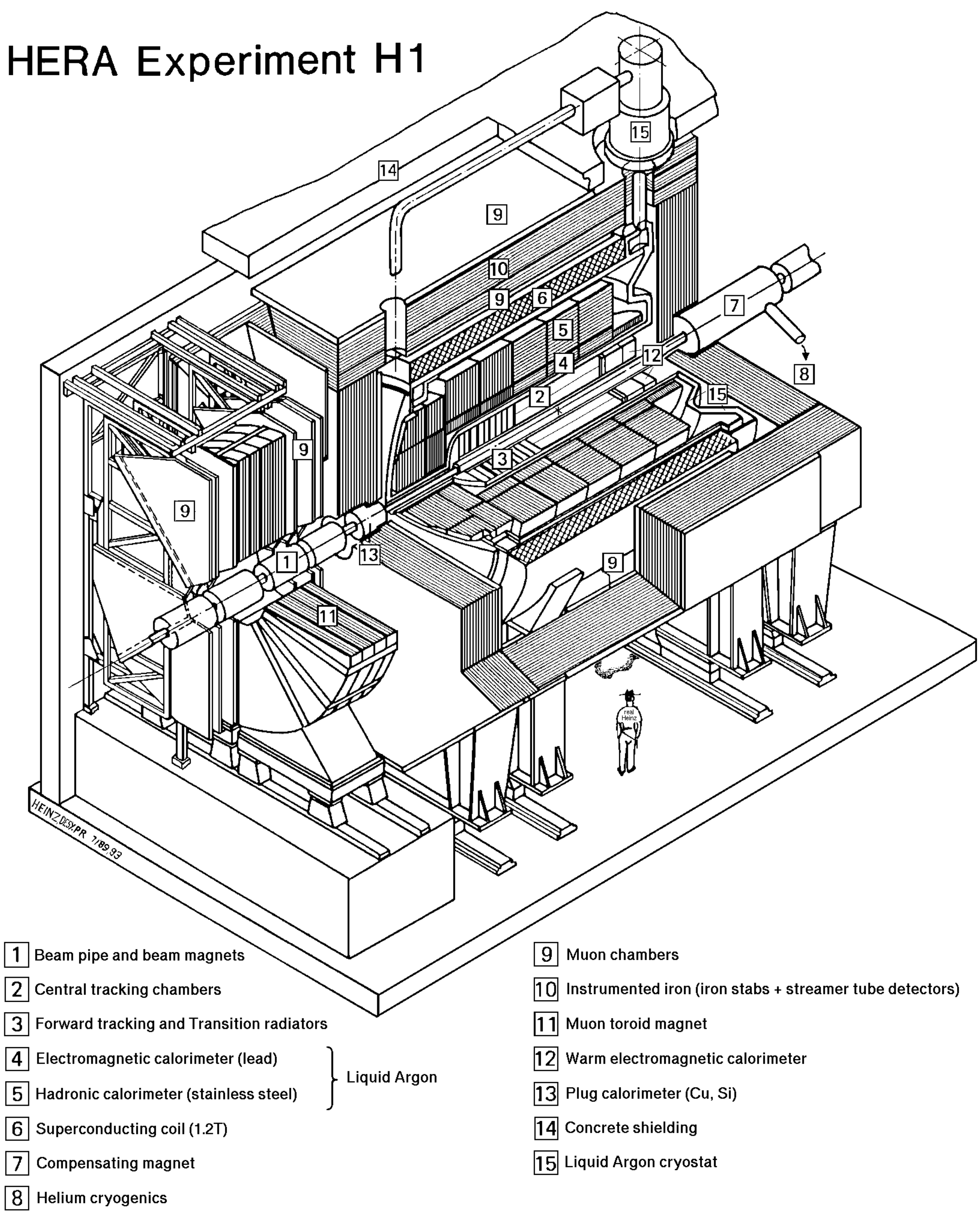

Figure 2.1: *An overview of the H1 experiment, showing the main detectors around the interaction point.*

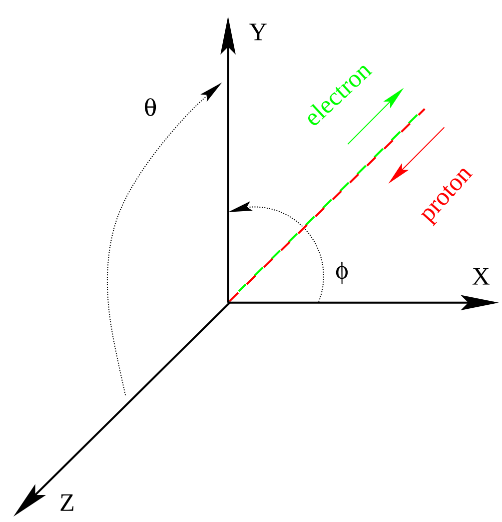

Figure 2.2: *Coordinates and angles in the H1 reference frame.*

with the collision time.
In addition to the detectors, the outgoing particles have to pass zones of 'dead material', like the beam pipe tube, the supports and the electronics, in which interactions may occur, too.
The heart of H1 is inside a magnetic field of 1.16 T, parallel to the z axis, provided by a superconducting magnet.
H1 has been taking data since 1992. From 2000 on it underwent a big upgrade before the high luminosity running of HERA started in 2002. In 1997 it had the configuration showed in figure 2.1.

2.1 The tracking chambers

The central part of the tracking system (fig. 2.3) consists of two MWPC (CIP and COP[1]) and four drift chambers (CIZ, CJC1, CJC2, COZ[2]); this system covers the angular range from 15° to 165°.
In the forward region, the tracking detector (FTD[3]) is made by three groups of chambers, each consisting of a set of planar drift chambers, MWPCs, transition radiators and radial drift chambers. The angular range of this detector is from 7° to 25°.
The backward tracking detector (BDC[4]) is a drift chamber covering the angular range

[1]respectively, *Central Inner Proportional Chamber* and *Central Outer Proportional Chamber*
[2]respectively, *Central Inner Z-Chamber, Central Jet Chamber 1* and *2, Central Outer Z-Chamber*
[3]*Forward Tracking Detector*
[4]*Backward Drift Chamber*

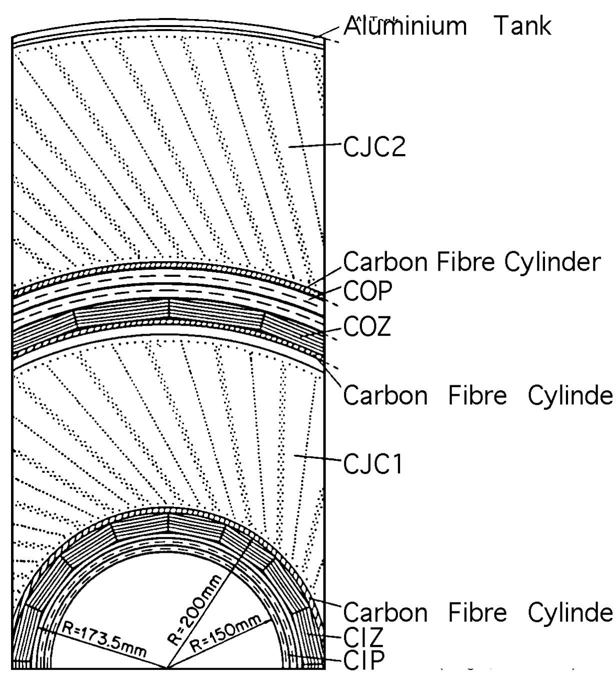

Figure 2.3: *The transverse section of the central tracking system. The CST is situated inside the CIP.*

from 153° to 177.5°.

2.2 The silicon detectors

The central silicon tracker (CST) is made of two silicon layers (fig. 2.4); it covers the region from 30° to 150°. This detector plays a fundamental role in the identification of tracks coming from the decays of long lived particles, like the b-hadrons. A detailed description can be found in [21].

The trajectory of a particle crossing both layers can be determined with a very good resolution (the single hit resolution is $\simeq 12$ μm in the (r,ϕ) plane plane and $\simeq 22$ μm along the z coordinate). Figure 2.5 a) shows the residuals the (r,ϕ) plane, from which the intrinsic resolution in the (r,ϕ) plane is obtained. Fig. 2.5 b) shows the inferred intrinsic resolution in the z-coordinate as a function of the incident angle of the particle; for small angles only one or two stripes collect the ionization of the crossing particles, for larger angles more stripes are involved and the measurement is improved, for very large angles the ionization is spread onto a very large area and only the border stripes contribute to the measurement.

The tracks are reconstructed from the hits in the tracking detectors; the resolution of

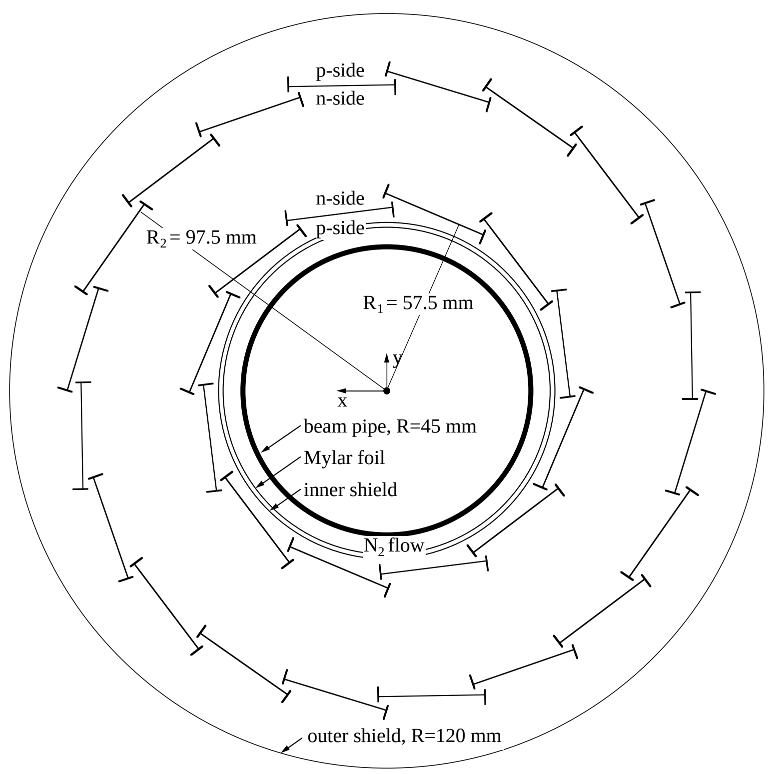

Figure 2.4: *The transverse section of the CST.*

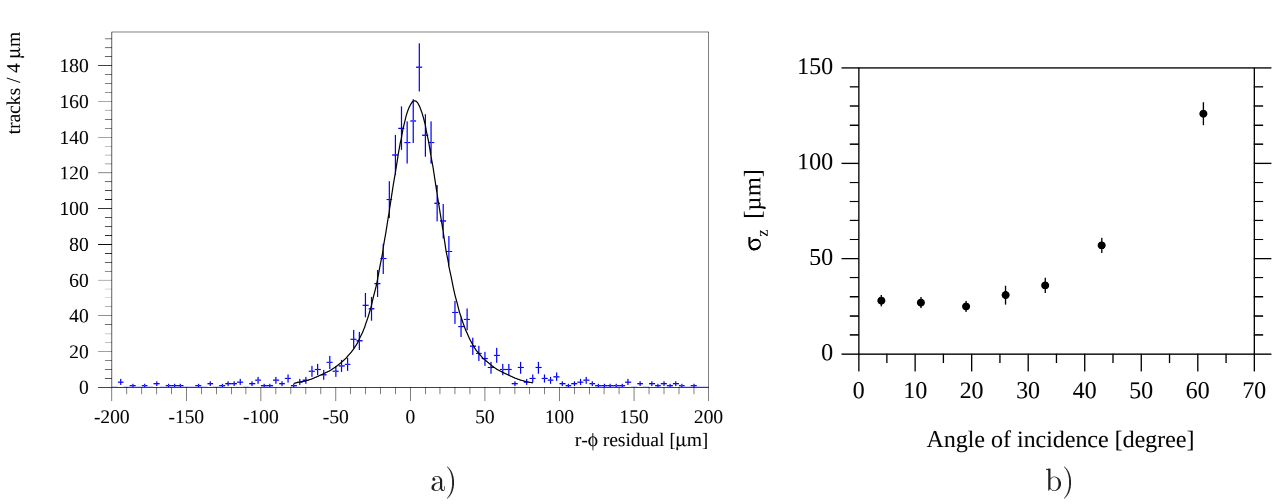

Figure 2.5: *a): Residuals in the (r,ϕ) plane b): Intrinsic resolution of the CST in z (from the n-side).*

central tracks is improved with the informations from the CST. In the (r,ϕ) plane a track is approximated as a circular arc, since the magnetic field is orthogonal to the (r,ϕ) plane; the sign of the curvature is the sign of the charge of the particle. The arc is extrapolated up to the point where the distance to the interaction point is minimal. This *distance of closest approach* (DCA) is a crucial parameter in our analysis and its sign is taken to be that of the vector product $\overrightarrow{DCA} \times \overrightarrow{P}$, where $\overrightarrow{P}$ is the momentum of the particle. The tangent to the arc drawn in the point of minimal distance intersects the x axis and defines the angle ϕ. Figure 2.6 shows how DCA is defined.

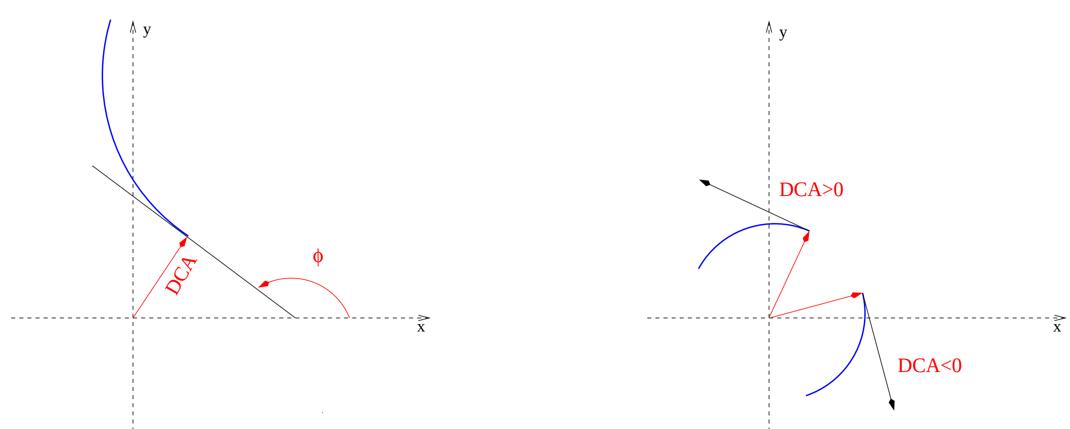

Figure 2.6: *Definition of DCA.*

The resolution of DCA gets contributions from the CST + CJC resolution, from the beam spot size and from the multiple scattering. These contributions can be made evident plotting the DCA resolution as a function of the angle of the track. The (r,ϕ) plane is divided in a number of ϕ sectors and DCA of tracks belonging to each sector is plotted. Then, each distribution is fitted with a gaussian: the value of σ is taken as the error on DCA. Plotting the values of σ against ϕ gives an estimation of the DCA resolution.
Around $\phi = \pm\, \frac{\pi}{2}$ it is made up by the track resolution plus the x component of the beam spot resolution (150 μm) and is therefore maximal, while around $\phi = 0$ and $\phi = \pm\, \pi$ it is made up by the track resolution plus the y component of the beam spot resolution (30 μm) and is therefore minimal. Figure 2.7 shows the distribution for tracks with a high momentum in the (r,ϕ) plane ($P_\perp > 4$ GeV), so that the multiple scattering is negligible. The distribution is fitted with the function

$$\sigma^2 = \sigma_0^2 + \sigma_x^2 sin^2\phi + \sigma_y^2 cos^2\phi \tag{2.1}$$

The CST intrinsic DCA resolution results $\sigma_0 = 54$ μm.
The backward silicon tracker (BST) is made of eight silicon disks; it covers the region from 162° to 176°.

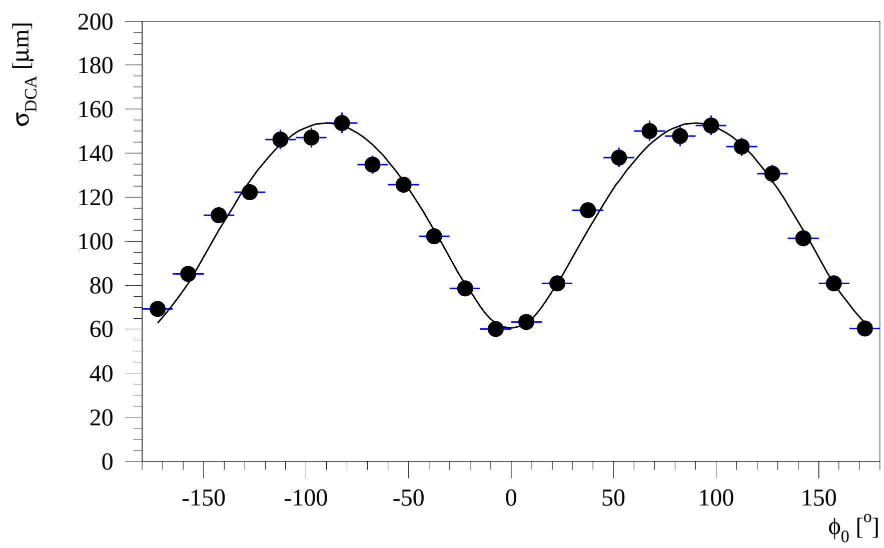

Figure 2.7: *DCA resolution (obtained with the CST) as a function of the track direction in the transverse plane.*

2.3 The calorimeters

The biggest calorimeter in H1 is the Liquid Argon Calorimeter (LAr). It covers the central and partially the backward region (angular range from 3° to 154°). It has an electromagnetic and an hadronic part (respectively, EMC and HAC in fig. 2.8), and is a sampling, non compensating calorimeter. The active material is liquid argon, the absorber medium is lead in the EMC and steel in the HAC. The energy resolution for electrons (in the EMC) is $\frac{\sigma(E)}{E} = \frac{0.12}{\sqrt{E}} + 0.01$ (here and below, E in GeV); for hadrons (EMC and HAC) it's $\frac{\sigma(E)}{E} = \frac{0.50}{\sqrt{E}} + 0.02$. To give an example of the resolution, figure 2.9 shows the $\gamma\gamma$ invariant mass from π^0 decay.
The forward region of H1 is closed with the Plug calorimeter; it alternates sampling sheets of silicon and copper absorber layers. The angular range is from 0.3° to 4°; the energy resolution is $\frac{\sigma(E)}{E} = \frac{1.50}{\sqrt{E}}$.
In the backward region there is the Spaghetti Calorimeter (SpaCal). The active material is scintillator fibers, inserted into grooves inside the lead absorber. It also consists of an electromagnetic and an hadronic part. The acceptance of the EM part is from 153° to 177.5°, with an energy resolution of $\frac{\sigma(E)}{E} = \frac{0.07}{\sqrt{E}} + 0.01$; the hadronic part has a resolution of $\frac{\sigma(E)}{E} = \frac{0.50}{\sqrt{E}}$.
Outside the liquid argon is the tail catcher (TC); it is constructed within the iron of the return yoke of the magnet. It measures the energy of hadronic showers not fully contained in the LAr and muons, in the angular range from 4° to 172°, which includes

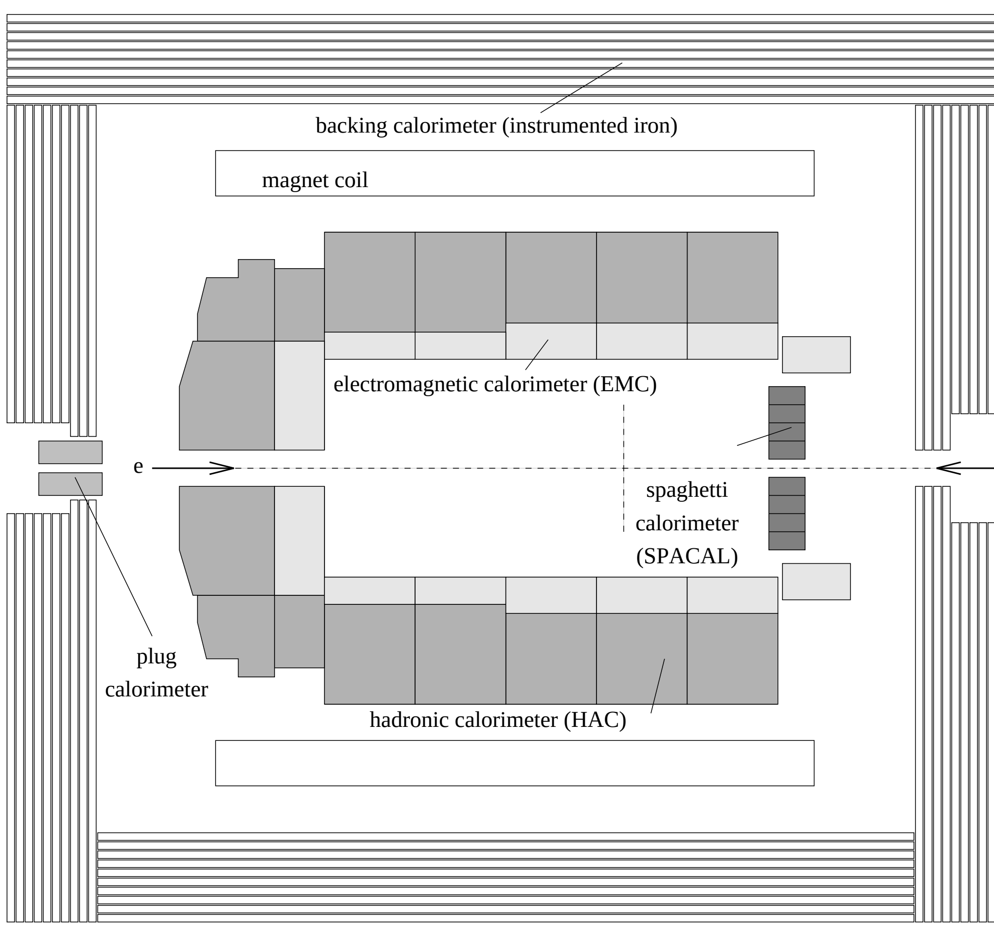

Figure 2.8: *The longitudinal section of the H1 calorimetry system, with the main calorimeters.*

also the backward part.

2.4 The muon detector

In the central region, the muon detector is built by layers in the instrumented iron of the magnet in addition to layers of streamer tubes. The resolution on the track position is of the order of a few mm. For muons with $P_\perp > 2$ GeV the efficiency is around 90 %.
In the forward region, the muon detector is made of six drift chambers and covers the angle from 3° to 17°.

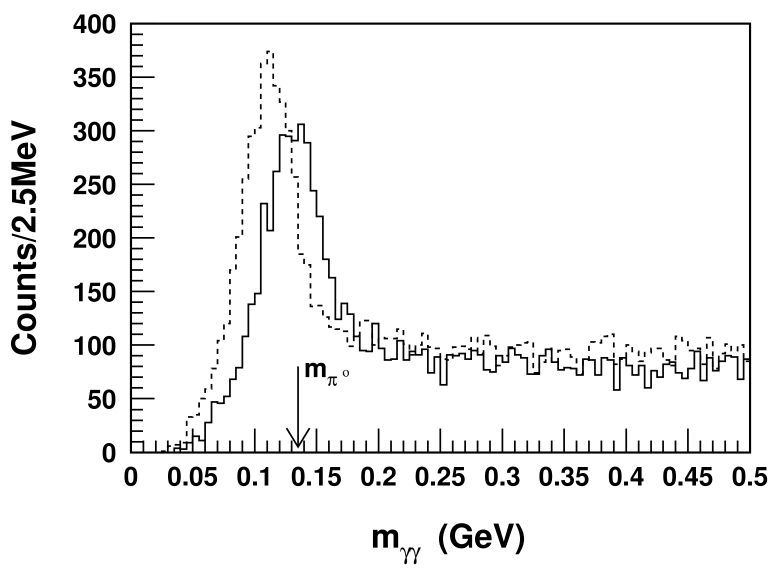

Figure 2.9: *The $\gamma\gamma$ invariant mass from π^0 decay. The spectrum before the correction for the dead material is also shown (dashed line).*

2.5 The trigger scheme

The trigger system of H1 has to pick up the rare eP interaction events out of the huge amount of non interesting events, in the order of 1 *physics* event out of 10^4 *background* events (the background events are mainly interactions with the residual beam gas and cosmics). The decision has to be fast, to keep up with the bunch collision rate of $\frac{1}{96\ ns} \simeq 10$ MHz.

The trigger is divided in 4 levels. The first level (L1) is made by subtriggers. Simple algorithms make decisions from different detectors signals, coincidences and threshold requirements; 128 subtriggers are available, which can be distinguished in three different groups:

- Vertex triggers - the tracking systems and the ToF scintillators determine if the event vertex is inside the fiducial interaction volume (defined around the collision point).
- Calorimetric trigger
- Muon trigger

The signals are fed to pipelines; this allows to keep signals from several (24) subsequent events, in order to give the trigger enough time to analyze the event. This system has zero dead time. Each subtrigger is *prescaled*, to take into account the different time needed by each detector to give an output signal (for a fast detector only a fraction of interaction events is kept, otherwise the corresponding part of the event that is being triggered would overlap to the next one). The subtriggers are put in logical OR. If one trigger gives the ok, then the event is kept for the next step, otherwise the whole event is dropped from the pipeline. Once an event is triggered from L1, a readout signal is sent to the front-end electronics of the systems, and the primary dead time starts to accumulate.

The second level trigger (L2) combines the L1 informations to perform a first reconstruction of the event, during the readout dead time; it gets additional informations from most of the subdetectors. According to the subsystems involved, different algorithms are run. A decision whether to keep the event or not is made within 20 μs.

A third level trigger (L3) was, in 1997, under construction, with a decision time of $\simeq$ 200 μs.

The output of the L2 (L3) trigger is fed to the level four (L4) farm filter, which analyzes all the detector informations, makes a decision and does the reconstruction of the selected events. Once the informations are given to L4, the dead time ends. The L4 filter works asynchronously, and writes events to tape at a rate of $\simeq$ 10 Hz. The overall dead time is 1-2 ms.

Further on, a full event reconstruction is performed (L5), using the energy scales and tracking parameters calculated for each run; applying appropriate cuts, the event is classified and written out for physics analysis.

Chapter 3

Event selection

The identification of beauty events at HERA demands the selection of a very small number of *good* events among a huge background or other eP events. The cross section for the production of b quarks is very small ($\sim$ 100 times smaller than for the charm production, which, in turn, is $\sim$ 10 times smaller than for the light quarks production). The identification of the B hadron can be done using the lifetime signature - the long life is visible in the *impact parameter* spectrum and in the existence of secondary vertices - and the big mass - visible in the $P_\perp$ spectrum of the decay products with respect to the jet axis. The large mass allows many decay channels for B hadrons, each having a small branching ratio; large mass also allows decays with many tracks, not easily distinguishable from the background.

At present, only the electron and muon channels has been studied (see, for example, [22, ..., 32]). The development of an inclusive method (as used at Tevatron and LEP [33]) to tag the secondary vertices from the decay of long lived particles promises an a priori gain in statistics of a factor 10, at the expense of an increased impurity in the non muonic channel, which has to be handled. A more efficient inclusive tag would hopefully allow to make measurements at higher $P_\perp$ and higher Q^2, and open the possibility to search for exotic channels.

The idea of this work is to test such a tag on a sample where the b content is already known and can be enriched: a sample of dijet events with an identified muon.

3.1 Selection on data

This work makes use of photoproduction data collected by H1 in 1997 (cfr. [24]). The data selection has been developed in two steps. First the sample has been reduced requiring at least two jets and a central muon. Only runs where the main systems of the detector (CST, CJC, MWPC, BDC, digital muon chambers, LAr calorimeter, SpaCal, ToF and luminosity systems) were fully operational are taken; the luminosity is $\mathcal{L} \simeq 11.6$ pb^{-1}. Then, a tighter jet and muon selection is applied, in addition to a track selection. Rejection cuts against DIS events are applied. The subtrigger S19 is required: it is a logical combination of trigger elements that recognize central muons with high $P_\perp$, and reads

$$zVtx\_sig > 1 \wedge Mu\_Bar \wedge DCRPh\_Tc \wedge DCRPh\_TNeg \wedge DCRPh\_THig$$

where $zVtx\_sig > 1$ means a successful identification of the primary vertex, Mu_Bar means a muon identified in the barrel, the $DCRPh$ trigger elements mean a successful reconstruction of tracks in the CJC.
The selection cuts are summarized in table 3.1. After the preliminary selection, the muon track undergoes the general track selection. The muon is also required to be within one of the selected jets.

EVENT	MUON	JETS	TRACKS
$0.1 \leq$ y ≤ 0.8 $E_{SpaCal} \leq 8$ GeV			
	$P_\perp \geq 2$ GeV $35^o \leq \theta \leq 130^o$ $l_{CJC} \geq 22$ cm $r_{0\ CJC} \leq 35$ cm $r_{VX} < 100$ cm $\Delta_{z_0} < 100$ cm ≥ 2 iron layers $I_{first} \leq 5$ $I_{last} \geq 2$ $p_{comb} \geq 0.001$		
		$P_\perp \geq 5$ GeV $10^o < \theta < 170^o$ $\#\mathrm{tr}_{CSPRIM} \geq 2$	
			$P_\perp \geq 1$ GeV $35^o \leq \theta \leq 135^o$ $\#\mathrm{hits}_{CST} \geq 2$ $l_{CJC} \geq 50$ cm $r_{0\ CJC} \leq 45$ cm $p_{fit} \geq 0.1$ $\mid \mathrm{DCA}' \mid \leq 0.3$ cm $\mid \Delta_{DCA'} \mid \leq 0.013$ cm V0 rejection

Table 3.1: *Cuts applied in the event, muon, jet and track selection.*

The selection of photoproduction events is made through the cut on the inelasticity y and

the rejection of DIS events, in which the scattered electron is identified by an electromagnetic cluster in the SpaCal with an energy (E_{SpaCal}) above 8 GeV.
In order to achieve a good measurement of the physical quantities, in particular to benefit from the high resolution of the CST, the jets and the tracks have to be central. The cuts on $P_\perp$ and on the polar angle θ select central particles.
The quality of the reconstruction of the primary vertex is guaranteed by the requirement of a minimum number of tracks ($\#\mathrm{tr}_{CSPRIM}$) used for the vertex fit (see section 4.1).
The quality of the track reconstruction is guaranteed by the cut on the minimum number of hits in the CST ($\#hits_{CST}$), by the quality of the reconstruction in the CJC (length l_{CJC}) and by the cut on the minimum probability for a successful linking of the CJC and the CST track informations (p_{fit}). The requirement on the starting radius ($r_{0\ CJC}$) of the CJC track is made to recover from a non sensitive zone in the CIP.
The muon track undergoes a special selection/identification. Tighter cuts on $P_\perp$ and on the CJC starting radius and track length are applied. The distance to the primary vertex, both in the (r,ϕ) plane (r_{VX}) and along z (Δ_{z_0}), has to be smaller than 1 m. The identification in the instrumented iron is made by the requirement of a minimum number of fired layers and on the index of the first and the last layer (respectively, I_{first} and I_{last}); a minimum probability for the successful combination of the CJC and iron informations (p_{comb}) is also required.
The quality of the vertex and track reconstruction is further controlled by cutting on the value and the error of DCA$'$, the distance of closest approach recalculated with respect to the reconstructed vertex (see section 4.1).
Particles like K's and Λ's have a quite long lifetime and can make secondary vertices when decaying. The algorithm V0FIND[1] maps the tracks originating in secondary vertices coming from the decay of a neutral particle. These tracks are rejected.
Two jets are used for the tag: the one that contains the muon (*muon jet*) and the one with highest $P_\perp$ among the others (*other jet*). At least one good track (to be used for the tag) is requested in the event; $\sim$ 20 % of the events have at least one good track in the *other jet*.
The selection produces a sample of 636 events.

3.2 Background contributions and simulated signal

Events which pass the jet and track selection can pass the muon selection as well, if one of the tracks has high $P_\perp$ and a low ionization in the calorimeters. Typically, such tracks belong to hadrons that are identified as muons, mainly light mesons.
To quantify this background, a Monte Carlo sample of light quarks events has been built (extracted out of an allflavour Monte Carlo generated using PYTHIA, version 61).
The main source of background is given by $c\bar{c}$ events. The $c\bar{c}$ events will pass the same

[1]code belonging to the H1 software

jet, track and muon selection, and can be separated from $b\bar{b}$ events only on the basis of the differences in the quarks mass and lifetime.
B mesons decay to D mesons in some 7% to 8%; thus, such a percentage of charm events is really present in the data and cannot be rejected as background.
Two Monte Carlo samples of $b\bar{b}$ and $c\bar{c}$ events have been used (generated with AROMA, version 22); the data selection has been repeated on these samples.
We also use our primary data sample with the jet and track selection applied, but with a veto on subtrigger S19 and a muon rejection. The resulting sample receives mainly light hadrons and will be called *hadron data*. With identical cuts applied to the allflavour Monte Carlo, a *hadron Monte Carlo* sample is obtained. These samples are used for checks on the quality of the reconstruction of the physical quantities and on some of the variables used to build the tag.

Efficiency	Selection			
	muon	jet	track	S19
ϵ_{μ}	0.40			
$\epsilon_{\mu J}$	0.12			
$\epsilon_{\mu JTr}$	0.076			
$\epsilon_{Trigger}$				0.81±0.02
$\epsilon_{analysis}$	0.062±0.002			

Table 3.2: *Selection and trigger efficiencies.*

Sample	Selection				#selected events (scaled to $\mathcal{L}$=11.6 pb^{-1})
	muon	jet	track	S19	
dijet + μ	$\surd$	$\surd$	$\surd$	$\surd$	636
hadron data	*vetoed*	$\surd$	$\surd$	*vetoed*	1535
$b\bar{b}$ MC	$\surd$	$\surd$	$\surd$	$\surd$	2887 (17)
$c\bar{c}$ MC	$\surd$	$\surd$	$\surd$	$\surd$	7412 (543)
uds MC	$\surd$	$\surd$	$\surd$	$\surd$	1203 (192)
hadron MC	*vetoed*	$\surd$	$\surd$	*vetoed*	9022

Table 3.3: *Data and Monte Carlo samples, and numbers of selected events scaled to the luminosity of the data sample.*

Tab. 3.2 shows the efficiencies of the muon, jet and track selections, one on top of the other, and the trigger efficiency (on top of the selection efficiencies); the jet efficiency includes the requirement that the muon is found within one of the selected jets, the track efficiency includes the requirement that the combined probability has a finite value. All these numbers are combined to give the overall efficiency of the analysis.
Tab. 3.3 summarizes the samples used in this work and shows the number of selected events scaled to the luminosity $\mathcal{L}$=11.6 pb^{-1}. The b-tag will be built on these events.

Chapter 4

Development of the b-tagging tool

The general idea of the tagging developed in this work is to separate $b\bar{b}$ events from $c\bar{c}$ and light quarks events through their long lifetime. A statistical approach is used: no explicit reconstruction of the secondary vertices is made, but the overall *presence* of secondary vertices in the event is estimated by calculating the probability that a group of selected tracks all come from the same primary vertex. The method consists of the following steps:

1. Reconstruction of the primary vertex and the tracks - these quantities are the basic building blocks of the event. It is crucial to make use of the highest possible precision and to control the resolution, since it influences directly the significance of the lifetime information we're trying to extract. Once the primary vertex and the tracks are reconstructed, the distance of closest approach is recalculated with respect to the fitted vertex. The new distance of closest approach will be called DCA$'$.

2. Determination of the impact parameter δ, which is equal to DCA$'$ if the track crosses the jet axis, and equal to $-$DCA$'$ otherwise.

3. Track quality - both the reconstructed vertex coordinates and the track parameters have errors, and the error on δ gets contributions from both. The larger the error on δ is, the less significant is the track.

4. Single track probability - the distribution of the significance has to be parametrized to calculate the probability that each single track comes from a given (primary) vertex; this step takes into account the weight of each track, and makes it known to the next step.

5. Combined track probability - it is the goal of the procedure: after this step, an observable is available to estimate the b content of the sample.

4.1 The reconstruction of the primary vertex and the tracks

In a real event, the tracks and the vertex are measured with a limited precision.
The reconstruction of tracks gets the highest precision combining the informations from the CST and the other trackers. The reconstruction of the vertex uses informations from several events (determination of the beam spot) and is further improved using the information from the tracks in an event (fit with the CSPRIM routine[1]). The errors on the tracks and the vertex give contributions of similar sizes to the error of DCA′; they have to be combined together. A good description in the Monte Carlo is demanded.
The track resolution alone is visible in the DCA and DCA′ pull plot (fig 4.1).

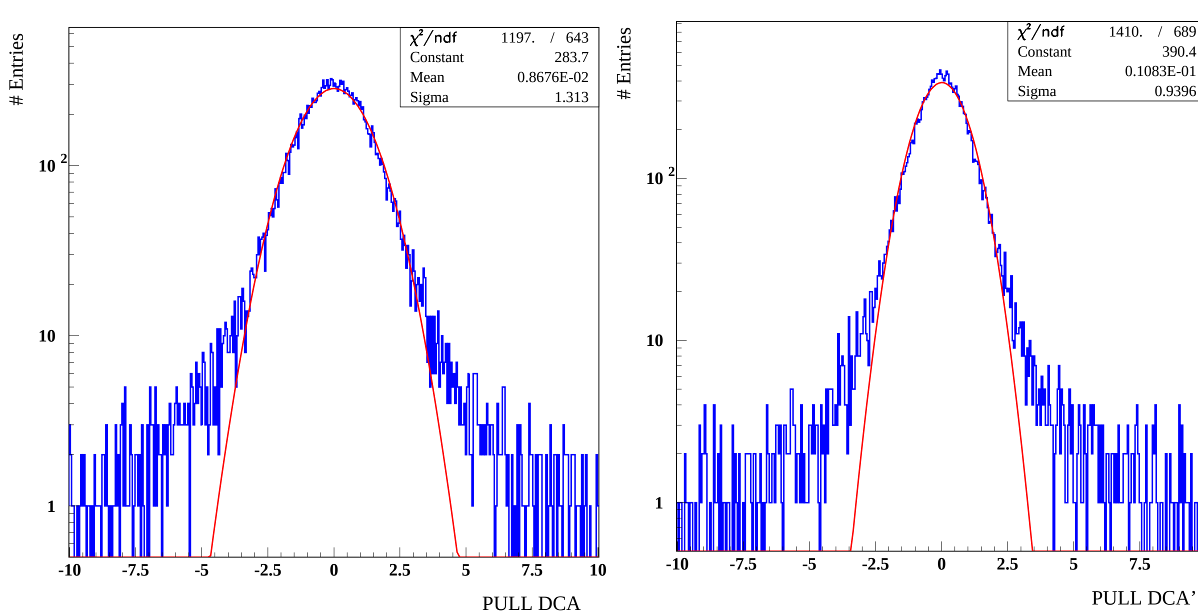

Figure 4.1: *Left: pull plot for DCA in the hadron Monte Carlo sample. Non gaussian tails are evident; the gaussian core has a width of $\simeq$ 1.3. Right: pull plot for DCA′ (same sample).*

The primary vertex is obtained from the CSPRIM fit; the calculation is repeated for each jet, excluding from the loop the *good* tracks in the jet itself, such that they do not bias the vertex reconstruction. From now on, the reconstructed primary vertex is thus a jet depending quantity. Fig. 4.2 shows the number of jets per event in the different samples. The errors on the x and y coordinates of the vertex are shown in fig 4.3, for hadron data (background) and Monte Carlo (hadron Monte Carlo sample). The agreement is good; in the Monte Carlo, there are cuts on the maximal values given by the widths of the beam

[1]code belonging to the H1 software

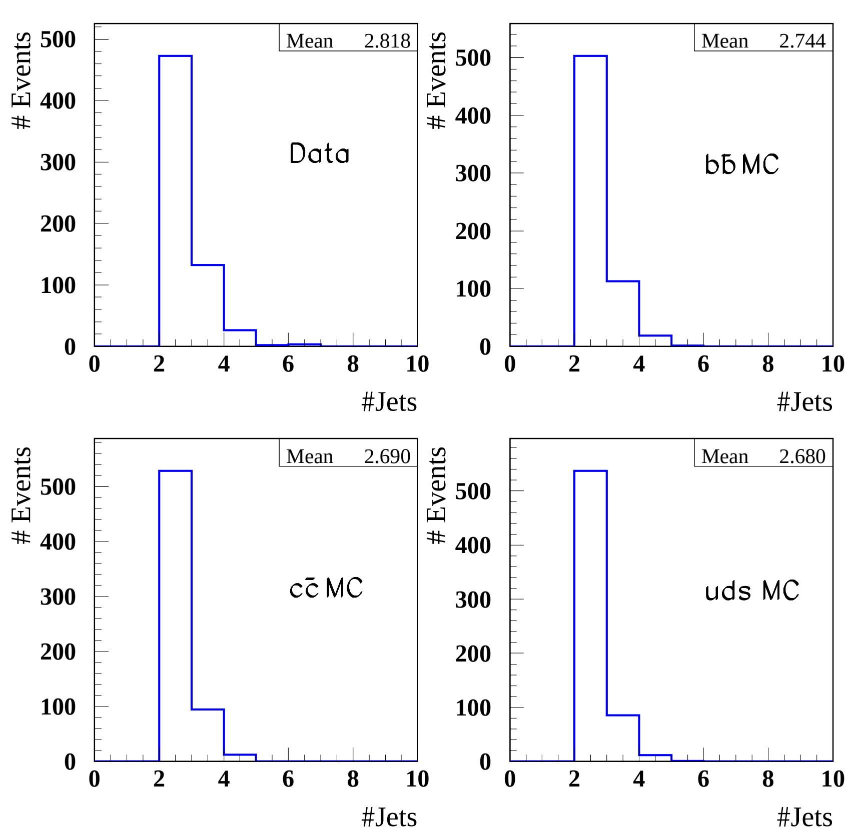

Figure 4.2: *Number of jets per event in data and Monte Carlo samples.*

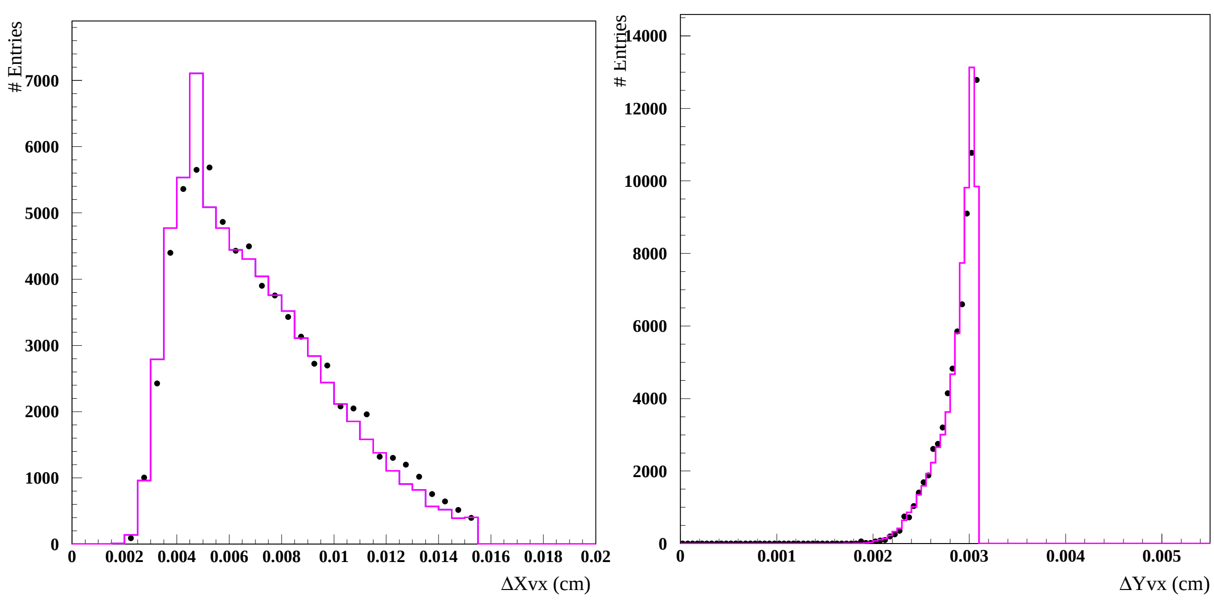

Figure 4.3: *Error on the X and Y coordinates of the CSPRIM vertex. Dots represent hadron data, the line is hadron Monte Carlo.*

spot.
The errors on the tracks and the vertex can be observed together if the DCA resolution is plotted against the ϕ angle of the tracks (see section 2.2). Fig. 4.4 shows the plot obtained for reconstructed tracks and the first-step-reconstructed vertex (the center of the beam spot); all tracks with $P_{\perp} \geq 1$ GeV are used. Fig 4.5 shows the same plot, but now the vertex is calculated with CSPRIM (one vertex per event). The improvement on the DCA$'$ resolution is noticeable, expecially for vertical tracks ($\phi = \pm\, \pi$), where the vertex error was much bigger. The good agreement between data and Monte Carlo for horizontal and vertical tracks shows that the beam spot, the event vertex and the track resolution are well under control.
Fig 4.6 illustrates the calculation of DCA$'$. The error on DCA$'$ results from the combination of the errors on the vertex (x, y component plus the correlation term) and on the track:

$$\Delta^2_{DCA'} = \Delta^2_{DCA} + \Delta^2_{VX_x} \times sin^2\phi + \Delta^2_{VX_y} \times cos^2\phi + \Delta^2_{VX_{xy}} \times sin2\phi \tag{4.1}$$

where ϕ is the angle of the track. There is a good agreement between the $\Delta_{DCA'}$ distribution in data and Monte Carlo (fig. 4.7).

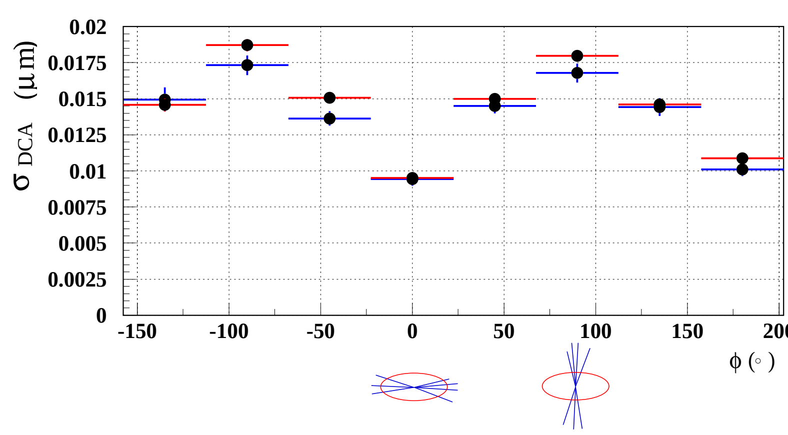

Figure 4.4: *Reconstructed tracks with the center of the beam spot taken as primary vertex. Red dots mean hadron data, blue dots mean hadron Monte Carlo.*

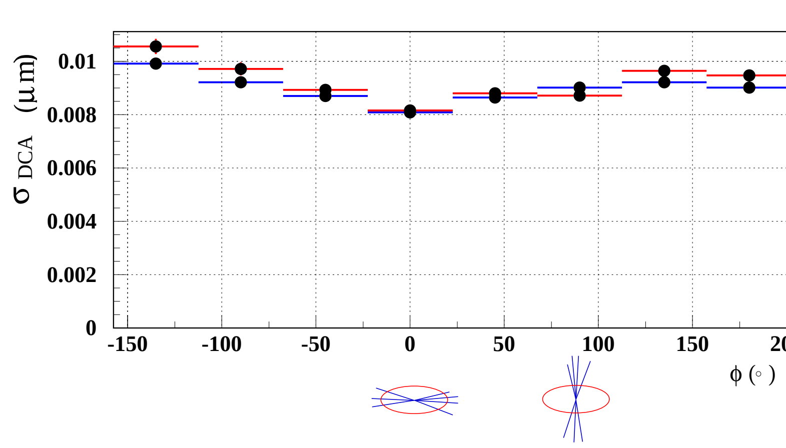

Figure 4.5: *Reconstructed tracks with a fully reconstructed primary vertex (CSPRIM). Red dots mean hadron data, blue dots mean hadron Monte Carlo.*

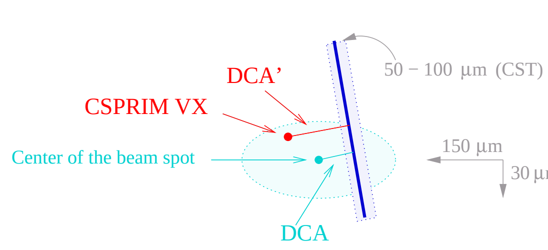

Figure 4.6: *Calculation of DCA of a reconstructed track w.r.t the reconstructed vertex from CSPRIM.*

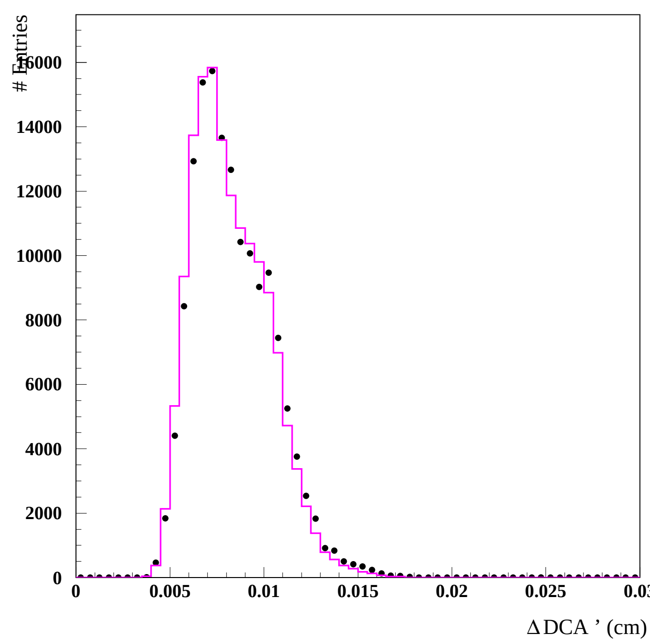

Figure 4.7: *The error on DCA'; the blue curve shows the distribution from the hadron data sample, the dots show the distribution from the hadron Monte Carlo.*

4.2 The impact parameter and its significance

The quantity |DCA'| indicates how far from the primary vertex the track originated, in the projection on the (r,ϕ) plane; δ is defined as

$$|\delta| = |DCA'|, \qquad \Delta_\delta = \Delta_{DCA'}$$

To calculate the sign of δ, the jet axis is also projected on the (r,ϕ) plane, as a vector with the origin in the primary vertex (see fig. 4.8); the track is approximated with a straight line. If the track intersects the projection of the jet axis then $\delta > 0$, if the tracks intersects the prolongation of the jet axis then $\delta < 0$. In other words, if $\delta > 0$ the track belongs to the jet, if $\delta < 0$ it doesn't. The significance of δ is calculated as

$$S(\delta) = \frac{\delta}{\Delta_\delta} \tag{4.2}$$

If a particle has a short life, its track is expected to originate in the primary vertex; the finite resolution of the detector will appear in a non zero value of δ and in a non zero error Δ_δ, both not depending on the geometry if the detector is isotropic. Thus, the distribution of S will be symmetric around 0. In the light quark Monte Carlo, where no lifetime effects are expected, the distribution of $S(\delta)$ can in principle be used to determine

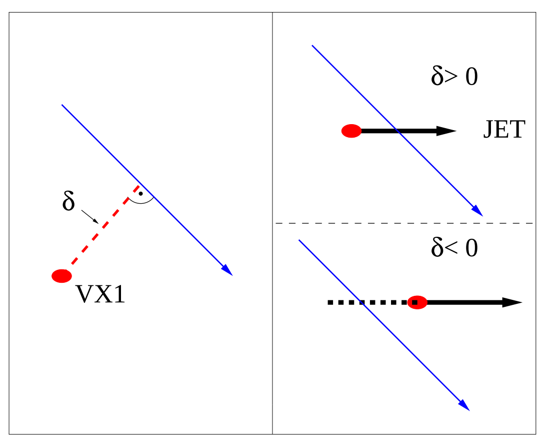

Figure 4.8: *Calculation of δ*

the detector resolution. A fit on the negative part of the distribution is expected to give the resolution function. In the sample available for this analysis, the limited statistics makes it impossible to describe the tails of the distribution. It will be assumed that the detector resolution is also well described by the distribution of $S(\delta)$ in the hadron samples. Fig. 4.9 shows the negative part of $S(\delta)$ in signal and Monte Carlo samples. The detector resolution will be determined from the hadron Monte Carlo sample (fig. 4.10). The function

$$R(S) = \Sigma_{i=1}^{3} C_i e^{-(\frac{(S-\mu_i)}{\sigma_i})^2} \tag{4.3}$$

is well suited (see fig 4.10). The fit yields the following parameters:

$$C_{1,2,3} = 9059.5\ (\pm 146.)\ ,\ 1058.\ (\pm 157.)\ ,\ 13.45\ (\pm 1.3)$$
$$\mu_{1,2} = 0.\ ,\ 0.\ ,\ 0.\ \text{(fixed values)}$$
$$\sigma_{1,2,3} = 1.1084\ (\pm 0.012),\ 2.0355\ (\pm 0.07),\ 12.343\ (\pm 0.65)$$
$$\chi^2 = 1.2$$

This parametrization has the additional benefit, that it allows a fast calculation of the probability.
The central part of the S distribution is, to a good approximation, gaussian; it represents particles from the primary vertex, where the tracks were well reconstructed. On the contrary, the symmetric tails represent particles from the primary vertex where the tracks were not well reconstructed or the error underestimated; the asymmetric tails represent well reconstructed tracks of particles from secondary vertices (for example K's).

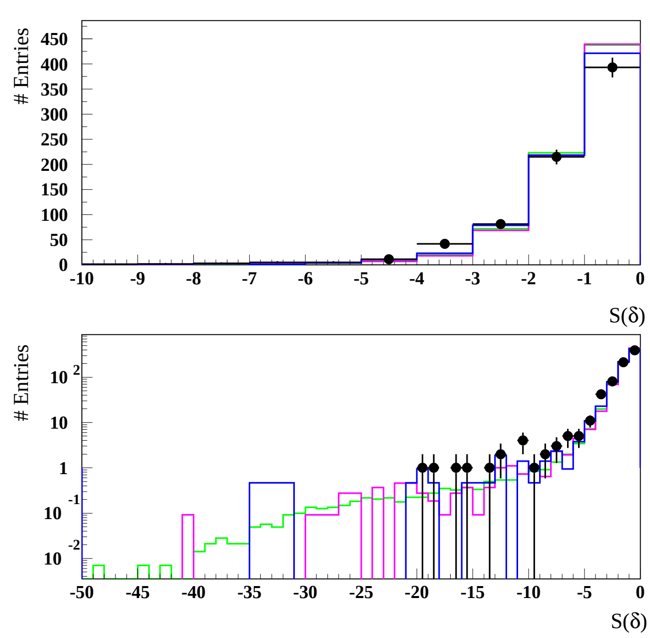

Figure 4.9: *Negative part of the significance of δ. The distribution in the signal data (dots), signal uds (blue curve), hadron data (pink) and hadron Monte Carlo (blue) are shown; the same plot is repeated in log vertical scale, to see the tail.*

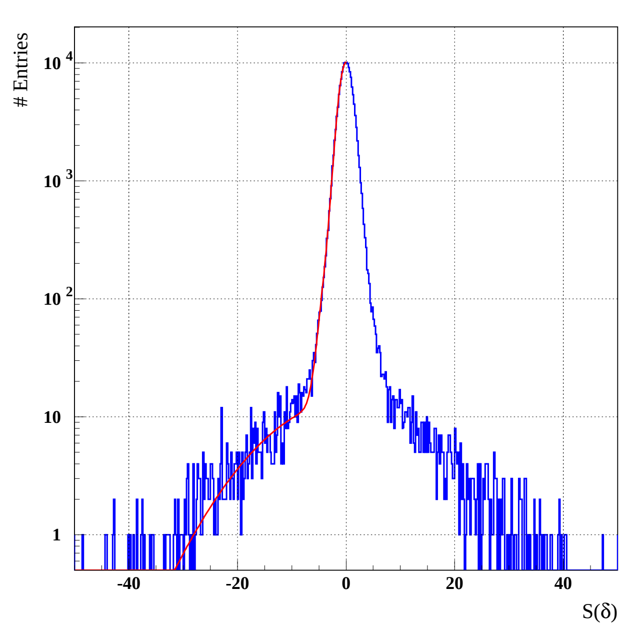

Figure 4.10: *Significance of δ, in the hadron Monte Carlo sample; a fit with three superimposed gaussians is performed on the left side of the distribution.*

4.3 The single track probability

The probability that a track comes from the primary vertex (the one associated to the jet the track belongs to) is calculated as:

$$SP(S_0) \;=\; \frac{\int_{|S_0|}^{\infty} R(S)dS}{\int_0^{\infty} R(S)dS} \tag{4.4}$$

In our case, it can be simplified to the following decomposition:

$$SP(S_0) \;= \Sigma_{i=1}^{3} C_i \; prob \; ((\frac{(S-\mu_i)}{\sigma_i})^2, 1) \tag{4.5}$$

It is shown in fig 4.11, respectively for the light quarks sample, for the $c\bar{c}$ and for the

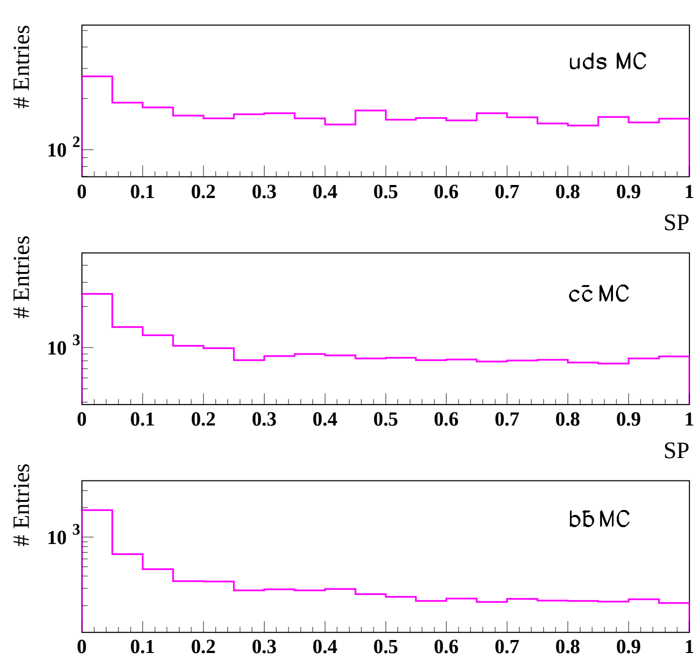

Figure 4.11: *Single track probability; respectively the distribution in the uds, $c\bar{c}$, $b\bar{b}$ Monte Carlo.*

$b\bar{b}$ Monte Carlo. In the $c\bar{c}$ and $b\bar{b}$ samples an excess of tracks not coming from the primary vertex is more and more evident.
The available informations from each single track are then combined to get the maximum efficiency for the tag.

4.4 The combined track probability

In the next analysis step the probabilities for each track are combined to define the tagging variable. Only tracks with a positive value of S are used:

$$CP = Q\ \Sigma_{i=0}^{n-1} \frac{(-lnQ)^i}{i!} \qquad\qquad Q = \Pi_{i=1}^{n}\ SP_i \tag{4.6}$$

where n is the number of tracks used by the tag and SP_i is the probability of each track as calculated above.

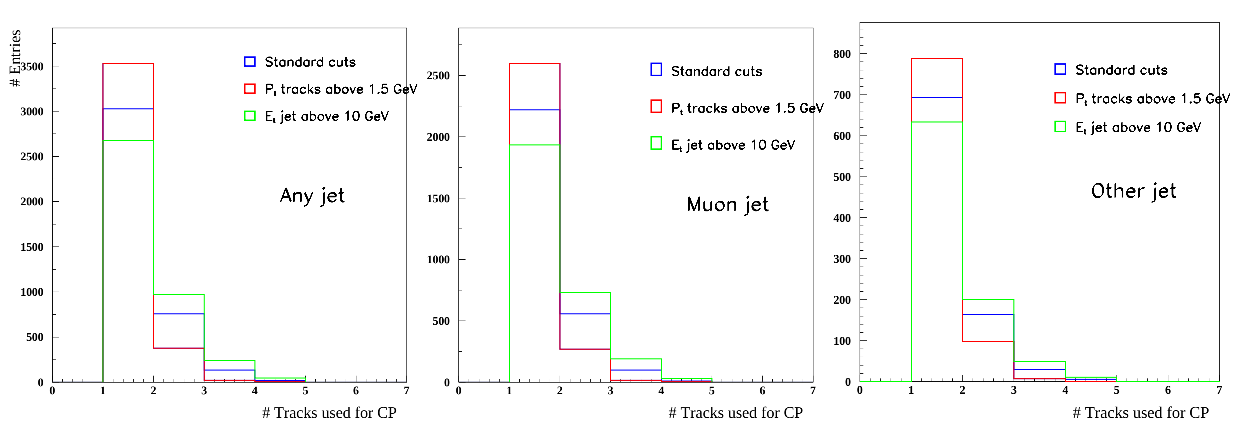

Figure 4.12: *Number of tracks selected for the tag with different selection cuts (respectively: in any jet, in the muon jet and in the other jet); the sample under consideration is the $b\bar{b}$ Monte Carlo.*

The number of tracks used by the tag is extremely small (see fig. 4.12, obtained from the $b\bar{b}$ Monte Carlo with different selection cuts; the number of tracks per jet is shown, for the two jets used); nevertheless, it is possible to build up a variable which can be used to separate the b and c content from the background: fig. 4.13 and 4.14 show the combined probability in the uds, $c\bar{c}$ and $b\bar{b}$ samples.
Fig. 4.15 shows a comparison between the hadron data and the hadron Monte Carlo; the agreement is good.

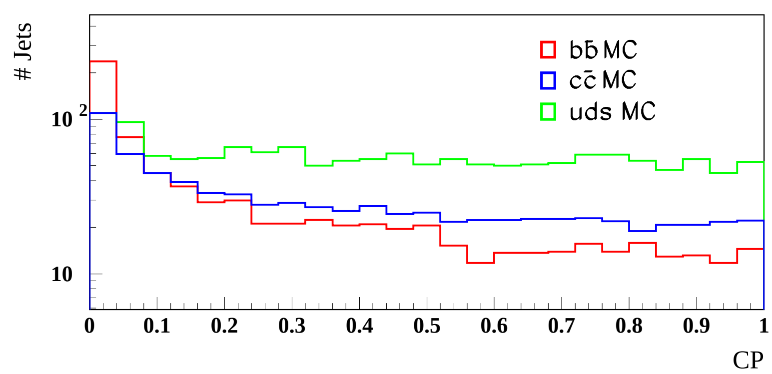

Figure 4.13: *Combined probability for any jet in the $b\bar{b}$, $c\bar{c}$ and uds sample. The distributions are normalized, to compare the relative shapes.*

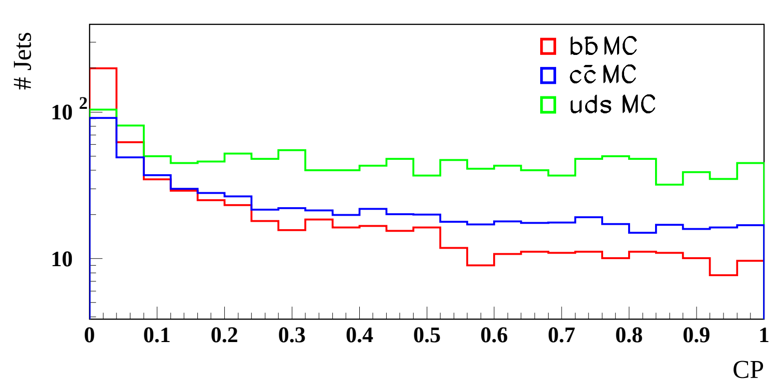

Figure 4.14: *Combined probability for the muon jet in the $b\bar{b}$, $c\bar{c}$ and uds sample. The distributions are normalized, to compare the relative shapes.*

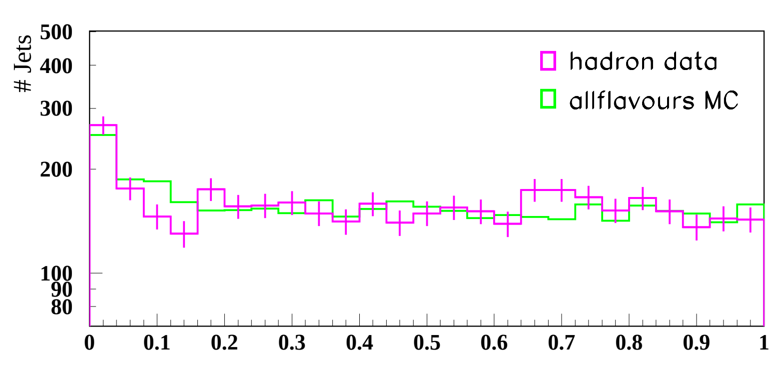

Figure 4.15: *Combined probability in the hadron Monte Carlo and in the hadron data. The distributions are normalized, to compare the relative shapes.*

Chapter 5

The beauty content in the data sample

Once the events have been selected, it's important to determine the composition of the sample. This is done on a statistical basis, comparing the distribution of a variable in the data, signal and background samples.

5.1 The Barlow-Beeston fit

This method calculates the contributions of different sources to the number of events observed in the data. For a full description and explanation, see [35].
A physical observable is chosen and its values are distributed in a number of bins (n). This is done for the data and Monte Carlo samples (signals and background, m sources in total); in each bin, the number of observed data is given by

$$d_i, \qquad \Sigma_{i=1}^{n} d_i = N_D \tag{5.1}$$

and the number of observed Monte Carlo data from source j is given by

$$a_{ji}, \qquad \Sigma_{i=1}^{n} a_{ji} = N_j \tag{5.2}$$

The predicted number of events in bin i is given by

$$f_i(P_1, ...P_m) = N_D \Sigma_{j=1}^{m} \frac{P_j a_{ji}}{N_j} \tag{5.3}$$

where P_j is the actual proportion of the jth Monte Carlo source, or, equivalently:

$$f_i = \Sigma_{j=1}^{m} p_j a_{ji}, \qquad p_j = \frac{N_D P_j}{N_j}, \qquad \Sigma_j P_j = 1 \tag{5.4}$$

In case d_i and f_i are small - which is the actual case, given the spread of the entries into several bins - the probability to observe a value d_i is described by the Poisson distribution

$$e^{-f_i} \frac{f_i^{d_i}}{d_i}! \tag{5.5}$$

Given the finite size of the Monte Carlo sample - which may be comparable to the data sample - fluctuations on a_{ji} have also to be considered; this is done introducing the expected number of Monte Carlo events from source j in bin i:

$$A_{ji} \tag{5.6}$$

A_{ji} gives the observed a_{ji} through a binomial distribution; in case $A_{ji} \ll N_j$ - which is the actual case - it can be approximated with a Poisson distribution.
So, the solution of the problem is the maximization of the total likelihood

$$lnL = \Sigma_{i=1}^{n} d_i ln f_i - f_i + \Sigma_{i=1}^{n} \Sigma_{j=1}^{m} a_{ji} ln A_{ji} - A_{ji} \tag{5.7}$$

which is the combined probability to observe d_i and a_{ji}; the first part accounts for the fluctuations in d_i and the uncertainties on p_j, the second part accounts for the fluctuations in a_{ji} (no uncertainties on A_{ji}).
To maximize the likelihood, $ln(L)$ is differentiated with respect to p_j and A_{ji} and the derivatives are set to zero; a system of $m \times (n+1)$ simultaneous equations (non linear and coupled) arises. The solution is done via numerical integration, and provides p_j and A_{ji}; p_j is the one which is interesting for the physics analysis. From the algebra comes:

$$\Sigma_{i=1}^{n} A_{ji} = \Sigma_{i=1}^{n} a_{ji} \qquad \forall j \tag{5.8}$$

which means that the expected shape of the Monte Carlo source j may differ from the observed one, but the total number of entries is conserved; as for p_j, one gets the normalization:

$$N_D = \Sigma_{j=1}^{m} p_j Nj \tag{5.9}$$

This clarifies the meaning of p$_j$: it is the factor which scales the number of events from the source j to the luminosity of the data sample.
To give an estimation of the error on each fraction, p_j is varied around the solution; the variation that makes L change by a given amount - set according to the number of free p_j and the required confidence level - is taken as error on p_j. In this work, the change of L is set to 1σ (corresponding to a confidence level of 0.5); the number of degrees of freedom is given by the number of bins with at least one entry in the data sample.

5.2 The determination of the beauty, charm and light quarks content of the data

The method described above is now used to determine the fractions f_b, f_c and f_{uds} of, respectively, $b\bar{b}$, $c\bar{c}$ and light quarks events in the data sample; the observables used are the muon δ and $P_\perp$ relative to the jet axis ($P_\perp^{rel}$), and their combination in a 2D distribution. The results are shown in tab. 5.1.

	$P_\perp^{rel}$	δ	Combined
f_b	0.30 ± 0.06	0.35 ± 0.11	0.29 ± 0.05
f_c	0.42 ± 0.18	0.16 ± 0.28	0.36 ± 0.11
f_{uds}	0.28 ± 0.22	0.49 ± 0.21	0.35 ± 0.13
corr. *b-c*	0.62	-0.84	0.24
corr. *b-uds*	-0.75	0.67	-0.48
corr. *c-uds*	-0.97	-0.95	-0.91

Table 5.1: *Fractions of $b\bar{b}$, $c\bar{c}$ and light quarks events calculated from the fit on the muon δ, $P_\perp^{rel}$ and their combination.*

Fig. 5.1 shows the distributions of $P_\perp$, θ, $P_\perp^{rel}$ and δ of the muon after the Monte Carlo samples have been summed according to the fractions obtained from the fit on the 2D distribution.
To get a feeling about the sensitivity of the fit to the binning of the distributions, different binnings have been tried; the fractions fluctuate within the errors. No relevant dependence of the fit on the starting values has been observed, even in case the strength of the $b\bar{b}$ and $c\bar{c}$ signals was set to zero.
The fit on the combined distribution gives smaller errors and is less binning dependent [1]; it is in good agreement with the fit on the muon δ and $P_\perp^{rel}$ distributions alone.

[1] These are direct consequences of the higher number of degrees of freedom

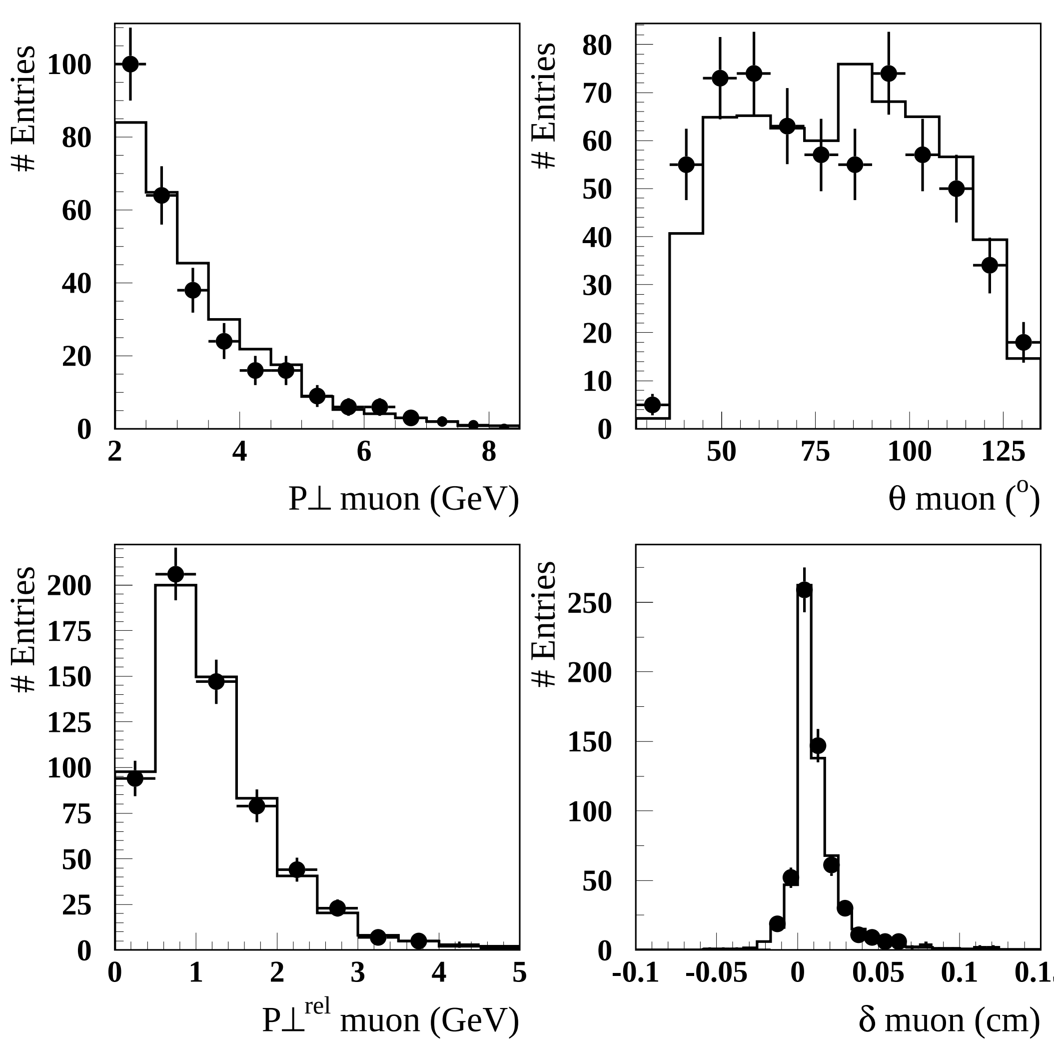

Figure 5.1: *$P_\perp$, θ, $P_\perp^{rel}$ and δ of the muon track in data (dots) and Monte Carlo, where the $b\bar{b}$, $c\bar{c}$ and light quarks components have been summed according to the fitted fractions.*

Chapter 6

The combined probability as b-tagging variable

6.1 Results on data

The combined probability is now an observable which can be fitted to determine the fraction of $b\bar{b}$, $c\bar{c}$ and light quark events in the sample (respectively, f_b, f_c and f_{uds},). If CP is a good observable to distinguish $b\bar{b}$ from $c\bar{c}$ from uds, a fit on it should give similar results to the previous fit using δ.
It is difficult to fit on the distribution of the combined probability itself, since there are bins with a much higher population than others, and also because there are only a few bins where the shape of the distribution in the different samples is significantly different. A better choice is the function:

$$L_{CP} = ln(1 - ln(CP)) \tag{6.1}$$

which is more evenly distributed and for which the three components differ in most of bins (see fig. 6.1 and 6.2). Fig. 6.2 seems to indicate that L_{CP} should be able to resolve the flavours.
The fit has been performed on the muon jet and then on any jet. The results are shown in table 6.1. The fit has been repeated with different binnings, to get a feeling about the fluctuations one has to expect. It has also been tried to repeat the fit on L_{CP} (in the muon jet) in combination with $P_\perp^{rel}$ of the muon track; the results are shown in table 6.2. The results of the fits on L_{CP} and L_{CP} combined with the muon $P_\perp^{rel}$ are compatible with what obtained in section 5.2. The results also clearly show that more statistics would be needed to confirm more precisely the results themselves.

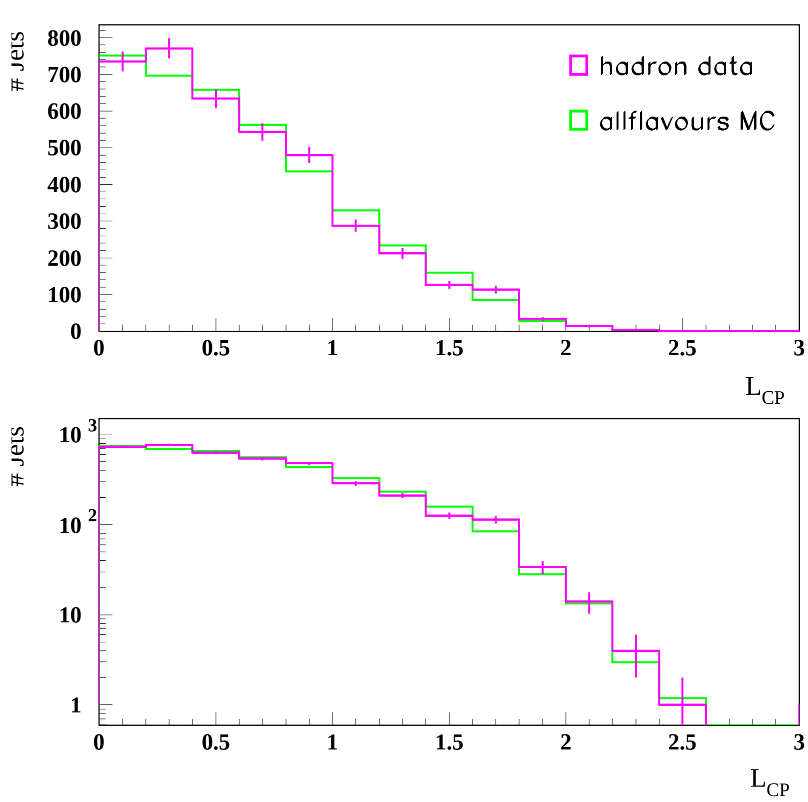

Figure 6.1: *L_{CP} in the hadron Monte Carlo and in the hadron data. The distributions are normalized, to compare the relative shapes. The plots with lin and log vertical scale are shown.*

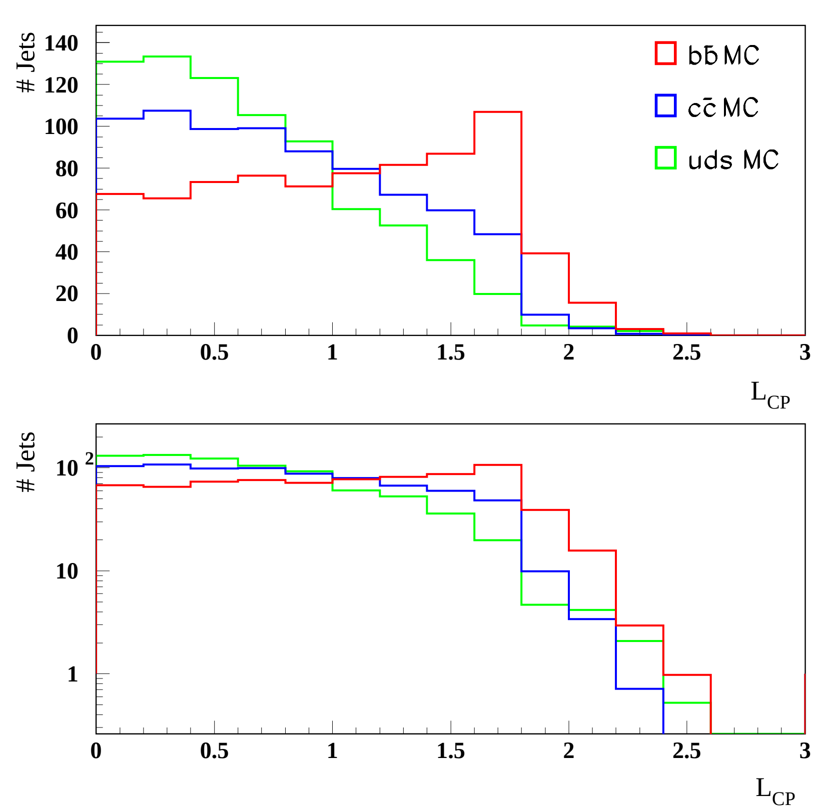

Figure 6.2: *L_{CP} in the $b\overline{b}$, $c\overline{c}$ and light quarks sample (respectively, linear and log scale). The distributions are normalized, to compare the relative shapes.*

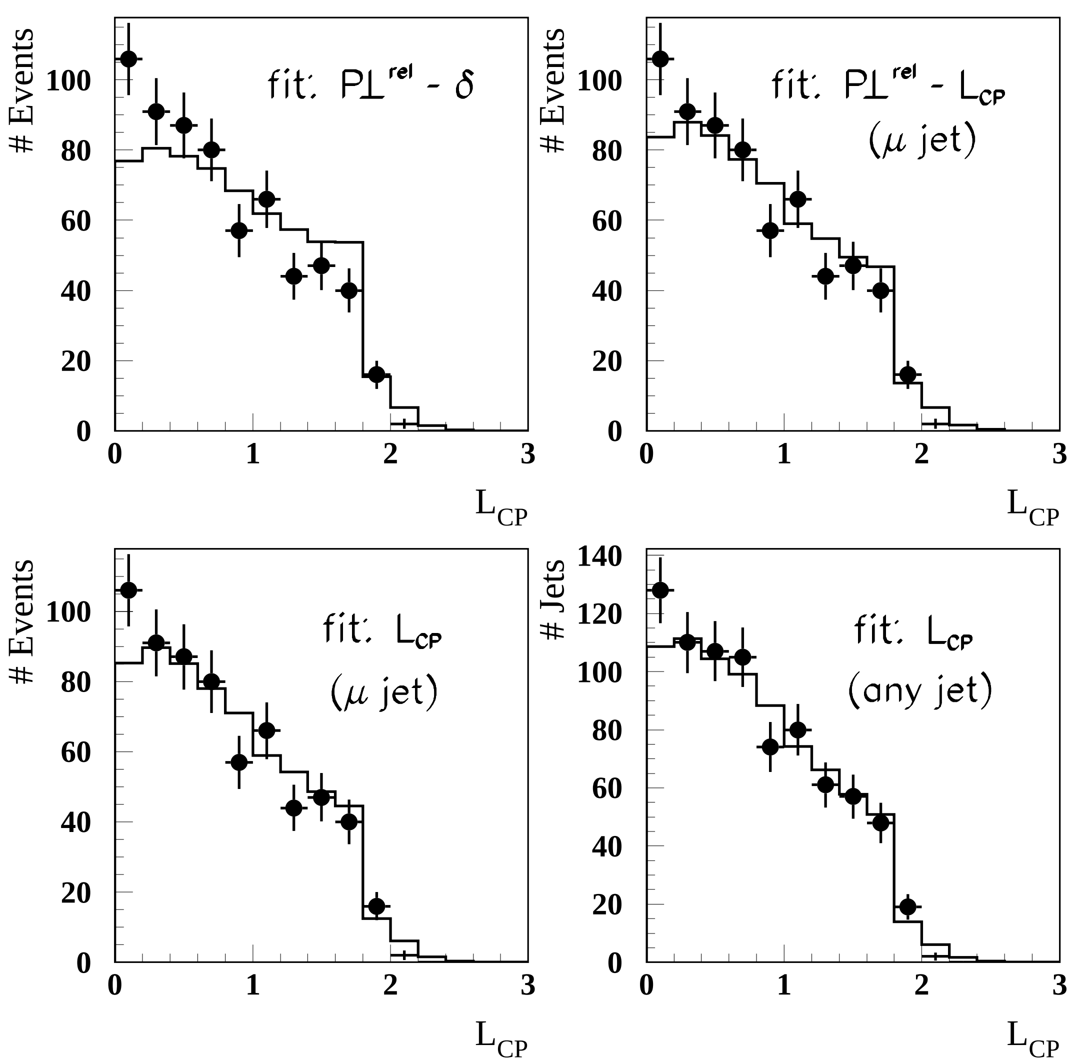

Figure 6.3: *L_{CP} in data (dots) and in Monte Carlo, where the $b\bar{b}$, $c\bar{c}$ and light quarks components have been summed according to the fitted fractions. The upper plots show L_{CP} in the muon jet, with the fractions obtained, respectively, from the fit on the muon $P_{\perp}^{rel} \otimes \delta$ and $P_{\perp}^{rel} \otimes L_{CP}$; the bottom plots show L_{CP} in the muon jet and in any jet, with the fraction obtained from the fit on L_{CP}.*

	30 bins	15 bins	10 bins
ANYJET			
f_b	0.17 ± 0.11	0.12 ± 0.12	0.17 ± 0.08
f_c	0.45 ± 0.31	0.58 ± 0.33	0.40 ± 0.19
f_{uds}	0.38 ± 0.23	0.30 ± 0.24	0.43 ± 0.14
corr. b-c	-0.91	-0.93	-0.81
corr. b-uds	0.81	0.85	0.57
corr. c-uds	-0.98	-0.98	-0.92
μJET			
f_b	0.20 ± 0.09	0.20 ± 0.09	0.19 ± 0.09
f_c	0.26 ± 0.23	0.22 ± 0.25	0.21 ± 0.23
f_{uds}	0.54 ± 0.17	0.58 ± 0.19	0.59 ± 0.17
corr. b-c	-0.80	-0.82	-0.81
corr. b-uds	0.60	0.65	0.60
corr. c-uds	-0.94	-0.95	-0.94

Table 6.1: *Fractions of $b\bar{b}$, $c\bar{c}$ and light quarks events calculated from the fit on L_{CP} for any jet and the muon jet; different binnings have been used.*

	30 bins in L_{CP}	15 bins in L_{CP}	10 bins in L_{CP}
f_b	0.22 ± 0.04	0.23 ± 0.04	0.23 ± 0.04
f_c	0.19 ± 0.11	0.23 ± 0.11	0.23 ± 0.11
f_{uds}	0.59 ± 0.12	0.54 ± 0.13	0.53 ± 0.12
corr. b-c	-0.06	0.17	0.06
corr. b-uds	-0.31	-0.41	-0.32
corr. c-uds	-0.90	-0.91	-0.90

Table 6.2: *Fractions of $b\bar{b}$, $c\bar{c}$ and light quarks events calculated from the fit on $L_{CP} \otimes P_{\perp}^{rel}$ (muon track).*

6.2 Check of the fit with a toy Monte Carlo sample

To test the quality of L_{CP} as discriminating variable with higher statistics, and to check the stability of the Barlow fit, a *toy* Monte Carlo sample with known components was employed. To generate this sample, the shape of the distribution of L_{CP} was taken from the $b\bar{b}$ sample and a new distribution with high statistics was generated according to this shape; the generation of the new distribution is based on the extraction of random numbers. The same was done on the $c\bar{c}$ and on the light quarks samples. After this, the three distributions are summed together, weighted with the fractions f_b, f_c and f_{uds}. The fractions f_b, f_c and f_{uds} are taken from the combined fit on the data sample, shown in tab. 5.1 (f_b=28.54%, f_c=36.37%, f_{uds}=35.09%); the number of events is 1500 $\times$ larger than that of the data sample. The results of the fit on L_{CP} are shown in tab. 6.3, 6.4; due the higher statistics, the errors on the fractions result much smaller, and allow to check for any bias or other effects. The input fractions are well reproduced by the fit.

	30 bins	15 bins	10 bins
ANYJET			
f_b	0.286 $\pm$ 0.003	0.287 $\pm$ 0.003	0.284 $\pm$ 0.003
f_c	0.363 $\pm$ 0.006	0.359 $\pm$ 0.006	0.367 $\pm$ 0.006
f_{uds}	0.351 $\pm$ 0.004	0.353 $\pm$ 0.005	0.349 $\pm$ 0.004
corr. *b-c*	-0.79	-0.79	-0.79
corr. *b-uds*	0.54	0.56	0.55
corr. *c-uds*	-0.92	-0.93	-0.92
μJET			
f_b	0.284 $\pm$ 0.002	0.285 $\pm$ 0.002	0.283 $\pm$ 0.002
f_c	0.366 $\pm$ 0.006	0.365 $\pm$ 0.006	0.370 $\pm$ 0.006
f_{uds}	0.350 $\pm$ 0.004	0.349 $\pm$ 0.005	0.347 $\pm$ 0.004
corr. *b-c*	-0.73	-0.74	-0.73
corr. *b-uds*	0.45	0.49	0.47
corr. *c-uds*	-0.91	-0.93	-0.92

Table 6.3: *Fractions of $b\bar{b}$, $c\bar{c}$ and light quarks events calculated from the fit on L_{CP} for any jet and the muon jet; different binnings have been used. The sample is the toy Monte Carlo with high statistics.*

The fit on the distributions of the muon quantities (δ, $P_{\perp}^{rel}$ and their combination) has also been repeated on the toy Monte Carlo; the results are given in tab. 6.5. The toy Monte Carlo results confirm with better statistics what has already been observed in data, namely that the fit on the combined distribution $P_{\perp}^{rel} \otimes \delta$ gives the most precise determination of f_b, f_c and f_{uds}, but that using the combined probability as a measure consistent results are obtained.

	30 bins in L_{CP}	15 bins in L_{CP}	10 bins in L_{CP}
f_b	0.277 ± 0.001	0.277 ± 0.001	0.277 ± 0.001
f_c	0.370 ± 0.003	0.370 ± 0.003	0.367 ± 0.003
f_{uds}	0.353 ± 0.003	0.353 ± 0.003	0.356 ± 0.003
corr. b-c	0.03	0.06	0.07
corr. b-uds	-0.31	-0.33	-0.34
corr. c-uds	-0.88	-0.89	-0.89

Table 6.4: *Fractions of $b\bar{b}$, $c\bar{c}$ and light quarks events calculated from the fit on L_{CP} combined with $P_{\perp}^{rel}$. The sample is the toy Monte Carlo with high statistics.*

	$\mathrm{P}_{\perp}^{rel}$	δ	Combined
f_b	0.287 ± 0.002	0.284 ± 0.003	0.281 ± 0.001
f_c	0.366 ± 0.006	0.363 ± 0.006	0.378 ± 0.003
f_{uds}	0.347 ± 0.007	0.353 ± 0.004	0.341 ± 0.003
corr. b-c	0.66	-0.74	0.06
corr. b-uds	-0.77	0.44	-0.35
corr. c-uds	-0.97	-0.91	-0.89

Table 6.5: *Fractions of $b\bar{b}$, $c\bar{c}$ and light quarks events calculated from the fit on the muon δ, $P_{\perp}^{rel}$ and their combination. The sample is the toy Monte Carlo with high statistics.*

Furthermore, it seems obvious that the fluctuations in the fractions observed in the data sample are due to the poor statistics of the sample.
To assert this we assembled a second toy Monte Carlo sample, with the same statistical size as the data sample; the composition of the sample has again been taken to be identical to that obtained from the combined fit on the data sample. The results are shown in tab. 6.6, 6.7 and 6.8. Large fluctuations are observed with respect to the input and to the values obtained with the toy Monte Carlo with high statistics, similar to what has been seen in the data sample.

6.3 Using CP as cut variable

After we have proven, that the combined probability may be used reliably to determine the b content of a data sample, we will investigate in this chapter to what extent a cut on this quantity may enrich a sample in $b\bar{b}$ events.
It is well evident (fig. 4.13, 4.14 and 6.2) that towards small values of CP (large values of L_{CP}) the signals from $b\bar{b}$, $c\bar{c}$ and light quarks events tend to separate. Figure 6.4 shows the purities predicted after different cuts on L_{CP}, in the toy Monte Carlo with high statistics;

	30 bins	15 bins	10 bins
ANYJET			
f_b	0.42 ± 0.05	0.40 ± 0.11	0.43± 0.06
f_c	0.0 ± 0.74	0.09 ± 0.51	0.0 ± 0.16
f_{uds}	0.58 ± 0.06	0.51 ± 0.22	0.57 ± 0.06
corr. b-c	-0.21	-0.89	-0.08
corr. b-uds	-0.61	0.78	-0.74
corr. c-uds	-0.36	-0.97	-0.16
μJET			
f_b	0.38 ± 0.09	0.35 ± 0.08	0.40 ± 0.06
f_c	0.11 ± 0.22	0.21 ± 0.22	0.0 ± 0.17
f_{uds}	0.51 ± 0.18	0.43 ± 0.17	0.60 ± 0.06
corr. b-c	-0.83	-0.77	-0.08
corr. b-uds	0.66	0.57	-0.73
corr. c-uds	-0.95	-0.95	-0.18

Table 6.6: *Fractions of $b\bar{b}$, $c\bar{c}$ and light quarks events calculated from the fit on L_{CP} for any jet and the muon jet; different binnings have been used. The sample is the toy Monte Carlo with low statistics.*

	30 bins in L_{CP}	15 bins in L_{CP}	10 bins in L_{CP}
f_b	0.27 ± 0.03	0.31 ± 0.04	0.31 ± 0.03
f_c	0.18± 0.09	0.22 ± 0.09	0.16 ± 0.08
f_{uds}	0.55 ± 0.10	0.47 ± 0.10	0.53 ± 0.09
corr. b-c	0.07	0.15	-0.01
corr. b-uds	-0.30	-0.37	-0.22
corr. c-uds	-0.90	-0.91	-0.89

Table 6.7: *Fractions of $b\bar{b}$, $c\bar{c}$ and light quarks events calculated from the fit on L_{CP} combined with $P_{\perp}^{rel}$. The sample is the toy Monte Carlo with low statistics.*

	$P_\perp^{rel}$	δ	Combined
f_b	0.33 $\pm$ 0.05	0.27 $\pm$ 0.07	0.32 $\pm$ 0.03
f_c	0.38 $\pm$ 0.17	0.39 $\pm$ 0.15	0.34 $\pm$ 0.08
f_{uds}	0.29 $\pm$ 0.20	0.34 $\pm$ 0.12	0.34 $\pm$ 0.09
corr. b-c	0.65	-0.63	0.11
corr. b-uds	-0.76	0.27	-0.36
corr. c-uds	-0.97	-0.88	-0.88

Table 6.8: *Fractions of $b\bar{b}$, $c\bar{c}$ and light quarks events calculated from the fit on the muon δ, $P_\perp^{rel}$ and their combination. The sample is the toy Monte Carlo with low statistics.*

the purities are calculated as explained in section 6.4, with the efficiencies given in table 6.9. A good separation of the three components is expected. Thus, different cuts on CP have been applied on data, and after each cut the fit on the combined distribution $P_\perp^{rel} \otimes \delta$ has been performed.
Table 6.10 summarizes the efficiencies of the tag after different cuts on the combined probability in the different samples, and the fractions of $b\bar{b}$, $c\bar{c}$ and light quarks events obtained from the fit on data; of course, only the muon jet has been used in this test. Even if with large uncertainties, it is observable that f_b tends to grow when one cuts more and more on CP, while f_c seems to decrease. Unfortunately the limited statistics makes it difficult to perform a good fit and separate the three components. The similarity of the distributions for the uds and the $c\bar{c}$ contributions makes the separation increasingly more difficult for hard cuts on the combined probability, where the statistics is low. For the b content of the sample we find consistent and stable answers. We further elaborate on that in the next section.
Fig. 6.5 shows $P_\perp^{rel}$ and δ of the muon, after a cut on CP smaller than 0.18 has been applied. The Monte Carlo still describes the data.

6.4 Calibration of the b-tag

As it was just shown, cutting on the combined probability is a way to enrich the b content in the sample. The Monte Carlo samples can be used to determine the efficiency of the tag. The purity, as seen in section 6.3, can be obtained from a fit on the data ($P_\perp^{rel} \otimes \delta$); it can also be calculated from the decomposition of the sample, according to the following formula (given for f_b^{MC}; f_c^{MC} and f_{uds}^{MC} are given by the same formula, after replacing the numerator with, respectively, $f_{c100} \times \epsilon_c$ and $f_{uds100} \times \epsilon_{uds}$) :

$$f_b^{MC} = \frac{f_{b100} \times \epsilon_b}{f_{b100} \times \epsilon_b + f_{c100} \times \epsilon_c + f_{uds100} \times \epsilon_{uds}} \tag{6.2}$$

CP <X(%)	100	90	80	70	60	50	40	30	25	22	20	18	16	14	12	10	8	6	4	2	1
ϵ ($b\bar{b}$) (%)	100	96	92	87	83	79	72	65	61	59	57	55.5	53	51	48	45	42.5	38	32	24	15
ϵ ($c\bar{c}$) (%)	100	93	87	79	72	65	56	48	44	40.5	38	36	34	31	29	26	23	19	15	09	04
ϵ (uds) (%)	100	92	83	74	65	56	47	38	32	30	27	25	23	21	19.5	17.5	15	11.5	09	05	03

Table 6.9: *Efficiencies of the tag on the different samples after different cuts on CP. The sample is the toy Monte Carlo with high statistics.*

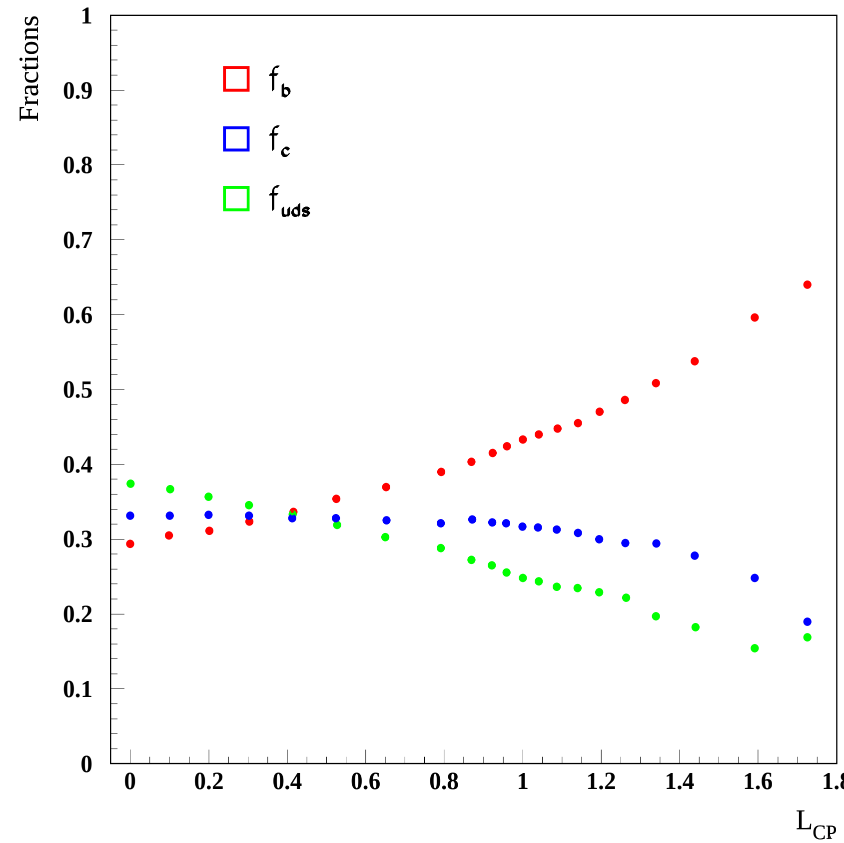

Figure 6.4: *Purities predicted after different cuts on CP: for each value of L_{CP}, the fractions are predicted for L_{CP} greater than the value. The statistical errors are extremely small in most of cases. The sample is the toy Monte Carlo with high statistics.*

Figure 6.4: *Purities predicted after different cuts on CP: for each value of L_{CP}, the fractions are predicted for L_{CP} greater than the value. The statistical errors are extremely small in most of cases. The sample is the toy Monte Carlo with high statistics.*

$CP < X(\%)$	100	50	30	25	18	12	8
ϵ_{tag} (dijet + μ)	1.	0.61	0.435	0.40	0.34	0.26	0.22
ϵ_{tag} ($b\bar{b}$)	1.	0.79	0.65	0.61	0.56	0.48	0.425
ϵ_{tag} ($c\bar{c}$)	1.	0.65	0.48	0.44	0.36	0.29	0.23
ϵ_{tag} (uds)	1.	0.56	0.38	0.32	0.25	0.195	0.15
$f_b \pm \Delta_{f_b}$ (%)	29±5	36±5	39±6	40±6	40±7	46±8	47±9
$f_c \pm \Delta_{f_c}$ (%)	36±11	64±6	61±7	60±7	60±7	54±8	53±9
$f_{uds} \pm \Delta_{f_{uds}}$(%)	35±13	0±63	0±12	0± 8	0± 9	0±7	0±4

Table 6.10: *Efficiencies of the tag on the different samples, and fractions of $b\bar{b}$ and $c\bar{c}$ events after different cuts on CP.*

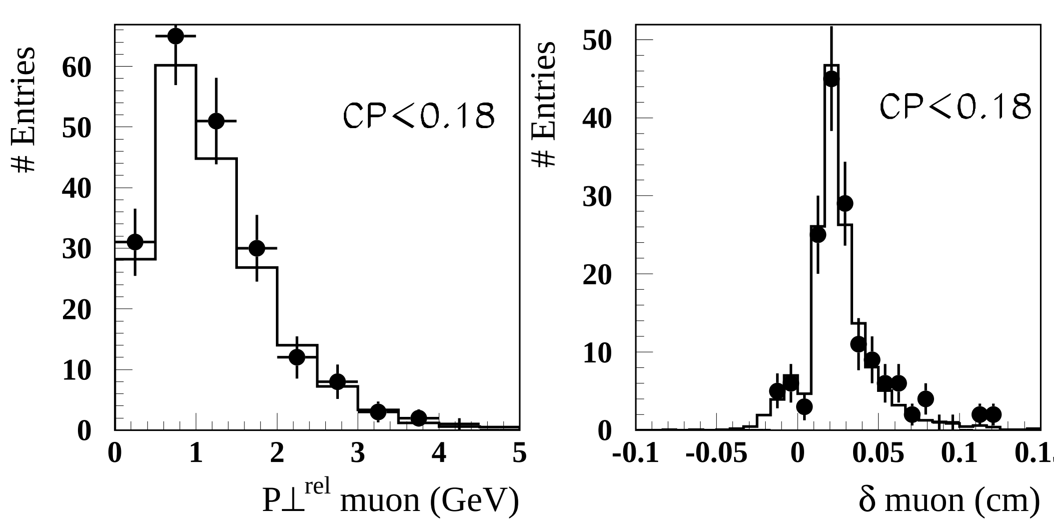

Figure 6.5: *$P_{\perp}^{rel}$ and δ of the muon track in data (dots) and Monte Carlo, where the $b\bar{b}$, $c\bar{c}$ and light quarks components have been summed according to the fitted fractions; a cut on $CP < 0.18$ has been applied.*

Here, f_{b100}, f_{c100}, f_{uds100} are the fractions of $b\bar{b}$, $c\bar{c}$ and light quarks events calculated with $CP \leq 1$; the efficiencies ϵ_b, ϵ_c and ϵ_{uds} are obtained from the number of events left after each cut on CP; the resulting f_b^{MC} is the fraction of $b\bar{b}$ events to be expected after each cut. Fig. 6.6 shows the purities obtained from the fit after different cuts on L_{CP}, together with the calculated values. The plots show clearly that the tag not only does a good job in the enrichment of the sample in $b\bar{b}$, but also that the value for f_b found from the data agrees well with what is expected from the above formula. Here, even if we could only

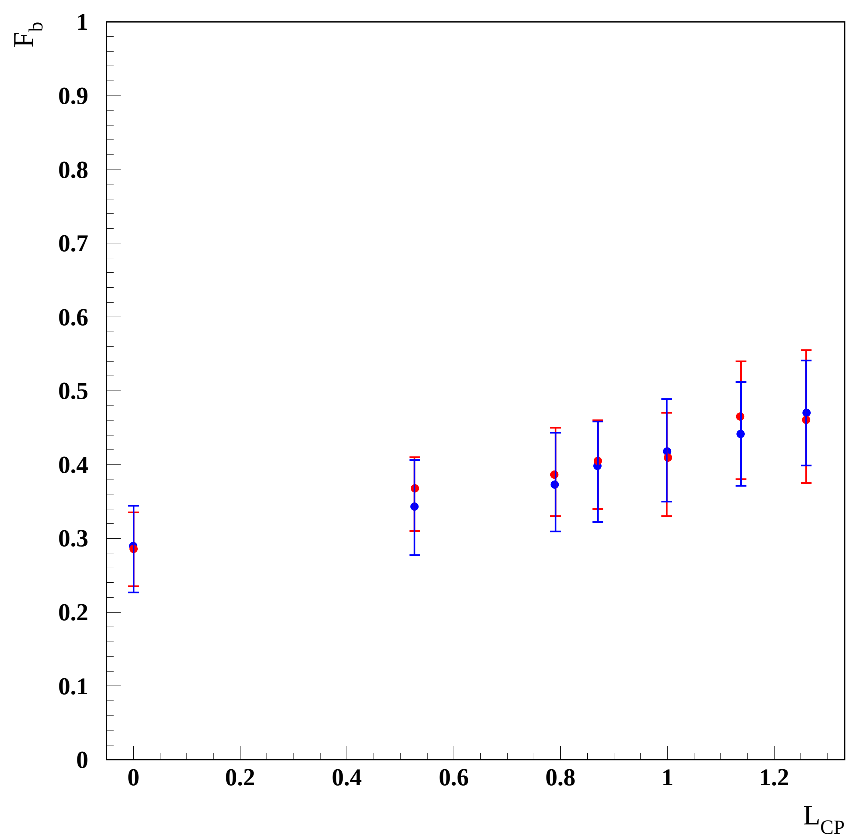

Figure 6.6: *Purities obtained from the combined fit on the muon track after different cuts on L_{CP}. The red dots represent the fitted values for f_b, the blue ones represent the predicted ones (cfr. comments to fig. 6.4). The sample is the data sample.*

test the method with a low statistics sample, the data efficiencies are well represented by the Monte Carlo simulation.

Chapter 7

Efficiency and purity in an inclusive sample

Once the tag is constructed, it can be applied to an inclusive sample. To obtain the correspondent data set, the same jet and track selection is applied; the subtrigger S19 is no longer requested and no muon is explicitly required.
In the all flavour Monte Carlo, it is possible to separate the $b\bar{b}$ and $c\bar{c}$ components and get the efficiencies of the tag after different cuts on the combined probability. These quantities have been calculated for any jet. The results are shown in tab. 7.1, together with the percentage of $b\bar{b}$, $c\bar{c}$ and light quarks events. Here, the a priori ratios $b\bar{b}$:$c\bar{c}$:uds are taken from the Pythia Monte Carlo.
Having dropped the muon requirement leads to a larger sample, but the percentage of heavy quarks events is smaller. It's well evident that also in this inclusive sample the combined probability can be used to enrich the b quarks content. The purity of the $b\bar{b}$ sample goes from 2.4% to 11.5% when cutting on CP <2%, and the rejection power of uds (defined as $\frac{\epsilon_b}{\epsilon_{uds}}$) goes up to 7.6 for CP <2%.

$CP < X(\%)$	100	50	30	25	18	12	4	2
ϵ (all flavours)	1.0	0.53	0.34	0.29	0.22	0.16	0.07	0.04
ϵ ($b\bar{b}$)	1.0	0.72	0.59	0.55	0.48	0.41	0.26	0.19
ϵ ($c\bar{c}$)	1.0	0.60	0.42	0.37	0.295	0.23	0.10	0.06
ϵ (light quarks)	1.0	0.50	0.30	0.25	0.185	0.13	0.05	0.025
$b\bar{b}$ perc. (%)	2.4	3.3	4.15	4.5	5.2	6.1	9.3	11.5
$c\bar{c}$ perc. (%)	27.0	30.5	33.1	34.2	35.9	37.9	41.8	42.7
uds perc. (%)	70.6	66.2	62.75	61.3	58.9	56.0	48.9	45.8

Table 7.1: *Efficiencies of the tag on the different samples, and purities after different cuts on CP.*

Chapter 8

The b cross section

The visible cross section for the production of $b\bar{b}$ events is calculated from the equation

$$\sigma^{vis} = \frac{N_\mu f_b}{2\mathcal{L}\epsilon} \tag{8.1}$$

where N_μ is the number of selected muons, $\mathcal{L}$ is the luminosity and ϵ is the efficiency of the analysis, which comes from the selection plus the trigger efficiencies. On top of the selection + trigger requirements, which give together an efficiency of 0.076, a further step is made when a cut on CP is applied. The efficiency ϵ_{tag}, given by the number of events left after the cut divided by the number before any cut, has to be included.
The effects of possible systematic contributions have also to be considered. The ones which result quantitatively relevant are listed in tab. 8.1.
With the value of f_b obtained from the fit on the muon quantities (f_b=0.29±0.05, N_μ=636) the cross section without any cut on CP results

$$\sigma^{vis} = (129 \pm 22(stat.)^{+22}_{-23}(syst.))pb$$

This result is in agreement with the one obtained in [24]:

$$\sigma^{vis} = (132 \pm 16(stat.)^{+26}_{-29}(syst.))pb$$

Fig. 8.1 shows the comparison of measured values of σ^{vis} to the theoretical predictions, including the result from [24] mentioned above.
The cross section has been calculated after different cuts on CP (cfr. table 6.10); the results are shown in table 8.2. The values of the visible cross section, calculated after different cuts on CP, are stable within the errors. It has to be consider that the results are highly correlated, thus the consistency of the values cannot be evaluated from the errors.

Source of uncertainty	contribution to $\Delta_{\sigma^{vis}}$ (%)
choice of the Monte Carlo generator	8
trigger efficiency	5
reconstruction of the event kinematics	2
hadronic energy measurement in the LAr calorimeter	8
muon reconstruction	8
jet reconstruction	+5, -10
(reconstruction of a good track)	(6)
reconstruction of the impact parameter	6
luminosity measurement	1.5

Table 8.1: *Systematic contributions to the error on σ^{vis}. The contribution from the reconstruction of a good track arises when the second jet is considered; it does not contribute to the cross section measurement presented in this work.*

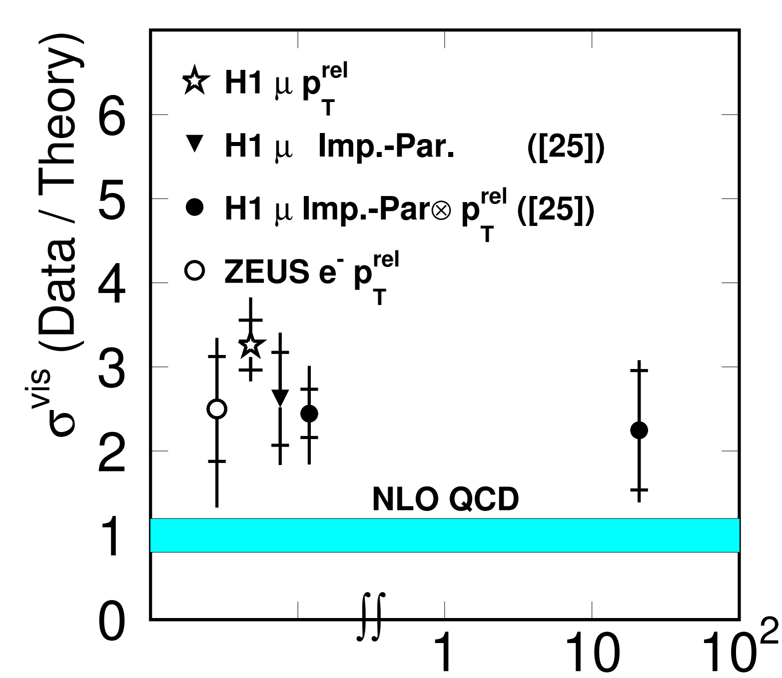

Figure 8.1: *Ratio of the measured to predicted values of the visible beauty production cross section. The result with $\sigma^{vis} = (132 \pm 16(stat.)^{+26}_{-29}(syst.))pb$ is the full dot in the low Q^2 region.*

$CP < X(\%)$	100	50	30	25	18	12	8
$\epsilon_{tag}\ (b\bar{b}\ )$	1.	0.79	0.65	0.61	0.555	0.48	0.425
N_μ	636	389	277	253	215	163	140
$f_b \pm \Delta_{f_b}$ (%)	29±5	36±5	39±6	40±6	40±7	46±8	47±9
$\sigma^{vis} \pm \Delta^{(stat.)}_{\sigma^{vis}}$ (pb)	129±22	124±17	116±18	114±17	106.5±19	109±19	106±21
$\Delta^{(syst.)}_{\sigma^{vis}}$ (pb)	${}^{+22}_{-23}$	${}^{+21}_{-22}$	${}^{+20}_{-21}$	${}^{+19}_{-20}$	${}^{+18}_{-19}$	${}^{+19}_{-19}$	${}^{+18}_{-19}$

Table 8.2: *Values of the visible cross section after different cuts on CP have been applied.*

Chapter 9

Conclusions

This work has developed the formalism for an inclusive tag using the combined probability CP as main variable. It fulfills the purpose of building a b-tagging method that can be applied to any process involving the production and decay of B hadrons. The combined probability CP results to be a good b-tagging variable, and a good variable to enrich the sample in $b\bar{b}$ events. This was expected from the estimates on Monte Carlo and has been confirmed by the data, though it is clear that larger statistics would be needed to make this statement stronger. A detailed study on systematic effects would also be needed to quantify the enrichment in $b\bar{b}$ events that can be obtained. The comparison of measured to estimate purities gives anyway a successful cross calibration of the method. A first study on an inclusive Monte Carlo sample indicates that the tag works well after the muon requirement has been dropped.
As final cross check of the analysis procedure the visible cross section for open b photoproduction has been determined, resulting:

$$\sigma^{vis} = (129 \pm 22(stat.)^{+22}_{-23}(syst.))pb$$

in good agreement with previous measurements. The cross section estimations after different cuts on CP indicate once more that a deeper investigation on the systematic effects will be demanded to perform accurate cross section measurements.
The efficiency of the tag drops down quite dramatically as harder cuts on CP are applied; correspondingly, the purity of the sample increases. With higher statistics, harder cuts could be applied and higher purities be available. The discrimination power at HERA is worst than what reached at LEP (see e.g. 13, where a purity above 50 % is achieved in the muon channel), mainly due to smaller track multiplicities and more multiple scattering, which pulls down the track momenta, and due to a smaller signal-to-background ratio. This work has provided a first indication about how far one can go at HERA.
Beauty production is still an open task in QCD; even the more recent maesurements of the beauty cross section have given results that do not always agree with QCD predictions. A comparison of recent HERA results to theoretical predictions is shown in fig. 9.1.

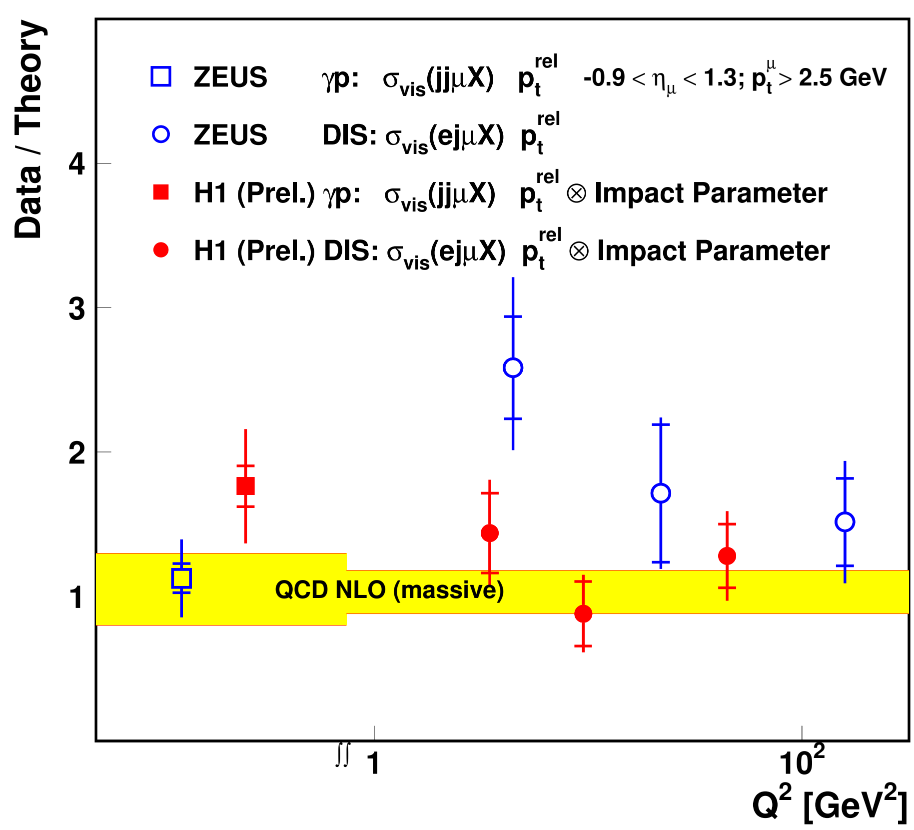

Figure 9.1: *Ratio of the measured to predicted values of the visible beauty production cross section. The measurements have been performed on data collected after 1997.*

Working with data samples of bigger statistical size, the *multi track method* for b-tagging might become an useful tool for high Q^2 and high $P_\perp$ analysis and searches for exotic processes. After the luminosity upgrade HERA is delivering a much higher luminosity; the integrated luminosity $\mathcal{L}$ is expected to be in the order of 100 pb^{-1} per year.
The forward silicon vertex detector FST[1] will give the possibility to extend this b-tagging method to a larger phase space, and study heavy quark production at higher X_P and lower X_γ and search for heavier, more forward going particles.

[1] *Forward Silicon Tracker*, included in H1 during the luminosity upgrade

List of Tables

List of Figures

Bibliography

[1] F. Halzen and A. D. Martin, *Quarks and Leptons: Introductory Course in Modern Particle Physics*, John Wiley & Sons, New York, 1984.

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

[3] R. K Ellis, W. J. Stirling and B. R. Webber, *QCD and Collider Physics*, Cambridge University Press, Cambridge (1996).

[4] J. D. Jackson , *Classical Electrodynamics*, second edition, John Wiley & Sons, Inc., New York-London-Sydney, 1975.

[5] V. N. Gribov and L. N. Lipatov , *Deep Inelastic Scattering in Perturbation Theory*, Sov. J. Nucl. Phys. 15 (1972), 438.
V. N. Gribov and L. N. Lipatov , *e^+e^- Pair Annihilation and Deep Inelastic ep Scattering in Perturbation Theory*, Sov. J. Nucl. Phys. 15 (1972), 675.
G. Altarelli and G. Parisi , *Asymptotic Freedom in Parton Language*, Nucl. Phys. B 126 (1977), 298.
Y. L. Dokshitzer , *Calculation of Structure Functions of Deep Inelastic Scattering and e^+e^- Annihilation by Perturbation Theory in Quantum Chromodynamics*, Sov. Phys. JETP 46 (1977), 641.

[6] J. J. Sakurai, *Theory of Strong Interactions*, Ann. Phys. (N. Y.) 11 (1960), 1.
Y. Nambu and J. J. Sakurai, *Rare Decay Modes of the ω (η) Meson*, Phys. Rev. Lett. 8 (1962) 79, 191 E.
M. Gell-Mann, D. Sharp and W. Wagner, *Decay Rates of Neutral Mesons*, Phys. Rev. Lett. 8 (1962), 261.

[7] R. K. Ellis and P. Nason, *QCD Radiative Corrections to the Photoproduction of Heavy Quarks*, Nucl. Phys. B 312 (1989), 551.

[8] C. F. Weizsäcker, *Ausstrahlung bei Stößen sehr schneller Elektronen*, Z. Phys. C 88 (1934), 612.
E. J. Williams, *Nature of the High Energy Particles of Penetrating Radiation and Status of Ionization and Radiation Formulae*, Phys. Rev. 45 (1934), 729.

[9] T. Sjöstrand, *High-Energy Physics Events Generation with PYTHIA 5.7 and JETSET 7.4*, Comp. Phys. Comm. 82 (1994), 74.

[10] S. Catani et al., *Longitudinally Invariant K(t) Clustering Algorithms for Hadron-Hadron Collisions*, Nucl. Phys. B 406 (1993), 187.

[11] C. Peterson et al.,*Scaling Violations in Inclusive e^+e^- Spectra*, Phys. Rev. D 27 (1983), 105.

[12] A. Meyer, Proc. XII Int. Workshop on Deep-Inelastic Scattering, Stavia-Lesna (Slovakia), April 2004.

[13] M. Acciarri et al. [L3 Collaboration], *Measurements of the Cross Sections for Open Charm and Beauty Production in γ γ Collisions at $\sqrt{s}$=189-202 GeV*, Phys. Lett. B 503 (2001), 10.

[14] F. Abe et al. [CDF Collaboration], *Measurement of bottom quark production in 1.8 TeV p anti-p collisions using muons from b-quark decays*, Phys. Rev. Lett. 71 (1993), 2396.
F. Abe et al. [CDF Collaboration], *Measurement of Correlated $\mu\bar{b}$ Jet Cross Section in $p\bar{p}$ Collisions at $\sqrt{s}$=1.8 TeV*, Phys. Rev. D 53 (1996), 1051.
S. Abachi et al. [DØCollaboration], *Inclusive m and b-Quark Production Cross Sections in $p\bar{p}$ Collisions at $\sqrt{s}$ = 1.8 TeV*, Phys. Rev. Lett. 74 (1995), 3548.
S. Abachi et al. [DØCollaboration], *J/ψ Production in $p\bar{p}$ Collisions at $\sqrt{s}$ = 1.8 TeV* , Phys. Lett. B 370 (1996), 239.

[15] M. Glück, E. Reya and A. Vogt, , Phys. Rev. D 45 (1992), 3945.

[16] C. Lai et al., *Global QCD Analysis of Parton Structure of the Nucleon: CTEQ5 Parton Distributions*, Eur. Phys. J. C 12 (2000), 375.

[17] S. Frixione et al. , *Total Cross Sections for Heavy Flavor Production at HERA*, Phys. Lett. B 348 (1995), 633.

[18] S. Frixione et al. , *Heavy Quark Correlations in Photon-Hadron Collisions*, Nucl. Phys. B 412 (1994), 225.

[19] G. Ingelman, J. Rathsman, G.A. Schuler *AROMA 2.2 - A Monte Carlo Generator for Heavy Flavour Events in ep Collisions*, Comp. Phys. Comm. 101 (1997), 135.

[20] I. Abt et al. [H1 Collaboration], *The H1 Detector at HERA*, Nucl. Instr. and Meth. A386 (1997), 310; ibid. 348.

[21] D. Pitzl et al., *The H1 Silicon Vertex Detector*, Nucl. Inst. Meth. A 454 (2000), 334.

[22] S. Lüders, *A Measurement of the Beauty Production Cross Section via $B \to J/\psi X$ at HERA*, Dissertation ETH Zürich (2001).

[23] M. Hilgers, *Untersuchung von Beauty-Ereignissen bei HERA an hand ihres myonischen Zerfalls und atrahlenharte Ausleseelektronik für den Silizium-Vertex-Detektor des H1-Experiments*, Dissertation ETH Zürich (2001).

[24] J. Kroseberg, *A Measurement of Beauty Production in High-Energy Positron-Proton Scattering* , Dissertation University of Zürich (2002).

[25] U. Langenegger, *A Measurement of the Beauty and Charm Production Cross Section at ep Collider HERA*, Dissertation ETH Zürich (1998).

[26] C. Adloff et al. [H1 Collaboration], *Measurement of Open Beauty Production at HERA*, Phys. Lett. B 467 (1999); Erratum ibid. B 518 (2001), 331.

[27] Adloff et al. [H1 Collaboration], *Measurement of the Beauty Production Cross Section at HERA Using Lifetime Information*, ICHEP2000, Osaka, contributed paper 979, 982 (2000).

[28] Adloff et al. [H1 Collaboration], *Measurement of Open Beauty Production at HERA Using Semi-Muonic Decays*, EPS03, Aachen, abstract 117.

[29] J. Breitweg et al. [ZEUS Collaboration], *Measurement of Open Beauty Production in Photoproduction at HERA*, Eur. Phys. J C18 (2001), 625.

[30] J. Breitweg et al. [ZEUS Collaboration], *Measurement of Open Beauty Production at HERA Using a D^* + muon Tag*, EPS03, Aachen, abstract 575.

[31] J. Breitweg et al. [ZEUS Collaboration], *Measurement of Beauty Production in Deep Inelastic Scattering at HERA*, EPS03, Aachen, abstract 561.

[32] J. Breitweg et al. [ZEUS Collaboration], *Measurement of Beauty Photoproduction at HERA*, ICHEP02, Amsterdam, abstract 785.

[33] M. Frank, *Eine Przisionsmessung des Verzweigungsverhältnisses in der e^+e^--Annihilation auf der Z0-Resonanz*, Dissertation Technische Universität München (1994).

[34] D. Buskulic et al. [ALEPH Collaboration], *Measurement of the Effective b Quark Fragmentation Function at the Z Resonance*, Phys. Lett. B 357 (1995), 699.

[35] R. Barlow, and C. Beestone, *Fitting Using Finite Monte Carlo Samples*, Comp. Phys. Comm. 77 (1993), 219.

I would like to thank all the people that walked with me during the time invested in this work. With their teaching, help, support, friendship and love they made this piece of research possible and worthwhile.
Thanks to all of you. Grazie a tutti voi.

Curriculum Vitae

Personal data:

Name:	Ilaria Foresti
Birth:	March 30th, 1970, in Firenze, Italia
Nationality:	Italian

Studium:

1976 - 1981	Primary school in Torino, Italia
1981 - 1984	Intermediate school in Firenze, Italia
1984 - 1989	Liceo specializing in scientific studies in Firenze, Italia
1989	School-leaving examination
1989 - 1997	Physics study at the University of Florence, Italia
1997	Degree in physics
1997 - 1998	Working in physics research with the L3 experiment, at the University of Florence, at CERN and DESY Zeuthen
1998 - 2004	Working in physics research with the H1 experiment, at the University of Zürich. Dissertation by Prof. Dr. P. Truöl